12-2023

The Private Pilot Check Ride: Applying the Spacing Effect Theory To Predict Time To Proficiency For the Practical Test

Michael Scott Harwin
Florida Institute of Technology - Melbourne, mharwin2019@my.fit.edu

The Private Pilot Check Ride: Applying the Spacing Effect Theory to Predict Time to Proficiency for the Practical Test

by

Michael Scott Harwin

Bachelor of Science
Aeronautical Science
Embry Riddle Aeronautical University
1986

Bachelor of Science
Airway Science
Embry Riddle Aeronautical University
1988

Juris Doctorate in Law
Rutgers Law School–Newark
1994

Master in Tax Law
New York University School of Law
1999

Master of Science
Aeronautics
Embry Riddle Aeronautical University
2018

Master of Science
Aviation Human Factors
Florida Institute of Technology
2021

A dissertation submitted to the College of Aeronautics
of Florida Institute of Technology
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Aviation Sciences

Melbourne, Florida
December, 2023

We the undersigned committee
hereby approve the attached dissertation

The Private Pilot Check Ride:
Applying the Spacing Effect Theory
to Predict Time to Proficiency for the Practical Test

by

Michael Scott Harwin

Michael A. Gallo, Ph.D.
Professor Emeritus
College of Aeronautics
Major Advisor

Deborah S. Carstens, Ph.D.
Professor
College of Aeronautics

Victoria L. Dunbar, Ph.D.
Graduate Faculty
College of Aeronautics

Thomas Marcinkowski, Ph.D.
Professor
Mathematics and Systems Engineering

John E. Deaton, Ph.D.
Professor and Dean
College of Aeronautics

Abstract

TITLE: The Private Pilot Check Ride: Applying the Spacing Effect Theory to Predict Time to Proficiency for the Practical Test

AUTHOR: Michael Scott Harwin

MAJOR ADVISOR: Michael A. Gallo, Ph.D.

This study examined the relationship between a set of targeted factors and the total flight time students needed to become ready to take the private pilot check ride. The study was grounded in Ebbinghaus's (1885/1913/2013) forgetting curve theory and spacing effect, and Ausubel's (1963) theory of meaningful learning. The research factors included (a) training time to proficiency, which represented the number of training days needed to become check-ride ready; (b) flight training program (Part 61 vs. Part 141); (c) organization offering the training program (2- or 4-year college/university vs. FBO); (d) scheduling policy (mandated vs. student-driven); and demographical variables, which consisted of (e) biological sex assigned at birth (female vs. male), (f) age, (g) race/ethnicity, and (h) marital status. Convenience/snowball sampling strategies were used to solicit flight students from various flight schools and pilots from United Airlines who provided the same data as flight students based on their recollection of when they were flight students in a PPL training program. The primary data collection instrument was a researcher-developed questionnaire designed to capture participants' self-reported factual data related to the targeted variables, and the sample size used to test the study's hypotheses was $n = 164$ participants.

Preliminary data screening eliminated all except three variables and required a natural logarithm transformation for training time to proficiency. A follow-up hierarchical multiple regression analysis indicated that training time to proficiency, organization, and scheduling policy were strong predictors of flight hours needed to become check-ride ready and collectively explained 19% of the variance in the DV. The results also supported Ebbinghaus's (1885/1913/2013) forgetting curve theory and confirmed that as the number of training days increase, the number of flight hours needed to become check-ride ready was asymptotic at 70 hours. An application of the prediction model suggests flight students enrolled in a Part 61 program at an FBO with a student-driven schedule need approximately 30 days of flight training to become check-ride ready.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I take this opportunity to thank Dr. Michael Gallo for agreeing to be my major advisor for this dissertation and for teaching all my Ph.D. courses. I do not know anyone who exhibited more professionalism in preparing his classes, teaching his classes, and as a major advisor to his Ph.D. students. Not only did Dr. Gallo agree to be my major advisor and provide me with invaluable and indispensable guidance, but it was at a time he was in the process of retiring. To agree to be a student's major advisor during his retirement shows a level of commitment I have not experienced. There are no words to describe his extraordinary selflessness and generosity. I am fortunate to have had the opportunity to work with him and experience his professionalism firsthand. I am proud to have him listed as my major advisor for this dissertation.

My committee members also are worthy of recognition for being generous with their time and patience. Thank you to committee member Dr. Debbie Carstens, who provided me with valuable information concerning aviation human factors, which was needed for this dissertation. Thank you to committee member Dr. Victoria Dunbar, who provided me with valuable insight into the process of obtaining a private pilot license, which was indispensable for this dissertation. Thank you to outside committee member Dr. Thomas Marcinkowski, who took the time to provide me with valuable educational insight into the theory of meaningful learning, which was a significant element of this dissertation.

I would be remiss not to thank my classmates who sat with me for a year and a half through the required Ph.D. graduate courses. I appreciate the many conversations and discussions I had with them, which gave me ideas and insights to complete the Ph.D. classwork and this dissertation.

My parents instilled in me the knowledge that education is essential. I could only have accomplished this Doctorate degree in aviation sciences with their support. Thank you for teaching me that education is a lifelong pursuit. They have supported all my educational goals and encouraged me to continue growing in each of my professions. I am very thankful for their support.

Dedication

My parents have always encouraged and supported me to pursue my education. For this reason, I dedicate this dissertation to them: Bruce Harwin and Sandra Rosenfelt. They have strongly influenced my life and instilled in me a professional work ethic that enabled me to pursue my education degrees and complete this doctorate in aviation sciences.

Chapter 1

Introduction

Background and Purpose

Background

The primary focus of the current study was to examine the relationship between various flight- and demographic-related factors and the number of flight hours students needed to complete before their certified flight instructor (CFI) considered them proficient to take the private pilot practical test. From a practical perspective, the research problem was to develop a model that could be used to predict the number of flight hours needed to become check-ride proficient. However, from a research perspective, the research problem was grounded in the dearth of past studies that examined the simultaneous effect of classroom instruction and motor skills learning.

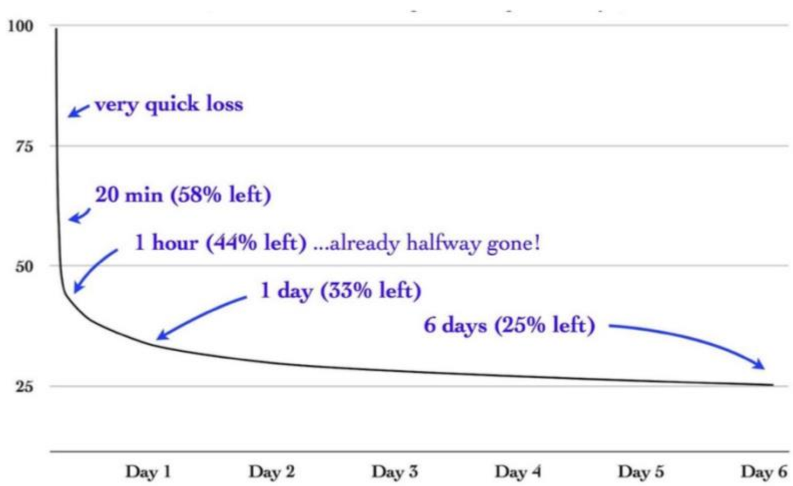
In accordance with Federal Aviation Regulations (FAR, 2021a), the Federal Aviation Administration (FAA) requires a minimum of 40 flight hours before a pilot can qualify to take the private pilot practical test, which is more commonly referred to as the "check ride." The minimum number of flight hours is reduced to 35 hours, however, for students receiving flight instruction from a Part 141 school (FAR, 2021b). This 5-hour reduction in minimum flight hours presumes that flight instruction at a Part 141 school is more effective than flight instruction at a Part 61 school. Regardless of the FAA's reason for this policy, though, few student pilots are deemed proficient by their CFIs to take the check ride at 40 hours, let alone 35 hours. According to the most recent data from the FAA (2006), the national average is nearly twice the minimum hours needed. For example, most flight students in the U.S. require, on average, between 60 and 75 hours to become proficient to take the private pilot check ride. This difference between the FAA's minimum requirement and national average is similar to a doctoral program that requires—independent of all other required coursework—a minimum of 18 semester credit hours of dissertation coursework: Although it might be possible to complete the program with exactly 18 hours of dissertation, additional hours most likely will be required. Recognizing the difference between minimum requirements and what actually is needed is important for planning and budgeting purposes because of the corresponding additional time and expense of a doctoral degree program. The same can be said for flight students. Based on the national average, achieving check ride proficiency most likely will require more time, and the associated cost will be more than if the minimum requirements were met. Furthermore, the cost can be substantial because not only does the student have to pay for the cost of the airplane rental, but also for the CFI's time. By focusing on the time interval or spacing between flights, it might be possible to reduce the number of flight hours and concomitantly achieve substantial savings.

Although flight instruction is a unique discipline, it still is considered an academic endeavor and therefore can be examined relative to various learning theories. When applied to the current study, these theories include Ebbinghaus's (1885/1913/2013) forgetting curve theory and spacing effect theory, and Ausubel's (1963) theory of meaningful learning. For example, Ebbinghaus observed that frequent repetitions are central to recalling information and that with a sufficient number of repetitions, mastery is obtained, and by additional practice over time, the information can be readily recalled. According to Ebbinghaus, to be able to recall previously learned information, though, the information must be grounded together by study and reviewed later. This phenomenon is reported in the literature as the forgetting curve theory or retention curve theory (Murre & Dros, 2015). Ebbinghaus examined this forgetting curve theory relative to cognitive related tasks such as learning vocabularies, discourses, and poems, and reported that if the information is not reviewed periodically, then the information is difficult to recall at a later date. As illustrated in Figure 1.1, the distribution of Ebbinghaus's experimental data resulted in a logarithmic-like relationship between the percentage of information lost and time (in days). Within 1 day, approximately 70% of the information cannot be recalled if no attempt is made to remember it.

However, Ebbinghaus (1885/1913/2013) also posited that the forgetting curve could be mitigated by a sufficient number of repetitions and later practice. Some authors have referred to this repetition as overlearning (Shrestha, 2017). The theory is that if students studied more than what is required to memorize something, overlearning occurs. This means that the information is now easier to

Figure 1.1

Graphical Representation of Ebbinghaus's Forgetting Curve



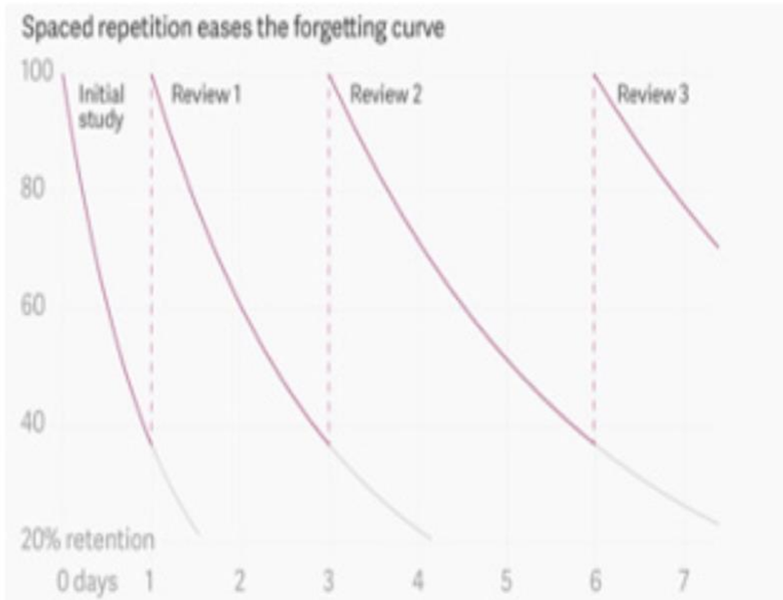
Note. The percentage of information lost (vertical axis) as a function of days if there is no effort made to remember it by practice or study. The curve is logarithmic in form and is asymptotic (levels off) at around 25% of lost information. Adapted from Sonnad (2018).

recall as time progresses, and the effect of the forgetting curve for overlearned information is shallower (Shrestha, 2017). As illustrated in Figure 1.2, this phenomenon is known as the spacing effect and supports the notion that distributed practice is more effective than compressed, or massed, practice (Ormrod, 2012, p. 208). As a result, reviewing newly acquired information in a distributed format continually slides the forgetting curve to the right, thereby delaying the effects of not being able to recall this information.

The forgetting curve and spacing effect also are consistent with Ausubel's (1963) theory of meaningful learning. In part, Ausubel reported that the most critical factor students bring to the learning table is the knowledge they currently possess that relates to what they are learning, which is their prior knowledge

Figure 1.2

Illustration of the Spacing Effect on the Forgetting Curve



Note. This figure demonstrates that review and practice slide the forgetting curve toward the right, thereby ameliorating the forgetting curve's effects. The graph shows that after 1 day, a student's retention percentage has decreased from 100% to 40%. After two review and practice sessions, the student maintains a 40% retention percentage for up to 3 days. With three review and practice sessions, a student maintains a 40% retention percentage for up to 6 days. *Source:* Sonnad (2018).

(Ausubel et al., 1978). By spacing out the information, students are connecting and retrieving previously learned information with new information. This spaced information provides learners with more opportunities to relate new knowledge with prior knowledge, which in turn, stimulates meaningful learning.

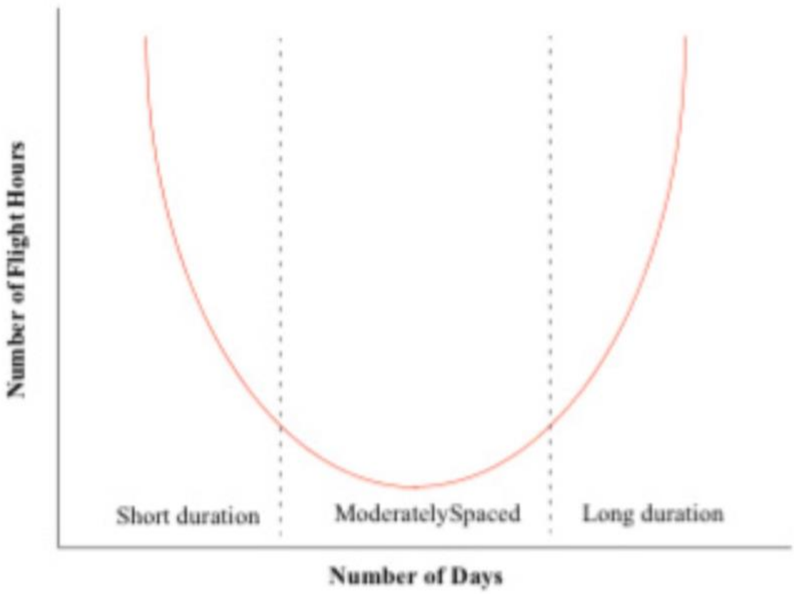
When applied to the context of the current study, Ebbinghaus's (1885/1913/2013) and Ausubel's (1963) respective theories suggest that the number of hours students need to become proficient to take the private pilot check ride should vary based on the time between their first and last flights before the check ride. Ebbinghaus's forgetting curve theory would posit that scheduling flights too close together would be considered cramming (i.e., massed practice), and students would need more flight hours to become proficient before their check ride than students who spaced out their flights (i.e., distributed practice). Complementing this perspective, Ausubel's theory would posit that scheduling flights too close together would not provide students with enough time to process newly acquired information by linking it to previously acquired information. More concretely, for flight students to retain their previous knowledge from prior flight lessons, they must link it with the new information learned in their current flight lesson. For example, flight students cannot land an airplane until they know how to control airspeed, prepare a descent, and maneuver the airplane, which all represent prior knowledge acquired from previous lessons. Thus, students would need more flight hours to become proficient before their check ride than students who appropriately spaced out their flight lessons. As noted by Ormrod (2012):

> … learning is sometimes a bit *slower* when it's spread out over time (but its) benefits are most clearly seen when we look at *long-term retention* rather than speed of initial learning … If we really want to remember information for the long haul, then, we should review it periodically at lengthy intervals. (p. 208)

On the other hand, if flight lessons are spaced too far apart, then this would lead to an increase in the number of flight hours needed for check ride proficiency. This is

Figure 1.3

Changes in the Number of Flight Hours as a Function of Number of Days Before Check Ride



Note. This figure illustrates the prohibitive effect on the number of flight hours needed to acquire proficiency for the check ride relative to the spacing of flight lessons. If flight lessons are spaced too close together (short duration), students will not have sufficient time to process new knowledge. On the other hand, if flight lessons are spaced too far apart (long duration), they will forget what they learned previously, which will increase the number of flight hours needed for proficiency.

because—as Ebbinghaus (1885/1913/2013) warned—repetition is required to attain proficiency, and by spacing flights too far apart, repetition would be minimized, and proficiency lost. This relationship is illustrated in Figure 1.3.

Prior studies (e.g., Bahrick et al., 1993; Dempster, 1991; Dempster & Farris, 1990) involving classroom or cognitive learning reported that long-term memory is enhanced when learning events are spaced apart in time rather than close together, thus giving support to Ebbinghaus's (1885/1913/2013) and

Ausubel's (1963) respective theories. More recently, Gallo and Odu (2009) reported that periodically reviewing classroom material within a given time enhances recall, and that by reviewing learned information periodically over time, the information was easier to recall from long-term memory than when the information was learned all at once. There also is an abundance of literature indicating that these theories have been found to support enhanced motor skills learning. For example, Baddeley and Longman (1978) reported that postmen with no keyboarding experience were more efficient at learning to use a typewriter and had higher retention when training was distributed over time than massed. Shea et al. (2000) reported a similar finding in their study, which examined participants' learning of various motor skills across days (i.e., distributed practice) as opposed to within days (i.e., massed practice): "The results indicated that spacing practice sessions over relatively long intervals (days) resulted in the enhancement of performance during the remaining practice sessions and enhanced learning as assessed by the delayed retention test" (Baddeley & Longman, p. 737). Although dated, in a study that involved teaching students basic maneuvers in a rudimentary flight simulator, Farr et al. (1956) reported that students who were provided distributed practice had superior flight performance than those who were not provided with distributed practice.

The reader will note that in the studies cited in the foregoing paragraph, the focus was on either cognitive skills or motor skills, but not both. Other than a few exceptions, there is a dearth of literature that examined the simultaneous effect of classroom instruction coupled with learning a motor skill, which is needed to obtain a private pilot license. One such exception is Mengelkock (1971), which involved teaching participants with no flying experience how to fly an aircraft. Mengelkock provided participants with 4 hours of academic training (cognitive skills) and one familiarization trial in a simulator. He then randomly assigned participants to one of two groups: One group was given five training trials in the simulator, and the second group was given 10 training trials in the simulator. These training trials, which reflected participants' motor skills, involved "a structured 50-min. mission of maneuvers and procedures from starting the engine and takeoff to landing and shut down of the engine" (Mengelkock, p. 397). Four months after the last training trial, both groups participated in four additional training trials. Mengelkock reported that although both groups had a loss of cognition between 16.5% and 20.1% with respect to academic classroom retention, the difference between groups was not statistically significant. However, the cognitive retention loss plots showed a logarithmic distribution, as predicted by Ebbinghaus's (1885/1913/2013) forgetting curve, which theorized a logarithmic loss of information without continual practice. Furthermore, because the second group was provided with distributed simulator practice while the first group was not (5 trials vs. 10 trials), the second group had superior flight performance, which supports Ausubel's (1963) meaningful learning theory as well as Ebbinghaus's (1885/1913/2013) forgetting curve and spacing effect theories.

As presented throughout this background section, the application of Ebbinghaus's (1885/1913/2013) and Ausubel's (1963) respective theories to flight instruction is appropriate from both theoretical and practical perspectives. With respect to the former, the theories may be used to deduce and test hypotheses to help explain the relationship between the number of days and number of flight hours it takes flight students to become proficient to take the private pilot check ride. The results of this analysis could then be used by flight students and CFIs for planning and budgeting purposes to estimate how long, on average, it will take flight students to become proficient to take their check ride.

Purpose

The purpose of the current study was to determine the relationship between a set of targeted factors and the number of flight hours flight students needed to become check-ride proficient as determined by their CFI. In the context of the current study, flight students were defined as students who were enrolled in a FAA approved flight training program that led to a private pilot license (PPL). A CFI was defined as a person who had a current FAA-issued flight instructor certificate and was authorized by the FAA to provide flight instruction and issue endorsements within the limits of the person's certificate and ratings. (FAR, 2021c). Examples of the types of training and endorsements CFIs could provide are provided in FAR (2021d) and include, but not limited to, those that would lead to a student pilot certificate, pilot certificate, and flight instructor certificate. Check-ride proficiency was defined as a CFI's professional judgment that a flight student has acquired the skills necessary to take the practical test and knowledge test as stipulated in the Federal Aviation Regulations (FAR, 2021e, 2021f, 2021g). The practical test consists of a check ride in which flight students fly an aircraft accompanied by an FAA-authorized examiner who assesses students' competency relative to the skills required for certification. As a result, the practical test is commonly referred to as the check ride and is considered the final examination that flight students must pass to either receive or maintain pilot certification. The knowledge test consists of a written multiple-choice examination encompassing general aviation knowledge areas.

The targeted factors included time to proficiency, student demographics, and flight school characteristics. Time to proficiency was defined as the number of flight training days it took students to become check-ride proficient. This time interval began the day students took their first flight to the day their CFI determined they were proficient to take the check ride, where 1 day was defined as 24 hours, starting from 12:00 a.m. to 11:59 p.m. Student demographics included biological sex assigned at birth (female or male), age, race, ethnicity, and marital status. These factors were considered is part to address the selection threat to internal validity and provided information relative to sample representativeness.

Flight school characteristics referred to the type of flight training program from which students received their private pilot training. According to the FAA (FAR, 2021a, 2021b), there are two types of flight training programs: (a) FAA-certificated (approved) and (b) non-FAA approved. FAA-approved pilot programs are certificated with respect to Title 14, Code of Federal Regulations Part 141 (i.e., 14 CFR Part 141), and organizations offering programs compliant with 14 CFR Part 141 are commonly referred to as Part 141 schools. The FAA requires all Part 141 schools to use a structured training program and syllabus, and they also must meet specific standards relative to equipment, facilities, and personnel. Most colleges and universities with an aviation program offer pilot training that is compliant with 14 CFR Part 141. Non-FAA-approved pilot training programs are described under 14 CFR Part 61, and organizations offering such programs are commonly referred to as Part 61 schools. Unlike Part 141 schools, the FAA does not certificate Part 61 schools as pilot schools and therefore the training environment is not as strict. Most flight training programs offered at airports are compliant with 14 CFR Part 61.

Although both types of training programs follow the same FAA practical test standards, there are key differences between the programs. Relative to the current study, one key difference is the minimum number of hours to PPL required by the FAA: 35 hours for Part 141 vs. 40 hours for Part 61. This difference, though, as noted later in this chapter, might not be relevant because most flight students require 60 to 75 hours of training to obtain a PPL. Another consideration is scheduling. Some flight schools, regardless of their FAA status, require students to maintain a rigid flight training schedule whereas others permit students to schedule their training at their convenience. For the current study, I collected data on the type of flight training program (Part 141 vs. Part 61), the type of organization that offers the program (public/private 2- or 4-year colleges/universities vs. fixed-base operators [FBOs], which offer flight training at airports), and the scheduling policy (mandated vs. student-driven).

The dependent variable was flight time to proficiency, which was defined as the number of flight hours needed before a CFI determines a student was check-ride proficient. A flight hour was defined as when the engine was started to when the engine was shut down as recorded by the aircraft's Hobbs meter, which measures the time an aircraft is in use. The reader will note that the current study was not concerned with whether a student passed the check ride, but instead is focused primarily on understanding the effect key variables had on the number of flight hours needed before CFIs declared flight students proficient to take the practical and knowledge tests.

Definition of Terms

In the context of the current study, the corresponding key terms and phrases are operationally defined as follows:

1. *2- or 4-year college/university* was defined as a post-secondary educational organization that delivered educational programs leading to an associate of science (A.S.), associate of arts (A.A.), and/or baccalaureate degrees.
2. *Certified flight instructor* (*CFI*) was defined as a person who had a current FAA-issued flight instructor certificate and was authorized by the FAA to provide flight instruction and issue endorsements within the limits of the person's certificate and ratings (Part 61.1(b), FAR, 2021c).
3. *Check-ride proficient* was defined as a CFI's professional judgment that a flight student had acquired the skills necessary to take the practical test and knowledge test as stipulated in the Federal Aviation Regulations (FAR, 2021e, 2021f).
4. *Fixed-based operator (FBO)* referred to any organization located at an airport that was given consent by the airport to provide various services such as aircraft fueling, rental, and maintenance, as well as flight instruction.
5. *Flight hour* was defined as the total number of hours a pilot accrued flying an aircraft either physically or in a simulator. This was measured by the aircraft's Hobbs meter, which records the time an aircraft is in use from when the aircraft engine is started to when the aircraft engine is shut down.
6. *Flight student* was defined as a student enrolled in a flight training program leading to a private pilot license (PPL). See also *Student pilot*.

7. *Flight time to proficiency* was defined as the number of flight hours needed before a CFI determines a student is proficient to take the check ride and knowledge test pursuant to Title 14 Code of Federal Regulations (CFR) Part 61, et seq. and Part 141 et seq. (FAR, 2021e, 2021f).
8. *Flight training organization*. See *Type of organization*.
9. *Flight training program* was defined by either 14 CFR Part 61 or 14 CFR Part 141. See also Part 61 program and Part 141 program.
10. *Flight training scheduling policy* referred to whether students followed a prescribed training schedule (*mandated*) or had the flexibility to schedule their flight training at their convenience (*student-driven*).
11. *Knowledge test* was defined as a test containing the areas of aeronautical knowledge listed in Title 14 Code of Federal Regulations (CFR) Part 61.105(b) (FAR, 2021f).
12. *Mandated schedule* was defined as a flight training schedule where the flight training organization imposes a required schedule that flight students must follow. Contrast with *Student-driven schedule.*
13. *Organization*. See *Type of organization*.
14. *Part 61 program* referred to FAA's flight training program as described in Title 14 Code of Federal Regulations (CFR) Part 61, et seq. (FAR, 2021a). Organizations that offer Part 61 programs are commonly called Part 61 schools and are not required to follow an FAA-approved curriculum.

15. *Part 141 program* referred to FAA's flight training program described in Title 14 Code of Federal Regulations (CFR) Part 141 (FAR, 2021b). Organizations that offer Part 141 programs are commonly called Part 141 schools and are required to follow an FAA-approved curriculum.
16. *Pilot* was defined as an individual with an FAA airline transport pilot (ATP) rating who worked for a U.S. airline as either captain or co-captain. Although the current study's sample was partially comprised of pilots in addition to flight students, pilots provided the same data as flight students except the pilot data were a recollection of when pilots were enrolled as flight students in a flight training program leading to a PPL. Thus, in this context, pilots were considered to be part of the general term "flight students."
17. *Practical test* was defined as the areas of operation listed in Title 14, Code of Federal Regulations (CFR) Part 61.107(b) (FAR, 2021e).
18. *Private pilot license (PPL)* referred to an FAA certificate that permits an individual to fly different types of aircraft. This is the most common type of pilot certificate, and it is not to be confused with a "student pilot certificate," which is for pilots in training.
19. *Proficiency*. See *Check-ride proficient.*
20. *Student-driven schedule* was defined as a flight training schedule where the flight training organization permits students to schedule their flight training at their convenience.

21. *Student pilots* were defined as individuals who had a student pilot license (SPL), which enabled them to fly a registered aircraft under the supervision or authority of a CFI. This term is used interchangeably with “flight student.”

22. *Training time to proficiency* was defined as the number of days it took students to become check-ride proficient and included both classroom instruction and flight time, although the current study focused on the latter. This time interval began the day students took their first flight to the day their CFI determined they were proficient to take the check ride, where 1 day was defined as 24 hours, from 12:00 a.m. to 11:59 p.m.

23. *Type of organization* referred to whether the organization that provided flight training was a public or private 2- or 4-year colleges/university or a fixed-base operator (FBO) that provided flight training at airports.

Research Questions and Hypotheses

Research Questions

The primary research questions that guided the current study were as follows:

RQ 1. What is the relationship between training time to proficiency and flight time to proficiency?

RQ 2. What is the relationship between flight school characteristics and flight time to proficiency?

RQ 3. What is the relationship between flight students’ demographics and flight time to proficiency?

Research Hypotheses

The corresponding research hypotheses were as follows:

Hyp 1. There will be a curvilinear relationship (logarithmic in form) between training time to proficiency and flight time to proficiency.

Hyp 2a. Flight students who receive flight training via a Part 141 program will require fewer hours of flight time to become check-ride proficient than flight students who receive flight training via a Part 61 program.

Hyp 2b. Flight students who receive flight training at a 2- or 4-year college/university will require fewer hours of flight time to become check-ride proficient than flight students who receive flight training at an FBO.

Hyp 2c. Flight students who receive flight training under a mandated training schedule will require fewer hours of flight time to become check-ride proficient than flight students who receive flight training under a student-driven schedule.

Hyp 3. Flight students' demographics will have a relationship with flight time to proficiency (although this is not expected).

Study Design

The research methodology/design that best fit the current study relative to answering its research questions was a combination of predictive correlational (RQ

1 and parts of RQ 2) and ex post facto (parts of RQ 2 and RQ 3). A correlational method with a predictive design was appropriate because it is used to assess relationships and patterns of relationships among variables in a single group of subjects (Ary et al., 2010). If the two variables are correlated, then one variable could be used to predict the other. In the current study, RQ 1 and parts of RQ 2 involve a single group (flight students) with multiple measures.

An ex post facto effects-type design was appropriate for studies involving pre-existing groups, in the absence of any intervention, and where the grouping is on the independent variable, which is the case for parts of RQ 2 and RQ 3. In RQ 2, group membership involved type of flight school, type of organization, and scheduling policy, and in RQ 3, group membership involved gender, race/ethnicity, and marital status. These groups were formed from participants' self-reported responses to the respective items on the questionnaire.

Significance of the Study

The study's primary significance was demonstrating the application of Ebbinghaus's (1885/1913/2013) forgetting curve theory, Ebbinghaus's spacing effect theory, and Ausubel's (1963) theory of meaningful learning to the context of flight instruction. As reported in the literature, these theories have been applied to many different contexts and were shown to be applicable and appropriate. However, they never have been applied to an aviation context, specifically for developing a prediction model to help flight students and CFIs assess the number of hours students need to achieve proficiency.

An additional significance is that if the study data provide evidence in support of the targeted theories, then flight students can use the prediction model as a guide to estimate the costs associated with obtaining a PPL if they know the hourly rate for flight instruction. The prediction model also can be examined via subsequent studies for other pilot training applications such as obtaining an instrument rating or commercial pilot's license, and the theories could be used by airlines to assess their training costs for newly hired line pilots. Lastly, the study added to the corresponding literature by demonstrating the simultaneous application of the targeted theories to an activity that involved both cognitive and motor skill learning.

Study Limitations and Delimitations

Limitations

The limitations of a study include circumstances, conditions, and events the researcher could not control but could limit the generalizability of the study's findings. Following is a brief discussion of the current study's limitations.

1. Sample Demographics. I had no control over the personological characteristics of the current study's participants, including their biological sex assigned at birth, age, race, marital status, and ethnicity. As a result, similar studies that involve samples with personological characteristics different from those of the current study might get different results.

2. Source of Study. The current study was a non-funded Ph.D. dissertation research project from a student in the College of Aeronautics of an independent Ph.D. granting university in the southeastern United States. Therefore, if a similar study were to be conducted by a federal agency, such as the FAA or the National Transportation Safety Board, or via funded research, the results might be different.

3. CFI Training. I had no control over the instructional experiences or teaching effectiveness of CFIs. Thus, similar studies that include specific CFI experience factors such as years held a CFI certificate, number of dual hours, and number of dual hours in the past 90 days might get different results.

4. Tower vs. Non-Tower Airports. I had no control over whether students received flight instruction at an airport with or without a control tower. As a result, similar studies that include this information as a study factor might get different results.

5. CFI Oversight. I had not control over what criteria CFIs used to personally judge or decide when they believed a student was ready to take the private pilot check ride. As a result, similar studies that include this information as a study factor might get different results.

Delimitations

A study's delimitations include circumstances, conditions, and events that the researcher imposes to make the implementation of the research more feasible but can further limit the generalizability of the study's findings. Following is a brief discussion of the current study's delimitations.

1. **Civilian Flight Students.** The participants of the current study were limited to those who received civilian flight training. Therefore, similar studies might get different results if their samples consist of participants who received military flight training or a mix of both civilian- and military-trained participants.
2. **Flight Training Programs.** The current study targeted participants who received their flight training from either a Part 61 or a Part 141 flight training program. Therefore, similar studies that restrict study implementation exclusively to a Part 61 or a Part 141 program (but not both) might get different results.
3. **Flight Training Organization.** The current study asked participants to specify the type of organization from which they received their flight training (2- or 4-year college/university or an FBO). Because the presence or absence of this factor can have an impact on the results of data analysis, similar studies that either do not include this factor or

provide more detailed information (e.g., the geographic location of the organization) might get different results.

4. **Category and Class of Airplane.** The current study's participants obtained their PPL in the airplane category and single engine land class as defined by the Federal Aviation Regulations (FAR, 2021h). Therefore, similar studies that do not include this restriction might get different results.
5. **Sampling Strategy.** The current study acquired data using convenience and snowball sampling strategies. Therefore, similar studies that use a different sampling strategy such as simple random or cluster random sampling might get different results.
6. **Targeted Participants.** Participants of the current study included flight students who were currently pursuing their PPL as well as pilots who previously earned their PPL earlier in their life and self-reported what they recalled when they were flight students. Therefore, similar studies that target different participants such as those who are currently pursuing their PPL, recently minted pilots with their PPL, or more generally, different age cohorts, might get different results.
7. **Measurement of the Dependent Variable.** The dependent variable of the current study was the number of flight hours participants accrued when their CFIs declared them to be check-ride proficient. Thus, similar

studies that use a different metric such as the flight hours students accrued at the time they received their PPL might get different results.

8. **Theoretical Grounding.** The current study was grounded in Ebbinghaus's (1885/1913/2013) forgetting curve theory and spacing effect, and Ausubel's (1963) theory of meaningful learning. Therefore, similar studies with a different theoretical grounding might get different results.

9. **Transformation of Training Time to Proficiency.** In the current study X_1 = Training time to check-ride proficiency (in days) was transformed using the natural logarithmic function (ln). Therefore, similar studies that either do not transform this variable or use a different transformation (e.g., log base 10) might get different results.

10. **Outliers.** In the current study, an outlier analysis was conducted using Jackknife distances, which flagged 15 rare cases and all 15 cases were included in the final data set. Therefore, similar studies that either use a different outlier analysis approach or do not include rare case outliers might get different results

Chapter 2

Review of Related Literature

Introduction

This chapter is organized into three sections: The first section summarizes and describes the theoretical foundation on which the current study was grounded. These theories, which were the basis for the predictive nature of the study, included Ebbinghaus's (1885/1913/2013) forgetting curve theory and spacing effect, and Ausubel's (1963) theory of meaningful learning. The second section reviews the past literature that applied Ebbinghaus's and Ausubel's respective theories to academic and motor skills learning relative to aviation and flight training. The last section provides a summary of the related literature and a discussion of its implications to the current study.

Overview of Underlying Theories

Ebbinghaus (1885/1913/2013) observed that one could not expect to recall vocabularies, discourses, and poems by learning them via a single repetition, and that the majority of information was soon forgotten after a single repetition. These observations are often referred to in the literature as the forgetting curve or retention curve theory (Murre & Dros, 2015). As depicted in Chapter 1, Figure 1.1, a plot of Ebbinghuas's experimental data resulted in a relationship that was logarithmic in form between the percentage of information lost and time. For example, within 1 day, approximately 70% of the information a person initially learned cannot be recalled if no attempt is made to remember it.

Ebbinghaus (1885/1913/2013) found that frequent repetitions facilitated recalling information, and through additional practice over time, the information readily could be recalled. In other words, continued practice over time allowed information to be recalled easier and effectively mitigated the effects of the forgetting curve. This was illustrated in Chapter 1, Figure 1.2, and the literature has dubbed this observation as the "spacing effect," which posits that distributed practice is more effective than massed practice (Ormrod, 2012, p. 208). This implies that new information received in a distributed format will slide the forgetting curve to the right, thereby delaying the effects of not being able to recall the information.

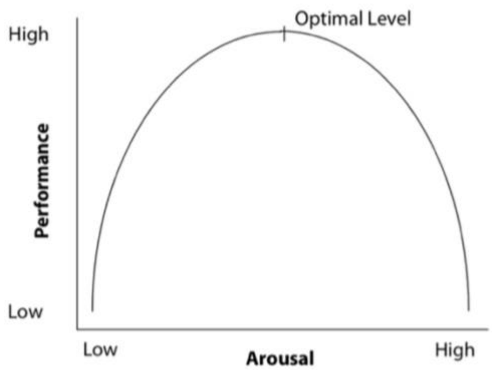
The concepts of the forgetting curve and spacing effect are inherent in Ausubel's (1963) theory of meaningful learning, which involves the acquisition of new knowledge. According to Ausubel (1968, p. 37), the single most important attribute learners can bring to the learning table is their prior knowledge. This is because any learning activity requires students to first link new knowledge to their prior knowledge and then to rehearse this link to facilitate recall. Key to this process, though, is that (a) students' prior knowledge must be correct and void of any misconceptions or misunderstanding, and (b) the links they establish between prior and new knowledge must be done so in a deliberate and non-arbitrary manner.

This is the heart of what Ausubel (1963) refers to as meaningful learning because the associations students make between their prior and new knowledge are meaningful to them. Ausubel focused his theory on expository teaching settings, which involve teachers presenting new information to students and students receiving this new information and incorporating (i.e., subsuming) it into their existing knowledge base. As a result, Ausubel's theory of meaningful learning is often referred to as reception learning. When considered collectively, Ausubel's (2008) and Ebbinghaus's (1885/1913/2013) respective theories can facilitate meaningful learning because they require learners to link new knowledge to prior knowledge and then space out the learning of this new knowledge to help mitigate the forgetting curve.

When applied to the context of the current study, Ebbinghaus's (1885/1913/2013) forgetting curve suggests that the number of flight hours students need to become proficient to take the private pilot check ride will vary based on the time between their first and last flights prior to the check ride. For example, flight lessons scheduled too close together would run afoul of Ebbinghaus's warning regarding cramming (i.e., massed practice), and result in students accumulating additional flight hours before becoming proficient for the private pilot check ride. By scheduling flight lessons too close together students are not able to process new knowledge adequately, which then prevents them from storing what they learned in long-term memory to be recalled during subsequent flight lessons. Because flight lessons are not taught in isolation, the failure to recall prior knowledge makes it more difficult for students to perform effectively on each successive flight.

On the other hand, students who spaced out their flight lessons (i.e., distributed practice) would benefit from Ebbinghaus's (1885/1913/2013) observations that repetition could mitigate the forgetting curve, and they (in theory) would require fewer hours to become proficient to take the private pilot check ride. The overriding question, then, is what is the optimal spacing? In other words, how many days apart should students schedule their flight lessons that would result in an optimal number of lessons needed for proficiency? In practice, the answer to this question is unknown. As noted above, spacing lessons too close together would be analogous to cramming, but it also could be equally detrimental if lessons were spaced too far apart. For example, applying the theories of Ebbinghaus (1885/1913/2013) and Ausubel (1963), if flight lessons were spaced too far apart, then students would forget what they learned previously because the lack of repetition would prevent them from linking new knowledge with prior knowledge and adequately rehearsing such links. This lack of sufficient time to process new knowledge would have the same effect as spacing lessons too close together, namely, there would be an increase in the number of flight hours needed for check ride proficiency. This was illustrated in Chapter 1, Figure 1.3 as a U-shaped curve to highlight the effect of spacing flight lessons too close together or too far apart. This relationship between the number of flight hours needed for proficiency and the

Figure 2.1
Yerkes-Dodson Curve



Note. Source: Yerkes and Dodson (1908).

number of days students should schedule their flight lessons is similar to the Yerkes–Dodson curve (Yerkes & Dodson, 1908), which posits a relationship between performance and arousal. Although the Yerkes–Dodson curve is depicted as an inverted U (see Figure 2.1), it demonstrates that in theory a certain amount of arousal is necessary for peak performance. Too little or too much arousal, though, will have a negative impact on performance. In practice, though, what would be considered optimal arousal is unknown.

With respect to Ebbinghaus (1885/1913/2013) and Ausubel (1963), one would expect the same curvilinear relationship that applied to Ebbinghaus's forgetting curve and spacing effect theory to apply to Ausubel's theory of meaningful learning: The number of hours students need to become proficient to take a private pilot check ride should vary based on the time between their first flight and their last flight before the check ride. The reader will note that the current study did not endeavor to determine the optimal number of flight lessons needed across a particular time span. Instead, the current study endeavored to apply Ebbinghaus's and Ausubel's theories to help develop a prediction model that flight students and CFIs could use to determine this number.

Review of Past Research Studies

This section contains a review of selected prior studies that helped inform the current study and is partitioned into four subsections: The first contains a review of studies that involved the spacing effect with respect to meaningful learning, the second focuses on motor skills learning, the third examines the spacing effect relative to meaningful learning combined with motor skills acquisition, and the last subsection highlights key demographic factors that have been shown to impact the spacing effect. The reader is advised that throughout this presentation the term *spacing effect* is used generically and refers to the respective theories of Ebbinghaus (1885/1913/2013) and Ausubel (1963).

The Spacing Effect and Academic Achievement

Gallo and Odu (2009). The robustness of the spacing effect with respect to academic achievement has been demonstrated to hold in many different disciplines and across various school levels (e.g., Dempster, 1988; Grote, 1995; Krug et al., 1990; Lu, 1978; Ruch, 1928; and Smith & Rothkopf, 1984). One related application relevant to the current study was class scheduling, and a representative study of this application was that of Gallo and Odu who investigated the effect of class scheduling on student achievement in college algebra at a 2-year college.

Gallo and Odu (2009) focused on three different class schedules: 3 days per week where classes met on Monday, Wednesday, and Friday (MWF) for 50 minutes per session; 2 days per week where classes met on Monday and Wednesday (MW) or Tuesday and Thursday (T-TH) for 75 minutes per session; and 1 day per week where classes met on Saturday for 165 minutes with a 15-minute break. Thus, regardless of schedule all students received exactly 150 minutes of instruction per week. Gallo and Odu also targeted other factors, including student attributes (gender, grade level, attitudes toward mathematics, prerequisite knowledge, learning styles, hours worked per week) and teacher attributes (gender and years teaching at the college level).

The targeted population was all community college students taking college algebra in Florida, the accessible population was all such students at the targeted 2-year college, and the final sample was comprised of $N = 166$ students. The corresponding schedules included two 3-day per week classes ($n = 20$), four 2-day per week classes ($n = 79$), one Saturday-only class ($n = 17$). All classes followed the same curriculum as defined by the State of Florida, used the same textbook, completed the same assessment protocols, and met for the same number of hours before each unit exam and final exam. Student assessments included: (a) an intermediate algebra final examination administered during the 2nd week of classes to confirm group equivalency with respect to prerequisite knowledge; (b) a researcher-constructed multicomponent questionnaire that consisted of a set of demographic questions, Tapia and Marsh's (2004) 40-item Attitudes Toward Mathematics Inventory (ATMI), and Kolb's (2005) Learning Style Inventory–Version 3.1 (KLSI–3.1) administered during the 4th week of classes; (c) four-unit exams consisting of 20 open-ended questions selected from homework items and administered at the end of a specific set of course work and instructional time throughout the semester; and (d) a comprehensive final exam that consisted of 25 selective-response items administered during final exam week. Gallo and Odu (2009) implemented their study as a modified quasi-experimental design involving intact classes of students and used a multiple regression strategy by hierarchically regressing college algebra final exam scores on Set A = Class Schedules, Set B = Student Attributes, and Set C = Teacher Attributes.

Gallo and Odu (2009) reported that students who took college algebra 1-day per week had significantly lower final examination scores than students who took college algebra 2-days per week or 3-days per week. More specifically, the 1-day per week group averaged 17 points lower on the final examination than the 3-day per week group, and approximately 13 points lower than the 2-day per week group. However, although the 2-day per week group averaged 4.4 points lower than the 3-day per week group, this difference was not statistically significant. Gallo and Odu also reported that the 3-day per week group had the highest overall mean final exam score (M = 72.5) with the lowest standard deviation (SD = 8.0), and the 1-day per week group had the lowest overall mean score (*M* = 55.5) with the highest standard deviation (*SD* = 19.8).

Gallo and Odu (2009) opined that a 1-day-a-week schedule ran afoul of Ausubel's theory of meaningful learning because students could not make meaningful connections with just 1 day of instruction per week. Because college algebra is not taught in isolation, meaningful connections are necessary to link new knowledge with prior knowledge. Thus, new knowledge was not retained because students could not transfer what they learned from short-term memory to long-term memory in a weekly class. This rationale is consistent with what Smith and Rothkopf (1984) reported in a study involving teaching statistics. They indicated that distributing lessons over 4 days was more effective than a massed lesson over 1 day. Similar to Gallo and Odu, Smith and Rothkopf speculated that the effect of a massed 1-day per week lesson prevented students from moving new knowledge from short-term memory to long-term memory. Gallo and Odu also surmised that students in the 1-day per week group might have been overwhelmed with too much material at one time, which would be tantamount to cramming and consistent with Ebbinghaus's (1885/1913/2013) position that cramming escalates information loss and prevents new knowledge from being retained.

When applied to the context of the current study, learning to fly an airplane is not taught in isolation. The lessons and their corresponding concepts are additive. Students need to apply the knowledge they acquired from previous flight lessons to their current lessons, and they must be able to move what they are currently learning from short-term memory to long-term memory so it can be recalled on subsequent flights. By equating a 1-day per week schedule with flying 1-day per week and equating a 2- and 3-day per week schedule with flying 2 or 3 days per week, Gallo and Odu's (2009) results support the proposition that students should be able to complete their private pilot license with less flight time by flying 2 or 3 days a week than by flying 1 day per week.

Although Gallo and Odu (2009) helped inform the proposed study, their findings only provided partial support of Ebbinghaus's (1885/1913/2013) and Ausubel's (1963) respective theories. For example, Gallo and Odu provided distributed *instruction*, not distributed *practice*. College classes usually do not provide practice for previously learned material, but each instructional lecture contains exclusively new information that builds on previous concepts. Students are left on their own to practice and study the information. In the context of learning to fly an airplane, though, pilots spend time practicing what they have learned in previous lessons until they attain proficiency. Thus, the instruction provided by a CFI can be described as distributed practice as opposed to distributed instruction. It also should be noted that Gallo and Odu provided *fixed* distributed instruction, which implies that instruction for all groups was identical or was for the same period, namely, a single semester. Thus, Gallo and Odu varied the instructional period to apply the spacing effect. In contrast, because flight lessons are usually 1.5

hours long (sans cross-country flights), to observe the spacing effect lessons must be based on a compressed schedule, which allows distributed practice to reduce the number of days to become proficient for the private pilot check ride. Finally, Gallo and Odu's study dealt exclusively with meaningful learning (i.e., academic achievement) whereas learning to fly an airplane involves the simultaneous application of reception and motor skill learning.

Sheldon and Durdella (2010). A study by Sheldon and Durdella provides additional support for Ebbinghaus's (1885/1913/2013) and Ausubel's (1963) respective theories with respect to the proposed study. Unlike Gallo and Odu (2009) who examined the effect of a fixed distributed instruction schedule, Sheldon and Durdella examined the effect of a compressed distributed instruction schedule. Data were collected from enrollment records of a large community college from spring 1998 through fall 2001. The targeted population was native or continuing community college students who enrolled in at least one developmental English, reading, or mathematics course offered in either a compressed or regular-length format. The accessible population was all students at the targeted 2-year college, and the final sample comprised $N = 21{,}165$ students. The corresponding schedules included a compressed schedule with a 5- to 6-week course and an 8- to 9-week course ($n = 3{,}360$), and a regular length course offered in the standard 15- to 18-week format ($n = 17{,}805$). Although Sheldon and Durdella did not document how many days a week students attended class, lecture time for a 15- to 18-week-semester is usually 3 hours per week; for an 8- to 9-week semester, 6 hours per week; and for a 5- to 6-week semester, 9 hours per week.

Because of the ex post facto nature of their study, it was not possible to determine whether each class contained the same instructional content. However, Sheldon and Durdella (2010) focused on developmental education courses as opposed to vocational or transfer courses. New, transfer, and dual-enrolled students were excluded from the study to further control for variation. Only those records that identified students as native or continuing students enrolled in courses designated as developmental were included in the study. Due to the study's exploratory ex post facto nature and the available data, Sheldon and Durdella used a descriptive and chi-square analysis. They also used a percent difference to assess the potential strength of the association between the variables because of the limitations of Cramer's *V* to evaluate large sample sizes.

Sheldon and Durdella (2010) reported that: (a) students in the 8- to 9-week English courses had a higher completion rate (86.90%) and were more likely to succeed than students in either the 5- to 6-week English courses (75.80%) or 15-18-week English courses (56.70%), $\chi^2 = 195.175$, $p = .000$; (b) students in the 8- to 9-week mathematics courses had a higher completion rate (65.35%) and were more likely to succeed than students in either the 5- to 6-week mathematics courses (57.91%) or 15- to 18-week mathematics courses (51.15%), $\chi^2 = 69.553$, $p < .001$. Sheldon and Durdella also reported that students in the 5- to 6-week courses could have had difficulty retaining the information because the schedule was too compressed, causing them to cram information and thereby violating Ebbinghaus's (1885/1913/2013) warning that cramming would prevent retention. Students might have been overwhelmed with too much material at one time, which is equivalent to cramming. Students in the 15- to 18-week courses might not have been as successful at making meaningful connections because the information was too attenuated, running a fowl of Ausubel's (1963) theory impeding students' ability to make meaningful connections that are necessary to link new knowledge with prior knowledge.

Applying these results to the context of the current study, a 5- to 6-week semester (9 hours per week) would be akin to taking six 1.5-hour flight lessons per week, which would be considered cramming, and students would become overwhelmed with too much material at one time. Similarly, a 15- to 16-week semester (3 hours per week) would be the equivalent of two 1.5-hour flight lessons per week and would hinder a student from making meaningful connections under Ausubel's theory. An 8- to 9-week semester (9 hours per week) would be comparable to flying 3 to 4 times a week and most likely facilitate obtaining a private pilot license in the least amount of time. The results of Sheldon and Durdella provide support to the current study's conjecture that the number of flight hours students need to become proficient to take the private pilot check ride will vary based on the time between their first and last flights before the check ride. As a result, time to proficiency, which represents the time in days from a flight student's first flight to when a CFI determines the student is proficient to take the private pilot check ride, was targeted as an independent variable.

The reader is cautioned, though, that applying the results of Sheldon and Durdella (2010) to the current study is problematic for the same reasons Gallo and Odu's (2009) study was problematic. First, Sheldon and Durdella focused on distributed instruction, not distributed practice. As described previously, taking flight lessons involves distributed practice, a combination of instruction and practice during each flight. Second, although Sheldon and Durdella's study involved compressed distributed instruction, the total instructional period was probably the same for each group. Thus, Sheldon and Durdella's classes had to vary the instructional period to apply the spacing effect. Total semester hours cannot be equated with total flight time because the current study seeks to determine if distributed practice will reduce flight time, whereas semester hours in the above studies were constant. Finally, Sheldon and Durdella exclusively dealt with meaningful learning, whereas the current study involved the simultaneous application of reception and motor skill learning.

Flexman et al. (1972). Flexman et al. studied whether practice in a flight simulator could reduce the flight time it takes to become proficient in an airplane. Although they did not refer to Ebbinghaus's (1885/1913/2013) or Ausubel's (1963) respective theories, Flexman et al. were cognizant of the basic principles of these theories. Consistent with Ausubel, Flexman et al. noted, “All learning is based upon a foundation of prior learning” (p. 1). To determine whether practice in a flight simulator could reduce the flight time it takes to become proficient in an airplane, Flexman et al. measured transfer of learning*,* defined as “terms that refer to the dependence of new learning on old” (p. 1), which is consistent with Ausubel.

Flexman et al.’s (1972) study consisted of a transfer group and a control group. The transfer group was provided with flight instruction in a simulator and the airplane, and the control group was only provided instruction in the airplane. The accessible population comprised $N = 12$ male students at the University of Illinois at Urbana-Champaign, and ages ranged from 21 to 36 years old. No students had any previous flight instruction. Although Flexman et al. did not specifically identify the corresponding target population, based on the context of the study it appears that it was all flight students in the United States. The 12 participants were assigned to the control and transfer groups by matching based on the Bennett-Fry Mechanical Comprehension Test results. The U.S. Navy used this test to select Naval Aviation Cadets during World War II, and the test was known to correlate with the successful completion of flight training in the United States Navy. Although matching was implemented, it was not considered in the statistical analysis because it was unknown whether the result of the Bennett-Fry Mechanical Comprehension Test would correlate with the ability to perform the flight maneuvers and procedural maneuvers participants performed in the study.

The simulator was a replica of a North American T-6 trainer used by the U.S. Air Force and U.S. Navy. The simulator's cockpit devices were salvaged from a destroyed T-6 aircraft, so the simulator's cockpit layout and switches were identical to the T-6. Participants were trained in four procedural tasks and nine flight maneuvers and were considered proficient when they performed the procedures and flight maneuvers three times without error, as determined by their instructor. Flexman et al. evaluated the number of trials, errors, and time it took to complete the procedures and maneuvers. To evaluate the metrics (trials, errors, and time), Flexman et al. measured transfer of training by percentage of transfer (PT) and transfer effectiveness ratio (TER). Independent-samples t tests were used to determine differences between the groups for both PT and TER metrics.

Procedural Task 3 involved the aircraft engine starting procedure. Because of the battery drain and the possibility of overheating the starter motor during the engine starting sequence, the control group could only practice the starting procedure once for every flight lesson. However, the transfer group was able to practice the procedure numerous times in the flight simulator in addition to once in the airplane for every flight lesson. Flexman et al. (1972) found a significant difference in the amount of time, the number of trials, and the number of errors it took for the groups to become proficient at performing the engine starting procedure. The transfer group performed significantly better ($p < .001$) than the control group in both PT and TER.

Procedural Task 4 involved performing an aircraft run-up prior to takeoff to verify that the engine and associated systems were functioning within normal limits. The transfer group repeatedly practiced the run-up procedure in the simulator and once every flight, while the control group was only able to practice the procedure in the aircraft once during every flight. Flexman et al. (1972) found a significant difference in the amount of time, the number of trials, and the number of errors it took the groups to become proficient to perform the task. The transfer group performed significantly better ($p < .001$) than the control group in both PT and TER.

Both procedural task results are consistent with Ebbinghaus's (1885/1913/2013) and Ausubel's (1963) respective theories because the additional practice provided to the transfer group allowed the group to retain the information by repetition and reduced the time it took to attain proficiency in the airplane. The additional practice was tantamount to allowing the transfer group to experience distributed practice while the control group ran afoul of Ebbinghaus's warning that repetition is required to obtain mastery. Applying the procedural results to the context of the current study, students who fly more often are expected to retain procedural skills more readily than those whose flights are spread too far apart. By retaining the procedural skills, students presumably will be able to reduce the flight time needed to become proficient for the private pilot check ride. Flexman et al.'s (1972) study provided support to include time to proficiency, representing the time in days from a flight student's first flight to when a CFI determines the student is proficient to take the private pilot check ride, as an independent variable in the proposed study.

It should be noted that Flexman et al.'s (1972) results do not necessarily support Ebbinghaus's (1885/1913/2013) and Ausubel's (1963) respective theories relative to motor skills acquisition, which also is needed when learning to fly an airplane. Although Flexman et al. provided evidence that practice in a flight simulator can substantially reduce the number of flight hours it takes to attain proficiency in an airplane, the transfer group required nearly a full hour more to attain proficiency than the control group—12.8 hours vs. 11.9 hours. This finding is opposite to what Ebbinghaus and Ausubel would posit. A possible explanation for this anomaly is that the most challenging flight maneuvers were not part of Flexman et al. For example, it takes a substantial amount of time to learn to land an aircraft and even more time to learn to land an aircraft in a crosswind condition. Additional maneuvers that require substantial practice, such as short and soft field landings, were not included in Flexman et al. These maneuvers are the types of maneuvers that repetitive practice is required to attain proficiency, and repetition reduces the time to proficiency. Thus, Flexman et al.'s results neither supported nor discounted Ebbinghaus's or Ausubel's respective theories relative to motor skill learning.

The Spacing Effect and Motor Skills Learning

Baddeley and Longman (1978). With the advent of the mechanical letter-sorting machine, it became necessary for the British Post Office to train up to 10,000 postal workers (postman and postman higher grade) to use a standard typewriter keyboard. To train its postal workers, the British Post Office needed to know the optimum training schedule. To determine such a schedule, Baddeley and Longman used a 60-hour course to train 72 postal workers who were partitioned into four separate groups: Group 1 trained for 1 hour each day for 12 weeks, Group 2 trained for 2 hours each day for 6 weeks, Group 3 trained for 1 hour twice a day for 6 weeks, and Group 4 trained for 2 hours twice a day for 3 weeks. The reader will note that this design was similar to Sheldon and Durdella (2010) in that it involved a compressed rather than fixed schedule.

The target population was all operators (postmen and postman higher grade) of the British Post Office, and the accessible population was postal volunteers from the Croydon sorting office ($N = 72$). Each of the four groups had an equal number of postal workers ($n = 18$). To mitigate variation in the random assignment, matching was employed based on age, which ranged from 19–46 years. Metrics involved the average number of hours of training required to learn the whole keyboard for each group (time), the average speed of performance by employing test runs for the four groups (speed), and the error rate for each group (error rate).

An analysis of variance (ANOVA) was performed to determine any significant differences between groups.

With respect to time, Baddeley and Longman (1978) reported Group 1 (1 session per day for 1 hour) took the least amount of time to learn the keyboard as a function of the training schedule ($M = 34.9$) compared to Group 2 (2 sessions a day for 1 hour each session, $M = 42.6$), Group 3 (1 session a day for 2 hours, $M = 43.2$), and Group 4 (2 sessions a day for 2 hours each session, $M = 49.7$). The fastest participant in Group 4 was slower than the slowest participant in Groups 1 and 2. Baddeley and Longman did not determine if differences in time were significant.

With respect to speed, Baddeley and Longman (1978) reported Group 1 consistently had the greatest keystroke speed as a function of time ($M = 79.31$) compared to the other three groups: $M_{G2} = 73.43$, $M_{G3} = 71.12$, and $M_{G4} = 64.78$. ANOVA results ($p < .01$) also showed that (a) 1-hour sessions (Groups 1 and 3) were significantly better than 2-hour sessions (Groups 2 and 4), and (b) one session per day (Groups 1 and 2) was significantly better than two sessions per day (Groups 3 and 4). No interaction was present between the length of session and the frequency of session. Group comparisons via an independent samples t test showed that Group 1 was faster than Group 4 ($p < .01$) and Group 2 ($p < .05$, 1 tail), and Group 3 was faster than Group 4 ($p < .05$).

For uncorrected errors, there was a significant difference between Group 4 ($M = 2.06$) and the three other groups: $M_{G1} = 1.09$, $M_{G2} = 1.14$, and $M_3 = 1.41$.

ANOVA results showed that the length of session frequency of session and interaction between length and frequency were significant ($p < .001$). Baddeley and Longman opined that the interaction was a product of the high error rate found in Group 4, which produced significantly more uncorrected errors than any of the other groups ($p < .01$ in each case).

Baddeley and Longman (1978) concluded that the 1-hour sessions groups (Groups 1 and 3) produced better results than the 2-hour sessions groups (Groups 2 and 4), which they surmised were too long. They also believed that one session per day rather than two sessions per day was slightly more effective, though not significantly. Baddeley and Longman's results support Ebbinghaus's (1885/1913/2013) and Ausubel's (1963) respective theories that cramming too much training into a single session (massed practice) impedes the performance and reception of motor skill development. Baddeley and Longman helped inform the current study because the results provided evidence that the spacing effect theory can be applied to motor skill development, which is required for learning to fly an airplane. Because Baddeley and Longman involved a discrete motor skill, a positive finding in the current study could provide evidence that the spacing effect theory is beneficial to learning a complex motor skill such as learning to fly an airplane. Their study also provided support to the current study because they evaluated distributed practice as opposed to distributed instruction and compared a compressed schedule as opposed to a regular schedule. The application of this variant of the spacing effect could be described as compressed distributed practice, which was the focus of the current study.

Donovan and Radosevich (1999). In a meta-analysis, Donovan and Radosevich examined the extent to which the spacing effect could be applied to simple and complex tasks. One designated complex task was airplane control simulation, which they classified as having a high overall complexity, requiring high mental and physical requirements. Also included within this category were an air traffic control simulation, milk pasteurization simulation, hand movement memorization, puzzle box task, music memorization, and performance. Donovan and Radosevich reported that this complex category had an effect size of $d = 0.07$. To determine which of the three task dimensions (mental requirements, physical requirements, task complexity) influenced the effectiveness of distributed practice, they calculated semipartial correlations among the three task dimensions and reported a significant negative correlation between overall complexity and effect sizes ($r = -.25$, $p < .05$). They also reported that the mental and physical requirements of the task were not significantly correlated with the effect sizes ($p = .35$). Thus, the overall complexity of the task appeared to be a key factor in determining whether spaced (distributed) practice is superior to massed practice.

Donovan and Radosevich's (1999) findings supported the position in the current study that distributed practice is superior to massed practice for complex motor skills, which are the type of skills pilots must utilize when learning to fly an airplane. Although Donovan and Radosevich's findings are consistent with Ebbinghaus's (1885/1913/2013) and Ausubel's (1963) respective theories relative to complex motor skill acquisition, the tasks they categorized as highly complex involved studies with spacing intervals ranging from a few seconds to 24 hours. Complementing Donovan and Radosevich, though, Smith and Scarf (2017) identified cases where the spacing effect could benefit complex tasks when spacing intervals are much greater than 24 hours and cited several studies reporting the benefits of the spacing effect while learning surgical skills (De Win et al., 2013; Gallagher et al., 2012; Kang et al., 2015; Moulton et al., 2006; Spruit et al., 2014; and Verdaasdonk et al., 2007). Although flight instruction is not equivalent to surgical training, both are indeed complex tasks that require more than 24 hours of training. As a result, Donovan and Radosevich's (1999) findings provided support to the current study, but it is noteworthy to point out that any observed reduction in flight hours might not be practically significant.

Moulton et al. (2006). Moulton et al. examined the spacing effect from the perspective of teaching microvascular anastomosis to junior surgical residents. Although Moulton et al. did not identify the target population, it was presumed to be surgical residents attending medical school in Ontario, Canada, and the accessible population was 1st-, 2nd-, and 3rd-year surgical residents at the University of Toronto. The sample consisted of n = 38 junior surgical residents (5 females) who were randomly assigned equally to either a massed practice group, which participated in four training sessions in 1 day, or a spaced (distributed) practice group, which participated in one training session each week for 4 weeks. Because the surgical residents were in different years of their residency (Year 1, $n = 16$; Year 2, $n = 15$; Year 3, $n = 7$), they were stratified according to their postgraduate year of training prior to random assignment.

Both groups were administered four assessments: (a) a pretest to assess residents' basic level of skill given during the first training session, (b) a posttest that served as the final assessment given on the fourth training session, (c) a retention test using the microsurgical drill for three tasks (time, number of hand movements, and expert global ratings), and (d) a live rat anastomosis (transfer test) performed 1 month following the last residents' training session. Results were measured by expert and computer-based methods. Interrater reliability was measured using Cronbach's alpha by the two expert examiners and varied between $\alpha = .67$ and $\alpha = .89$. Statistical analysis involved the Kolmogorov-Smirnov Goodness of Fit test. Moulton et al. reported no significant differences between the massed and spaced groups on the posttest results for the expert and computer-based measures. However, residents in the distributed practice (spaced) group performed significantly better ($p < .05$) on the retention test in most outcome measures and in performing an anastomosis on anesthetized rats in all expert-based measures (global ratings, checklist score, final product analysis, competency for operating room).

Moulton et al. (2006) helped inform the current study from the perspective of flight students who take a break in their training. Generalizing Moulton et al.'s findings to the current study, one would expect that students who receive distributed flight training practice and resume their training after a retention period will have higher retention of previously learned motor skills than those who did not receive distributed practice. This in turn, theoretically, could reduce the number of hours needed to become check-ride proficient. Moulton et al. also helped inform the current study because the *process* of learning to perform surgery is akin to learning to fly an airplane. Both skills require reliance on meaningful learning in the development of motor skills. A surgeon must apply previously learned information during the development of the motor skills necessary to perform surgeries. Similarly, a pilot also must apply previously learned information during the development of the motor skills necessary to land an airplane successfully, among many other flight maneuvers. Moulton et al. findings also supported Ausubel's (1962) theory because students must retrieve key pieces of meaningful learning from memory to apply the motor skill being learned: the more deeply the learned information is encoded into a student's memory, the more available it is for recall.

Spruit et al. (2015). Spruit et al. examined the spacing effect relative to the perspective of teaching laparoscopy surgery to medical students. Spruit et al. randomly assigned n = 38 (13 male) medical students without prior experience in laparoscopy surgery to either a massed (n =18) or spaced practice group (n =20).

No information about the target and accessible populations was disclosed. Both groups received three laparoscopy training sessions of 75 minutes per session. The massed practice group received all three sessions in 1 day, while the spaced group received one training session per week for 3 weeks. Training consisted of learning five laparoscopic tasks (rubber band, pipe cleaner, beads, circle cutting, and intra-corporeal suturing). The training was videotaped and accessed by Spruit et al. for completion times. Accuracy for each task was evaluated based on principles of metrics by Gallagher and O'Sullivan (2012) by creating an accuracy scoring tool. Accuracy and completion times were documented at four different stages: at the end of the first and third training sessions, at the end of a 2-week retention interval, and at the end of a 12–14-month retention interval. Differences between groups were statistically analyzed using Mann-Whitney at each stage.

At the end of first training session, the spaced group performed significantly better on four of the five training tasks: rubber band ($ES = 0.76$, $p < .001$), pipe cleaner ($ES = 0.73$, $p < .001$), beads ($ES = 0.65$, $p <.001$), and circle cutting ($ES = 0.36$, $p < .05$). The spaced group also had significantly higher accuracy rates for elastic band ($ES = 0.63$, $p < .001$), pipe cleaner ($ES = 0.57$, $p < .001$), and beads ($ES = 0.48$, $p <.05$). At the 2-week retention period, the spaced group performed significantly better in pipe cleaner ($ES = 0.42$, $p < .05$) and spaced beads ($ES = 0.45$, $p < .01$), and had significantly higher accuracy rates for pipe cleaner ($ES = 0.36$, $p < .01$) and elastic band ($ES = 0.16$, $p <. 05$).

For the fifth task, intra-corporeal suturing, Mann-Whitney tests showed substantial effects for completion times at the end of the third training session, and at the end of the short and long retention terms: $U = 71.5$, $ES = 0.58$, $p = .001$; $U = 85.4$, $ES = 0.53$, $p = .002$; and $U = 4$, $ES = 0.77$, $p = .014$, respectively. Mann-Whitney tests also showed substantial effects for accuracy scores at the end of training and at the end of the short and long retention terms: $U = 77$, $ES = 0.55$, $p = .002$); $U = 87.5$, $ES = 0.51$, $p = .003$; and $U = 0$, $ES = 1.0$, $p = .002$, respectively.

Spruit et al. (2015) helped inform the current study because it demonstrated that the spacing effect is applicable to students learning a complex skill who stop training for a substantial time, which is analogous to a retention period, or continue with their training in the absence of a retention period. Thus, Spruit et al. provided evidence that distributed practice can benefit motor skills development involving complex tasks. As a result, one of the independent factors I targeted was training time to proficiency, which represented the time in days from a flight student's first flight to when a CFI determines the student is check-ride proficient.

Farr et al. (1956). Farr et al. evaluated the ability of 27 male college students to maintain straight and level flight in a rudimentary flight simulator. Although no target or accessible population was identified, the study was performed for the Department of the Air Force and thus the target population was presumed to be all pilots considered for a United States Air Force commission. It also was presumed that the objective of the study was to determine the best method to train pilots to fly aircraft for the United States Air Force.

The simulator contained a control yoke and rudder pedals but had no visual screen. Instead, a model airplane was mounted on top of the simulator. Inputs from the control yoke and rudder moved the model airplane in three dimensions. Eighteen participants were given massed practice, and nine participants were given distributed practice. Farr et al. (1956) documented the amount of time participants were able to keep the model aircraft straight and level and reported that the distributed practice group performed significantly better ($p < .05$) at this task than the massed practice group.

Although Farr et al.'s (1956) study involved motor skill acquisition in a simulator whereas the current study involved the performance of complex flight maneuvers in an airplane, Farr et al. nevertheless helped inform the current study because of the progression that occurs during flight training. Students do not start with complex maneuvers but initially start learning maneuvers like those studied by Farr et al.—that is, learning to fly an airplane straight and level. Other simple maneuvers such as turning, climbing, and descending, although not considered complex maneuvers, are necessary to master so they can be built on to develop complex maneuvers. Because simple maneuvers are necessary to learn before attempting complex flight maneuvers, the spacing effect theory would apply to those simple maneuvers involved in the flight training process.

The Spacing Effect and the Simultaneous Application of Meaningful Learning and Motor Skills Learning

The studies presented thus far focused on the application of the spacing effect with respect to either meaningful learning or motor skills learning separately. Learning to fly an airplane, however, involves the simultaneous application of reception and motor skill learning. Students must apply what they learned in ground school to the development of motor skills while flying the airplane. Motor skill development cannot be taught in a vacuum. Although students could develop motor skills without applying information from meaningful learning, this would be a recipe for disaster. For example, trying to land an aircraft without having the cognitive ability to identify and procedurally react to wind shear during final approach could be catastrophic. Following is a summary of targeted studies that involved the application of spacing effect to simultaneously reception and motor skills learning that helped inform the current study.

Mengelkoch et al. (1971). Mengelkoch et al. studied meaningful learning and motor skill acquisition in the context of learning to fly an airplane. The study was carried out in a flight simulator with students learning to fly by reference to instruments, without any outside visual references. Thus, the study was similar to Farr et al. (1956) but included a meaningful learning component. Mengelkoch et al. did not identify a target or accessible population. However, the study's purpose was to ascertain the amount of forgetting that occurs in pilot proficiency skills. Hence, the target population could be assumed to be all pilots in the United States.

The study involved 33 participants who had no prior flight training or planned to take flight training lessons. The participants were divided into two groups: Group 5 and Group 10. Group 5 was given five trial simulator sessions, and Group 10 was given 10 trial simulator sessions. Both groups were given a 4-hour academic course on flight procedures. Each simulator session lasted for approximately 50 minutes and included climbs, descents, and turns to headings among other maneuvers. Participants' performance was evaluated once after their respective trials and a second time after a 4-month retention period.

Mengelkoch et al. (1971) reported that the final training trial before the retention period showed Group 10 had a very high flight and procedural performance while Group 5 attained only an intermediate level of procedural performance proficiency. Although no statistical analysis was performed on the initial evaluation, Mengelkoch et al. described the two groups' performance as follows:

> Observations of the subjects of Group 10 revealed that they had acquired smooth, coordinated control of the simulator and that procedural sequence were performed positively and almost error-free …. Group 5 had half the number of training trials given Group 10 and attained a satisfactory but intermediate level of proficiency. (Mengelkoch et al., 1971, p. 401)

Mengelkoch et al.'s opinion was consistent with their plotted data and showed that at the end of Group 10's training sessions, participants performed the procedures 95% error-free while Group 5's performance was approximately 78% free of error. Performance of the flight maneuvers in all categories showed similar trends. These reported findings supported Ebbinghaus's (1885/1913/2013) and Ausubel's (1963) respective theories that distributed practice is superior to massed practice. These findings also provided support for the current study because, in theory, students who use distributed practice should become proficient to take the private pilot check ride with fewer hours than students who are provided massed practice.

After the 4-month retention period, Mengelkoch et al. (1971) reported that both groups were given four retention trials, and their procedural and flight performances were evaluated. Both groups had a significant loss of procedural performance: Group 10 had a 16.5% loss, and Group 5 had a 20.1% loss. A *t* test established that loss of performance between groups was significant ($p < .01$). This finding supported Ebbinghaus's theory by clearly portraying a forgetting curve when it comes to procedures. Without continual practice, information is lost and much more difficult to be recalled. Although Mengelkoch et al. did not statistically analyze the performance differences between groups during the retention period, graphical representations of the data showed that Group 10 had better performance at the end of their training in all maneuvers (controlling altitude, bank angle, airspeed, turns to headings, and in leveling off at an assigned altitude) than Group 5.

Mengelkoch et al.'s (1971) findings provided a foundation for the current study with respect to understanding what happens if there is a gap in training. Based on their findings, the presumption is that students who take substantial time off from flight training can expect to accumulate more flight hours to obtain proficiency before taking the private pilot check ride than those who do not take time off from their flight training. This is because students must relearn information that was lost during their down time to attain the level of performance required to qualify for the check ride. Spruit et al. (2015) confirmed these results when they reported findings that were similar to those of Mengelkoch et al. The primary difference, though, is that Mengelkoch et al.'s study involved learning to fly an airplane. Mengelkoch et al. also was beneficial to the current study with respect to both procedural maneuvers and flight performance. In the current study, students are required to have 3 hours of maneuvering an aircraft solely by reference to instruments, which includes straight and level flight, constant airspeed climbs and descents, and turns to headings (Part 61.109(a)(3), (FAR 2021a), which are the same maneuvers involved in Mengelkoch et al.

Caligan, Jr. (2012). Caligan, Jr. studied whether the amount of flight time students needed to solo an aircraft could predict of the amount of flight time needed to obtain a PPL. The target and accessible populations were all pilots who held a private, commercial, or airline transport pilot (ATP) certificate for airplanes in 2008, and he collected data from civilian pilots who obtained their FAA PPL ($N =$

273). Caligan, Jr. conducted a simultaneous regression analysis involving the number of hours needed to solo versus the number of hours needed to earn the private pilot's license. He also regressed the number of hours needed to solo against the number of hours needed to earn the private pilot's license by school type (Part 61 v. Part 141), by the total number of instructors each student had, and by the average number of hours flown per week. Pilots self-reported the number of hours flown, organization type, and the flight time they had when they soloed the aircraft.

Caligan, Jr. (2012) found that the more hours a student flew per week, the more predictive solo time was to forecast the number of hours needed to earn a private pilot's license. Those who flew greater than or equal to 3 hours per week could predict the amount of time needed to earn a private pilot's license better than those who trained less frequently ($R^2 = .824$, $p < .001$). Those who flew between 2 and 2.9 hours per week could predict the amount of time needed to earn a private pilot's license better than those who trained less frequently ($R^2 = .568$, $p < .001$). Those who flew between 1 and 1.9 hours a week could predict the amount of time needed to earn a private pilot's license better than those who trained less frequently ($R^2 = .628$, $p < .001$).

Caligan Jr.'s (2012) findings helped inform the current study because, in theory, the more hours students fly per week, the more days they would fly per week, considering flight lessons are usually 1.5 hours long (excluding cross country flights). As a result of these findings, I targeted the independent variable, *training time to proficiency*, which represents the time in days it takes a student to become proficient to take the private pilot check ride. Caligan, Jr. also reported that the amount of time it took a student to become proficient to solo an airplane was a better predictor of total flight time to complete the PPL for students who trained in a Part 141 school ($R^2 = .798$, $p < .001$) than students who attended training under a Part 61 school ($R^2 = .490$, $p < .001$). These findings supported Ebbinghaus's (1885/1913/2013) and Ausubel's (1963) respective theories because students generally fly more times per week when attending a Part 141 school than a Part 61 school. These results also prompted the inclusion of *flight training program* (Part 141 and Part 61 schools) as an independent variable in the current study.

Absent from Caligan Jr. (2012) were analyses involving the organization offering the flight training program (e.g., college/university or private flight school), and the type of scheduling policy for flight lessons (e.g., mandated or student-driven scheduling). When Caligan Jr.'s findings are couched with Ebbinghaus's (1885/1913/2013) and Ausubel's (1963) respective theories, it is reasonable to conjecture that students attending a college/university would require less flight time to become proficient to take the private pilot check ride than students attending a private flight school. Similarly, students subjected to a mandated schedule would require less flight time to become check-ride proficient than students who create their own schedules. Given the absence of these data from

Caligan Jr., I targeted these two factors as independent variables for the current study to examine these conjectures. Thus, in addition to including Part 141 college/university schools, the current study also included Part 61 private flight schools in the form of fixed based operators (FOBs), a comparison between Part 141 and Part 61 schools, and a comparison between mandated vs. student-driven scheduling policies.

Graham (2017). In his dissertation research, Graham (2017) examined the effect of a compressed distributed practice flight schedule in conjunction with pursuing a private pilot's certificate. As a result, Graham's study provided the missing link that Gallo and Odu (2009) and Sheldon and Durdella (2010) could not provide because these latter studies focused on instruction and not on practice. Graham's study also involved the simultaneous application of reception and motor skill development in learning to fly an airplane.

The target and accessible populations were flight students enrolled in the Utah Valley University flight program during or after the spring 2014 semester, and who completed their PPL prior to July 18, 2017. The sample consisted of $N = 149$ students who were assigned to one of three groups, A, B, or C. Group A ($n = 23$) consisted of flight students who met with their instructor twice a week, individually and as a cooperative group. Group B ($n = 22$) was given the option to meet with their instructor as a cooperative group once a week. Group C ($n = 104$) was assigned to individualized flight instruction, which did not involve weekly meetings with a flight instructor. Because he lacked administrative control of the study, Graham (2017) could not randomly assign students to a group. He also did not conduct any inferential statistical analyses but instead provided a descriptive analysis by comparing means for the three separate semesters. Among his findings, Graham reported that Group A averaged the fewest number of days and flight hours to complete their PPL, followed by Group B and Group C.

Graham (2017) concluded that cooperative group meetings with students' CFIs reduced the time and flight hours needed to obtain a PPL. However, a closer look at the data revealed an alternative explanation. In all three semesters, Group A flew more hours per week than Group B, and Group B flew more hours per week than Group C. Group A also took fewer days to complete the private pilot certificate than Group B, and Group B took fewer days than Group C. Thus, the confounding variable *flight hours per week* could have been responsible for reducing the number of days and flights hours to obtain a private pilot license. If flight hours per week were a confounding variable, then Ebbinghaus's (1885/1913/2013) and Ausubel's (1963) theories would be a plausible explanation for the reduction in flight hours and, consequently, reducing the number of days to obtain a PPL.

Graham (2017) provided strong support for the current study because it also involved examining the spacing effect theory with respect to a Part 141 flight training program administered at a college or university. Based on the results from

Caligan, Jr. (2012), which were presented earlier, it was surmised that the spacing effect theory would apply to settings other than just a Part 141 college/university. Thus, as noted for Caligan Jr., the current study essentially extended Graham's study by examining the extent to which Ebbinghaus's (1885/1913/2013) and Ausubel's (1963) theories applied to Part 61 flight training programs and scheduling policies in addition to Part 141 programs.

Demographic Factors That Could Impact the Spacing Effect

It is doubtful that the application of Ebbinghaus's (1885/1913/2013) and Ausubel's (1963) respective theories would discriminate based on demographics such as biological sex at birth, age, or race/ethnicity. However, there are mixed findings reported in the literature. For example, Gallo and Odu (2009) reported that biological sex at birth (female or male), age, and race/ethnicity were not correctly specified with respect to the assumptions of regression and were removed from their analysis, but Sheldon and Durdella (2010) reported the opposite. With respect to biological sex at birth, Sheldon and Durdella indicated that although differences were not statistically significant, $\chi^2 = 1.348$, $\alpha \leq .51$, there was practical significance: Both females and males performed better in the compressed format than in a regular format. Sheldon and Durdella also reported statistical and practical significance of higher success rates in compressed format college courses across ethnicity (Asian/Pacific Islander, African American, Latino, White, Other), $\chi^2 = 214.667$, $\alpha \leq .001$, as well as age (25 and under, versus over 25), $\chi^2 = 10.785$, $\alpha \leq$

.005. As a result, the current study included these demographic factors as independent variables.

Summary and Study Implications

Although not exhaustive, the studies reviewed in this chapter provide examples of how Ebbinghaus's (1885/1913/2013) and Ausubel's (1963) respective theories have been studied in connection with meaningful learning and motor skill acquisition. For example, there is strong support that the spacing effect applies to meaningful learning for fixed and compressed distributed instruction as demonstrated by Gallo and Odu (2009) and Sheldon and Durdella (2010). The results from Flexman et al. (1972) also provide some support that Ebbinghaus's and Ausubel's respective theories apply to fixed distributed practice for meaningful learning.

There is less literature examining motor skill learning, though. The literature is segregated on a continuum between simple and complex tasks, with most of the reported research examining simple tasks such as Baddeley and Longman's (1978) research involving compressed distributed practice to learn the typewriter keyboard. Nevertheless, studies have shown that the spacing effect in the form of distributed practice for a complex task such as learning surgical skills can facilitate reception and motor skill learning (Moulton et al., 2006; Verdaasdonk et al. 2007). Mengelkoch (1971) demonstrated that the spacing effect theory could be applied to meaningful learning by showing distributed practice was superior to mass practice in learning to fly an airplane by reference to instruments. Although Mengelkoch's results did not support the spacing effect theory for motor skill acquisition in learning to fly an airplane, Graham's (2017) results demonstrated that applying the spacing effect with a compressed distributed practice flight schedule can facilitate meaningful learning and motor skill acquisition in learning to fly an airplane. Graham's results also were consistent with those of Caligan, Jr. (2012), who reported that flying more often can make flight time to solo an aircraft a better predictor in the amount of time needed to obtain a private pilot's license than flying less often. Because Caligan, Jr. documented the type of flight training (Part 61 school vs. Part 141 school), he reported that attending a Part 141 allowed the time to solo a better predictor of flight time to obtain a private pilot's license than attending a Part 61 school. Although Caligan, Jr. did not report the type of organization (college/university v. private flight school) or type of scheduling policy (mandated v. student-driven), the expectation is that type of organization and scheduling policy will both be correlated with hours flown per week.

The next logical step is to determine whether the spacing effect exists exclusive of cooperative group learning that was studied by Graham (2017) and to confirm if there are any significant differences between the type of training (Part 61 school v. Part 141 school), the type of organization (college/university vs. private flight school), and the type of scheduling policy (mandated vs. student-driven). Although the literature provides evidence that the spacing effect forms a curvilinear relationship, the studies cited here did not report a curvilinear relationship between days and flight hours to become proficient to take the private pilot check ride. The current study endeavored to identify the type of relationship that exists among the targeted variables with respect to learning to fly an airplane, and to generate a model that could be used to determine the optimum training schedule to reduce flight training costs.

Chapter 3

Methodology

Population and Sample

Population

The target and accessible populations were identical and consisted of all flight students pursuing their private pilot's license (PPL) in the United States. Because these two populations were the same, they are referred to singularly henceforth as the parent population. Furthermore, as presented in the definitions section of Chapter 1, flight students also are referred to as student pilots.

According to the FAA (2021) database, the number of student pilots in the United States has steadily increased since 2010. For example: in 2010, there were 119,119 student pilots; in 2015, there were 122,729 student pilots; and in 2019, there were 197,665 student pilots. These figures demonstrate there was an approximate 3% increase between 2010 and 2015, but an approximate 60% increase between 2015 and 2019. Furthermore, of the 197,665 student pilots in the United States in 2019, 27,255, or 13.78%, were women. Other demographics such as age, race, ethnicity, and marital status, however, remain unknown because the FAA does not collect these data for student pilots.

Sample

The sampling strategy I used was a combination of convenience and snowball sampling. To acquire the sample, I contacted flight school representatives and CFIs by telephone, in person, email, and text, and asked them to disseminate a flyer to their flight students. This flyer contained information about the study, a link to the questionnaire, and my contact information. I also personally solicited individual flight students and pilots and requested that they complete the online questionnaire. Because I solicited pilots employed by United Airlines, a supermajority of the sample most likely contained pilots employed by United Airlines. (*Note:* The reader is reminded from Chapter 1's Definitions section that pilots provided the same data as flight students except the pilot data were a recollection of when pilots were enrolled as flight students in a flight training program leading to a PPL. Thus, in this context, pilots were considered to be part of the general term "flight students.")

As reported in Table 3.1, of the $N = 167$ participants who comprised the final sample, 89.2% ($n = 149$) were male and 10.8% ($n = 18$) were female. Furthermore, 17 of the 18 females (94.4%) were single, and 128 of the 149 males

Table 3.1

Summary of Participants' Marital Status by Biological Sex at Birth

			Marital Status [a]			
			Single		Married	
Group [a]	***N***	**%**	***N***	**%**	***N***	**%**
Female	18	10.8	17	94.4	1	5.6
Male	149	89.2	128	85.0	21	14.0
Overall	167	100.0	145	86.8	22	13.2

Note. $N = 167$.
[a] The base for all percents is the overall sample size $N = 167$. All percents were rounded up and therefore the total percents might be a function of round-off error.

(85%) were married. The reader also will note that the male–female ratio of the sample was consistent with that of the parent population reported earlier.

As reported in Table 3.2, the overall mean age of the sample was $M = 20.2$ years ($SD = 5.8$), and the range was from 13 years to 53 years. There also was very little difference in the mean ages between female and male flight students, $M_F =$ 19.2 years ($SD = 2.8$) and $M_M = 20.3$ years ($SD = 6.0$).

Table 3.3 contains a summary of participants' flight instruction characteristics. Focusing on the type of flight training program, 69 participants overall received training under a Part 61 program and 98 received training under a Part 141 program. When disaggregated by sex, of the 18 females, 12 received training under a Part 61 and 6 received training under a Part 141 program, and of the 149 males, 57 received training under a Part 61 program and 92 received training under a Part 141 program. As for the type of organization, 52 participants overall reported receiving their flight instruction at a 2- or 4-year college/university, and 115 reported receiving their flight instruction at an FBO. When disaggregated by sex, four females received their training at a 2- or 4-year

Table 3.2
Summary of Participants' Age by Biological Sex

Group	N^a	M	SD	Range
Female	18	19.2	2.8	16–25
Male	149	20.3	6.0	13–53
Overall	167	20.2	5.8	13–53

Note. $N = 167$. Reported ages are when participants began taking flight instruction, not when they became check-ride proficient.

Table 3.3

Summary of Participants' Flight Instruction Characteristics by Biological Sex at Birth

		Flight Training				Organization				Scheduling Policy [b]			
		Part 61		Part 141		College		FBO		Student		Mandated	
Group	***N***	***N***	**%**	***N***	**%**	***N***	**%**	***N***	**%**	***N***	**%**	***N***	**%**
Female	18	12	7	6	4	4	2	14	8	15	9	3	2
Male	149	57	36	92	55	48	29	101	61	110	66	39	23
Overall	167	69	41	98	59	52	31	115 [a]	69	125	75	42	25

Note. $N = 167$. The base for all percents is the overall sample size of $N = 167$. All percents were rounded to nearest percent.
[a] 41 participants selected "Other," but their accompanying response was interpreted as FBO. [b] Four participants selected "Other," but based on their accompanying responses, three were interpreted as Mandated and one was interpreted as Student.

college/university and 14 attended an FBO, and 48 males received their training at a 2- or 4-year college/university and 101 attended an FBO. With respect to scheduling policy, 125 participants overall reported that their scheduling policy was student-driven, whereas 42 reported a mandated schedule. When disaggregated by sex, 15 female flight students reported their scheduling policy was student-driven and three reported it was mandated, and 110 male flight students reported a student-driven policy whereas 39 reported it was mandated.

Table 3.4 contains a summary of participants' race, which reflected the categories of the 2020 U.S. Census. As reported in Table 3.4, of the 167 participants, 135 (80.9%) were White, 6 (3.6%) were Black or African American, 8 (4.8%) were Asian, 2 (1.2%) were Native Hawaiian or Pacific Islander, and 3 (1.8%) reported "Other." Of the 17 female flight students who responded to this item, 14 (8.4%) were White, 2 (1.2%) were Native Hawaiian or Pacific Islander, and 1 reported "Other." Of the 137 male flight students who responded to this item,

Table 3.4

Summary of Participants' Race by Biological Sex at Birth

		White		B/AA [a]		Asian		H/PI [b]		Other	
Group	***N***	***N***	**%**	***N***	**%**	***N***	**%**	***N***	**%**	***N***	**%**
Female	18	14	8.4	0	0.0	0	0.0	2	1.2	1	0.6
Male	149	121	72.5	6	3.6	8	4.8	0	0.0	2	1.2
Total	167	135	80.9	6	3.6	8	4.8	2	1.2	3	1.8

Note. N = 167. The base for all percents is the overall sample size of N = 167. All percents were rounded to nearest percent.
[a] B/AA = Black or African American. [b] H/PI = Native Hawaiian or Other Pacific Islander.

121 (72.5%) were White, 6 (3.6%) were Black or African American, 8 (4.8%) were Asian, and 2 (1.2%) reported "Other." The reader will note that because the sample was predominately White, I reconstituted this factor into a dichotomy (White vs. Nonwhite) prior to conducting inferential statistics. This is discussed in Chapter 4.

In addition to their race, participants also were asked to respond to an ethnicity question, which reflected the categories of the 2020 U.S. Census. Of the 167 participants, 12 reported their ethnicity as Hispanic or Latino, 150 reported their ethnicity as not Hispanic or Latino, 4 reported "Other," and 1 did not respond to this item. Given that 90% of the sample specified their ethnicity as "not Hispanic or Latino," I considered this variable a constant for inferential statistics and did not include it as part of the final data set. This is discussed in Chapter 4.

Sample Representativeness

Based on the most recent data from the FAA (2021) database, the sample composition was expected to be 86% male and 14% female. Despite using convenience and snowball sampling, the sample composition reflected the parent population with the males accounting for 89.2% and females 10.8% of the sample (see Table 3.1). Although I reported other demographics such as age, race, ethnicity, and marital status, as noted earlier, I was unable to determine whether these personological characteristics were representative of the parent population because these data are not included in the FAA database. Nevertheless, the corresponding descriptive statistics as summarized in Tables 3.1–3.4 and described in the accompanying narrative will facilitate readers' assessment of the current study's sample representativeness.

Power Analysis

To determine the minimum sample size needed in advance of conducting the current study, I conducted an a priori power analysis using *G*Power* (Faul et al., 2007, 2009). Following Cohen et al.'s (2003) guidelines, I set power to .80 (β = .20) and alpha to $\alpha = .05$, and I set the effect size to $f^2 = 0.50$, which was derived from Caligan Jr. (2012). This analysis indicated that I would need an overall sample size of at least $N = 45$ participants. Subsequent a priori power analyses for detecting various effect sizes for each stage of a hierarchical regression strategy indicated that minimum sample sizes of $n = 29$, 77, and 295 for the three sets of variables, and $n = 395$ for any single research factor would be needed. The results of this a priori power analysis are summarized in Table 3.5.

Table 3.6 contains a summary of the results of a post hoc power analysis, which was based on the actual sample and effect sizes derived from the current

Table 3.5

A Priori Power Analysis and Estimated Minimum Sample Sizes (N)

Sets / Variables [a]	***k*** [b]	***ES*** [c]	**Estimated *N***
Overall	11	0.50	45
Set A = Time Interval (X_1)	1	0.30	29
Set B = Flight School Characteristics (X_2, X_3, X_4)	3	0.15	77
Set C = Flight Student Demographics (X_5, X_6, X_7, X_8, X_9)	7	0.05	295
Any single research factor	1	0.02	395

Note. All calculations are with respect to $\alpha = .05$ and power = .80 ($\alpha = .20$) as recommended by Cohen et al. (2003). See also Table 3.6.
[a] Overall represents the collective relationship the 11 initially targeted variables were estimated to have to the dependent measure, Y = Flight Time to Proficiency. X_1 = Number of days to check-ride proficiency. X_2 = Part 61 vs. Part 141 flight training programs. X_3 = 2- or 4-year college/university vs. FBO. X_4 = Student-driven vs. Mandated scheduling for flight instruction. X_5 = Male vs. Female flight students. X_6 = Age. X_7 = Race. X_8 = Not Married vs. Married flight students. [b] k = Total number of independent variables. [c] Effect sizes based on Caligan (2012), Cohen et al. (2003), and Graham (2017).

Table 3.6

Post Hoc Power Analysis Summary

Sets / Variables [a]	R^2	ΔR^2	***k*** [b]	***ES*** [c]	**Power**
Overall	.190	.190	3	0.233	> .99
Set A (X_1)	.143	.143	1	0.167	> .99
Set B (X_3 and X_4) in presence of Set A	.190	.047	2	0.058	.79

Note. N = 164. All calculations are with respect to $\alpha = .05$.
[a] Overall = The collective relationship the three IVs had with the dependent measure Y = Flight time to proficiency. X_1 = Number of days to check-ride proficiency. X_3 = 2- or 4-year college/universities vs. FBO. X_4 = Student-driven vs. Mandated scheduling for flight instruction. [b] k = Number of IVs under discussion. [c] ES = Effect size of the increment.
$^*p < .05$.

study's data. Of the 11 initially targeted independent variables, 3 remained in the final model after conducting preliminary data screening: X_1 = Training time to proficiency, which was the number of days to check-ride proficiency; X_3 = Organization offering flight training program, which compared 2- or 4-year colleges/universities vs. FBOs; and X_4 = Scheduling policy, which compared

Student-driven vs. Mandated schedules for flight instruction. (*Note:* For a discussion pertaining to why 8 of the 11 variables were removed from the final model, the reader is directed to the Preliminary Analyses section of Chapter 4.) X_1 was in Set A = Time Interval, X_3 and X_4 were in Set B = Flight School Characteristics, but no variables remained in Set C = Flight Student Demographics. As a result, the corresponding post hoc power analysis was conducted from a two-step hierarchical regression analysis perspective.

As reported in Table 3.6, the overall power of the study was greater than .99, and the respective powers for each of the components was .99 for Set A and .79 for Set B. Although the estimated a priori effect sizes were not consistent with the actual post hoc effect sizes, the actual sample size of the current study was sufficiently large to detect the smaller effect for Set B. As a result, the current study's data set provided adequate power per Cohen's et al. (2003) guidelines.

Instrumentation

The primary data collection instrument was a researcher-developed questionnaire that consisted of items designed to capture self-reported factual data. These data included the number of flight hours and days it took participants to achieve check-ride proficiency, the type of flight training program (Part 61 vs. Part 141), the type of organization from which participants received their flight training (2- or 4-year college/university vs. FBO), flight school scheduling policy (mandated vs. student-driven), participants' biological sex assigned at birth (female vs. male), age, race, marital status (Not married vs. Married), and ethnicity. A copy of the questionnaire is provided in the appendix.

The reader will note that the questionnaire did not measure any psychological constructs such as attitudes, motivation, or self-efficacy, and therefore the concepts of instrumentation validity and reliability were not applicable. As an example of why this is the case, consider the concept of reliability, which is a measure of response consistency. When participants respond to the item asking them to specify their age, I have to assume that their response will be an honest one, and that they will not report a radically different age (e.g., 49 and 31) if I were to administer the questionnaire a second time within a few days or weeks. As a result, I had no control over the veracity of participants' responses to these fact-based items and considered this a limitation to the current study.

With respect to instrumentation validity, the items were considered standard fare and designed to measure what they were intended to measure. For example, an item that asks participants to report their age is indeed designed to measure how old they are. Nevertheless, I gave attention to face and content validity by asking my advisor and a group of CFIs to review the items relative to their format, structure, and grammar (face validity), and to ensure that they were useful and relevant with respect to the research questions (content validity). Subsequent to this review, I performed a preliminary study by asking 10 pilots to complete the questionnaire to verify that the items were clear and understandable.

Procedures

Research Methodology

The research methodology/design that best fit the current study relative to answering its research questions was a combination of predictive correlational (RQ 1 and parts of RQ 3) and ex post facto (RQ 2 and parts of RQ 3). A correlational method with a predictive design was appropriate because it is used to assess relationships and patterns of relationships among variables in a single group of subjects (Ary et al., 2010). If two variables are correlated, then one variable can be used to predict the other. In the current study, RQ 1 and parts of RQ 3 involved a single group (flight students) with multiple measures. An ex post facto design was appropriate for studies involving pre-existing groups, which was the case for RQ 2 where group membership involved the targeted flight school characteristics, and parts of RQ 3 where group membership involved the targeted demographics of biological sex assigned at birth, age, race, and marital status.

Human Subjects Research

Because the current study involved collecting data from individuals, it was considered human subject research. As such, I submitted an application to Florida Institute of Technology's Institutional Review Board (IRB). Because the study involved the use of educational tests and survey procedures, it qualified for exempt status.

Study Implementation

The current study was implemented during the summer 2022 semester. The primary data collection instrument as described in the Instrumentation section presented previously was prepared and administered in electronic format through *Qualtrics*, an online survey platform. The instrument was also distributed in hardcopy flyer format by soliciting pilots to complete the survey. As described earlier in this chapter, I prepared a flyer that contained information about the study, a link to the questionnaire, and my contact information. I acquired a list of flight schools in Florida from the Aircraft Owners and Pilots Association (AOPA) and requested that these schools' representatives disseminate this flyer to their flight students. I also personally solicited individual flight students and pilots and requested that they complete the online questionnaire. At the end of the data collection period, all data initially were entered directly into *Microsoft Excel* and then transferred into the *JMP* statistical program for analysis.

With respect to IRB issues, I removed any identifying information from the data set, and I deleted all the data stored on Qualtrics. Furthermore, access to the data was password protected and only accessible to my major advisor and me. Although I also advised the survey respondents they could submit their e-mail address so the results of the study and dissertation can be mailed to them should they request one, I reminded them that doing so could compromise their anonymity.

Threats to Internal Validity

Internal validity refers to the extent to which observed changes in a dependent variable are attributed directly and solely to the independent variable and not to some uncontrolled factors (Ary et al., 2010). These uncontrolled factors are considered threats to internal validity because they could impact the dependent variable, which in turn could lead to spurious results. Campbell and Stanley (1963) initially identified eight threats to internal validity. Subsequent to Campbell and Stanley's seminal publication, several additional threats have been identified in the literature. Although the concept of threats to internal validity initially was applied to experimental research designs, some also are relevant to observational studies. As a result, this section is partitioned into two parts that present those threats that were and were not relevant to the current study.

Threats Relevant to the Current Study. As noted earlier in this chapter, the current study's design involved both ex post facto and correlational methodologies. According to Fraenkel et al. (2003), the relevant threats to these methodologies are selection, location, instrumentation, and mortality. A discussion of these threats—including their definitions, how they could affect the proposed study's results, and what I did to mitigate or control them—follows.

Selection. A selection threat, which also is referred to as a subject characteristics threat, occurs when there are established differences between groups before the study begins (Campbell & Stanley, 1963). With respect to ex post facto studies, groups are intact prior to the study, which means that participants cannot be randomly assigned. Thus, if the groups are not equivalent before the study begins, then the researcher cannot know if differences on the dependent variable are due to the treatment or to the pretreatment differences between the groups. For the ex post facto component of the current study, I did not have the opportunity to assign participants randomly to particular groups. As an example, consider RQ 2, which examined differences in flight time to proficiency between various flight school characteristics, one of which was a comparison between Part 141 vs. Part 61 flight training programs. If the mean age of the former group was significantly younger than that of the latter group, and the former group also had a shorter flight time to proficiency, then one could argue that it was because of their age and not the type of training program.

With respect to correlational studies, it is possible that a third variable could be the reason for a particular relationship. As an example, consider RQ 1, which examined the relationship between training time to proficiency and flight time to proficiency. If a significant relationship were found, it could be that this relationship was actually due to a third variable such as participants' age or biological sex assigned at birth.

To mitigate the selection threat, I collected key demographical data from participants and confirmed there were no significant differences between groups with respect to these demographical factors (ex post facto), and that these factors were not the reason for any observed relationship between the study's targeted variables (correlational).

***Location*.** A location threat refers to changes in the location where a study is implemented such that it provides an alternative explanation for the results (Ary et al., 2019). In the current study, a location threat is relevant to both the ex post facto and correlational components because data were collected from flight students who received their flight instruction from different flight schools. Because the sampling strategy involved soliciting pilots who received training throughout the United States, there were differences in flight schools' facilities and resources. For example, with respect to the ex post facto component, a Part 141 flight school might have more resources than a Part 61 flight school, and with respect to the correlational component, the location of a flight school could be the reason for any relationship that is found among the targeted variables. Thus, it was conceivable that differences in resources could impact students' flight time to proficiency. To help control this threat, I chose independent variables that reflected differences in flight schools. For example, I reported the type of training school (Part 61 vs. Part 141), the type of organization (2- 4-year colleges vs. fixed base operator (FBO), and the type of scheduling policy (student-driven scheduling vs. mandated scheduling).

Instrumentation. An instrumentation threat involves any changes made to the instruments used to measure the dependent variable (Campbell & Stanley,

1963) and is associated with three facets: instrument decay, data collector characteristics, and data collector bias. Instrument decay, which refers to modifying the nature of an instrument during the implementation of a study (e.g., changing from selective- to constructive-response items), is not relevant to the current study because a single instrument was administered one time. However, data collector characteristics and data collector bias were relevant to the current study because flight time to proficiency was determined by different CFIs. With respect to the former, it was possible that CFIs' characteristics such as their age, biological sex assigned at birth, or race/ethnicity could impact their decision in deciding when a student was check-ride proficient. For example, older CFIs might be more conservative in their training regimen than younger CFIs and require students to complete more training than their younger counterparts. Similarly, it also is possible for a CFI to unconsciously favor one student over another. For example, an older male CFI might unconsciously favor male flight students over female flight students. These factors could be the reason one group (e.g., Part 141 vs. Part 61) has a lower (or higher) flight time to proficiency than the other group or be the reason for the relationship between the targeted factors. To control for this threat, I intended to have participants provide information about their CFIs' characteristics but based on the design of the study this was not feasible.

Mortality. Experimental mortality, or attrition, refers to the loss of participants during the implementation of a study (Campbell & Stanley, 1963).

With respect to ex post facto studies, if participants in one group were to be eliminated because they submitted incomplete questionnaires, then this could have an effect of the results. As an example, consider RQ 2, which examined differences in flight time to proficiency between various flight school characteristics, one of which was a comparison between Part 141 vs. Part 61 flight training programs. If participants in the Part 61 group had higher flight times to proficiency than those in the Part 141 group, and the Part 61 group also had a high attrition because they did not submit fully completed questionnaires, then the loss of these data could result in no differences in flight time to proficiency between the groups. Similarly, with respect to correlational studies, it is possible that the loss of these data could impact whether or not there is a relationship between the targeted variables. To control for this threat, I monitored the data collection process and increased the minimum sample size to account for any possible incomplete questionnaires. I also, where appropriate, followed Cohen et al.'s (2003) guidelines for handling missing data.

Threats Not Relevant to the Current Study. Because the current study did not involve any type of intervention, many of the threats to internal validity were not applicable. These included history, testing, maturation, implementation, attitudes of subjects, regression threats, diffusion, and selection-maturation interaction. A brief discussion follows.

History. A history threat refers to any extraneous event that occurs at the same time a treatment is being administered, and which could result in the same observed outcome regardless of the treatment. Because a history threat did not refer to past events that might have occurred, and because the current study was not administering any treatments, this threat was not applicable.

Testing. A testing threat is relevant when participants are administered the same assessment prior to and after treatment. The concern here is that scores on the post-assessment, which tend to be higher, might not be the result of treatment but instead are due to participants' exposure to the items on the pre-assessment: they learned the material from the pretest, they developed strategies to take the posttest, or they might not be as anxious to taking the posttest. Because the current study did not administer any pre- and post-assessments, this threat was not applicable.

Maturation. A maturation threat refers to biological and/or psychological changes in the results of an intervention that are due to factors associated with the passage of time or changes in participants' characteristics rather than the intervention. In the current study, the reader will note from Table 3.2 that participants' ages ranged overall from 13 to 53 years, which means that some of the participants were likely to undergo considerable biological and psychological changes. However, because (a) the study period window was short (a few weeks), (b) flight instruction generally takes less a 1 year to complete, and (c) there was no intervention, I did not consider this threat applicable to the current study.

Implementation. An implementation threat, which also is called an experimenter effect, refers to the effect researchers might have on a study by the way they implement the study. For example, if a researcher is implementing a study personally, then it is possible the researcher might unconsciously favor the treatment group over the control group (or vice versa), which could then impact the results. This threat is not relevant to observational studies because there was no intervention that could lead to different implementations.

Attitudes of Subjects. This threat, which also is referred to as subject effects, involves attitudes participants develop in response to the research situation, and these attitudes could influence the results of a study (Ary et al., 2019). The subject effects threat is generally manifested in one of three ways: (a) the Hawthorne effect, which often occurs in the treatment group, and involves participants doing whatever they can to help the researcher because they feel honored that they were selected to be part of the study; (b) the John Henry effect, or compensatory rivalry, which is when participants in the control group adopt an attitude of competition and try to outperform their treatment group counterparts; and (c) compensatory demoralization, which is when participants in either group feel they are being neglected or treated differently and may become demoralized and exert less effort into the experiment than they should exert. This threat was not applicable because the current study did not involve any type of intervention.

Statistical Regression. This threat refers to the situation where the results of a study are due to determining group membership based on extreme scores from a pre-assessment. By focusing only on participants with either low or high pre-assessment scores prior to treatment, it is possible that their post-assessment scores will regress to the mean regardless of the intervention. This threat was not applicable to the current study because no pre-assessment was administered, there was no intervention, and group membership with respect to the ex post facto component of the study was preexisting.

Diffusion. Diffusion occurs when one group, usually the treatment group, communicates with the control group regarding the treatment that is being administered (Ary et al., 2019). This sharing of information in turn affects the control group's behavior, which can influence the dependent measure. This threat was not applicable to the current study because no treatment was administered.

Interaction of Selection and Maturation. Selection and maturation may interact to create an effect on the dependent variable that is mistakenly attributed to the treatment (Campbell & Stanley, 1963). This can occur in quasi-experimental or ex-post facto designs where the groups involved in the experiment are preexisting rather than being randomly assigned. By chance, one group might mature faster than the other group, and the increased maturation rate could account for the change in the dependent variable rather than the treatment effect. A selection-maturation interaction threat was initially not thought to be a serious threat in the context because of the current study because the time to obtain a private pilot license rarely takes longer than 1 year and most likely is fewer than 6 months. Hence, if there were maturation effects because of preexisting conditions, they would be ameliorated by the short time to obtain a private pilot license. However, there were a considerable number of participants who took longer than 1 year to obtain their PPL. The selection-maturation threat was mitigated, however, because participants' ages were not a significant factor in predicting the DV. In fact, as indicated earlier, none of the participants' personological characteristics had any effect on the dependent variable, and therefore I considered this threat to be non-applicable.

Treatment Verification and Fidelity

In research studies that involve the administration of a treatment, it is important for researchers to (a) verify that the treatment was administered as proposed or planned and (b) provide evidence that fidelity to the treatment was maintained throughout the study's implementation (Moncher & Prinz, 1991). Giving attention to treatment verification enhances the integrity of a study because its focus is on ensuring that the manipulation of the independent variables was consistent with the researcher's intention. This verification process is required to make valid interpretations of a study's results and to adequately discuss issues related to external validity. Furthermore, by providing empirical evidence of a treatment's fidelity enhances both internal and external validity (Shaver, 1983).

Because the current study was not experimental and hence there was no manipulation of an independent variable, the concepts of treatment verification and fidelity did not apply here in a traditional sense. However, before issues associated with external validity—particularly those related to ecological generalizability such as replication studies—could be properly discussed, it is critical that details of a study's independent variables be clearly presented. As a result, and following Shaver's (1983) guidelines: (a) Based on the literature review and theory presented in Chapter 2, I confirmed that the targeted IVs of the current study were appropriate with respect to influencing the dependent variable, flight time to proficiency; (b) I provided a detailed description of the current study's targeted IVs in Table 3.7; (c) I documented in the Study Implementation section of this chapter the procedures I followed; and (d) I provided a description (see next section) of the data analysis strategies I used to test the hypotheses and answer the corresponding research questions.

Data Analysis

The current study's data were analyzed using both descriptive and inferential statistical strategies. Descriptive statistics were used in the current chapter to summarize participants' demographical data and included measures of central tendency (mean and median), variability (range and standard deviation), and percentages. I reported the results of these analyses in both narrative and table forms (see Tables 3.1–3.4). I also reported additional descriptive statistics with respect to the DV in Chapter 4.

With respect to inferential statistics, the primary strategy was hierarchical multiple regression, which was used to examine the relationship among the targeted variables and predict which factors influenced the dependent variable, which was Y = Flight time to proficiency. This statistical strategy was appropriate because the DV was a single continuous variable, and the targeted factors were either continuous or numerically coded categorical variables. The current study initially targeted 11 independent variables, which I partitioned into three functional sets: Set A = Time Interval, Set B = Flight School Characteristics, and Set C = Flight Student Demographics. The initial set-entry order was A–B–C, but Set C was eliminated as a result of preliminary data analyses (see Chapter 4) and therefore was excluded from the final data set. The use of functional sets prevented or reduced inflated alpha levels and enhanced power and decrease standard errors (Cohen et al., 2003). A summary of these sets/variables is given in Table 3.7, and a brief description follows.

Set A = Time Interval. Set A consisted of the single variable, X_1 = Number of days to check-ride proficiency, which represented the time from a flight student's first flight to when a CFI determined the student was proficient to take the private pilot check ride.

Set B = Flight School Characteristics. Set B consisted of three variables: X_2 = Flight training program, which compared Part 61 vs. Part 141; X_3 = Organization offering flight training, which compared 2- or 4-year colleges/universities vs. fixed-based operation (FBO) programs; and X_4 = Scheduling policy, which mandated vs. student-driven flight instruction scheduling.

Table 3.7

Summary and Description of Independent and Dependent Variables

Sets/Variables	Description
Set A = Time Interval	
X_1 = Training time to proficiency (in days)	X_1 was continuous and represented the time from a flight student's first flight to when a CFI determined the student was proficient to take the private pilot check ride.
Set B = Flight School Characteristics	
X_2 = Flight training program	X_2 was dichotomous and represented a comparison between Part 61 vs. Part 141. The variable was dummy coded with Part 141 as the reference group.
X_3 = Organization offering flight training program	X_3 was dichotomous and represented a comparison between flight training programs offered at a 2- or 4-year college/university vs. those offered at a Fixed-Base Operator (FBO) flight school. The variable was dummy coded with FBO as the reference group.
X_4 = Scheduling policy	X_4 was dichotomous and represented a comparison between Mandated scheduling vs. Student-driven scheduling. The variable was dummy coded with Student-driven scheduling as the reference group.
Set C = Flight Student Demographics	
X_5 = Biological sex assigned at birth	X_5 was dichotomous and represented a comparison between Female vs. Male flight students. The variable was dummy coded with Male as the reference group.
X_6 = Age	X_6 was continuous and measured flight students' age in years.
X_7 = Race	X_7 was categorical and represented a comparison among four races as defined by the 2020 U.S. Census: White, Black or African American, Asian, and Hawaiian or Pacific Islander. The variable was dummy coded with White as the reference group where X_{7a} = Black/African American, X_{7b} = Asian, and X_{7c} = Hawaiian or Pacific Islander. Due to disparate sample sizes, this factor was reduced to a dichotomy of White vs. Nonwhite.
X_8 = Marital status	X_8 was dichotomous and represented a comparison between Not married vs. Married flight students. The variable was dummy coded with Married as the reference group. For participants who self-reported Not Married, they were asked to qualify their status as Single, Separated, Divorced, or Widowed.
X_9 = Ethnicity	X_9 was dichotomous and represented a comparison between "Hispanic or Latino" vs. "Not Hispanic or Latino." Due to disparate sample sizes, this factor became a constant.
Set D = Dependent Variable	
Y = Flight time to proficiency	Y was continuous and represented the total hours of flight time flight students accrued prior to being declared check-ride proficient by their CFI. These hours included both physically flying an aircraft as well as simulator time. Hours of the former were measured by the aircraft's Hobbs meter.

Set C = Flight Student Demographics. Set C consisted of four variables: X_5 = Biological sex assigned at birth, which compared female vs. male flight students; X_6 = Age of participants given in years; X_7 = Race, which represented a comparison among White, Black or African American, Asian, and Native Hawaiian or Pacific Islander; and X_8 = Marital Status, which compared Not married vs. Married flight students.

Set D = Dependent Variable. Set D consisted of the single dependent variable, Y = Flight time to proficiency, which was defined as the number of flight hours needed before a CFI determined a student was check-ride proficient.

For a detailed discussion of the respective data analyses and corresponding results, the reader is directed to Chapter 4.

Chapter 4

Results

Introduction

This chapter is organized into three sections. The first section contains a summary of the descriptive statistics related to the independent and dependent variables. The section includes summaries of the results from examining the single variable of Set A (X_1 = Training time to proficiency in days) with respect to: (a) the factors of Set B (X_2 = Flight training program, X_3 = Flight training organization, and X_4 = Scheduling policy), and (b) the factors of Set C (X_5 = Biological sex assigned at birth, X_7 = Race, and X_8 = Marital status). The section also includes a summary the results from examining the dependent variable, Y = Flight time to proficiency in hours, with respect to the training-related factors of Set B and the demographic factors of Set C.

The second section contains a summary of the inferential statistics results and is apportioned into two subsections: preliminary analyses and primary analysis. The first subsection contains a discussion of the various analyses I conducted as part of data screening. These included modifications I made to the data set to prepare it for analysis, missing data analysis, outlier analysis, multicollinearity analysis, and an analysis to verify that the data set was compliant with the corresponding statistical strategy, which was multiple regression. The primary analysis subsection contains a summary of the results of hierarchal regression. The last section presents the results of hypothesis testing, which are aligned to the three research questions and corresponding research hypotheses as set forth in Chapter 1.

Descriptive Statistics

The primary data collection instrument was a researcher-developed questionnaire that consisted of items where participants self-reported factual data related to their flight training experiences as well as key demographics. The former included the number of training days and flight hours it took participants to achieve check-ride proficiency, their flight training program (Part 61 vs. Part 141), the type of organization from which they received their flight training (2- or 4-year college/university vs. FBO), and their flight school's scheduling policy (Mandated vs. Student-driven). The latter included participants' biological sex assigned at birth (female vs. male), age, race (White vs. Nonwhite), and marital status (Married vs. Not Married). A summary of participants' responses to these items is presented here. The reader is reminded that participants' personological characteristics were presented in Chapter 3 (Tables 3.1–3.4) with respect to sample representativeness.

Summary of Flight Training Program, Organization, and Scheduling Policy

As reported in Table 4.1, of the 167 participants, 98 were trained under a Part 61 program, 59 were trained under a Part 141 program, and 10 reported "Other," which I assigned these responses to either Part 141 or Part 61 as reflected in subsequent tables. Thus, there was an approximate 60–40 split between Part 61 and Part 141 training programs in favor of Part 61.

Table 4.1

Summary of Descriptive Statistics for Type of Flight Training Organization and Scheduling Policy by Fight Training Program

Flight Training Program		Organization [b]					Scheduling Policy [d]				
		C	F	O	Overall [c]		S	M	O	Overall [e]	
	N [a]	N	N	N	N_T	%	N	N	N	N_T	%
Part 61	98	8	63	27	98	62.4	87	10	2	99	62.2
Part 141	59	42	9	8	59	37.6	32	26	2	60	37.8
Total	157	52	74	39	165	100.0	124	39	4	167	100.0

Note. N = 167.

[a] 10 participants did not report all the requested information. [b] C = College (i.e., 2- or 4-year college/university, F = Fixed-Base Operator, and O = "Other." [c] N_T = The total of the previous three columns. The Total of 165 is greater than the sum of the previous cells (98 + 59 = 157) because eight participants reported "Other" for their flight training program or for their organization, or both. This is reconciled in Table 4.2. The % is based n = 157, which is the total reported. [d] S = Student-driven, M = Mandated, and O = Other. [e] N_T = The total of the previous three columns. The Total of 167 is greater than the sum of the previous two cells (99 + 60 = 159) because eight respondents reported "Other" for their flight training program. This is reconciled in Table 4.3. The % is based on n = 159, which is the total reported.

When these data were examined with respect to the type of flight training organization, 52 attended a 2- or 4-year college/university, 74 attended a fixed-base operator (FBO) organization, and 39 reported "Other." The reader is apprised that after reviewing the comments participants provided with respect to the "Other" category, I assigned these responses to either FBO or College (i.e., 2- or 4-year college/university). This reassignment is reflected in subsequent tables. Lastly, when examined with respect to scheduling policy, 124 operated under a student-driven schedule, 39 operated under a mandated schedule, and four reported "Other." Similar to Organization, I reassigned the "Other" category to either Mandated or Student-driven, which also is reflected in subsequent tables.

Training Time to Proficiency (in Days) with Respect to Flight Training Program, Organization, and Scheduling Policy

I next examined X_1 = Training time to proficiency with respect to X_2 = Flight Training Program (Part 61 vs. Part 141), X_3 = Flight Training Organization (2- or 4-year college/university vs. FBO), and X_4 =Scheduling Policy (Mandated vs. Student-driven) separately. A summary of the results is provided in Table 4.2, and a brief discussion with respect to each of the three factors follows.

Table 4.2

Summary of Descriptive Statistics for X_1 = Training Time to Proficiency (in Days) with Respect to X_2 = Flight Training Program, X_3 = Flight Training Organization, and X_4 = Scheduling Policy

Group [a]	***N***	M_{Days}	***SD***	***Mdn***	**Range**	**Approximate** M_{Months} [b]	**Skewness**
			X_2 = Flight Training Program				
Part 61	101	516	852	275	56–6748	17.2	5.08
Part 141	63	285	243	226	12–1359	9.5	2.23
			X_3 = Flight Training Organization				
FBO	113	435	562	253	12–3946	14.5	3.81
College	51	409	927	236	51–6748	13.6	6.65
			X_4 = Scheduling Policy				
Mandated	41	291	236	236	12–1063	9.7	1.46
Student-driven	123	472	785	249	54–6748	15.7	5.48
Overall	164	427	693	247	12–6748	14.9	6.10

Note. N = 167. Three cases were excluded from all calculations. All data rounded to nearest whole number.

[a] FBO = Fixed-base operator and College = 2- or 4-year college/university. [b] The approximate mean number of months was calculated by dividing the mean number of days by 30 and rounding to one decimal place.

Training Time to Proficiency Based on Flight Training Program. As reported in Table 4.2, the overall mean number of training days participants required before they were check-ride proficient was $M = 427$ days ($SD = 693$), the median was $Mdn = 247$ days, and the range was from 12 days to 6,748 days. When these data were disaggregated by the type of flight training program, Part 61 participants had a mean of $M_{61} = 516$ days ($SD = 852$), the median was $Mdn = 275$ days, and the range was from 56 days to 6,748 days. Part 141 participants had a mean of $M_{141} = 285$ days ($SD = 243$), the median was $Mdn = 226$ days, and the range was from 12 days to 1,359 days. More concretely, Part 61 participants took nearly twice as long as Part 141 participants (17.2 months vs. 9.5 months), on average, to be check-ride proficient. However, given the skewness of the data, when comparing the respective medians, Part 61 participants required approximately 9 months (275 days) to become check-ride proficient and Part 141 participants took approximately 7.5 months (226 days).

Training Time to Proficiency Based on Organization. Focusing on the type of flight training organization (2- or 4-year college/university vs. FBO), as reported in Table 4.2, there was an approximate 70–30 split (113 vs. 51) between the two organizations where participants reported receiving their flight training, favoring FBOs. The mean number of training days it took participants to become check-ride proficient at an FBO was $M_{\text{FBO}} = 435$ days ($SD = 562$), the median was $Mdn = 253$ days, and the range was from 12 days to 3,946 days. For those at a 2- or

4-year college/university, the mean number of training days it took participants to become check-ride proficient was M_{College} = 409 days (*SD* = 927), the median was *Mdn* = 236 days, and the range was from 51 days to 6,748 days. Regardless, if comparing means or medians, there was very little difference between the two types of organizations with respect to the number of training days it took participants to become check-ride proficient: FBO-trained participants took on average no more than 1 month longer than college-trained participants.

Training Time to Proficiency Based on Scheduling Policy. Focusing on scheduling policy (Mandated vs. Student-driven), as reported in Table 4.2, there was a 75–25 split (123 vs. 41) between the two schedules, favoring student-driven. Under a mandated schedule, the mean number of training days it took participants to become check-ride proficient was M_{Mandated} = 291 days (*SD* = 236), the median was *Mdn* = 236 days, and the range was from 12 days to 1,063 days. With respect to student-driven schedules, the mean number of training days it took participants to become check-ride proficient was M_{Srudent} = 472 days (*SD* = 785), the median was *Mdn* = 249 days, and the range was from 54 days to 6,748 days. Thus, with respect to group means, the training time required to become check-ride proficient for participants under a student-driven schedule was 6 months longer compared to those on a mandated schedule. However, given the skewness of both groups, this difference was 2 weeks when compared to the groups' respective medians.

Training Time to Proficiency (in Days) with Respect to Demographics

I next examined X_1 = Training time to proficiency with respect to participants' reported demographics, which included X_5 = Biological sex assigned at birth, X_7 = Race, and X_8 = Marital status. The reader will note that because 82% of the sample was White, I merged the other racial groups into a single Nonwhite group, which consisted of Black or African American (n = 6), Asian (n = 8), Native Hawaiian or Pacific Islander (n = 2), and Other (n = 3). The reader also will note that with respect to ethnicity, 90% (n = 150) of the initial sample reported "not Hispanic or Latino." As a result, I treated this factor as a constant. Lastly, 87% of the sample reported being not married, which consisted of single, separated, divorced, and widowed. Rather than treat these as separate groups, I maintained a dichotomy for this factor: Married vs. Not Married. A brief discussion of these analyses follows, and a summary of the results is provided in Table 4.3.

Training Time to Proficiency Based on Biological Sex Assigned at Birth. As reported in Table 4.3, there was a near 90–10 split between the two sexes with males accounting for the vast majority of the sample. The mean number of training days female participants required before they were check-ride proficient was M_{Female} = 265 days (SD = 202), the median was Mdn = 179 days, and the range was from 76 days to 866 days. The mean number of training days male participants required before they were check-ride proficient was M_{Male} = 447 days (SD = 729), the median was Mdn = 247 days, and the range was from 12 days to 6,748 days.

Table 4.3

Summary of Descriptive Statistics for X_1 = Training Time to Proficiency (in Days) with Respect to X_5 = Biological Sex Assigned at Birth, X_7 = Race, and X_8 = Marital Status

Group [a]	***N***	M_{Days}	***SD***	***Mdn***	**Range**	**Approximate** M_{Months} [b]	**Skewness**
				X_5 = Biological Sex Assigned at Birth			
Female	18	265	202	179	76–866	8.8	1.81
Male	146	447	729	247	12–6748	14.9	5.82
				X_7 = Race			
White	135	447	751	247	51–6748	14.9	5.76
Nonwhite	29	333	295	257	12–1454	11.1	2.21
				X_8 = Marital Status			
Married	21	547	1432	186	75–6748	18.2	4.47
Not Married	143	409	510	247	12–3946	13.6	4.15

Note. N = 167. Three cases were excluded from all calculations. All data rounded to nearest whole number.

[a] Nonwhite = Black or African American (n = 6), Asian (n = 8), Native Hawaiian or Pacific Islander (n = 2), and Other (n = 3). Not Married = Single, Separated, Divorced, and Widowed, but sample sizes were not recorded. [b] The approximate mean number of months was calculated by dividing the mean number of days by 30 and rounding to one decimal place.

More concretely, male participants took nearly twice as long as female participants (14.9 months vs. 8.8 months), on average, to be check-ride proficient. However, given the skewness of the data, when comparing the respective medians, male participants required approximately 8 months (247 days) to become check-ride proficient whereas female participants took approximately 6 months (179 days).

Training Time to Proficiency Based on Race. Focusing on race, as reported in Table 4.3, the mean number of training days it took White participants to become check-ride proficient was M_{White} = 447 days (*SD* = 751), the median was *Mdn* = 247 days, and the range was from 51 days to 6,748 days. The mean number of training days it took Nonwhite participants to become check-ride proficient was

M_{Nonwhite} = 333 days (*SD* = 295), the median was *Mdn* = 257 days, and the range was from 12 days to 1,454 days. More concretely, White participants took nearly 4 months longer than Nonwhite participants (14.9 months vs. 11.1 months), on average, to be check-ride proficient. However, given the skewness of the data, when comparing the respective medians, White participants required 10 fewer days than Nonwhite participants (247 days vs. 257 days).

Training Time to Proficiency Based on Marital Status. Focusing on marital status, as reported in Table 4.3, the mean number of training days it took Married participants to become check-ride proficient was M_{Married} = 547 days (*SD* = 1,432), the median was *Mdn* = 186 days, and the range was from 75 days to 6,748 days. The mean number of training days it took Not Married participants to become check-ride proficient was $M_{\text{Not Married}}$ = 409 days (*SD* = 510), the median was *Mdn* = 247 days, and the range was from 12 days to 3,946 days. More concretely, Married participants took 4.6 months longer than Not Married participants (18.2 months vs. 13.6 months), on average, to be check-ride proficient. However, given the skewness of the data, when comparing the respective medians, Married participants required 61 fewer days than Not Married participants (186 days vs. 247 days).

Flight Time to Proficiency (in Hours) with Respect to Flight Training Program, Organization, and Scheduling Policy

I next examined *Y* = Flight time to proficiency with respect to factors X_2 = Flight training program (Part 61 vs. Part 141), X_3 = Flight training organization (2-

Table 4.4

Summary of Descriptive Statistics for Y = Flight Time to Proficiency (in Hours) with Respect to X_2 = Flight Training Program, X_3 = Flight Training Organization, and X_4 = Scheduling Policy

Group	*N*	M_{Hours}	*SD*	*Mdn*	Range	Skewness
			X_2 = Flight Training Program			
Part 61	101	62.1	14.1	60.9	40.0–107.0	0.57
Part 141	63	60.7	18.3	55.0	35.0–116.8	0.92
			X_3 = Flight Training Organization			
FBO	113	62.2	14.7	60.0	36.6–107.0	0.63
College	51	60.3	18.1	55.0	35.0–116.8	0.97
			X_4 = Scheduling Policy			
Mandated	41	64.4	18.6	58.0	35.2–116.8	0.61
Student-driven	123	60.6	14.7	59.7	35.0–107.0	0.74
Overall	164	61.6	15.8	59.4	35.0–116.8	0.75

Note. *N* = 167. Three cases were excluded from all calculations.
[a] FBO = Fixed-base operator and College = 2- or 4-year college/university.

or 4-year college/university vs. FBO), and X_4 = Scheduling policy (Mandated vs. Student-driven) separately. As reported in Table 4.4, the overall mean flight time participants accrued before they were check-ride proficient was M = 61.6 hours (SD = 15.8), the median was Mdn = 59.4 hours, and the range was from 35 hours to 116.8 hours. A brief discussion of these results relative to each factor follows.

Flight Time to Proficiency Based on Flight Training Program. Focusing on the type of flight training program, Part 61 participants had a mean of M_{61} = 62.1 hours (SD = 14.1), the median was Mdn = 60.9 hours, and the range was from 40 hours to 107 hours. Part 141 participants had a mean of M_{141} = 60.7 hours (SD = 18.3), the median was Mdn = 55 hours, and the range was from 35 hours to 116.8

hours. Thus, Part 61-trained participants required on average at most 2 hours more than Part 141-trained participants to become check-ride proficient.

Flight Time to Proficiency Based on Organization. Focusing on the type of flight training organization (2- or 4-year college/university vs. FBO), as reported in Table 4.4, the mean flight time to proficiency for participants trained at an FBO was M_{FBO} = 62.2 hours (*SD* = 14.7), the median was *Mdn* = 60 hours, and the range was from 36.6 hours to 107 hours. For participants trained at a 2- or 4-year college/university, the mean flight time to proficiency was M_{College} = 60.3 hours (*SD* = 18.1), the median was *Mdn* = 55 hours, and the range was from 35 hours to 116.8 hours. Thus, FBO-trained participants required on average at most 2 hours more than college-trained participants to become check-ride proficient.

Flight Time to Proficiency Based on Scheduling Policy. Focusing on the type of scheduling policy (Mandated vs. Student-driven), as reported in Table 4.4, the mean flight time to proficiency for participants operating under a mandated schedule was M_{Mandated} = 64.4 hours (*SD* = 18.6), the median was *Mdn* = 58 hours, and the range was from 35.2 hours to 116.8 hours. For participants operating under a student-driven schedule, the mean flight time to proficiency was M_{Student} = 60.6 hours (*SD* = 14.7), the median was *Mdn* = 59.7 hours, and the range was from 35 hours to 107 hours. Thus, participants under a mandated schedule required on average approximately 4 hours more than those under a student-driven schedule to become check-ride proficient.

Table 4.5
Summary of Descriptive Statistics for Y = Flight Time to Proficiency (in Hours) with Respect to X_5 = Biological Sex Assigned at Birth, X_7 = Race, and X_8 = Marital Status

Group [a]	*N*	M_{Hours}	*SD*	*Mdn*	Range	Skewness
			X_5 = Biological Sex Assigned at Birth			
Female	18	63.2	14.7	59.4	44.1–99.0	1.30
Male	146	61.4	15.9	59.4	35.0–116.8	0.72
			X_7 = Race			
White	135	61.0	15.4	58.7	35.0–116.8	0.79
Nonwhite	29	64.3	17.4	64.7	40.0–100.0	0.55
			X_8 = Marital Status			
Married	21	61.8	15.8	60.0	40.0–100.0	0.78
Not Married	143	61.5	15.8	59.0	35.0–116.8	0.75

Note. N = 167. Three cases were excluded from all calculations. All data rounded to nearest whole number.
[a] Nonwhite = Black or African American (n = 6), Asian (n = 8), Native Hawaiian or Pacific Islander (n = 2), and Other (n = 3). Not Married = Single, Separated, Divorced, and Widowed, but sample sizes were not recorded. [b] The approximate mean number of months was calculated by dividing the mean number of days by 30 and rounding to one decimal place.

Flight Time to Proficiency (in Hours) with Respect to Demographics

The last set of descriptive statistics I performed involved examining Y = Flight time to proficiency with respect to X_5 = Biological sex assigned at birth, X_7 = Race, and X_8 = Marital status. A brief discussion of these analyses follows, and a summary of the results is provided in Table 4.5. The reader is reminded that race and marital status were treated as dichotomies, and ethnicity was treated as a constant.

Flight Time to Proficiency Based on Biological Sex Assigned at Birth. As reported in Table 4.5, the mean number of flight hours for female participants prior to becoming check-ride proficient was M_{Female} = 63.2 hours (SD = 14.7), the median was *Mdn* = 59.4 hours, and the range was from 44.1 hours to 99 hours. For male participants, the mean was M_{Male} = 61.4 hours (*SD* = 15.9), the median was *Mdn* = 59.4 hours, and the range was from 35 hours to 116.8 hours. Thus, female participants required, on average, at most 2 flight hours more than male participants to become check-ride proficient.

Flight Time to Proficiency Based on Race. Focusing on race, as reported in Table 4.5, the mean flight time to proficiency for White participants was M_{White} = 61 hours (*SD* = 15.4), the median was *Mdn* = 58.7 hours, and the range was from 35 hours to 116.8 hours. For Nonwhite participants, the mean flight time to proficiency was M_{Nonwhite} = 64.3 hours (*SD* = 17.4), the median was *Mdn* = 64.7 hours, and the range was from 40 hours to 100 hours. Thus, Nonwhite participants required, on average, approximately 3 more flight hours than White participants to become check-ride proficient.

Flight Time to Proficiency Based on Marital Status. Focusing on marital status, as reported in Table 4.5, the mean flight time to proficiency for Married participants was M_{Married} = 61.8 hours (*SD* = 15.8), the median was *Mdn* = 60 hours, and the range was from 40 hours to 100 hours. For Not Married participants, the mean flight time to proficiency was $M_{\text{Not Married}}$ = 61.5 hours (*SD* = 15.8), the median was *Mdn* = 59 hours, and the range was from 35 hours to 116.8 hours. Thus, on average, there was little difference in the number of flight hours needed to become check-ride proficient between Married and Not Married participants.

Inferential Statistics

Overview

The overall goal of the current study was to determine the relationship between a set of targeted factors and the number of flight hours students need to complete before their CFI considers them proficient to take the private pilot check ride. Following Cohen et al.'s (2003) guidance, the targeted variables were partitioned into functional sets as follows:

- Set A = Training Time to Proficiency consisted of one variable: X_1 = Training time to proficiency (in days).
- Set B = Flight School Characteristics consisted of three factors: X_2 = Flight training program, which was defined as Part 141 vs. Part 61; X_3 = Flight training organization, which was defined as 2- or 4- year college/university vs. fixed-base operator (FBO) flight school; and X_4 = Scheduling policy, which was defined as mandated vs. student-driven.
- Set C = Flight Student Demographics consisted of four variables: X_5 = Biological sex assigned at birth, which was defined as female vs. male; X_6 = Age; X_7 = Race, which initially examined four races as defined by the 2020 U.S. Census but was delimited to White vs. Nonwhite; X_8 = Marital status, which was defined as Married vs. Not Married; and X_9 = Ethnicity, which was defined by the 2020 U.S. Census as "Hispanic or Latino" vs. "Not Hispanic or Latino," but was declared a constant.

The primary objective was to perform a hierarchical multiple regression analysis using the set entry order A–B–C to determine the cumulative and unique contributions each set made in explaining the variance in the dependent variable, which was Y = Flight time to proficiency (in hours).

Preliminary Analyses

Prior to conducting the hierarchical multiple regression analysis, I performed several preliminary analyses. These included: (a) data set modifications, (b) missing data analysis, (c) outlier analysis, (d) multicollinearity analysis, and (e) an analysis to verify that the data set was compliant with the assumptions of regression. A summary of the results of these analyses follows.

Data Set Modifications. Several modifications to the data set needed to be performed to make the data set clean for descriptive and inferential analyses. A brief summary follows:

- I included a case number column to maintain numerical order relative to the raw data for subsequent manipulations.
- I changed the data types of the nominal variables to continuous variables where appropriate.
- I dummy coded the categorical variables of Set B (X_2, X_3, X_4) and Set C (X_5, X_7, X_8).
- I used an online application (Date Calculator, 2022) to determine the number of days relative to participants' responses to Items B1 and B2 of

the questionnaire. In B1, participants entered the date of the first flight they took toward their PPL, and in B2, participants entered the date when their CFI signed their logbook indicating they were check-ride proficient and ready to take the private pilot check ride. I then entered the number of days into the data set for X_1 = Training time to proficiency.

- I deleted all superfluous data reported by *Qualtrics*, including response ID, IP address, time stamp, device data, sequence number, external references, respondent email, and email list columns.
- I reconciled all "Other" responses based on the information participants provided in the corresponding textboxes. This information consisted of descriptions of participants' flight training program, organization, or scheduling policy. For example, with respect to flight training, 10 participants described their training as opposed to specifying the type of program. Based on their descriptions, I was able to identify whether their training program was Part 61 or Part 141. I performed a similar analysis for type of organization and scheduling policy.

At this juncture, there were $N = 278$ cases.

Missing Data. Missing data occurs when participants forget or opt not to respond to an item, if they responded to an item incorrectly, or if they responded to an item unclearly (Cohen et al. 2003). Methods to resolve missing data issues depend on whether the variable is a DV or IV, and whether the IVs are nominal or continuous.

With respect to missing data on the DV, where Y = Flight time to proficiency (in hours), there were: (a) 99 cases where participants did not respond to this item; (b) seven cases where participants entered a flight time that was less than 35 hours, which is not legally possible; (c) two cases where participants entered a flight time that was not consistent with their documented flight training program (Part 61 or Part 141); and (d) three cases where participants' training did not qualify for the study: two obtained their PPL through the military, and one participant did not obtain his PPL in the United States. Following Cohen et al.'s (2003) and Allison's (2009, p. 84) guidance, I deleted these 111 cases, which reduced the sample size to $N = 167$.

In addition to missing data on the DV, several IVs also had missing data. Unlike the former, though, there are various ways to address the latter other than deleting the cases with missing data. In the current study, data were missing from three IVs: X_1 = Training time to proficiency (in days) had 27 cases (16% of the sample) where participants did not report this information, X_4 = Scheduling policy had one case where the participant indicated that the policy was both mandated and student-driven, and X_6 = Age had 10 cases (6% of the sample) where participants did not report their age. Following Cohen et al.'s (2003) guidance for handling missing data on an IV, I first confirmed that the data were missing randomly for all cases relative to the three IVs. I then plugged the missing data for X_1 with the median because X_1 was not normally distributed (Acock, 1997; El-Masri & Fox-Wasylyshyn, 2005, p. 166), and I plugged the missing data for X_4 and X_6 with their respective means (Cohen et al. 2003). Thus, the sample size remained at $N = 167$.

Outlier Analysis. Outliers are extreme observations that lie an abnormal distance relative to the data points in a sample. Cohen et al. (2003) classifies outliers as either rare cases or contaminants. Rare cases are true values but extreme. For example, a 13-year-old student pilot who is studying to become a hot air balloon or glider pilot might be considered an outlier among the population of all student pilots' ages regardless of aircraft type, but this age is perfectly acceptable. Similarly, an air transport pilot (ATP) with 40,000 flight hours might seem extreme among the majority of ATPs but is reasonable if the pilot has flown mostly international routes. Contaminated cases, however, involve incorrect/inaccurate data that could result from a date entry error. For example, a CFI who reports 2,000 hours dual given in the past 90 days might have inadvertently entered an extra zero.

I conducted an outlier analysis using two different approaches. I first conducted a visual examination of the data to see if anything unusual was reported and identified what I presumed to be were three contaminated cases: Case Number 16 had a flight time to check-ride proficiency of 266.9 hours in 67 days, which I considered unrealistic; Case Number 166 had a flight time to check-ride proficiency of 220 hours in 8,824 days (≈ 24 years), which I theorized was not appropriate for someone pursuing a PPL; and Case Number 241 included a mandated scheduling policy but required 3,788 days (≈ 10 years) to become check-rider proficient, which is unreasonable. As a result, I deleted these three cases, which reduced the data set to $N = 164$.

I then performed an outlier analysis using Jackknife distances, which flagged 15 cases. After reviewing these cases, I concluded they were rare cases and not contaminants. To determine the effect of these outliers, I ran two separate simultaneous regression analyses—one in the presence and one in the absence of the outliers. With outliers present, $R^2 = .09$, $R^2_{\text{adjusted}} = .04$, $F(8, 155) = 1.93$, $p = .0587$., $RMSE = 15.44$, $RMSE_{Press} = 16.47$, $R^2_{\text{Cross Validated}} = 0$, and the significant IVs were X_1 = Training time to proficiency and X_4 = Scheduling policy. With outliers absent, $R^2 = .14$, $R^2_{\text{Adjusted}} = .09$, $F(8, 140) = 2.86$, $p = .0056$. $RMSE = 14.19$, $RMSE_{Press} = 14.70$, $R^2_{\text{Cross Validated}} = .04$, and the significant IVs were X_1 = Training time to proficiency and X_4 = Scheduling policy. Thus, it appeared that the outliers were masking significance.

Because the cross-validated R^2 with outliers present was 0 (and it was negative depending on the number of decimal places the calculation was carried out), I suspected an incorrect model fit. After additional examinations of the data set—which included a Johnson Su transformation and a Box-Cox transformation (Frost, 2019, p. 240)—I discovered there was a logarithmic relationship between the DV and X_1 = Training time to proficiency. As a result, I transformed X_1 using the natural logarithm and then once again performed two simultaneous regression analyses to determine the effect of the outliers, but this time I used the transformed X_1. With outliers present, $R^2 = .22$, $R^2_{\text{adjusted}} = .18$, $F(8, 155) = 5.55$, $p < .0001$., $RMSE = 14.27$, $RMSE_{\text{Press}} = 14.77$, $R^2_{\text{Cross Validated}} = .14$, and the significant IVs were X_1 = Training time to proficiency (transformed to the natural logarithm) and X_4 = Scheduling policy. With outliers absent, $R^2 = .19$, $R^2_{\text{Adjusted}} = .14$, $F(8, 140) = 4.04$, $p = .0002$. $RMSE = 13.80$, $RMSE_{\text{Press}} = 14.21$, $R^2_{\text{Cross Validated}} = .095$, and the significant IVs were again X_1 = Training time to proficiency (transformed to the natural logarithm) and X_4 = Scheduling policy. Based on the results of the four outlier analyses, I decided that the logarithmic version of X_1 with outliers present was the best model not only statistically but also because it was more representative of the target population. Thus, at this stage, the sample size remained at $N = 164$, but X_1 was now transformed using the natural logarithm (*ln*) function.

Multicollinearity Analysis. According to Cohen et al. (2003), a predictor's variable inflation factors (*VIF*) "provides an index of the amount that the variance of each regression coefficient is increased relative to a situation in which all the IVs are uncorrelated" (pp. 421–422). To determine this amount of increased variance, the square root of the *VIF* is examined. For example, if an IV's $VIF = 9$, then this indicates that the corresponding standard error would be 3 times as high than it would be if the IV was not correlated with any of the other IVs in the model.

There is considerable disagreement among statisticians regarding the acceptable size of *VIF*s. For example, Cohen et al. (2003) indicated a $VIF > 10$ suggests there is a multicollinearity problem, Keith's (2015, p. 202) threshold is $VIF > 6$, and Allison's (1999, pp. 141-142) threshold is $VIF > 2.5$. For the current study, the *VIF*s among the eight IVs varied between $VIF = 1.02$ for X_7 = Race and $VIF = 2.18$ for X_3 = Scheduling policy. Thus, multicollinearity was not an issue.

Regression Assumptions. According to Cohen et al. (2003), six assumptions must be satisfied before the data can be evaluated by using multiple regression techniques to determine relationships between the targeted IVs and the DV. The six assumptions are as follows: (a) multivariate linearity, which confirms that the form of the relationship between the multiple IVs and DV is linear; (b) correct specification of the IVs, which determines if the targeted IVs are appropriate for the context of a study; (c) perfect reliability, which examines the reliability coefficients of those IVs representing a psychological construct; (d) homoscedasticity of the residuals, which examines the extent to which the variances of the residuals around the regression line are constant for any value of the IVs; (e) independence of residuals, which confirms that the residuals of the observations are independent of one another; and (f) normality of residuals, which confirms that the residuals in the population are normally distributed. A brief discussion of the procedures I used to examine the data set for compliance to these assumptions follows.

Multivariate Linearity. To check for multivariate linearity, I conducted two separate residual analyses. The first involved a bivariate plot of the residuals (y-axis) vs. the predicted values (x-axis) from the results of a simultaneous multiple regression analysis where the DV was regressed against the IVs, which included the natural logarithm-based transformed version of X_1. I then included the 0-line and overlayed a Kernel smoother line to see if it followed the trend of the data. A smoothness alpha less than .80 confirmed that this was indeed the case. The second analysis involved a bivariate plot of the DV (y-axis) vs. the residuals (x-axis), which produced a strong linear pattern. Therefore, based on the results of these two plots, I concluded that the data set was compliant with the linearity assumption.

Correct Specification of the IVs. One of the a priori challenges to any study is to determine the appropriate IVs to target. For the current study, I relied on my 30 years of experience in the aviation industry, my experience as a CFI, the published literature, advice from my committee members, and theory. To confirm that the IVs I targeted were indeed correctly specified with respect to their relationship with the DV (Y = Flight time to proficiency in hours), I examined the leverage plots of each IV separately. These plots depicted the relationship between the residuals of the DV and the residuals of the respective IVs. The DV residuals represented that part of Y that was not associated with all IVs except the one under discussion, and the respective IV residuals represented that part of the IV under discussion that was not associated with the other IVs. Thus, both Y and the IV under discussion were freed of any relationship with all of the other IVs. These leverage plots flagged five incorrectly specified factors: X_2 = Flight training program (Part 61 vs Part 141), X_5 = Biological sex assigned at birth (Female vs. Male), X_6 = Age, X_7 = Race (White vs. Nonwhite), and X_8 = Marital status (Married vs. Not Married). Therefore, I removed these five IVs from the final model.

Perfect Reliability. This assumption is applicable to studies that use various instruments to measure psychological constructs such as attitudes, motivation, and self-efficacy. The key here is that these instruments must be reliable with little to no measurement error, which could lead to bias in the estimates of the regression coefficients and their standard errors as well as incorrect significance tests and confidence intervals (Cohen et al., 2003). For the current study, no such instruments were used, and no cognitively loaded questions were included on the questionnaire. Therefore, this assumption was not applicable.

Homoscedasticity of Residuals. To verify the equal variances assumption, I referred to the residual analysis involving the bivariate plot of residuals vs. predicted used for the multivariate linearity assumption. As noted previously, this assumption of multivariate linearity was met, and therefore the dataset was compliant with the homoscedasticity of residuals assumption.

Independence of Residuals. To verify this assumption, I examined the bivariate plot of the residuals vs. the case numbers and included the corresponding 0 line. When I overlayed the Kernel smoother line, it converged to the 0 line with a smoothness alpha of less than .60. Thus, the data set was compliant with this assumption.

Normality of the Residuals. To test this assumption, I examined the results of the Shapiro-Wilk Goodness of Fit test to determine the extent to which the residuals fitted a normal distribution. This test yielded a p value of $p = .0003$, which indicated that the data were not from a normal distribution. However, because the sample size of the current study ($N = 164$) was sufficiently large, the central limit theory was applied and therefore this assumption was satisfied.

Summary of Preliminary Analyses. Based on the preliminary data screening presented in this section, the initial data set was modified relative to sample size and number of variables. The initial sample size was $N = 278$ but was reduced to $N = 164$ because of (a) missing data on the DV, (b) spurious data entries or cases involving respondents who did not meet the current study's selection criteria, and (c) the deletion of outliers. The initial data set also consisted of eight IVs partitioned into three functional sets as presented earlier in the Overview. Five variables, however, were found to be incorrectly specified based on their respective leverage plots and therefore deleted. These included one IV in Set B and all four IVs in Set C. As a result, the final data set that was used to test the study's hypotheses consisted of three variables partitioned into two sets: X_1 = Training time to proficiency in days (Set A), and X_3 = Flight training organization and X_4 = Scheduling policy (Set B).

Primary Analysis

The statistical strategy used to test the study's hypotheses was hierarchical multiple regression in which Y = Flight time to proficiency in hours was regressed on two sets of independent variables using the set entry order A–B. As noted in the foregoing paragraph, Set A = Training time to proficiency consisted of the single variable X_1 = Training time to proficiency in days, which was transformed using the natural logarithm function, and Set B = Flight school characteristics consisted of two variables: X_3 = Flight training organization (2- or 4-year college/university vs. FBO), and X_4 = Scheduling policy (Student-driven vs. Mandated). A discussion of the results of this analysis follows, and a summary of these results is provided in Table 4.6.

Set A: Training Time to Proficiency. When the single variable of Set A entered the model, the contribution it made in explaining the variance in the DV was significant, $R^2 = .14$, $F(1, 162) = 26.98$, $p < .0001$, $R^2_{\text{Adj}} = .14$, $R^2_{\text{Press}} = .12$, $RMSE = 14.66$, $RMSE_{\text{Press}} = 14.77$, the effect size was $f^2 = 0.16$, and power = .9993. Thus, $\ln(X_1)$ = Training time to proficiency in days explained 14% of the variance in the number of flight hours needed to become check-ride proficient.

Set B: Flight School Characteristics. When the two variables of Set B entered the model in the presence of X_1, the overall model involving the three IVs was significant, $R^2 = .19$, $F(3, 160) = 12.49$, $p < .0001$, $R^2_{\text{Adj}} = .17$, $R^2_{\text{Press}} = .15$, $RMSE = 14.34$, $RMSE_{\text{Press}} = 14.66$, the effect size was $f^2 = 0.23$, and power = .9998.

Table 4.6

Hierarchical Regression Results for Y = Flight Time (Hours) to Check-Ride Proficiency

Variable [a]	B_i	95% CI [b] *LL*	95% CI [b] *UL*	*SE* B_i	b_i	R^2	DR^2
Step 1						.14***	.14***
Constant	24.14	9.73	38.55	7.29			
X_1	6.71	4.16	9.26	1.29	.38***		
Step 2						.19***	.05**
Constant	20.71	6.19	35.23	7.35			
X_1	7.23	4.71	9.76	1.28	.41***		
X_3	- 6.03	- 11.80	- 0.26	2.92	- .18*		
X_4	9.56	3.29	15.82	3.17	.26**		

Note. $N = 164$.
[a] X_1 = Training time (days) to check-ride proficiency. X_3 = Flight training organization (2- or 4-year college/university vs. FBO flight school). X_4 = Scheduling policy (Mandated vs. Student-driven). [b] CI = Confidence interval; *LL* = lower limit; *UL* = upper limit.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Thus, collectively, the three variables of $\ln(X_1)$ = Training time to proficiency in days, X_3 = Flight training organization, and X_4 = Scheduling policy explained 19% of the variance in the number of flight hours needed to become check-ride proficient.

Furthermore, the unique contribution of X_3 and X_4 from Set B also was significant, $sR^2 = .05$, $F(2, 160) = 4.94$, $p = .0083$, the effect size was $f^2 = 0.06$, and power = .80. Thus, the collective influence of these two variables uniquely accounted for an additional 5% of the variance in the number of flight hours needed to become check-ride proficient.

Final model. The final regression model was

$$Y' = B_1[\ln(X_1)] - B_3X_3 + B_4X_4 + 20.7$$

$$= 7.2[\ln(X_1)] - 6.0X_3 + 9.6X_4 + 20.7$$

As noted earlier, this model was significant, $R^2 = .19$, $F(3, 160) = 12.49$, $p < .0001$. Given a significant omnibus, an interpretation of the corresponding regression coefficients was warranted.

***B_1* = 7.2**. The regression coefficient for $\ln(X_1)$ was significant, $B_1 = 7.2$, $t(160) = 5.66$, $p < .0001$, and 95% CI = [4.7, 9.8]. The reader is reminded that X_1 was transformed using the natural logarithm function and therefore its interpretation must be reflective of this transformation. According to Ford (2018), the interpretation of a regression coefficient that was based on a natural logarithm transformation involves multiplying B by the natural log of 2, which is $\ln(2) \approx 0.69$. For the current study, $B_1 \times \ln(2) = (7.2)(0.69) = 4.968$. Thus, holding all other variables in the model constant, for every 100% increase in training days to proficiency, the flight time to proficiency increases by approximately 5 hours. For example, by increasing the number of training days to proficiency from 15 to 30 (a 100% increase), the flight time to proficiency increases by 5 hours. Similarly, by increasing the number of training days to proficiency from 30 to 60 (a 100% increase), the flight time to proficiency also increase by 5.0 hours.

***B_3* = –6.0**. The regression coefficient for X_3 = Organization (2- or 4-year college/university vs. FBO) was significant, $B_3 = -6.0$, $t(160) = -2.06$, $p = .0407$, and 95% CI = [–11.8, –0.3]. Thus, holding all other variables in the model constant, students who attended a 2- or 4-year college/university averaged 6 fewer hours to become check-ride proficient than students who attended an FBO flight school for flight training.

B_4 = 9.6. The regression coefficient for X_4 = Scheduling policy (Mandated vs. Student-driven) was significant, $B_4 = 9.6$, $t(160) = 3.01$, $p = .0030$, and 95% CI = [3.3, 15.8]. Thus, holding all other variables in the model constant, students who operated under a mandated flight training schedule averaged 9.6 more hours to become check-ride proficient than students who operated under a student-driven flight training schedule.

B_0 = 20.7. The regression constant, which is the direct result of entering 0 into the model for B_1, B_2, and B_3, was significant, $B_0 = 20.7$, $t(160) = 2.82$, $p = .0055$, and 95% CI = [6.2, 35.2]. Thus, flight students with 0 training days, who attend an FBO to receive their flight instruction, and who operate under a student-driven flight schedule will require, on average, approximately 21 hours to become check-ride proficient. Although, not realistic, this coefficient is nevertheless interpretable in the context of the current study.

Applying the Final Model: An Example. To apply the final regression model for a specific case, consider the following situation: A flight student is enrolled in a Part 61 program at an FBO with a student driven schedule and would like to know approximately how many hours of flight training are needed to become check-ride proficient for 30 days of flight training. In this example, $X_1 = 30$ days, $X_3 = 0$, and $X_4 = 0$. Therefore, the predicted number of flight hours is:

$$Y' = 7.2[\ln(X_1)] - 6.0X_3 + 9.6X_4 + 20.7$$

$$= 7.2[\ln(30)] - 6.0(0) + 9.6(0) + 20.7$$

$$= 7.2(3.40) + 0 + 0 + 20.7$$

$$= 24.48 + 20.7$$

$$= 45.18$$

The reader will note that if X_1 = 60 days, which is a 100% increase from 30 days, then this corresponds to an additional 5 hours of flight time that will be needed. This can be confirmed by substituting 60 for X_1 in the prediction equation.

Results of Hypothesis Testing

The current study's research hypotheses were set forth in Chapter 1. For testing purposes, these research hypotheses are restated here in null form. The decision to reject or fail to reject a null hypothesis was based on the respective primary analyses results reported above. The null hypotheses and a discussion of the decisions made with respect to each are provided below.

Null Hypothesis 1. There Will Be No Significant Curvilinear Relationship between Training Time to Proficiency and Flight Time to Proficiency.

As reported in Table 4.6, there was a significant predictive gain at Step 1 of the hierarchical regression model. When X_1 = Training time to proficiency, which represented the natural logarithmic transformation of this variable, the model was significant, $R^2 = .14$, $F(1, 162) = 26.98$, $p < .0001$. Because this variable was transformed from a linear relationship to a curvilinear relationship, Null Hypothesis

1 was rejected: There was a significant curvilinear (logarithmic) relationship between the number of flight training days needed to become check-ride proficient and the number of flight hours needed to become check-ride proficient.

Null Hypothesis 2a. There Will Be No Significant Difference in the Number of Hours of Flight Time Needed to Become Check-Ride Proficient between Flight Students Who Receive Flight Training via a Part 141 Program vs. Flight Students Who Receive Flight Training via a Part 61 Program.

As previously reported in the Regression Assumptions section, X_2 = Flight training program (Part 61 vs. Part 141) was deleted from the final data set because it was not correctly specified (i.e., it had no relationship with the DV). As a result, Null Hypothesis 2a was not rejected: There is no significant difference in flight time to proficiency between flight students trained under a Part 141 flight program versus those trained under a Part 61 flight program.

Null Hypothesis 2b. There Will Be No Significant Difference in the Number of Hours of Flight Time Needed to Become Check-Ride Proficient between Flight Students Who Receive Flight Training at a 2- or 4-Year College/University vs. Flight Students Who Receive Flight Training at an FBO.

As reported in Table 4.6, there was a significant predictive gain at Step 2 of the hierarchical regression model. When X_3 = Organization and X_4 = Scheduling policy entered the model together in the presence of X_1 = Training time to proficiency, the increment was significant, $sR^2 = .05$, $F(2, 160) = 4.94$, $p = .0083$.

Furthermore, when X_3 was examined individually in the final model and in the presence of X_1 and X_4, it too was significant, $B_3 = -6.03$, $t(160) = -2.06$, $p = .0407$. Because 2- or 4-year college/university was coded 1, this result indicates that students who received flight training from 2- or 4-year college/university required, on average, 6 *fewer* hours than students who received flight training at an FBO. As a result, Null Hypothesis 2b was rejected. The difference in flight time to proficiency between 2- or 4-year colleges/universities vs. FBOs was significant with flight students at the former organization requiring significantly fewer hours to become check-ride proficient.

Null Hypothesis 2c. There Will Be No Significant Difference in the Number of Hours of Flight Time Needed to Become Check-Ride Proficient between Flight Students Who Receive Flight Training Under a Mandated Schedule vs. Flight Students Who Receive Flight Training Under a Student-Driven Schedule.

As reported in Table 4.6, there was a significant predictive gain at Step 2 of the hierarchical regression model. When X_3 = Organization and X_4 = Scheduling policy entered the model together in the presence of X_1 = Training time to proficiency, the increment was significant, $sR^2 = .05$, $F(2, 160) = 4.94$, $p = .0083$. Furthermore, when X_4 was examined individually in the final model and in the presence of X_1 and X_3, it too was significant, $B_4 = 9.56$, $t(160) = 3.01$, $p = .0030$. Because a mandated scheduling policy was coded 1, this result indicates that students on a mandated schedule required on average 9.56 *more* hours to check-ride proficiency than students under a student-driven schedule. The reader will note that this finding appears to warrant a decision to reject the corresponding null hypothesis. However, the reader is reminded that the corresponding alternative hypothesis as expressed relative to Research Hypothesis 2c in Chapter 1, stated that students under a mandated schedule would require *fewer* hours to check-ride proficiency than students under a student-driven schedule. Thus, the result is in the opposite direction relative to the alternative hypothesis and hence Null Hypothesis 2c was not rejected: There is no significant difference in flight time to proficiency between mandated vs. student-driven scheduling.

Null Hypothesis 3. There Will Be No Significant Relationship between Flight Students' Demographics and Flight Time to Check-Ride Proficiency.

As previously reported in the Regression Assumptions section, all the IVs in Set C = Flight Student Demographics were not included in the final data set because they were not correctly specified (i.e., they had no relationship with the DV) or were treated as a constant. As a result, Null Hypothesis 3 was not rejected: Flight students' demographics of X_5 = Biological sex assigned at birth (Female vs. Male), X_6 = Age, X_7 = Race (White vs. Nonwhite), X_8 = Marital status (Married vs. Not Married), and X_9 = Ethnicity ("Hispanic or Latino" vs. "Not Hispanic or Latino") had no significant relationship with flight time (in hours) to check-ride proficiency.

Chapter 5

Conclusions, Implications, and Recommendations

Summary of Study

The purpose of the current study was to examine the relationship between a targeted set of factors and the number of flight hours needed to become proficient to take the private pilot check ride. The ultimate objective was to develop a model that flight students could use to determine the minimum number of hours needed to become check-ride proficient based on key flight and/or demographic-related factors. The study and model derived from data analysis were grounded in Ebbinghaus's (1885/1913/2013) forgetting curve theory and spacing effect, and Ausubel's (1963) theory of meaningful learning.

Nine independent variables were targeted initially and grouped into three functional sets. Set A = Time Interval consisted of a single variable, X_1 = Training time to proficiency, which represented the time in days from a flight student's first flight to when a CFI determined the student was check-ride ready. Set B = Flight School Characteristics consisted of three categorical variables, X_2 = Flight training program (Part 141 vs. Part 61), X_3 = Organization offering flight training program (2- or 4-year college/university vs. fixed-base operator [FBO]), and X_4 = Scheduling policy (mandated vs. student-driven). Set C = Flight Student Demographics consisted of five variables, X_5 = Biological sex assigned at birth (female vs. male); X_6 = Age of participants (in years); X_7 = Race, which represented a comparison among four races as defined by the 2020 U.S. Census—White, Black or African American, Asian, and Hawaiian or Pacific Islander; X_8 = Marital status (married vs. not married); and X_9 = Ethnicity, which represented a comparison between "Hispanic or Latino" vs. "Not Hispanic or Latino." The DV was flight time to proficiency, defined as the total time in hours flight students accrued prior to being declared check-ride proficient by their CFI.

The research methodology/design that best fit the current study relative to answering its research questions was a combination of predictive correlational (RQ 1 and parts of RQ 2) and ex post facto (parts of RQ 2 and RQ 3). With respect to the former, a correlational method with a predictive design was appropriate because it is used to examine relationships and patterns among variables in a single group (Ary et al., 2010). In the current study, RQ 1 and parts of RQ 2 involved a single group (flight students) with multiple measures, and the objective was to assess the extent to which these factors could be used to predict the DV. With respect to the latter, an ex post facto effects-type design is appropriate for studies involving pre-existing groups, in the absence of any intervention, and where the grouping is on the independent variable, which was the case for parts of RQ 2 and RQ 3. In RQ 2, group membership involved gender, race, marital status, and ethnicity, and in RQ 3, group membership involved the type of flight school, type of organization, and scheduling policy. The formation of these groups was based on participants' self-reported responses to the corresponding items on the questionnaire.

The target and accessible populations were all flight students pursuing their private pilot's license (PPL) in the United States and who received flight instruction from an FAA certified flight instructor. Because I solicited pilots employed by United Airlines, a supermajority of the sample most likely contained pilots employed by United Airlines. (*Note:* The reader is reminded from Chapter 1's Definitions section that pilots provided the same data as flight students except pilot data were obtained from their student pilot logbook created while they were enrolled as flight students in a flight training program leading to a PPL. Thus, in this context, pilots were considered to be part of the general term "flight students.") The initial sample size consisted of n = 278 participants. After preliminary data screening in advance of conducting inferential statistical analyses, the final sample size used to test the study's hypothesis was n = 164.

The primary data collection instrument was a researcher-developed questionnaire that consisted of items designed to capture participants' self-reported factual data related to the targeted variables specified earlier. The reader will note that the questionnaire did not measure any psychological constructs such as attitudes, motivation, or self-efficacy, and therefore the concepts of instrumentation validity and reliability were not applicable. Nevertheless, I gave attention to face and content validity by asking my advisor and a group of CFIs to review the items relative to their format, structure, and grammar (face validity), and to ensure that they were useful and relevant with respect to the research questions (content validity). I also performed a preliminary study by asking 10 pilots to complete the questionnaire to verify that the items were clear and understandable. Furthermore, because the collected data were fact-based, I presumed that response-reliability would be high. For example, it is unlikely that a participant would self-report 100 hours for X_1 = Training time to proficiency on one administration of the questionnaire, and then report something completely different (e.g., 150 hours) on a subsequent administration. A copy of the questionnaire is provided in the appendix.

Summary of Findings

To collect the sample data, I contacted flight school representatives and CFIs by telephone, in person, email, and text, and asked them to disseminate a flyer to their flight students. This flyer contained information about the study, a link to the questionnaire, and my contact information. I also personally solicited individual flight students and pilots and requested that they complete the online questionnaire.

At the conclusion of data collection, I performed several preliminary analyses for data screening purposes. These included: (a) data set modifications; (b) missing data analysis; (c) outlier analysis; (d) multicollinearity analysis; and (e) an analysis to verify that the data set was compliant with the assumptions of regression, which was the primary statistical strategy. The results of these analyses had the following effect on the final data set (see Chapter 4 for specific details):

- The final sample size was reduced from n = 258 to n = 164.

- The number of IVs was reduced from nine to three: X_1 = Training time to proficiency, X_3 = Organization offering flight training program (2- or 4-year college/university vs. FBO), and X_4 = Scheduling policy (mandated vs. student-driven).
- X_1 was transformed using the natural logarithm function (ln) because its relationship with the DV was logarithmic in nature.
- Set C = Flight Student Demographics, which contained five demographical variables, was eliminated.

After completing preliminary analyses, I then tested the study's hypotheses by performing a hierarchical multiple regression analysis with the set entry order of Set A, which consisted of the transformed version of X_1, followed by Set B, which consisted of X_3 and X_4. A brief summary of the findings follows, and a summary of the results of hypothesis testing is provided in Table 5.1.

As reported in Table 4.6 (Chapter 4), significant relationships at the preset alpha level of $\alpha = .05$ were found at each step of the analysis. At Step 1, Set A, which consisted of the single variable $\ln(X_1)$ = Training time to proficiency, was significant, $R^2 = .14$. Thus, 14% of the variance in the number of flight hours needed to become check-ride proficient was explained by the number of training days needed to become check-ride proficient. At Step 2, Set B, which consisted of X_3 = Flight training organization and X_4 = Scheduling policy, entered the model in the presence of $\ln(X_1)$. The overall result was significant, $R^2 = .19$, and thus these

Table 5.1
Summary of Results of Hypothesis Testing

Null Hypothesis		Decision
H_1:	There will be no significant curvilinear relationship between training time to proficiency and flight time to proficiency.	Reject
H_{2a}:	There will be no significant difference in the number of hours of flight time needed to become check-ride proficient between flight students who receive flight training via a Part 141 program vs. flight students who receive flight training via a Part 61 program.	Fail to Reject
H_{2b}:	There will be no significant difference in the number of hours of flight time needed to become check-ride proficient between flight students who receive flight training at a 2- or 4-year college/university vs. flight students who receive flight training at an FBO.	Reject
H_{2c}:	There will be no significant difference in the number of hours of flight time needed to become check-ride proficient between flight students who receive flight training under a mandated schedule vs. flight students who receive flight training under a student-driven schedule.	Reject
H_3:	There will be no significant relationship between flight students' demographics and flight time to proficiency.	Fail to Reject

Note. N = 164.

three IVs collectively explained 19% of the variance in the number of flight hours needed to become check-ride proficient. Furthermore, the unique contribution of Set B also was significant, $sR^2 = .05$, which means that X_3 and X_4 uniquely accounted for an additional 5% of the variance in the number of flight hours needed to become check-ride proficient. The final regression model was

$$Y' = B_1[\ln(X_1)] - B_3X_3 + B_4X_4 + 20.7$$

(See Chapter 4 for an example on how to apply this prediction model in practice.) Because the omnibus for both stages was significant, an independent examination of the regression coefficients was possible.

B_1 = 7.2. The regression coefficient for $\ln(X_1)$ was significant at the preset alpha level of $\alpha = .05$. Because X_1 was transformed using the natural logarithm function, its interpretation must be reflective of this transformation. According to Ford (2018), the interpretation of a regression coefficient that was based on a natural logarithm transformation involves multiplying B by the natural log of 2, which is $\ln(2) \approx 0.69$. For the current study, $B_1 \times \ln(2) = (7.2)(0.69) = 4.968$. Thus, holding all other variables in the model constant, for every 100% increase in training days to proficiency, flight time to proficiency increases by about 5 hours.

B_3 = – 6.0. The regression coefficient for X_3 = Organization (2- or 4-year college/university vs. FBO) was significant, at the preset alpha level of $\alpha = .05$. Thus, holding all other variables in the model constant, students who attended a 2- or 4-year college/university averaged 6 fewer hours to become check-ride proficient than students who attended an FBO flight school for flight training.

B_4 = 9.6. The regression coefficient for X_4 = Scheduling policy (mandated vs. student-driven) was significant at the preset alpha level of $\alpha = .05$. Thus, holding all other variables in the model constant, students who operated under a mandated flight training schedule averaged 9.6 more flight hours to become check-ride proficient than students who operated under a student-driven flight training schedule.

Conclusion and Inferences

This section contains a review of the study's findings with respect to answering the research questions presented in Chapter 1. As part of this presentation, each RQ is examined separately and includes a summary of the findings, an interpretation of the results in the context of the given research setting, and plausible explanations for the results.

Research Question 1. What Is the Relationship between Training Time to Proficiency and Flight Time to Proficiency?

To answer RQ 1, I regressed X_1 = Training time to proficiency (in days) on Y = Flight time to proficiency (in hours). As reported in Step 1 of Table 4.6 in Chapter 4, the result of this analysis was significant, and the corresponding regression coefficient was $B_1 = 6.71$. The reader will recall that X_1 was transformed using the natural logarithm function (ln). According to Ford (2018), to interpret the effect a natural log transformed IV has on the dependent variable: (a) the interpretation is from the perspective of a 100%-unit increase on the transformed IV, and (b) the quantitative effect on the DV is calculated by multiplying the corresponding regression coefficient by the natural log of 2, which is $\ln(2) \approx 0.69$.

When applied to the current study, the effect of X_1 on Y is interpreted as follows: As training time to proficiency increases by 100%, flight time to proficiency increases by $B_1 \times \ln(2) = (6.71)(0.69) = 4.633$ hours. Thus, in the absence of any other variable at Step 1 of the hierarchical regression analysis, flight students will need, on average, approximately 4.6 additional hours of flight time to become check-ride ready for every 100% increase in training time. For example, by increasing the number of training days to proficiency from 30 to 60, flight time to proficiency will increase approximately 4.63 hours.

At Step 2 of this analysis, the reader will note from Table 4.6 that (a) the overall model, which also is the final model, was significant; and (b) $B_1 = 7.23$ but is now in the presence of two other variables (X_3 and X_4). With $B_1 = 7.23$, the corresponding quantitative effect of X_1 on Y is now (7.23)(0.69) = 4.9887, and the interpretation of B_1 in the final model is as follows: Holding all other variables in the model constant, as training time to proficiency increases by 100%, flight time to proficiency increases by 5 hours.

A plausible explanation for this finding is related to the relatively large sample size of $n = 164$. As reported in Table 3.6 (Chapter 3), the corresponding effect size was $ES = f^2 = 0.16$, which according to Cohen et al. (2003) is a medium effect. To find a medium effect for a single IV via a regression analysis would require a sample size of $n = 54$. Because the current study's sample size was 3 times as large, it was not surprising that this factor was significant because of the relationship between sample size and effect size: larger sample sizes are able to detect smaller effect sizes.

A second plausible explanation for this result is related to the application of Ebbinghaus's (1885/1913/2013) forgetting curve as illustrated in Figure 1.1 and the parabolic relationship illustrated in Figure 1.3, which reflects the spacing effect with respect to training days and flight hours. If the number of training days were to double (i.e., a 100% increase), it is reasonable to assume that this increase will lead to flight lessons being spaced too far apart. This in turn, in theory, will result in students forgetting what they learned previously and concomitantly increase the number of flight hours needed to become check-ride proficient.

Research Question 2. What is the Relationship between Flight School Characteristics and Flight Time to Proficiency?

As initially proposed, RQ 2 focused on the effect of Set B = Flight School Characteristics, which had three IVs: X_2 = Flight training program (Part 61 vs. Part 141), X_3 = Flight training organization (2- or 4-year college/university vs. FBO), and X_4 = Scheduling policy (mandated vs. student-driven). A separate discussion of these IVs follows.

X_2 = Flight Training Program. As reported in Chapter 4, after preliminary data screening, X_2 was deleted from the final data set because it was not correctly specified: It had no relationship with the DV, which means there was no significant difference in flight time to proficiency between students trained under a Part 61 flight program vs. those trained under a Part 141 flight program.

A plausible explanation for this result is related to the ultimate objective of both types of flight schools. Although a Part 141 program is required to follow a strict FAA-approved curriculum whereas a Part 61 program is not mandated to follow such a curriculum, both flight programs provide instruction that will lead to a PPL. This infers that both flight programs will have similar syllabi because all students, regardless of program type, must pass the same FAA written exams and perform the same maneuvers for the check ride. Thus, it is reasonable to expect that the number of flight hours needed to become check-ride proficient would not differ that greatly between the two flight programs.

X_3 = Flight Training Organization and X_4 = Scheduling Policy. Because X_2 was eliminated from the final data set, this left Set B with only two variables: X_3 and X_4. As a result, to answer RQ 2, I regressed X_3 and X_4 on Y = Flight time to proficiency (in hours) in the presence of X_1. As reported in Step 2 of Table 4.6 in Chapter 4, the overall result of this analysis was significant, and the corresponding increment involving X_3 and X_4 also was significant. Follow-up analyses revealed that each IV also had a significant effect on Y.

2- or 4-Year College/University vs. FBO (X_3). As reported in Table 4.6, the corresponding regression coefficient for X_3 was $B_3 = –6.03$. Because X_3 was dummy coded with "college" as the reference group, this finding means that students who received flight instruction at a 2- or 4-year college/university required, on average, 6 *fewer* hours to become check-ride proficient than students who received flight instruction at an FBO, and this was statistically significant.

One plausible explanation for this result is related to the business model of the two organizations. Most colleges/universities are not-for-profit, and the cost of flight instruction generally is included with tuition. FBOs, however, generally are for-profit organizations and it is conceivable that by requiring more flight time of their students, this will increase the organization's bottom line. The reader is cautioned that this is not based on any type of concrete or anecdotal evidence, but instead is simply a conjecture as something that could happen.

A second plausible explanation is related to the type of CFIs employed at the two types of organizations. For example, many CFIs at colleges/universities are former flight students who previously pursued their PPL at that college/university. This is a common phenomenon because such students tend to remain at their college/university to seek additional ratings, including a CFI certificate, and therefore have acquired extensive, recent experiences. These students who become CFIs also have much more opportunity to hone their teaching skills because they are flying more often than they would at an FBO. This in turn translates to helping flight students become more proficient in fewer hours than those at FBOs. Flight students also tend to regard these CFIs as peers and are able to quickly develop a rapport, which also could lead to fewer hours needed to become check-ride proficient.

Mandated vs. Student-Driven Scheduling (X_4). The reader will recall that a mandated schedule is imposed by the flight training organization and flight students are required to follow it. For example, Student A's flight instructions are scheduled for MWF from 10 a.m. to 11 a.m. On the other hand, a student-driven schedule is established by students at their convenience. For example, Student A might have some free time on Thursday at 2 p.m. and requests a flight lesson at that time. As reported in Table 4.6, the corresponding regression coefficient for X_4 was B_4 = 9.56. Because X_4 was dummy coded with "mandated" as the reference group, this finding means that students who received flight instruction under a mandated schedule required, on average, 9.56 *more* hours to become check-ride proficient than students who received flight instruction under a student-driven schedule, and this was statistically significant.

A plausible explanation for this result is that under a mandated schedule, it is conceivable that some flight students might not be able to show up for their lessons as scheduled due to unexpected events that emerge (e.g., an exam, car trouble, lack of a babysitter, doctor's appointment, etc.). Given this presumption, it is reasonable to conclude that such "missed appointments" would lead to an increase in the number of hours needed to become check-ride proficient because of the spacing effect concept as reported earlier.

A second plausible explanation is related to the concept of a state of readiness. Under a mandated scheduling policy—which is common for students receiving flight instruction at a college/university—it is conceivable that these students are not mentally prepared for flight instruction because they are overloaded with homework or tests for their other courses. This could lead to a lack of concentration and unpreparedness causing maneuvers to be repeated, which would lead to an increase in the number of flight hours needed to become check-ride ready. Under a student-driven schedule, however, this is not necessarily the case. When students arrive for a flight lesson based on *their* schedule, they are mentally prepared for the lesson and are ready for action: they are situationally aware of what their lesson will involve, and they have prepared themselves for it. In one sense, they are “waiting in the wings” for the lesson. As a result, it is reasonable to conclude that such preparedness would lead to fewer hours needed to become check-ride proficient when compared to a mandated schedule.

Independent of these plausible explanations, this finding appears to be somewhat contradictory to the result associated with X_3 because mandated schedules generally are associated with Part 141 flight training programs, which are found mostly at colleges/universities. The conundrum is if students are requiring, on average, 6 fewer hours to become check-ride proficient at a college/university than at an FBO, then why are they requiring 9.56 more hours, on average, to become check-ride proficient under a mandated schedule, which is usually imposed at a college/university? The only plausible explanation related to this conundrum is related to sample size. For example, there were 3 times as many participants who reported training under a student schedule (n = 123) than a mandated schedule (n = 41). Furthermore, of the 113 participants who reported received flight training at an FBO, 103 were under a student-driven schedule, and of the 51 who reported receiving flight training at a college/university, 20 were under a student schedule. Thus, it is possible than these disparate sample sizes skewed the results.

Research Question 3. What is the Relationship between Flight Students' Demographics and Flight Time to Proficiency?

As initially proposed, RQ 3 focused on the effect of Set C = Flight Student Demographics, which had five IVs: X_5 = Biological sex assigned at birth (Female vs. Male); X_6 = Age; X_7 = Race, which represented a comparison among four races as defined by the 2020 U.S. Census (White, Black or African American, Asian, and Hawaiian or Pacific Islander); X_8 = Marital status (Not married vs. Married); and X_9 = Ethnicity (Hispanic or Latino vs. Not Hispanic or Latino).

As reported in Chapter 4, after preliminary data screening, X_5–X_8 were deleted from the final data set because they were not correctly specified—they had no relationship with the DV—and X_9 was treated as a constant because 90% of the initial sample reported "not Hispanic or Latino." Thus, flight students' demographics as defined by Set C had no significant relationship with flight time to check-ride proficiency. The reader will note that based on my professional aviation experiences, this finding was expected as I indicated in Chapter 1 with respect to the parenthetical comment associated with Hypothesis 3a. The primary purpose for collecting these data was to strengthen the sample description and sample representative discussions.

Implications

This section contains a discussion of the implications of the current study's results and is organized into three parts: (a) implications of the results relative to the study's theoretical grounding, (b) implications of the results relative to the prior research presented in Chapter 2, and (c) implications for aviation practice.

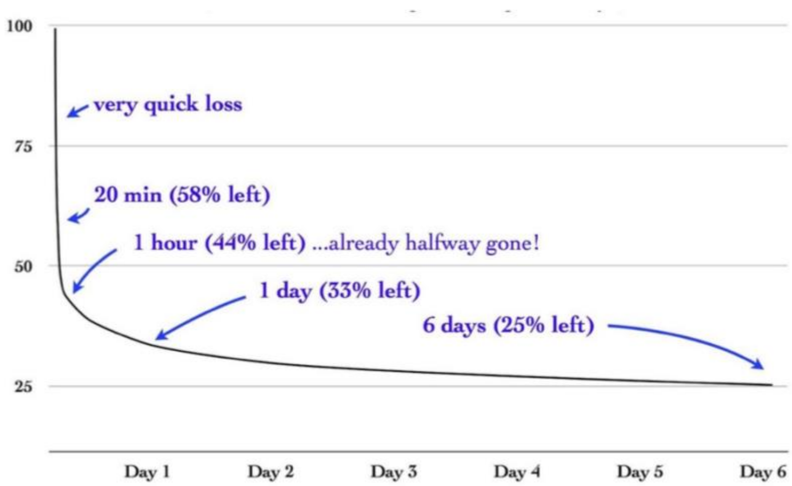
Implications Relative to Theory

Ebbinghaus' Forgetting Curve Theory. The current study was grounded in Ebbinghaus's (1885/1913/2013) forgetting curve theory and spacing effect. With respect to the former, Ebbinghaus theorized that if new information is not reviewed in a timely manner, then it will be difficult to recall at a later date and reported a logarithmic-like relationship between the percentage of lost information and number of days without review or practice. This relationship was illustrated in Figure 1.1 (Chapter 1) but is replicated here as Figure 5.1 for the reader's convenience. As shown in Figure 5.1, the percentage of lost information happens quickly with nearly 70% of newly learned information being lost after 1 day without any type of practice or study.

When applied to the current study, the variable I targeted relative to Ebbinghaus's (1885/1913/2013) forgetting curve theory was X_1 = Training time (in days) to proficiency. As discussed previously, the relationship between X_1 and Y = Flight time (in hours) to check-ride proficiency was logarithmic and X_1 was transformed using the natural logarithm function. The reader also will recall that

Figure 5.1

Replication of Figure 1.1: Graphical Representation of Ebbinghaus's Forgetting Curve



Note. The percentage of information lost (vertical axis) as a function of days if there is no effort made to remember it by practice or study. The curve is logarithmic in form and is asymptotic (levels off) at around 25% of lost information. Adapted from Sonnad (2018).

this relationship was significant (Table 4.6 / Chapter 4). As illustrated in Figure 5.2, when the sample data between these variables are examined graphically, the relationship is indeed nonlinear: As training days to proficiency increases, the number of flight hours to become check-ride proficient also increases but becomes asymptotic at approximately 70 hours. Furthermore, as depicted in Figure 5.3, by inverting the *y*-axis of Figure 5.2, this relationship now apes Ebbinghaus's theoretical forgetting curve of Figure 5.1: The relationship is logarithmic in form with a corresponding asymptote, and similar to Ebbinghaus, the greatest loss occurs relatively early—within the first 100 days of flight training—requiring additional

Figure 5.2

Ebbinghaus's (1885/1913/2013) Forgetting Curve as it Relates to the Current Study

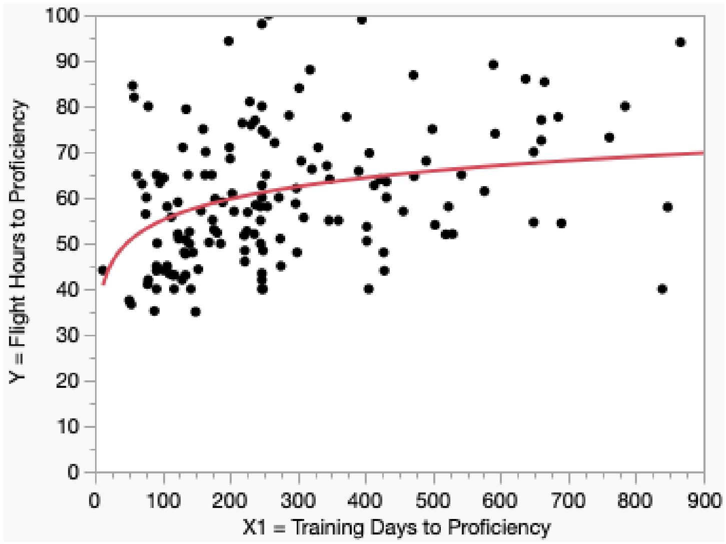
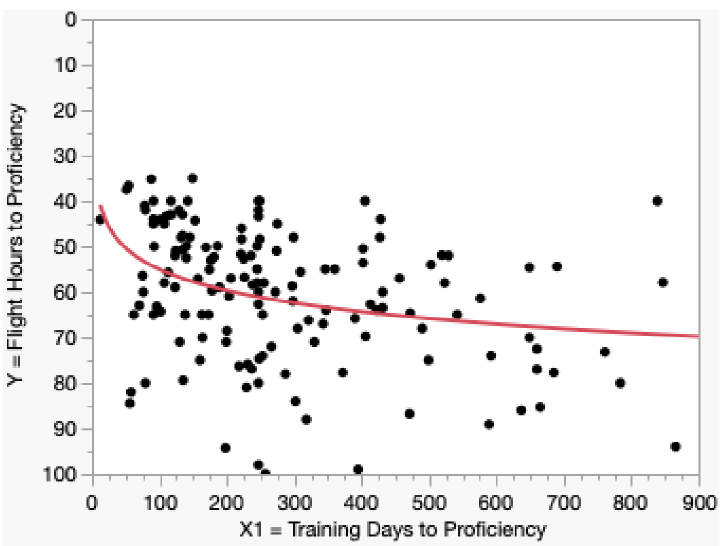


Figure 5.3

Ebbinghaus's (1885/1913/2013) Forgetting Curve with Figure 5.2's Vertical Axis Inverted for Practical Purposes



Note: By inverting Figure 5.2's vertical axis, the results and context of the current study can now be mapped directly to Ebbinghaus's forgetting curve shown in Figure 5.1. There is a loss of knowledge ("forgetfulness") as the number of training days to proficiency increases, which leads to an increase in the number of flight hours needed for check-ride proficiency.

flight hours to become check-ride proficient. Thus, this finding provides support to Ebbinghaus's forgetting curve theory when applied to this aspect of flight training.

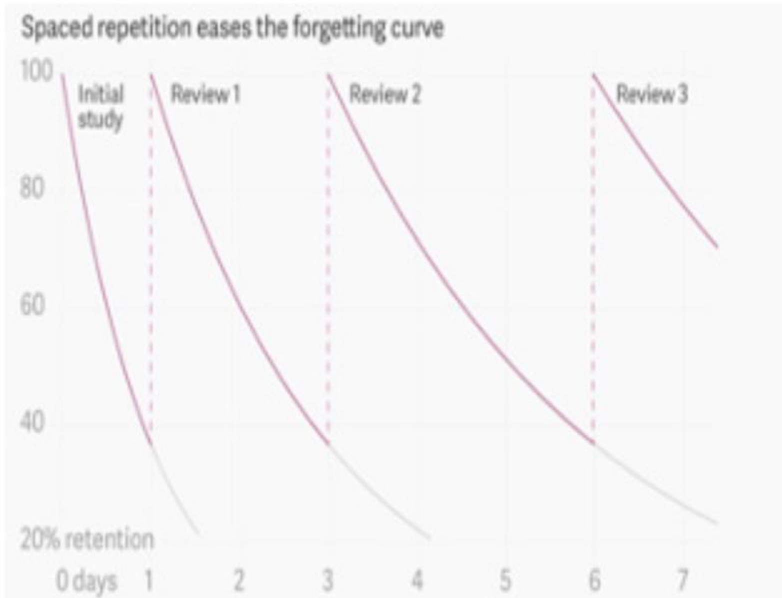
Although this finding is consistent with Ebbinghaus's (1885/1913/2013) forgetting curve theory, the corresponding implication to this theory is problematic. For example, Ebbinghaus's forgetting curve presents the amount of lost information relative to the number of days without practice or study as a percentage. In the current study, lost information is examined from the perspective of increased flight hours. Secondly, the reader will observe from Figures 5.2 and 5.3 that training time (x-axis) is extended to 900 days, which is approximately 2.5 years. Most flight students do not extend their flight training over the course of multiple years, and even if they did, the amount of information lost equates to approximately an additional 20 hours of flight time based on the corresponding asymptote. This infers that as training days increase from days to weeks to months to years, the amount of information flight students lose is not as pronounced as posited by Ebbinghaus. It also infers that the absence of periodic practice of flight maneuvers does not necessarily mean it will be difficult to recall these maneuvers at a later date. It also is noteworthy to observe the increase in flight hours between the minimum required by the FAA (35 hours for a Part 141 program and 40 hours for a Part 61 program) relative to the asymptote, which approaches 70 hours: As training days to proficiency increases, the number of flight hours approaches at least twice the minimum needed. However, the reader might recall from Chapter 1 that most flight students in the U.S. require, on average, between 60 and 75 hours to become check-ride proficient. This implies that with respect to the asymptote, the relationship between flight hours and training days found in the current study is consistent with practice. This also implies that extending the number of training days needed to become check-ride proficient—for example, from 100 days to 1 year—is not as cost-prohibitive as one might think because the number of flight hours will “bottom out,” on average, at approximately 70 hours.

As for the spacing effect, Ebbinghaus (1885/1913/2013) conjectured that continued practice over time allowed newly acquired information to be recalled easier, effectively mitigating the effects of the forgetting curve. This was illustrated in Figure 1.2 (Chapter 1) but is replicated here as Figure 5.4 for the reader’s convenience. The spacing effect is manifested in the comparison between distributed vs. massed practice, with the former being more effective, which implies that new information received in a distributed format will slide the forgetting curve to the right and hence delay not being able to recall the information.

Participants in the current study did not self-report their practice methods when pursuing their PPL. Thus, it is unknown, for example, if participants consistently received flight instruction several times per week (distributed practice), once per week (approaching massed practice), or less than once per week such as once or twice per month (massed practice). Similarly, it also is unknown if

Figure 5.4

Replication of Figure 1.2: Illustration of the Spacing Effect on the Forgetting Curve



Note. This figure demonstrates that review and practice slide the forgetting curve toward the right, thereby mitigating the forgetting curve's effects. The graph shows that after 1 day, a student's retention percentage has decreased from 100% to 40%. After two review and practice sessions, the student maintains a 40% retention percentage for up to 3 days. With three review and practice sessions, a student maintains a 40% retention percentage for up to 6 days. *Source:* Sonnad (2018).

participants prepared for the PPL written exam by "cramming" (massed practice) or by reviewing past concepts repeatedly several days or weeks prior to the exam (distributed practice). The absence of these data implies that the findings of the current study cannot be used to assess the effect of the spacing effect with respect to flight hours to check-ride proficiency. However, based on the sample data as depicted in Figure 5.3, one could infer that the longer training days were a function of massed practice. The reader is cautioned, though, that this is mere speculation because the delay could be related to other factors such as lack of finances.

Ausubel's Theory of Meaningful Learning. In addition to Ebbinghaus's (1885/1913/2013) forgetting curve theory and spacing effect, the current study also was grounded in Ausubel's (1963) theory of meaningful learning. Ausubel advanced the notion that the most critical factor that influences learning is what students currently know, that is, their prior knowledge. Given this belief, Ausubel developed a theory of meaningful learning, which is in contrast to rote learning, and is grounded in the concept that new knowledge must be related to prior knowledge in a manner that is meaningful to the learner. In other words, when introduced to new knowledge, learners must integrate and link this new knowledge in a nonarbitrary manner to their current cognitive structure. Ausubel focused his theory on expository teaching settings, which involve reception learning: Teachers present new information to students in a meaningful manner by considering students' prior knowledge and experiences, and students receive this new information and incorporate (i.e., subsume) it into their existing knowledge base.

The current study was partly grounded in Ausubel's (1963) meaningful learning theory because the theory provides a plausible explanation for Ebbinghaus's (1885/1913/2013) forgetting curve theory and spacing effect. By appropriately spacing out new information (i.e., distributed practice), students have more of an opportunity to relate new knowledge more efficiently with prior knowledge than they would with massed practice and this in turn would help mitigate the effect of the forgetting curve. In the current study, participants did not self-report any information related to Ausubel's theory, though. More specifically, the questionnaire was fact-based and absent of any cognitively loaded items that, for example, related to participants' study habits and learning strategies, the extent to which instructors considered participants' prior knowledge, and whether instructors used advance organizers to help link students' prior knowledge to new knowledge. This was not an omission of the current study but instead was by design because of the prospective participants. As anticipated, the sample primarily consisted of pilots working in the profession who responded to the fact-based items by consulting their records or from memory as opposed to flight students who had just completed their flight training. Thus, the current study did not collect sufficient data to determine the extent to which the findings supported Ausubel's theory directly. This implies that the appropriateness of using Ausubel's theory as the theoretical grounding for studies involving flight training is uncertain. However, one could conclude that the findings indirectly supported Ausubel's theory when examined in concert with Ebbinghaus's forgetting curve theory and spacing effect.

Implications Relative to Prior Research

This section contains a discussion of the current study's findings relative to the findings of the prior research presented in Chapter 2. The reader is reminded that part of the literature review focused on the spacing effect with respect to either meaningful learning or motor skills learning separately. Because flight students must apply what they learned in ground school to the development of motor skills when flying an airplane, the summary of prior research presented here is limited to the application of the spacing effect simultaneously to meaningful and motor skills learning. As part of this discussion, I provide a brief overview of the prior research, examples of where my results were or were not consistent with those of past studies, and plausible explanations for any differences.

Mengelkoch et al. (1971). Mengelkoch et al. studied meaningful learning and motor skill acquisition in the context of learning to fly an airplane. The study was implemented in a flight simulator with students learning to fly by reference to instruments, without any outside visual references. The study's purpose was to ascertain the amount of forgetting that occurs in pilot proficiency skills.

Participants were divided into two groups. Group 5 was given five trial simulator sessions, which were tantamount to massed practice, and Group 10 was given 10 trial simulator sessions, which equated to distributed practice. Each simulator session lasted for approximately 50 minutes and included climbs, descents, turns to headings, among other maneuvers. Both groups also completed a 4-hour academic course on flight procedures. Performance was evaluated once after the respective trials and a second time after a 4-month retention period. Mengelkoch et al. reported that after each respective group's last simulator session—prior to the retention period—Group 10's level of procedural performance proficiency was highly rated and "almost error-free" (p. 401), whereas Group 5's level was rated "satisfactory but intermediate" (p. 401). Performance of corresponding flight maneuvers in all categories showed similar trends.

Unlike Mengelkoch et al. (1971), the current study neither directly compared massed vs. distributed groups nor did it assess each group's performance proficiency via a simulator. However, because the current study examined the relationship between training days and flight hours needed to become check-ride ready, which is a form of proficiency, a parallel can be drawn to Mengelkoch et al. When examined from this perspective, the findings of the current study were consistent with those of Mengelkoch et al. (1971). For example, the findings of both studies supported Ebbinghaus's (1885/1913/2013) forgetting curve theory and spacing effect. For Mengelkoch et al., distributed practice (Group 10) was more effective than massed practice (Group 5), and in the current study, fewer flight hours to become check-ride proficient were needed when training days were fewer, which is consistent with the forgetting curve and presumes distributed practice. As the number of training days increased, though, flight students were taking flight lessons less often, which is analogous to massed practice, leading student to become less proficient and requiring additional flight hours to become check-ride ready.

The findings of the current study also are consistent with Mengelkoch et al.'s (1971) observations regarding the 4-month retention period. Mengelkoch et al. reported that after four retention trials, both groups had a significant loss of procedural performance: Group 10 had a 16.5% loss, and Group 5 had a 20.1% loss. In the current study, students who sustained a break in flight training—as evidenced by an increase in the number of training days—accumulated more flight hours to relearn information that had been lost to attain the level of proficiency required to take the private pilot check ride. As noted by Ebbinghaus (1885/1913/2013), without continual practice information is lost and much more difficult to retain. Furthermore, although Mengelkoch et al. did not statistically analyze the performance differences between the groups during the retention period, graphical representations of the data were logarithmic-like and showed that Group 10 had better performance at the end of its training in all maneuvers than Group 5. This is yet another area of consistency between the findings of the current study and those of Mengelkoch et al.

The reader will note the following differences between the current study and Mengelkoch et al. (1971): (a) time frame: 2023 vs. 1971; (b) sample: flight students/pilots vs. participants with no prior flight training; and (c) implementation: evaluation of simulator performance of flight maneuvers vs. fact-based data self-reported by participants. Given that the current study's findings were consistent with those of Mengelkoch et al. despite these differences, a corresponding implication is that the concepts of the forgetting curve and spacing effect are appropriate for examining flight performance proficiency and provides additional credibility to distributed practice for flight instruction.

Caligan, Jr. (2012). Caligan, Jr. examined the extent to which flight time needed to solo an aircraft could predict flight time needed to obtain a PPL and if time to solo could predict a successful first-try pass on the private pilot check ride. For the first objective, Caligan, Jr. conducted a simultaneous regression analysis involving the targeted IV and DV, but also included additional factors such as the type of flight training program (Part 61 vs. Part 141), the total number of CFIs students had, and the mean number of hours students flew per week. For the second objective, Caligan, Jr. performed a linear regression analysis (DV = "yes," the student passed the check ride on the first try vs. "no," the student did not pass the check ride on the first try). With respect to the first objective, Caligan, Jr. (2012) reported that as the mean number of hours students flew per week increased, time to solo became a stronger predictor of time to PPL. Caligan, Jr. also reported that this predictive relationship between time to solo and time to PPL was stronger for students training in a Part 141 program than a Part 61 program. With respect to the second objective, Caligan, Jr. reported that time to solo was not a significant predictor of a first-try pass on the check ride.

Although the current study was not a replicate of Caligan, Jr. (2012), there were some similarities with respect to Caligan, Jr.'s first objective (the second objective was not applicable to the current study). For example, both studies focused on PPL as the dependent variable, with Caligan, Jr. examining flight time needed to obtain a PPL and the current study examining flight time needed to become proficient to take the private pilot check ride, which is a precursor to obtaining a PPL. Similarly, both studies targeted “time to proficiency” as the primary predictor, with Caligan, Jr. examining time to solo and the current study examining training time (in days) to check-ride proficiency.

With these differences noted, the findings of the current study were partially consistent with those associated with Caligan, Jr.’s (2012) first objective. For example, “time to proficiency” in both studies was found to be a significant predictor to each study’s respective DV. However, unlike Caligan, Jr., the type of training program (Part 61 vs. Part 141) had no relationship with the DV in the current study. One plausible explanation for this difference is sample size: Caligan, Jr. reported a final sample size of n = 273 compared to n = 164 (40% increase) of the current study. A second plausible explanation is the manner in which Caligan, Jr. analyzed his data. For example, he reported that time to proficiency “showed high skewness and high kurtosis” (p. 34), which means that the data set did not satisfy the normality assumption of multiple regression. As a result, he should have transformed this factor as was done in the current study. Caligan, Jr. also coded the group membership variable for the type of flight training program using 1, 2, and 3. This coding strategy is contrary to acceptable standards such as dummy coding, which uses 0 and 1, and effects coding, which uses 0, 1, and –1, and hence does not provide accurate group comparisons. These differences notwithstanding, given that the current study’s findings were consistent with those related to Caligan Jr.’s

(2012) first objective, a corresponding implication is that “time to proficiency”—whether defined as hours to solo or training days to check-ride proficiency—is a robust predictor relative to factors surrounding the PPL. Furthermore, given that both study’s respective R^2 values relative to their prediction models were at most 50% implies there are other factors impacting the PPL-related outcome variables.

Graham (2017). Graham’s study focused on the concepts of distributed practice and cooperative group learning. Working with a sample of volunteer flight students enrolled in Utah Valley University’s flight program. Graham partitioned the sample into three groups: Students in Group A met with their CFI individually twice a week and also as a cooperative group. Students in Group B were given the option to meet with their CFI as a cooperative group once a week. Students in Group C were assigned to individualized flight instruction, which did not involve weekly meetings with their CFI. The dependent variable was time to PPL.

Instead of analyzing the data via an inferential statistical strategy such as a single-factor ANOVA, Graham (2017) instead simply reported that Group A’s mean time to PPL—with respect to total number of days and flight hours—was less than that of Group B’s, and Group B’s mean time to PPL was less than that of Group C’s. Based on the results of these descriptive statistics, Graham concluded that cooperative group meetings with students’ CFIs reduced the time and flight hours needed to obtain a private pilot license. However, because he relied on descriptive statistics, it was unclear if the respective differences in group means were statistically significant. Furthermore, the data also revealed that Group A flew more hours per week than Group B, and Group B flew more hours per week than Group C, which suggests “flight hours per week” might have been an alternative explanation for the results.

Although the current study did not involve a group membership variable, the findings were consistent with Graham (2017) if considered from the perspective that “flight hours per week” was an alternative explanation for Graham’s results. This is because flight hours per week is synonymous with training time to proficiency, which was defined in the current study as the number of days it took students to become check-ride proficient. A flight training schedule that consists of an increase in the number of flight hours per week will decrease training time to proficiency, which will then reduce the total number of flight hours needed to become check-ride proficient. This is because such a schedule promotes distributed practice and mitigates the effect of the forgetting curve. An implication of this consistency between the current study’s findings with those of Graham’s is that the key to minimizing flight time to become check-ride proficient is maintaining a consistent flight training schedule that involves working with a CFI more than 1 day per week.

Implications for Aviation Practice

In addition to the preceding implications relative to theory and prior research, the current study’s findings also have implications for aviation practice.

The first implication is relative to flight students' personological characteristics, including their biological sex at birth, age, race, marital status, and ethnicity. As reported in Chapter 4, these factors had no significant relationship with flight time to check-ride proficiency. This finding implies that CFIs are unbiased in their assessment of students' flight skills during training. It also implies that students themselves are not handicapped because of their respective demographics, which infers that the number of flight hours needed to become check-ride proficient are not a function of these attributes.

A second implication of the study's findings for practice is related to the different flight training programs. As reported in Chapter 4, there was no significant difference in number of flight hours needed to become check-ride proficiency between Part 61 vs. Part 141 programs. This finding implies that both programs are providing equivalent flight instruction even though their respective curriculums are different. It also implies that the 5-hour reduction in the minimum number flight hours students need before they can take the private pilot check ride that the FAA affords a Part 141 program might be misguided. With respect to this latter point and based on the most recent FAA (2006) data where the minimum hours needed to become check-ride proficient for the PPL is between 60 and 75 hours, this 5-hour reduction is really a moot point. This is because today's airspace and aircraft are more complex, which increases the number of hours needed to become check-ride proficient.

A third implication of the study's results for practice is related to the different flight training organizations. As reported in Chapter 4, when examined in the context of the final regression model and holding all other variables in the model constant, students attending a 2- or 4-year college/university averaged 6 fewer flight hours to become check-ride proficient than students attending an FBO. This finding implies that colleges and universities appear to be more efficient in their private pilot flight training than FBO flight schools. This finding also implies that a structured flight training curriculum, which is used by colleges and universities, is more beneficial than the less structured curriculum used by FBOs. So, although there was no significant difference in flight hours needed to check-ride proficiency between Part 61 and Part 141 training programs as reported earlier, it appears that a Part 141 program, which is predominately implemented by colleges and universities, expedites check-ride proficiency when compared to a Part 61 program, which is predominately implemented by FOBs.

A fourth implication of the study's results for practice is related to the different scheduling policies. As reported in Chapter 4, when examined in the context of the final regression model and holding all other variables in the model constant, flight students who operated under a mandated policy averaged 9.5 more hours to become check-ride proficient than flight students who operated under a student-driven policy. The reader is reminded that with a mandated policy the flight training organization establishes the training schedule and students are required to follow it, whereas with a student-driven policy the flight training organization permits students to schedule their flight training at their convenience. Although the current study did not collect data on check-ride pass rates, this finding implies that CFIs working under a mandated scheduling policy have higher first-attempt check ride pass rates than CFIs operating under a student-driven schedule because of the additional training time required of students.

A fifth implication of the study's results for practice is related to training time (in days) to check-ride proficiency. As reported in Chapter 4, training time had a nonlinear relationship, which was logarithmic-like, with the number of flight hours needed to pass the private pilot check ride and hence was transformed using the natural logarithmic function. This finding implies that using training days to proficiency to predict the number of flight hours needed to become check-ride proficient is problematic. With the benefit of hindsight, this is because training days as a single variable does not account for other imbedded factors such as the number of flight lessons conducted in a training day or the number of flight hours accrued during a training day.

Generalizability, Limitations, and Delimitations

Generalizability

Generalizability, which also is known as external validity, generically refers to the extent to which the findings of a study can be extended beyond the scope of the study. Generalizability commonly is considered from two perspectives:

population generalizability, which focuses on the degree sample results can be applied to the parent population, and ecological generalizability, which focuses on the degree sample results can be applied to other populations, settings, or conditions. The reader is reminded that detailed information about how the study was conducted and the corresponding results were provided in Chapters 3 and 4, respectively, of this dissertation.

As presented in Chapter 3, the target and accessible populations were identical and consisted of all flight students pursuing their private pilot's license (PPL) in the United States. The sample, which was acquired by convenience and snowball strategies, consisted initially of $N = 167$ participants, and later reduced to $N = 164$. Except for flight students' biological sex assigned at birth, the FAA does not maintain key statistics of flight students' demographics. This makes it challenging to assess the population generalizability of the current study. Nevertheless, because the current study's female–male ratio was consistent with that reported by the FAA, and because the targeted personological characteristics of flight students were not related to the dependent variable, it is reasonable to conclude that the current study's findings are generalizable to the parent population. This population generalizability, though, is restricted to flight students whose ethnicity is "not Hispanic or Latino," which reflected 90% of the initial sample. Readers interested in making their own determination of population generalizability are directed to Tables 3.1–3.4 (Chapter 3), which contain descriptive statistics of the sample.

With respect to ecological generalizability, the current study's results are limited to civilian flight training in the U.S. that leads to an FAA-approved private pilot license. This is because other FAA certificates or ratings such as commercial or airline transport pilot (ATP) certificates, as well as instrument, seaplane, and multi-engine ratings require different flight time requirements, which would cause the regression coefficients of the final model to have different weights. Furthermore, the required maneuvers for these other certificates might be more or less difficult than private pilot maneuvers thereby causing increased or decreased training time. As for different populations, it is conceivable that the current study's findings might be applicable to private pilot flight training programs based on the requirements of the International Civil Aviation Organization (ICAO), which would include countries outside the U.S. This is because ICAO's PPL requirements are similar to those of the FAA. The current study's findings, however, would not be applicable to military-based PPL training because of the military's culture and requirements.

Limitations and Delimitations

The current study and its results were bounded by several limitations and delimitations, which were presented initially in Chapter 1. These limitations and delimitations are replicated here as a courtesy to the reader so they can be easily accessible when reviewing the corresponding recommendations for future research relative to the study's limitations and delimitations.

Limitations. A study's limitations include circumstances, conditions, and events that the researcher cannot control but limit the study's generalizability. Following is a brief discussion of the current study's limitations, which were derived in part by reviewing similar studies and methodologies and guided by my own professional experiences.

1. Sample Demographics. I had no control over the personological characteristics of the current study's participants, including their biological sex assigned at birth, age, race, marital status, and ethnicity. As a result, similar studies that involve samples with personological characteristics different from those of the current study might get different results.

2. Source of Study. The current study was a non-funded Ph.D. dissertation research project from a student in the College of Aeronautics of an independent Ph.D. granting university in the southeastern United States. Therefore, if a similar study were to be conducted by a federal agency, such as the FAA or the National Transportation Safety Board, or via funded research, the results might be different.

3. CFI Training. I had no control over the instructional experiences or teaching effectiveness of CFIs. Thus, similar studies that include specific CFI experience factors such as years held a CFI certificate, number of dual hours, and number of dual hours in the past 90 days might get different results.

4. Tower vs. Non-Tower Airports. I had no control over whether students received flight instruction at an airport with or without a control tower. As a result, similar studies that include this information might get different results.

5. CFI Oversight. I had not control over what criteria CFIs used to personally judge or decide when they believed a student was ready to take the private pilot check ride. As a result, similar studies that include this information might get different results.

Delimitations. A study's delimitations include circumstances, conditions, and events that the researcher imposes to make the implementation of the research more feasible but can further limit the generalizability of the study's findings. Following is a brief discussion of the current study's delimitations.

1. Civilian Flight Students. The participants of the current study were limited to those who received civilian flight training. Therefore, similar studies might get different results if their samples consist of participants who received military flight training or a mix of both civilian- and military-trained participants.

2. Flight Training Programs. The current study targeted participants who received their flight training from either a Part 61 or a Part 141 flight training program. Therefore, similar studies that restrict study implementation exclusively to a Part 61 or a Part 141 program (but not both) might get different results.

3. Flight Training Organization. The current study asked participants to specify the type of organization from which they received their flight training (2- or

4-year college/university or an FBO). Because the presence or absence of this factor can have an impact on the results of data analysis, similar studies that either do not include this factor or provide more detailed information (e.g., the geographic location of the organization) might get different results.

4. Category and Class of Airplane. The current study's participants obtained their PPL in the airplane category and single engine land class as defined by the Federal Aviation Regulations (FAR, 2021b). Therefore, similar studies that do not include this restriction might get different results.

5. Sampling Strategy. The current study acquired data using convenience and snowball sampling strategies. Therefore, similar studies that use a different sampling strategy such as simple random or cluster random sampling might get different results.

6. Targeted Participants. Participants of the current study included flight students who were currently pursuing their PPL as well as pilots who previously earned their PPL earlier in their life and self-reported what they recalled when they were flight students. Therefore, similar studies that target different participants such as those who are currently pursuing their PPL, recently minted pilots with their PPL, or more generally, different age cohorts, might get different results.

7. Measurement of the Dependent Variable. The dependent variable of the current study was the number of flight hours participants accrued when their CFIs declared them to be check-ride proficient. Thus, similar studies that use a different metric such as the flight hours students accrued at the time they received their PPL might get different results.

8. Theoretical Grounding. The current study was grounded in Ebbinghaus's (1885/1913/2013) forgetting curve theory and spacing effect, and Ausubel's (1963) theory of meaningful learning. Therefore, similar studies with a different theoretical grounding might get different results.

9. Transformation of Training Time to Proficiency. The current study transformed X_1 = Training time to check-ride proficiency (in days) using the natural logarithmic function (ln). Therefore, similar studies that either do not transform this variable or use a different transformation (e.g., log base 10) might get different results.

10. Outliers. In the current study, an outlier analysis was conducted using Jackknife distances, which flagged 15 rare cases and all 15 cases were included in the final data set. Therefore, similar studies that either use a different outlier analysis approach or do not include rare case outliers might get different results

Recommendations for Future Research and Practice

This section contains four sets of recommendations arising from the findings of the current study. The first two sets of recommendations are made for future research relative to the study's limitations and delimitations, respectively. The third set of recommendations is for future research based on the implications to prior research and theory. The last set of recommendations is for future research based on the implications for practice.

Recommendations for Future Research Relative to Study Limitations

1. The current study's targeted personological characteristics, which were captured in Set C, included participants' biological sex assigned at birth, age, race, marital status, and ethnicity. None of the first four factors had a significant relationship with flight hours needed to become check-ride proficient, and the last factor was treated as a constant ("not Hispanic or Latino"). Therefore, a recommendation for future research is to examine other personological characteristics such as educational level, personality traits, and socioeconomic status (SES)/income.
2. The current study did not receive any internal or external funding from a college/university, or from a federal or state agency such as the FAA, or by an organization that had the support of a federal or state agency. Therefore, a recommendation for future research is to replicate the current study using resources supported by a funding organization to broaden the scope of the study and give it greater credibility.
3. The current study did not consider the instructional experiences or teaching effectiveness of CFIs. Examples include where CFIs acquired their training (e.g., civilian- or military-based), number of years they've held a CFI certificate, number of dual hours, number of dual hours in the past 90 days, and

success rate of students passing their check ride on their first attempt. Therefore, a recommendation for future research is to prepare and administer to flight students' CFIs a questionnaire that includes this information and then examines the effect these factors have on flight hours students needed to become check-ride proficient.

4. The current study did not consider whether students received flight instruction at a towered or nontowered airport. Therefore, a recommendation for future research is to have students self-report this information.
5. The current study did not have any oversight on CFIs' judgement on when a flight student was ready for a check ride. Therefore, a recommendation for future research is to collect oversight information via the questionnaire alluded to in Recommendation 3 above so the effect of this factor can be examined.

Recommendations for Future Research Relative to Study Delimitations

1. The current study's sample was restricted to participants who received civilian-based flight training for their PPL. Therefore, a recommendation for future research is to augment this sample by including flight students who received their PPL training in the military or to focus only on military-based PPL training.
2. The current study's sample included participants from both Part 61 and Part 141 flight training programs. Therefore, a recommendation for future research

is to focus on flight student who received their PPL training solely from a Part 61 program or solely from a Part 141 program.

3. The current study's focus on flight training organization was exclusively with respect to two organizations: 2- or 4-year colleges/universities vs. FBOs. Therefore, a recommendation for future research is to disaggregate the colleges/universities factor: (a) maintaining separate groups for 2-year colleges, 4-year colleges, and universities; (b) qualifying these organizations as public or private; and (c) including their geographical location.
4. The current study's sample was restricted to participants who obtained their PPL in the airplane category and single engine land class as defined by the Federal Aviation Regulations (FAR, 2021b). Therefore, a recommendation for future research is to remove this restriction and focus on other certificates such as Instrument Rating, Commercial Pilot Airplane Single-Engine Land Rating, CFI, CFII, Commercial Pilot Multi-Engine, and MEI.
5. The current study acquired data using convenience and snowball sampling strategies. Therefore, a recommendation for future research is to implement a different sampling strategy such as simple random or cluster random.
6. The current study's sample was restricted to flight students who were currently pursuing their PPL as well as pilots who previously earned their PPL earlier in their life and self-reported what they recalled when they were flight students. Therefore, a recommendation for future research is to replicate this study using

current flight students who are not pilots, recently minted pilots, or partitioning the participants into different age cohorts.

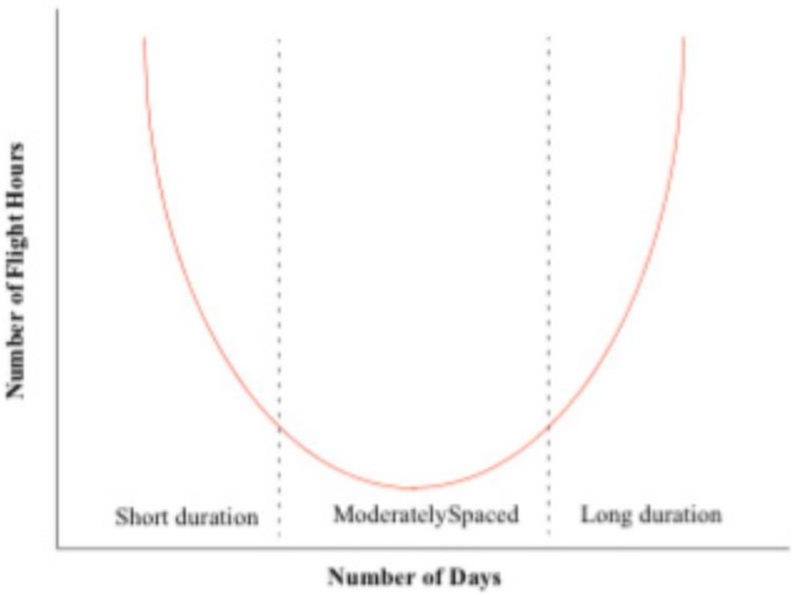
7. The dependent variable of the current study was the number of flight hours participants accrued when their CFIs declared them to be check-ride proficient. Therefore, a recommendation for future research is to use a different dependent variable such as the number of flight hours students accrued at the time they received their PPL, or the number of hours needed to solo (Caligan, 2012).
8. The current study was grounded in Ebbinghaus's (1885/1913/2013) forgetting curve theory and spacing effect, and Ausubel's (1963) theory of meaningful learning. Therefore, a recommendation for future research is to use a different theoretical grounding such as locus of control and attribution theory.
9. The current study transformed X_1 = Training time to check-ride proficiency using the natural logarithmic function (ln). Therefore, a recommendation for future research is to use a different transformation function such as log base 10.
10. The current study used a Jackknife distances strategy for its outlier analysis and included in the final data set the 15 rare-case outliers that were flagged. Therefore, a recommendation for future research is to use a different outlier analysis strategy or excluded any rare-case outliers from the final data set.

Recommendations for Future Research Based on Implications Relative to Theory and Prior Research

1. Although the current study's results were consistent with Ebbinghaus's (1885/1913/2013) forgetting curve theory—for example, a logarithmic-like relationship emerged between flight hours needed to become check-ride proficient and training time (in days) to check-ride proficiency—the implication was problematic. If most students schedule flight lessons close together, then Ebbinghaus would consider this as "cramming" or massed practice, requiring students to need more flight hours to become check-ride proficient than students who space out their flights (i.e., distributed practice). This scenario, in practice, would support a U-shaped distribution as presented in Figure 1.3 (Chapter 1)—and replicated here as Figure 5.5 for the convenience of the reader—rather than a logarithmic-like distribution. Therefore, a recommendation for future research is to replicate the current study but restrict the sample to flight students who become check-ride proficient in fewer days than what was reported in the current study (M = 426 days and Mdn = 247 days). One possible example is to restrict the time period to 10 weeks (70 days), which would equate to 75 flight hours to become check-ride proficient if students scheduled five 1.5-hour flight lessons per week. The reader is reminded from Chapter 1 that flight students in the U.S. require, on average, between 60 and 75 hours to become proficient to take the private pilot check

Figure 5.5

Replication of Figure 1.3: Changes in the Number of Flight Hours as a Function of Number of Days Before Check Ride



Note. This figure illustrates the prohibitive effect on the number of flight hours needed to acquire proficiency for the check ride relative to the spacing of flight lessons. If flight lessons are spaced too close together (short duration), students will not have sufficient time to process new knowledge. On the other hand, if flight lessons are spaced too far apart (long duration), they will forget what they learned previously, which will increase the number of flight hours needed for proficiency.

ride. This scenario might then reflect a quadratic relationship. As a point of information, when the current study's data set was restricted to at most 96 training days (n = 19 cases), a 4th-degree polynomial relationship emerged between flight hours needed to become check-ride proficient and training days to proficiency.

2. With respect to Ebbinghaus's (1885/1913/2013) spacing effect, the current study did not collect any data relative to how many flight lessons per day or per

week students completed and hence it was not possible to examine the role of the spacing effect directly. Therefore, a recommendation for future research, which ties into the previous recommendation, is to collect flight lesson frequency data. A second recommendation is to conduct a qualitative study to determine common characteristics of participants as they progress through their flight instruction.

3. As noted earlier, the current study initially was grounded in Ausubel's theory of meaningful learning because it provided a plausible explanation for Ebbinghaus's (1885/1913/2013) forgetting curve theory and spacing effect. However, the current study did not directly assess the appropriateness of Ausubel's theory to flight instruction because no cognitively loaded questions that focused on participants' study habits, learning strategies, prior knowledge, and the use of advance organizers were administered as part of data collection. Therefore, a recommendation for future research is to replicate the current study but include these factors to assess the extent to which Ausubel's theory can be applied to flight instruction. This could be done by focusing on both classroom and flight instruction so that researchers can examine the extent to which students are applying the concepts developed in the classroom to their flight instruction as well as the levels of knowledge that are being demonstrated.
4. The results of the current study were consistent with Mengelkoch et al. (1971) in that both studies supported in part Ebbinghaus's (1885/1913/2013) forgetting

curve theory and spacing effect. Therefore, a recommendation for future research is to conduct replication studies that focus on the concepts of spacing effect and distributed practice applied to flight instruction.

5. The results of the current study were consistent with those related to Caligan Jr.'s (2012) first objective, which focused on using time to solo as a predictor of time to PPL. Although the current study was not directly related to Caligan's, both studies demonstrated that "time to proficiency"—whether it was hours needed to solo or training days needed to become check-ride proficient—was a robust predictor for outcome variables related to time to PPL. Therefore, a recommendation for future research is to use the concept of "time to proficiency" as a key factor for studies that examine the various aspects of time to PPL.
6. Continuing with Caligan (2012), both the current study's and Caligan's respective prediction models suggested there are additional factors other than those targeted that have an effect on PPL-related outcome variables. Therefore, a recommendation for future research is to conduct preliminary studies in search of these other factors.
7. The current study's findings were consistent with Graham's (2017) if considered from the perspective that "flight hours per week" was an alternative explanation for Graham's results. However, neither the current study nor Graham examined students' flight training schedule that reflected the number

of flight lessons and hours students completed per week. Therefore, a recommendation for future research, which ties in with the second recommendation of this section, is to collect flight lesson frequency data to determine its effect on hours needed to become check-ride proficient.

Recommendations for Practice Relative to Study Implications

1. One implication to practice was that students did not appear to be handicapped in pursuit of their PPL because of their respective demographics. Therefore, a recommendation for practice is to ask students to respond to a question(s) that reflect their locus of control. An example of one such question might be: “Who do you think is responsible for your success in becoming ready for your private pilot check ride? (a) I am, (b) My CFI, (c) Both my CFI and me.”
2. The current study found no significant difference in flight hours needed to check-ride proficiency between Part 61 vs. Part 141 programs, which challenges FAA’s 5-hour reduction rule applied to Part 141 programs. Therefore, a recommendation for practice is to continue examining differences between these programs and if the results consistently show “no difference,” then a case can be made to the FAA to apply this reduction to Part 61 programs as well.
3. An implication relative to the finding that 6 fewer flight hours, on average, are needed to become check-ride proficient for students attending a 2- or 4-year college/university vs. an FBO is that flight training programs are more efficient

at colleges/universities. This could be related to their use of a structured flight training curriculum. Therefore, a recommendation for practice is to perform a content analysis of the respective PPL flight training curriculums used by colleges/universities and FBOs to get a more concrete understanding of their differences.

4. An implication relative to the finding that students under a mandated scheduling require 9.5 more flight hours, on average, to become check-ride proficient vs. a student-driven schedule is that CFIs operating under the former policy have higher first-attempt check ride pass rates than CFIs operating under the latter policy. Therefore, a recommendation for practice is to collect data on CFI private pilot check ride pass rates and compare differences between the two scheduling policies.
5. Because of the nonlinear relationship between training time and number of flight hours needed to pass the private pilot check ride, the predictor was transformed using the natural logarithmic function. However, in the context of the current study, training time was treated singularly as the total number of training days students completed prior to becoming check-ride ready without any consideration for additional factors that correspond to this variable. Therefore, a recommendation for practice is to capture specific data associated with what took place during a training day, including the number of flight

lessons completed, the number of flight hours accrued, and how much time students studied independent of their flight lessons.

6. The prediction model presented in Chapter 4 included an example of how it can be applied in practice to predict the number of flight hours students would need to become check-ride proficient. As a result, a recommendation for practice is to apply this model a priori and then compare the post hoc results to the predicted results to determine the strength and direction of the relationship.
7. The prediction model presented in Chapter 4 was based on a sample that consisted mostly of pilots who relied on their past experiences as flight students to self-report the requested information. As a result, a recommendation for practice is to refine this model by focusing solely on flight students from the time they begin flight lessons to the time their CFIs declare they are ready for the private pilot check ride.
8. The prediction model presented in Chapter 4 was based on flight student data and was absent any information about CFIs. As a result, a recommendation for practice is to refine this model by including key characteristics of CFIs, including their relevant experiences as a CFI as well as the maximum number of students they teach per day.
9. Given the relationship between the current study and Caligan (2012), a corresponding recommendation for practice is to examine the relationship between time to solo and training time to check-ride proficiency. If the

correlation is strong (e.g., $r > .85$), then apply the prediction model of the current study using time to solo as the predictor instead of training time to check-ride proficiency to see if the model holds. If the correlation is not strong, then include time to solo as another factor to help refine the model further.

10. A possible enhancement to the prediction model presented in Chapter 4 is to consider flight students' receptivity to instructional feedback (RIF). For example, it is conceivable that being receptive to CFI feedback could lead to fewer training days whereas not being receptive could lead to more training days. Therefore, a recommendation for practice is to examine this factor either anecdotally or via a formal instrument such as Lipnevich et al.'s (2021) RIF scale.

References

Acock, A. (1997). Working with missing data. *Family Science Review*, *1*(10), 76–102.

Allison, P. D. (2009). Missing data. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The Sage handbook of quantitative methods in psychology* (pp. 72–89). Sage Publications Ltd. https://doi.org/10.4135/9780857020994.n4

Ary, D., Lucy, J. C., & Sorensen, C. (2010). *Introduction to research in education* (8th ed.). Wadsworth.

Ausubel, D. P. (1963). *The psychology of meaningful verbal learning*. Grune & Stratton.

Ausubel, D. P. (1968). *Educational psychology: A cognitive view.* Holt, Rinehart, and Winston.

Ausubel, D. P., Novak, J. D., & Hanesian, H. (1978). *Educational psychology: A cognitive view* (2nd ed.). Holt, Rinehart, and Winston.

Baddeley, A. D., & Longman, D. J. A. (1978). The influence of length and frequency of training sessions on the rate of learning to type. *Ergonomics, 21*(8)*,* 627–635. https://doi.org/10.1080/00140137808931764

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science, 4*, 316–321. https://doi.org/10.1111/j.1467-9280.1993.tb00571.x

Caligan, Jr., R. J. (2012). Time to solo as a predictor of performance in civilian flight training. *The International Journal of Aviation Psychology, 22*(1), 30–40. https://dx.doi.org/10.1080/10508414.2012.635124

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research on teaching*, Houghton Mifflin Company. (Reprinted from "Handbook of research on teaching," 1963) (N. L Gage, Ed.). Rand McNally & Company.

Cohen, J., Cohen, P., West, S. G., & Aiken, S. L. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Lawrence Erlbaum Associates, Inc.

Date Calculator (n.d.). *Date calculator: Add to or subtract from a date*. https://www.timeanddate.com/date/dateadd.html

Dempster, F. N. (1991) Synthesis of research on reviews and tests. *Educational Leadership, 48*(7), 71–76. http://www.ascd.org/ASCD/pdf/journals/ed_lead/el_199104_dempster.pdf

Dempster, F. N. (1988). A case history in the failure to apply the results of psychological research. *American Psychologist, 43*, 627–634. https://doi.org/10.1037/0003-066X.43.8.627

Dempster, F. N., & Farris, R. (1990). The spacing effect: Research and practice. *Journal of Research and Development in Education, 23*(2), 97–101. ERIC. https://eric.ed.gov/?id=EJ407864

De Win, G., Van Bruwaene, S., De Ridder, D., & Miserez, M. (2013). The optimal frequency of endoscopic skill labs for training and skill retention on suturing: a randomized controlled trial. *Journal of Surgical Education, 70,* 384–393. https://doi: 10.1016/j.jsurg.2013.01.005

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology, 84*(5), 795–805. https://doi.org/10.1037/0021-9010.84.5.795

Ebbinghaus, H. (1885/1913/2013). Memory: A contribution to experimental psychology. *Annals of Neurosciences*, *20*(4), 155–156. (Original work published 1885, Tran. 1913). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4117135/

El-Masri, M. M., & Fox-Wasylyshyn, S. M. (2005). Missing data: An introductory conceptual overview for the novice researcher. *Canadian Journal of Nursing Research (37)*4, 156-171. https://www.researchgate.net/publication/7236506_Missing_data_An_introductory_conceptual_overview_for_the_novice_researcher

Farr, G. R., Key, M. K., & Bloch, E. (1956). The airplane control test: A compensatory pursuit task. *Perceptual and Motor Skills, 6*, 77–80. https://doi.org/10.2466/pms.1956.6.3.77

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*, 1149–1160. [Download software at http://www.gpower.hhu.de]. https://www.psychologie.hhu.de/fileadmin/redaktion/Fakultaeten/Mathematisch-Naturwissenschaftliche_Fakultaet/Psychologie/AAP/gpower/GPower31-BRM-Paper.pdf

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*, 175-191. [Download software at http://www.gpower.hhu.de]. https://www.psychologie.hhu.de/fileadmin/redaktion/Fakultaeten/Mathematisch-Naturwissenschaftliche_Fakultaet/Psychologie/AAP/gpower/GPower3-BRM-Paper.pdf

Federal Aviation Administration. (2021). *U.S. civil airmen statistics*. U.S. Department of Transportation. https://www.faa.gov/data_research/aviation_data_statistics/civil_airmen_statistics/

Federal Aviation Administration. (2006). *Student pilot handbook* (FAA-H-8083-27A). U.S. Department of Transportation. https://www.cpaviation.com/images/downloads/faa-h-8083-27a.pdf

Federal Aviation Regulations. (2021a). *Electronic code of federal regulations. Title 14, Chapter 1, Subchapter D, Part 61, Section 61.109: Aeronautical experience.* https://www.ecfr.gov/current/title-14/chapter-I/subchapter-D/part-61/subpart-E/section-61.109

Federal Aviation Regulations. (2021b). *Electronic code of federal regulations. Title 14, Chapter 1, Subchapter H, Part 141, Appendix B: Private Pilot Certification Course.* https://www.ecfr.gov/current/title-14/chapter-I/subchapter-H/part-141/appendix-Appendix%20B%20to%20Part%20141

Federal Aviation Regulations. (2021c). *Electronic code of federal regulations. Title 14, Chapter 1, Subchapter D, Part 61, Section 61.1(b): Applicability and definitions*. https://www.ecfr.gov/current/title-14/chapter-I/subchapter-D/part-61#61.1

Federal Aviation Regulations. (2021d). *Electronic code of federal regulations. Title 14, Chapter 1, Subchapter D, Part 61, Section 61.1 et seq.: Certification: Pilots, flight instructors, and ground instructors.* https://www.ecfr.gov/current/title-14/chapter-I/subchapter-D/part-61#part-61

Federal Aviation Regulations. (2021e). *Electronic code of federal regulations. Title 14, Chapter 1, Subchapter D, Part 61, Section 61.107: Flight proficiency.* https://www.ecfr.gov/current/title-14/chapter-I/subchapter-D/part-61/subpart-E/section-61.107

Federal Aviation Regulations. (2021f). *Electronic code of federal regulations. Title 14, Chapter 1, Subchapter D, Part 61, Section 61.105: Aeronautical knowledge.* https://www.ecfr.gov/current/title-14/chapter-I/subchapter-D/part-61/subpart-E/section-61.105

Federal Aviation Regulations. (2021g). *Electronic code of federal regulations. Title 14, Chapter 1, Subchapter D, Part 61, Section 61.103: Eligibility requirements: General. https://www.ecfr.gov/current/title-14/chapter-I/subchapter-D/part-61#61.103*

Federal Aviation Regulations. (2021h). *Electronic code of federal regulations. Title 14, Chapter 1, Subchapter D, Part 1, Section 1.1: General definitions.* https://www.ecfr.gov/current/title-14/chapter-I/subchapter-A/part-1

Flexman, R. E., Roscoe, S. N., Williams, A. C., Jr., & Williges, B. H. (1972). Aviation research monographs. *Studies in pilot training: The anatomy of transfer, 2*(1). Aviation Research Laboratory, Institute of Aviation University of Illinois at Urbana-Champaign. http://www.larryscanlan.com/Monographs/Flexman%20et%20al%20(1972)%20V2%20(1).pdf

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2019). *How to design and evaluate research in education* (10th ed.). McGraw-Hill Education.

Frost, J. (2019). *Regression analysis: An intuitive guide for using and interpreting linear models.* Statistics by Jim Publishing.

Gallagher, A. G., Jordan-Black, J. A., & O'Sullivan, G. C. (2012). Prospective, randomized assessment of the acquisition, maintenance, and loss of laparoscopic skills. *Annals of Surgery*, *256,* 387–393. https://doi.org/10.1097/SLA.0b013e318251f3d2

Gallo, M. A., & Odu, M. (2009). Examining the relationship between class scheduling and student achievement in college algebra. *Community College Review*, *36*(4), 299–325. http://doi.org/10.1177/0091552108330902

Graham, M. R. (2017). *A case study of instructional methods used for private pilot certification at Utah Valley University Flight School* [Unpublished master's thesis, Brigham Young University]. BYU Scholars Archive. https://scholarsarchive.byu.edu/etd/6601/

Grote, M. G. (1995). Distributed versus massed practice in high school physics. *School Science and Mathematics, 95*(2), 97–101. https://doi.org/10.1111/j.1949-8594.1995.tb15736.x

Kang, S. G., Ryu, B.J., Yang, K. S., Ko, Y. H., Cho, S., & Kang, S. H. (2015) An effective repetitive training schedule to achieve skill proficiency using a novel robotic virtual reality simulator. *Journal of Surgical Education*, *72,* 369–376. https://doi.org/10.1016/j.jsurg.2014.06.023

Krug, D., Davis, T. B., & Grover, J. A. (1990). Massed versus distributed repeated reading: A case of forgetting helping recall? *Journal of Educational Psychology, 82,* 366–371. https://doi.org/10.1037/0022-0663.82.2.366

Lipnevich, A.A., Gjicali, K., Asil, M., & Smith, J. K. (2021). Development of a measure of receptivity to instructional feedback and examination of its links to personality. *Personality and Individual Differences*, *169*, 1–4. https://doi.org/10.1016/j.paid.2020.110086

Lu, P. K. (1978). Three integrative models of kinetic structure in teaching astronomy. *Journal of Research in Science Teaching, 15,* 249–255. https://doi.org/10.1002/tea.3660150403

Mengelkoch, R. F. & Gainer, C. A. (1971). The forgetting of instrument flying skills. *Human Factors, 13*(5), 397–405. https://doi.org/10.1177/001872087101300502

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*, 247–266. https://doi.org/10.1016/0272-7358(91)90103-2

Moulton, C. E., Dubrowski, A., MacRae, H., Graham, B., Grober, E., & Reznick, R. (2006). Teaching surgical skills; What kind of practice makes perfect? A randomized, controlled trial. *Annals of Surgery*, *244*, (3). 400–409. https://doi.org/10.1097/01.sla.0000234808.85789.6a

Murre, M. J., & Dros, J. (2015). Replicating and analysis of Ebbinghaus' forgetting curve. *Plos One*, *10*(7). https://doi/10.1371/journal.pone.0120644

Ormrod, J. E. (2012). *Human learning* (6th ed.). Pearson.

Ruch, T. C. (1928). Factors influencing the relative economy of massed and distributed practice in learning. *Psychological Review*, *35*, 19–45. https://doi.org/10.1037/h0074423

Shaver, J. P. (1983). The verification of independent variables in teaching methods research. *Educational Research, 72*(8), 3–9. https://doi.org/10.3102/0013189X012008003

Shea, C. H., Lai, Q., Black, C., & Park, J. (2000). Spacing practice sessions across days benefits the learning of motor skills. *Human Movement Science, 19,* 737–760. https://doi.org/10.1016/S0167-9457(00)00021-X

Sheldon, C. Q., & Durdella, N. R. (2010). Success rates for students taking compressed and regular length developmental courses in the community college. *Community College Journal of Research and Practice, 34,* 39–54. https://doi/org/10.1080/10668920903385806

Shrestha, P. (2017, November 17). Ebbinghaus forgetting curve. *Psychestudy*. https://www.psychestudy.com/cognitive/memory/ebbinghaus-forgetting-curve

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction, 1984,* 1*(3)* 341–358. https://doi.org/10.1207/s1532690xci0103_4

Smith, C. D., & Scarf, D. (2017). Spacing repetitions over long timescales: A review and a reconsolidation explanation. *Frontiers in Psychology*, *8*, Article 962. https://doi.org/10.3389/fpsyg.2017.00962

Sonnad, N. (2018, March 29). You probably won't remember this, but the "forgetting curve" theory explains why learning is hard. *Quartz*. https://qz.com/1213768/the-forgetting-curve-explains-why-humans-struggle-to-memorize/

Spruit, E. N., Band, G. P., & Hamming, J. F. (2014). Increasing efficiency of surgical training: Effects of spacing practice on skill acquisition and retention in laparoscopy training. *Surgical Endoscopy*, *29,* 2235–2243. https://doi.org/10.1007/s00464-014-3931-x

United States Census Bureau. (2021). *State population by characteristics: 2010–2019.* https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-detail.html

Verdaasdonk, E. G. G., Stassen, L. P. S., Wijk, R. P. J., & Dankelman, J. (2007). The influence of different training schedules on learning of psychomotor skills for endoscopic surgery. *Surgical Endoscopy, 21,* 214–219. https://doi.org/10.1007/s00464-005-0852-8

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

Appendix A

Data Collection Instrument

Introduction

You are invited to participate in a research study designed to examine the relationship between a set of targeted factors (demographical and flight training experiences) and the number of hours flight students need to complete before their certified flight instructor (CFI) considers them proficient to take the private pilot practical test (check ride). As part of the study, I am requesting that you complete this questionnaire, which should take no more than 5 minutes. Before you begin responding to the items it is important that you understand the following:

1. There are no perceived risks involved. This study will be used for educational purposes only, as I seek to gain a better understanding of the targeted relationship.

2. Your responses will be treated as strictly confidential and will be accessible only by my dissertation advisor and me.

3. Your responses will remain completely anonymous. There will not be any information that will link your responses to you.

4. Participation in the study is strictly voluntary.

5. The study has been reviewed and approved by Florida Institute of Technology's Institutional Review Board (IRB).

6. By clicking on the link below, you are indicating that you are at least 18 years old and have agreed to voluntarily participate in the study.

7. If you have any questions about this research, you may to contact me at mharwin2019@my.fit.edu or my dissertation advisor, Dr. Michael Gallo, at gallo@fit.edu.

A. Background Information

1. Please enter your age at the time of your first flight ________.

2. Please specify your biological sex (assigned at birth).
   - ❑ Male
   - ❑ Female

3. Please specify the race with which you most closely identify:
   - ❑ White
   - ❑ Black or African American
   - ❑ Asian American
   - ❑ American Indian/Alaska Native
   - ❑ Native Hawaiian/Pacific Islander
   - ❑ Other ______________________________

4. Please specify the ethnicity with which you most closely identify:
   - ❑ Hispanic or Latino (e.g., Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin regardless of race)
   - ❑ Not Hispanic or Latino
   - ❑ Other ______________________________

5. Pleas enter your marital status at the time your first flight.
   - ❑ Married
   - ❑ Not Married

B. Flight Training

1. Please enter the date (Month/Day/Year) of your first flight toward your private pilot license. If you do not know the exact date then enter the approximate date (Month/Year) or (Year). ______________________________
2. Please enter the date (Month/Day/Year) your CFI signed your logbook entitling you to take the private pilot check ride (*not* the date you took the check ride).

   ____
3. Please enter the type of flight training program under which you predominantly trained for your private pilot license relative to the Federal Aviation Regulations (FARs).

   ❑ Part 61 ❑ Part 141 ❑ Military

   ❑ Other ______________________________
4. Please enter the type of organization under which you predominantly trained for your private pilot license.

   ❑ 2- or 4-year college/university ❑ Fixed-Base Operator (FBO)

   ❑ Other ______________________________
5. Please enter the type of training schedule under which you predominantly trained for your private pilot license.

   ❑ Mandated Scheduling (I had no choice over the number of days per week I could fly.)

   ❑ Student-Driven Scheduling (I was able to choose the number of days per week I could fly.)

   ❑ Other ______________________________
6. Please specify the number of hours (including simulator hours) you had when your CFI approved you for your check ride. __________________________
7. Please specify the number of hours (including simulator hours) you had when you actually took the check ride. ___________________________

Appendix B

Raw Data

Table A1

Raw Data

Row	Case[a]	FH[b] (*Y*)	DTP[c] (X_1)	FTP[d] (X_2)	Org.[e] (X_3)	SP[f] (X_4)	Sex[g] (X_5)	Age[h] (X_6)	Race[i] (X_7)	MS[j] (X_8)	Eth.[k] (X_9)
1	37	44.1	12	141	FBO	M	M	25	Not W	Not M	Not H or L
2	94	37.5	51	141	College	M	M	18	W	Not M	Not H or L
3	133	36.6	54	141	FBO	SD	M	22	W	Not M	Not H or L
4	101	84.5	56	61	FBO	M	M	23	W	Not M	Not H or L
5	24	82	58	141	FBO	M	M	19	W	Not M	Not H or L
6	65	65	62	61	FBO	SD	M	23	Not W	Not M	H or L
7	16	266.9	67	61	FBO	SD	M	22	W	Not M	Not H or L
8	172	63	70	61	FBO	SD	M	16	W	Not M	Not H or L
9	237	56.4	75	61	FBO	SD	M	19	Not W	M	H or L
10	88	60	76	61	FBO	SD	F	19	W	Not M	Not H or L
11	15	41	78	61	FBO	SD	M	31	W	M	Not H or L
12	103	80	79	61	FBO	SD	M	22	W	Not M	Not H or L
13	192	42	79	141	College	SD	M	20	W	Not M	Not H or L
14	206	35.2	88	141	College	M	M	18	W	Not M	Not H or L
15	45	65	91	61	FBO	SD	M	18	W	M	Not H or L
16	62	45	91	141	College	SD	M	19	Not W	M	Not H or L
17	161	40	91	61	FBO	SD	M	20	W	M	Not H or L
18	223	44	91	61	College	M	M	18	W	Not M	Not H or L
19	126	50	92	61	FBO	SD	F	21	Not W	M	Other
20	252	63.2	96	61	FBO	SD	M	24	W	M	Not H or L
21	278	44.1	101	61	FBO	SD	F	21	W	Not M	Not H or L
22	190	64.3	102	61	FBO	SD	F	23	W	Not M	Not H or L
23	60	58	107	61	FBO	SD	M	16	W	Not M	Not H or L
24	138	45	107	141	FBO	M	M	19	W	Not M	Not H or L
25	273	43.4	110	61	FBO	SD	M	27	Not W	Not M	Not H or L
26	202	55.7	113	141	College	M	F	18	Not W	Not M	Not H or L
27	191	43	115	141	College	M	M	20.2	W	Not M	Not H or L
28	92	40	117	61	FBO	SD	M	18	W	Not M	Not H or L
29	238	43	118	141	College	SD	M	17	W	Not M	Not H or L
30	159	52	123	141	FBO	SD	M	35	W	Not M	Not H or L

Note. [a]Case = Original case numbers. [b]FH = Flight Hours. [c]DTP = Days to Proficiency. [d]FTP = Flight Training Program. [e]Org. = Organization. [f]SP = Scheduling Policy where M = Mandated and SD = Student-driven. [g]Sex = Biological sex at birth where F = Female and M = Male. [i]Race: W = White and Not W = Not White. [j]MS = Marital Status where M = Married and Not M = Not Married. [k]Eth. = Ethnicity where H or L = Hispanic or Latino, Not H or L = Not Hispanic or Latino, Other = the participant selected "other" but did not specify, and an empty cell indicates the participant did not respond to this item.

Table A1 (Continued)

Raw Data

Row	Case[a]	FH[b] (Y)	DTP[c] (X_1)	FTP[d] (X_2)	Org.[e] (X_3)	SP[f] (X_4)	Sex[g] (X_5)	Age[h] (X_6)	Race[i] (X_7)	MS[j] (X_8)	Eth.[k] (X_9)
31	244	59	123	141	FBO	SD	M	19	W	Not M	Not H or L
32	131	51	124	141	College	M	M	18	Not W	Not M	H or L
33	177	42	129	141	College	M	M	18	W	Not M	Not H or L
34	64	71	130	61	FBO	SD	M	29	W	Not M	H or L
35	141	48	132	61	FBO	SD	M	20.2	W	Not M	Not H or L
36	117	50.8	134	141	College	M	M	19	W	Not M	Not H or L
37	217	47.6	134	141	FBO	SD	M	17	W	M	Not H or L
38	260	43	134	61	FBO	SD	M	17	Not W	Not M	Not H or L
39	236	79.4	135	61	FBO	M	M	24	W	Not M	Not H or L
40	104	65	138	61	FBO	SD	M	20.2	W	M	Not H or L
41	216	52.5	140	141	College	SD	M	19	W	Not M	Not H or L
42	219	49.9	140	141	FBO	SD	F	17	W	Not M	Not H or L
43	108	40	142	61	College	SD	M	23	W	Not M	Not H or L
44	137	48	145	61	FBO	SD	M	16	W	Not M	Not H or L
45	91	35	149	141	College	SD	M	53	W	Not M	Not H or L
46	95	44.3	153	61	FBO	SD	M	22	W	Not M	Not H or L
47	164	57.1	157	61	FBO	SD	M	24	W	Not M	Not H or L
48	222	75	160	61	FBO	SD	F	25	W	Not M	Not H or L
49	30	65	163	141	FBO	SD	M	17	W	Not M	Not H or L
50	49	70	164	61	FBO	SD	M	32	Not W	Not M	Other
51	271	50.2	169	61	FBO	SD	M	17	W	Not M	Not H or L
52	224	65	173	141	College	M	M	17	Not W	Not M	Not H or L
53	232	55.1	174	141	College	M	F	18	W	Not M	Not H or L
54	14	53	176	61	FBO	SD	F	22	W	Not M	Not H or L
55	213	59.7	178	61	FBO	SD	M	20.2	W	M	Not H or L
56	209	52.3	181	141	College	M	F	18	W	Not M	Not H or L
57	34	50	186	141	College	SD	M	18	W	M	Not H or L
58	146	49.9	186	61	FBO	SD	M	18	W	Not M	Not H or L
59	155	59	189	141	College	SD	M	18	W	Not M	Not H or L
60	230	94.3	198	141	College	M	M	17	W	Not M	Not H or L

Note. [a]Case = Original case numbers. [b]FH = Flight Hours. [c]DTP = Days to Proficiency. [d]FTP = Flight Training Program. [e]Org. = Organization. [f]SP = Scheduling Policy where M = Mandated and SD = Student-driven. [g]Sex = Biological sex at birth where F = Female and M = Male. [i]Race: W = White and Not W = Not White. [j]MS = Marital Status where M = Married and Not M = Not Married. [k]Eth. = Ethnicity where H or L = Hispanic or Latino, Not H or L = Not Hispanic or Latino, Other = the participant selected “other” but did not specify, and an empty cell indicates the participant did not respond to this item.

Table A1 (Continued)

Raw Data

Row	Case[a]	FH[b] (*Y*)	DTP[c] (X_1)	FTP[d] (X_2)	Org.[e] (X_3)	SP[f] (X_4)	Sex[g] (X_5)	Age[h] (X_6)	Race[i] (X_7)	MS[j] (X_8)	Eth.[k] (X_9)
61	231	71	199	141	College	M	M	18	W	Not M	Not H or L
62	21	68.5	200	141	FBO	M	M	25	W	Not M	Not H or L
63	9	60.9	203	61	FBO	SD	M	23	W	Not M	Not H or L
64	185	57	206	61	College	M	M	19	W	Not M	Not H or L
65	96	76.3	218	141	College	SD	M	20	W	Not M	Not H or L
66	143	51.7	221	141	College	SD	M	21	W	Not M	Not H or L
67	174	48.4	222	61	FBO	SD	M	17	W	Not M	Not H or L
68	267	46	222	61	FBO	SD	M	19	W	Not M	Not H or L
69	149	52.7	225	61	FBO	SD	M	20.2	W	Not M	Not H or L
70	262	56.8	226	141	College	SD	M	19	W	Not M	Not H or L
71	87	81	229	61	FBO	SD	M	42	W	M	Not H or L
72	53	75.9	231	61	FBO	SD	M	19	W	Not M	Not H or L
73	50	52	236	141	College	M	M	20.2	W	Not M	Not H or L
74	268	76.9	237	141	College	M	M	19	W	Not M	Not H or L
75	227	58.4	238	61	FBO	SD	M	18	W	Not M	Not H or L
76	218	49.9	245	141	College	M	M	17	Not W	Not M	H or L
77	249	58	245	61	FBO	SD	M	28	W	Not M	Not H or L
78	270	55	245	141	College	SD	M	25	W	Not M	Not H or L
79	25	98	247	141	FBO	SD	M	30	W	Not M	Not H or L
80	86	42	247	61	FBO	SD	M	18	W	Not M	Not H or L
81	119	74.7	248	141	College	M	M	18	W	Not M	Not H or L
82	142	40	247	61	FBO	SD	M	19	W	Not M	Not H or L
83	158	60	247	141	FBO	SD	M	18	Not W	Not M	Not H or L
84	165	60	247	61	FBO	SD	M	22	W	Not M	Not H or L
85	199	43.4	247	141	College	M	M	20	W	Not M	Not H or L
86	201	80	247	61	FBO	M	M	20.2	W	Not M	Not H or L
87	210	60	247	61	FBO	SD	M	51	W	M	Not H or L
88	233	62.7	247	141	College	SD	M	18	W	Not M	Not H or L
89	43	48.4	249	141	College	M	M	18	W	Not M	Not H or L
90	208	40	249	61	FBO	SD	M	25	W	Not M	Not H or L

Note. [a]Case = Original case numbers. [b]FH = Flight Hours. [c]DTP = Days to Proficiency. [d]FTP = Flight Training Program. [e]Org. = Organization. [f]SP = Scheduling Policy where M = Mandated and SD = Student-driven. [g]Sex = Biological sex at birth where F = Female and M = Male. [i]Race: W = White and Not W = Not White. [j]MS = Marital Status where M = Married and Not M = Not Married. [k]Eth. = Ethnicity where H or L = Hispanic or Latino, Not H or L = Not Hispanic or Latino, Other = the participant selected “other” but did not specify, and an empty cell indicates the participant did not respond to this item.

Table A1 (Continued)

Raw Data

Row	Case[a]	FH[b] (Y)	DTP[c] (X_1)	FTP[d] (X_2)	Org.[e] (X_3)	SP[f] (X_4)	Sex[g] (X_5)	Age[h] (X_6)	Race[i] (X_7)	MS[j] (X_8)	Eth.[k] (X_9)
91	47	65	253	61	FBO	SD	M	33	Not W	Not M	Not H or L
92	220	74	253	141	FBO	SD	M	18	W	Not M	
93	46	58	255	141	College	M	M	18	W	Not M	Not H or L
94	193	100	257	141	College	M	M	16	Not W	M	H or L
95	207	72	266	141	FBO	SD	M	30	Not W	Not M	Not H or L
96	250	60	272	61	FBO	SD	M	21	W	Not M	Not H or L
97	79	51	274	61	FBO	SD	M	21	W	Not M	Other
98	26	45	275	61	FBO	SD	M	20	W	Not M	Not H or L
99	61	78	287	141	College	M	M	20	W	Not M	Not H or L
100	247	58.7	297	61	FBO	SD	F	16	W	Not M	Not H or L
101	78	62	298	61	FBO	SD	F	20.2	W	Not M	Not H or L
102	52	48	299	61	FBO	SD	M	17	W	Not M	Not H or L
103	18	84	302	141	College	SD	M	18	Not W	Not M	Not H or L
104	251	68	305	61	FBO	SD	F	18	W	Not M	Not H or L
105	194	55.6	309	61	FBO	SD	M	15	W	Not M	Not H or L
106	157	88	318	141	College	SD	M	18	Not W	M	H or L
107	7	66.2	321	61	FBO	SD	M	20	W	M	Not H or L
108	144	71	330	61	FBO	SD	M	18	W	Not M	Not H or L
109	182	67	343	61	FBO	SD	F	16	W	Not M	Not H or L
110	66	55	346	61	FBO	SD	M	16	W	M	Not H or L
111	160	64	347	61	FBO	SD	M	16	Not W	Not M	Not H or L
112	106	55	360	141	College	SD	F	18	Not W	Not M	Not H or L
113	41	77.7	372	61	FBO	SD	M	18	W	Not M	Not H or L
114	187	65.8	390	61	FBO	SD	M	16	W	Not M	Not H or L
115	36	99	395	141	FBO	SD	F	16	W	Not M	Not H or L
116	107	53.6	402	141	College	M	M	20	W	Not M	Not H or L
117	272	50.5	402	61	FBO	SD	M	17	W	Not M	Not H or L
118	93	40	405	141	College	SD	M	21	Not W	Not M	H or L
119	136	69.75	406	61	College	M	M	21	W	Not M	Not H or L
120	239	62.7	413	141	FBO	SD	M	20.2	W	Not M	Not H or L

Note. [a]Case = Original case numbers. [b]FH = Flight Hours. [c]DTP = Days to Proficiency. [d]FTP = Flight Training Program. [e]Org. = Organization. [f]SP = Scheduling Policy where M = Mandated and SD = Student-driven. [g]Sex = Biological sex at birth where F = Female and M = Male. [i]Race: W = White and Not W = Not White. [j]MS = Marital Status where M = Married and Not M = Not Married. [k]Eth. = Ethnicity where H or L = Hispanic or Latino, Not H or L = Not Hispanic or Latino, Other = the participant selected "other" but did not specify, and an empty cell indicates the participant did not respond to this item.

Table A1 (Continued)

Raw Data

Row	Case[a]	FH[b] (Y)	DTP[c] (X_1)	FTP[d] (X_2)	Org.[e] (X_3)	SP[f] (X_4)	Sex[g] (X_5)	Age[h] (X_6)	Race[i] (X_7)	MS[j] (X_8)	Eth.[k] (X_9)
121	234	64	422	61	FBO	SD	M	17	W	Not M	Not H or L
122	121	48	427	61	FBO	SD	M	16	W	M	Not H or L
123	183	44	428	61	FBO	SD	M	18	Not W	Not M	H or L
124	44	60	431	61	FBO	SD	M	17	W	Not M	Not H or L
125	156	63.5	431	61	FBO	SD	M	16	W	Not M	Not H or L
126	125	57	456	61	FBO	SD	M	22	W	Not M	Not H or L
127	17	86.8	471	61	FBO	SD	M	16	W	Not M	Not H or L
128	269	64.7	472	61	FBO	SD	M	22	Not W	Not M	H or L
129	140	68	490	61	College	M	M	16	W	Not M	Not H or L
130	173	75	499	61	FBO	SD	M	17	Not W	Not M	Not H or L
131	169	54	503	141	College	M	M	22	W	Not M	Not H or L
132	256	51.9	519	61	FBO	SD	M	20.2	W	Not M	Not H or L
133	153	58	523	61	FBO	M	M	17	W	Not M	Not H or L
134	196	52	529	61	FBO	SD	M	16	W	Not M	Not H or L
135	242	65	542	61	FBO	SD	M	16	Not W	Not M	Not H or L
136	198	61.4	576	61	FBO	SD	M	18	W	Not M	Not H or L
137	23	89.1	589	61	FBO	SD	M	19	Not W	Not M	Not H or L
138	31	74	592	61	FBO	SD	F	16	W	Not M	Not H or L
139	127	86	637	61	College	M	M	16	W	Not M	Not H or L
140	254	70	649	61	FBO	SD	M	16	Not W	Not M	Not H or L
141	257	54.6	649	61	FBO	SD	M	15	W	Not M	Not H or L
142	245	72.5	660	141	FBO	M	M	23	W	Not M	Not H or L
143	264	77	660	61	FBO	SD	M	15	W	M	Not H or L
144	176	85.3	665	141	College	SD	M	19	W	Not M	Not H or L
145	28	77.7	685	141	College	M	M	33	W	M	Not H or L
146	170	54.4	690	61	FBO	SD	M	25	W	Not M	Not H or L
147	11	116.8	745	141	College	M	M	19	W	Not M	Not H or L
148	255	73.2	761	61	FBO	SD	M	16	W	Not M	Not H or L
149	67	107	768	61	FBO	SD	M	15	W	Not M	Not H or L
150	265	80	784	141	College	M	M	18	W	Not M	Not H or L

Note. [a]Case = Original case numbers. [b]FH = Flight Hours. [c]DTP = Days to Proficiency. [d]FTP = Flight Training Program. [e]Org. = Organization. [f]SP = Scheduling Policy where M = Mandated and SD = Student-driven. [g]Sex = Biological sex at birth where F = Female and M = Male. [i]Race: W = White and Not W = Not White. [j]MS = Marital Status where M = Married and Not M = Not Married. [k]Eth. = Ethnicity where H or L = Hispanic or Latino, Not H or L = Not Hispanic or Latino, Other = the participant selected "other" but did not specify, and an empty cell indicates the participant did not respond to this item.

Table A1 (Continued)

Raw Data

Row	Case[a]	FH[b] (Y)	DTP[c] (X_1)	FTP[d] (X_2)	Org.[e] (X_3)	SP[f] (X_4)	Sex[g] (X_5)	Age[h] (X_6)	Race[i] (X_7)	MS[j] (X_8)	Eth.[k] (X_9)
151	100	40	839	141	FBO	SD	M	24	W	Not M	Not H or L
152	179	57.9	847	61	FBO	SD	M	15	W	Not M	Not H or L
153	253	94	866	61	FBO	SD	F	23	Not W	Not M	H or L
154	248	50	913	141	College	SD	M	16	W	Not M	Not H or L
155	10	55	944	61	FBO	SD	M	15	W	Not M	Not H or L
156	81	86	1063	61	FBO	M	M	16	W	Not M	Not H or L
157	150	64	1160	61	FBO	SD	M	15	W	Not M	Not H or L
158	258	81	1271	61	FBO	SD	M	16	W	Not M	Not H or L
159	243	50.5	1359	141	FBO	SD	M	15	W	Not M	Not H or L
160	51	96.2	1454	61	FBO	SD	M	41	Not W	Not M	H or L
161	40	82	1851	61	FBO	SD	M	13	W	Not M	Not H or L
162	20	82	2635	61	FBO	SD	M	13	W	Not M	Not H or L
163	6	65	2934	61	FBO	SD	M	18	W	Not M	Not H or L
164	241	47	3788	61	College	M	M	18	W	Not M	Not H or L
165	211	66	3946	61	FBO	SD	M	19	W	Not M	Not H or L
166	180	61.7	6748	61	College	SD	M	15	W	M	Not H or L
167	166	220	8824	61	FBO	SD	M	28	W	M	Other

Note. [a]Case = Original case numbers. [b]FH = Flight Hours. [c]DTP = Days to Proficiency. [d]FTP = Flight Training Program. [e]Org. = Organization. [f]SP = Scheduling Policy where M = Mandated and SD = Student-driven. [g]Sex = Biological sex at birth where F = Female and M = Male. [i]Race: W = White and Not W = Not White. [j]MS = Marital Status where M = Married and Not M = Not Married. [k]Eth. = Ethnicity where H or L = Hispanic or Latino, Not H or L = Not Hispanic or Latino, Other = the participant selected “other” but did not specify, and an empty cell indicates the participant did not respond to this item.