Blended and Adaptive Learning in Ground School Instruction for Aviators

Patrick L. Craven
Lockheed Martin Corporation
Orlando, FL
patrick.craven@lmco.com

ABSTRACT

This paper presents a study investigating adaptive e-learning software as part of a blended learning approach in the Aerodynamics segment of a military flight training ground school. Thirty-four trainees participated in this investigation where existing educational materials were converted to an adaptive e-learning platform with the goal of keeping the instructional content equivalent. New questions, or probes, were developed for each learning objective, and responses to these probes served as input to an adaptive engine which personalized the learning. Students were exposed to the instructional content multiple times in seemingly randomized fashion until demonstrating both accuracy and confidence in each learning objective. An instructor then delivered an abbreviated lecture meant to expand on the foundational knowledge delivered via e-learning. The historical performance of approximately 200 students who were taught using only lectures was used as a control group. All students received a multi-hour exam following completion of the topic. Additionally, 21 instructor-led and 24 blended-learning students also received a surprise exam about 10 weeks after the initial exam in order to measure how the two instructional methods influence the rate of knowledge decay.

Initial exam performance was similar for both groups. However, when knowledge was measured 10 weeks later the blended learning group experienced only a 14.5% drop compared with the instructor-led group who experienced a 27.3% drop, which was a significant difference. Despite almost all students in both groups passing the initial exam, the practical impact was only 29% of students in the instructor-led group would have achieved a passing score at 10 weeks compared with 63% in the blended learning group. If any of the lost knowledge is important to later skill development, then these results suggest that blended learning may offer a path for improved knowledge retention that could benefit future aviators in subsequent training.

ABOUT THE AUTHOR

Patrick L Craven, Ph.D., is a human factors design engineer with Lockheed Martin where he focuses on developing human-centered technologies for both training and operations. His recent work has focused on both building data-driven algorithms for personalizing human-technology interaction as well as fostering improved manned-unmanned teaming. He has nearly two decades of professional experience investigating, designing, and evaluating physical, cognitive, and neurological interactions among humans and machines. Dr. Craven has a PhD in Psychology from The Pennsylvania State University, an MSc in Cognitive Science from the University of Edinburgh, and a BA is Cognitive Science from Dartmouth College. He is a board certified human factors professional who also serves as a director on the Board of Certification in Professional Ergonomics. In this role he serves as Secretary on the executive council as well as chairs the certification committee which oversees the process of defining assessment standards for applicants.

Blended and Adaptive Learning in Ground School Instruction for Aviators

Patrick L. Craven
Lockheed Martin Corporation
Orlando, FL
patrick.craven@lmco.com

INTRODUCTION

This paper presents the results of a study where a blended learning approach to teach Aerodynamics was compared to a traditional-instructor-led approach at a military flight training ground school in order to understand the efficacy of this approach in that environment. A growth in understanding how students learn has paved the way for advances in pedagogical practice, including blended learning which combines aspects of instructor-led with technology-based education. Benjamin Bloom (1984) famously quantified the benefits of tutoring, or one-to-one instruction, as two standard deviations better than conventional learning, which remains a common approach where students are given equal amounts of time to learn the material. In-between conventional learning and tutoring lies mastery learning, which focuses on each student reaching the same level of achievement, even if it takes different amounts of time to achieve the same level of skill. Mastery learning has been implemented in educational environments such as medical school (Barsuk, Cohen, Wayne, McGaghie, & Yudkowsky, 2018) which uses a method of students working through a simulator until a standard is achieved before working directly with patients, and has also been implemented through strategies already employed in K-12 education (Guskey, 2010; Zimmerman & Dibenedetto, 2008). A variety of pedogogical processes and e-learning tools have been used in an effort to move educational outcomes closer to those observed under tutoring conditions, and many of them include mastery learning because of the inherent benefits.

Technology has provided an exciting opportunity to seamlessly incorporate quantitative measures of student behaviors and learning, which is an important component for recognizing topic mastery. Although there are more than 23 different terms that have been used to describe technological instructional tools (Manuela, Fernando, & Tiago, 2016), *e-learning* is used in this paper to describe the use of technological tools to facilitate instruction. A specific type of e-learning called *adaptive learning* aims to replicate the individualized nature of one-to-one tutoring by using the scaling potential of software distributed among numerous computer workstations (Vandewaetere, Desmet, & Clarebout, 2011). Adaptive learning is comprised of a feedback loop with mechanisms for assessing the learner and then pairing that assessment with either immediate modifications within the learning experience (*micro-adaptation*), or modifications of the learning path at the topic level (*macro-adaptation*) (Durlach & Spain, 2014; Vandewaetere et al., 2011). Macro-adaptive instruction is focused on the pace or complexity of the learning tasks that are offered next, whereas micro-adaptive instruction provides a dynamic experience that supports the needs of learners within a task (Vandewaetere et al., 2011).

Although an adaptive e-learning platform might provide a powerful mechanism to reach students and provide some benefits of a human tutor, the human instructor retains a significant role in educating aviators. The research team partnered with an ab initio military flight training program to understand this dynamic, and it began with an analysis of current practice through conversations with the instructional staff and leadership. This revealed that a traditional lecture-based approach was used as the primary means to educate future pilots. For example, instructors used a combination of slides presented at the front of the class coupled with verbal explanations. Instructors worked through a defined sequence of 50-minute long topics from 08:00 to 17:30, and this approach raised two concerns. First, a near-continuous lecture like this has been shown to have detrimental effects on learning (Risko, Anderson, Sarwal, Engelhardt, & Kingstone, 2012). The second concern was the use of the instructors' expertise wasn't optimized as it could be. For example, it is challenging for an instructor to know how to enrich a lecture when there is uncertainty with how well the students prepared. For example, although students are provided a digital textbook, there is no assigned time to read it. When asked, students shared that they primarily use it in the evenings as part of preparing for taking exams. Thus, it was hypothesized that preparation before the lectures might provide better learning by allowing the instructor to elaborate on concepts of which students already had a basic grasp.

The research team hypothesized that the *flipped classroom* pedagogical technique would benefit learning outcomes (Chen et al., 2018). This technique is a type of blended learning, which can be one of many different combinations of e-learning and face-to-face instructional activities (Bonk, Kim, & Zeng, 2005). In a flipped classroom approach students engage in content learning before class which maximizes in-class time for active learning (DeLozier & Rhodes, 2017). With this approach students first focus on the foundational levels of Bloom's taxonomy, knowledge and comprehension (Furst, Hill, Krathwohl, Bloom, & Engelhart, 1956). Instructors could then focus their time on building towards application (level three) by using their flight experience to better illustrate that. As there wasn't a feasible way to control or measure the learning outside of the classroom environment, the favored approach was to modify the lecture schedule and create in-class time for students to work through the foundational knowledge material at an individual pace. Instructors would then lead the class through an abbreviated instructor-led session focused on the application of the knowledge. Thus, using adaptive learning with lectures created a blended instructional approach that was studied as part of four research objectives exploring the effects of this approach on student performance.

Objective 1: Determine the effect of blended learning on immediate knowledge acquisition

The purpose of this objective was to investigate how a blended learning approach could aid a learner in acquiring knowledge. Since blended learning was being contrasted with a traditional instructor-led only approach, the hypothesis was that a blended learning approach would provide an advantage and be observable in the exam scores.

Objective 2: Determine the effect of blended learning on knowledge retention

The purpose of this objective was to investigate how the decay rate of knowledge might differ between blended learning and instructor-led only. Similar to Objective 1, but the focus was on how much knowledge was retained 10 weeks after the topic finished. It was similarly hypothesized that blended learning would provide an advantage over the instructor-led approach. This is a key measure as there is a weeks-long delay in this flight training program between when a student first learns a concept and when that student gets to put the concept into skill-based practice. In order to investigate this there was a 10-week delay before an outgoing class was given a surprise exam that was used as comparison data for a similarly surprised blended learning group.

Objective 3: Determine predictive power of adaptive e-learning process measures

The purpose of this objective was to understand whether any of the built in measures that underlie the adaptive engine also provided predictive power for learning outcomes. It was hypothesized that the e-learning process measures would not only be useful as input to the adaptive engine but also as a predictor of student performance.

Objective 4: Determine students' and instructors' perceptions of blended learning and e-learning usability

The purpose of this objective was to examine how students and instructors, as users of the system, found the experience of interacting with the adaptive learning software platform. It was hypothesized that students and instructors would see more positive than negative aspect of the blended learning approach.

METHOD

Participants

A cohort of 34 ground school students (33 males and one female) began the 13-week long ground school in March 2020. Twenty-one were training to be pilots, three were training to be Weapon System Officers (WSO), and 10 were training to be uninhabited aerial vehicle (UAV) pilots, and they formed the blended learning (BL) group. The average student age was 22.1 years old with a range of 20 to 27 years, and four of these students had previously obtained their private pilot's license (PPL). This BL group was compared to an instructor-led (IL) group comprised of 202 students from 11 previous cohorts who were taught using a traditional lecture-based approach. A sub-set of 21 IL students were still in residence in February 2020 and they were given a surprise exam to test long-term retention.

Materials

Aerodynamics is the first subject taught at ground school and was selected as the subject that would be converted to an adaptive learning version. The content is organized into four groupings called lessons that each contain several topics (see Table 1). Traditionally, each of the 18 topics are instructor-taught during a 50-minute class period with a few requiring more or less time. Students are given copies of both the courseware, which are the instructors' slides, and a student study guide, which serves as an electronic textbook.

Table 1. Lessons (bolded) and underlying topics

AD01: Principles of Flight	**AD03: Basic Maneuvering**
AD101: Principles of Flight Overview	AD301: Stability
AD102: Structure of the Atmosphere	AD302: Aerodynamic Forces in Flight Maneuvers
AD103: Newton's Three Laws of Motion	**AD04: Aerodynamics of Flight II**
AD104: Bernoulli's Principle of Pressure	AD401: High Speed Flight
AD105: Airfoil Design	AD403: Stalls
AD02: Aerodynamics of Flight I	AD404: Spinning
AD201: Four Forces	AD406: Basic Propeller Principles
AD202: Wingtip Vortices	AD407: Load Factors
AD204: Axes of an Airplane	AD408: Effects of Weight & Balance
AD205: Moment and Moment Arms	

Traditional Courseware

Instructors use content developed in Adobe Flash that includes a combination of text along with static and animated images. A top-level menu provides access to all of the Aerodynamics topics. Once a topic is open, one uses the next or back commands to navigate through the topic in a manner very similar to advancing slides in Microsoft PowerPoint.

Adaptive Learning Courseware

The e-learning platform used in this study is called Rhapsode (Area9 Lyceum, n.d.). This internet-accessible platform is housed on an Amazon Web Services cloud server, and the software includes role-based access with dedicated interfaces for students, instructors, and courseware developers. A team of courseware developers converted the traditional Aerodynamics courseware into adaptive modules for each topic and made use of much of the same text and graphics from the traditional courseware. Topic modules are built around a series of learning objectives that are identified by the content developers. Typically, a module might have between about 10 and 30 learning objectives. Developers created novel *probes* or questions to test learners' knowledge of the content (see Figure 1). These probes can take a variety of forms ranging from multiple-choice, fill-in-the-blank, crossword, matching, and more. Novel probes were generated for this study and were compared with the exam questions to ensure that the question stems and answers were sufficiently unique. Additionally, *learning resources* were created for each of the learning objectives, and these resembled traditional slides as well as animations or resource documents. A virtual coach on the left side of the screen displays text and presents it auditorily by default with a text to speech synthesizer. In building the modules, the goal was to provide no educational advantage by presenting more detail or information in the adaptive courseware compared with the traditional courseware. Thus, only the courseware and the Student Study Guide (SSG) were used to create the adaptive courseware.

A student's performance on the knowledge probes, self-assessment of confidence, and any constraints written by developers provide input to a learning engine algorithm that determines the system behavior. Additionally, a user can set a self-assessment of knowledge on a five-point scale, and the participants were instructed to start that at the *advanced beginner* level. Typically, a learner is given a knowledge probe first to determine whether they understand a learning objective. Additionally, a learner must also rate their confidence on a four-part scale: a) I know it, b) Think I know it, c) Not sure, or d) No idea. The idea of rating one's confidence is rooted in earlier work (Finetti, 1965; Gardner-Medwin, 1995) demonstrating the merit of capturing both a response and one's self-assessment of confidence in that response. If they answer correctly and have high confidence, then the associated instructional content is usually skipped. Rhapsode uses a proprietary adaptive learning engine, but what was observed of its behavior is that it presents learning objectives in a semi-random order (unless constrained during module authoring). It samples a learner's knowledge across objectives and then presents instructional content when a learner's knowledge probe or confidence requires it. Thus, it's a form of mastery learning where time varies for each student based on individual need, and learners only complete a module once all objectives are satisfied with appropriate competence and confidence.

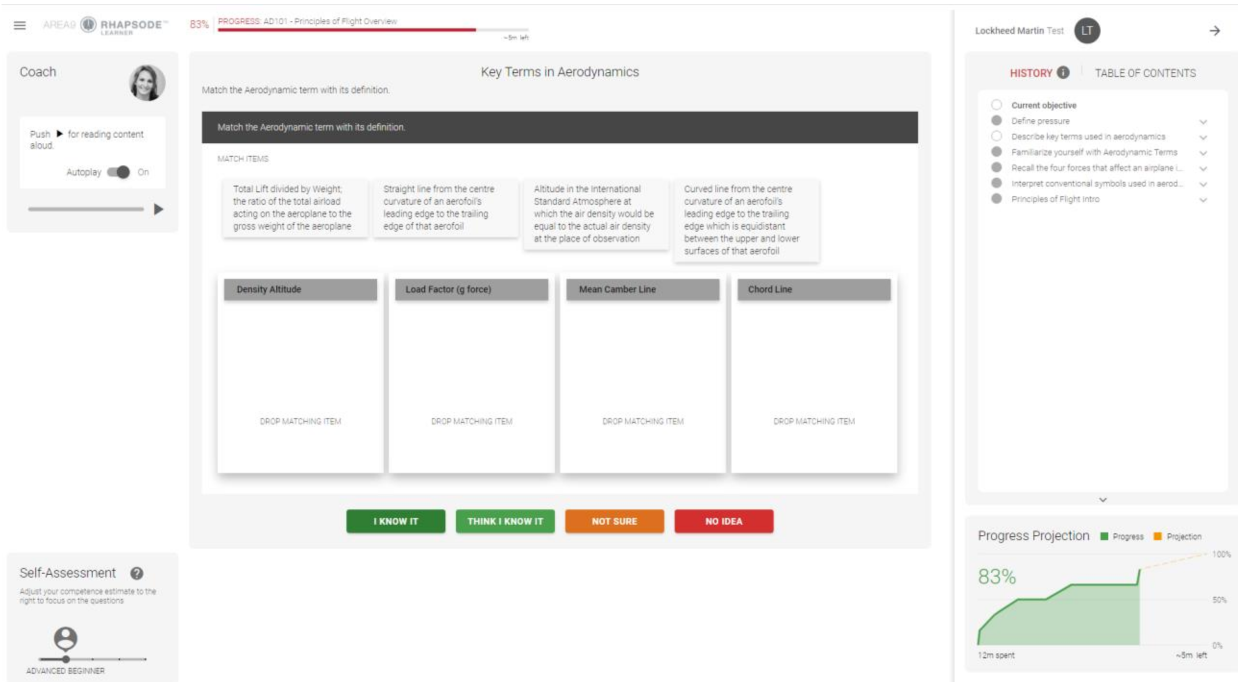

Figure 1. Adaptive e-learning platform user interface with a question probe

Student Study Guide

The SSG for Aerodynamics consists of four volumes organized by lesson (AD01, AD02, AD03, and AD04) and these volumes are 57, 52, 35, and 82 pages long respectively. The SSG plays a role similar to a textbook and typically includes more detail and description than the instructor-led courseware, but it lacks animations to illustrate concepts like roll, pitch, and yaw. Thus, the SSG served as another content source for the adaptive learning courseware development team. Additionally, it should be noted that students are provided with copies of the SSG on the laptops they are loaned for the course.

Laptops

All students are provided laptops loaded with a study guide and traditional courseware, and the BL group was told they could make use of them as needed. Docking stations for each student with an external monitor and wired ethernet for the adaptive courseware are provided in the classroom. Students use both the laptop and external monitors for increased screen real estate. Additionally, students were provided with audio headsets so that the text spoken by the coach would not disrupt other students.

Exam

The exam served as the measure of knowledge. Since there was a large volume of historical data available from previous classes, it was desirable to use the existing exam forms in order to directly compare the BL group with historical performance from the IL group. The exam is a computer-based assessment comprised of 20 multiple-choice questions (MCQ) worth two points apiece, 10 fill-in-the-blank (FIB) worth three points apiece, and five essay questions worth six points apiece for a maximum score of 100. Although the MCQ and FIB are automatically scored, the essay responses are hand scored by the instructors using a grading rubric. Therefore, a modified version of the exam that included only MCQ and FIB question were given to a sub-set of participants to measure knowledge decay.

System Usability Scale and Questionnaire

Students were asked to complete the System Usability Scale (SUS) (Brooke, 1986), which consists of ten questions that are answered using a 5-point Likert scale. Questions are clustered as positive questions (e.g., "I think I would like to use this system") for questions 1, 3, 5, 7, and 9. Conversely, questions 2, 4, 6, 8, and 10 are negatively phrased (e.g., I found the system unnecessarily complex). If students found the system positive, one would expect them to respond with scores greater than 3 for the positively phrased odd-numbered questions and less than 3 for the negatively phrased even-numbered questions. Students were also asked a series of additional questions to determine whether they held a private pilot license (PPL), asked to rate any previous Aerodynamics knowledge, and asked open-ended usability questions about what they liked the most and least in interacting with the adaptive courseware.

E-Learning Process Measures

The e-learning platform records data based on learner behavior and produces several measures that are used by the adaptive engine. These measures were exported from the system after the study in order to explore whether they are related to learning outcomes. The three principal measures explored were:

- **Metacognition**: percentages for the four possible combinations of correctness and confidence:
    - Conscious competent: answering correctly, and knowing that you did so
    - Unconscious competent: answering correctly, but uncertain you knew it
    - Conscious incompetent: answering incorrectly, but knowing you didn't know the correct answer
    - Unconscious incompetent: answering incorrectly, but believing you were correct
- **Time spent learning**: a measure of active time engaged with the e-learning platform
- **Number of incorrect probes**: Probes are the questions asked for each learning objective, so this is just a measure of the number of times probes were answered incorrectly.

Design and Procedure

The study was designed as a between-subjects design with the BL group performance being compared with historical student performance from the IL group. As there was an inability to run groups simultaneously, the traditional exam outcome measure was used as the primary means of performance. In order to measure knowledge retention and decay, a surprise exam was given 10 weeks after being taught Aerodynamics and only a few weeks before leaving ground school. These students had no prior knowledge that they would be receiving the exam, and therefore had no opportunity to prepare for it. Due to schedule differences, the UAV pilot participants did not participate in the surprise exam in either the BL or IL group.

This instruction was comprised of lectures given over nine periods from 08:00 to 17:30 each day. Figure 2 shows how the topics within Aerodynamics were taught during 1050 instructional minutes spread across three days. The gray boxes represent other activities which are most often other topics, but sometimes consist of other ground school activities such as physical training, medical exams, or testing G-force tolerance. In this model, the instructor lectures are typically the first time students see topic information. For example, on the first day instructors cover content from all of the AD01 topics and one AD02 topic which is covered across 78 pages in the SSG. Students wanting to read in advance of the second day topics would also need to read an additional 83 pages. Students are encouraged to take notes during lectures in their paper notebook or electronically on their student laptop. Instructors use the courseware to guide their presentation of the material. They use their subject matter expertise to elaborate on content by describing real-world examples or using hand-held aircraft models to illustrate concepts. Instructors encourage students to ask questions, but it is unclear how often students use that opportunity. Finally, instructors ask students direct questions in an effort to probe their understanding as well as foster greater interaction, but this can be done only one student at a time.

	PERIOD 1 0800-0850	PERIOD 2 0855-0945	PERIOD 3 1000-1050	PERIOD 4 1055-1145	PERIOD 5 1245-1335	PERIOD 6 1340-1430	PERIOD 7 1435-1525	PERIOD 8 1545-1635	PERIOD 9 1640-1730
Mon			AD101	AD102	AD102	AD103	AD104	AD105	AD201
Tue	AD201/AD202	AD204	AD205	AD301	AD301	AD302	AD302	AD401	AD401
Wed									
Thu									
Fri		AD403	AD404	AD406	AD407	AD408			
Mon							AD Revision		
Tue									
Wed	AD Exam								

Figure 2. Original Aerodynamics Schedule

The blended combination of adaptive learning followed by an abbreviated lecture was designed to fit within the same amount of time normally occupied by traditional instructor lectures. The time was split such that 65% of the time was allocated for e-learning, and 35% of the time for the lectures. Figure 3 shows the way in which e-learning (in red) preceded an abbreviated instructor-led lecture (marked as ILS). The BL group students were given a quick overview of the study which explained how the adaptive e-learning software worked, and students were provided with unique login information that kept their identities anonymous on the web server. They were asked to do their best to complete the modules indicated in the red boxes prior to the instructor-led sessions. Although not every student could complete the modules before the instructor began his session, most did. The class leader took responsibility for making sure that students completed any remaining modules between Monday evening and the start of the instructor-led session Tuesday, and between the end of Tuesday and the beginning of the instructor-led session on Friday morning. The class had access to and made use of the classroom in the evenings to catch up on any unfinished modules.

	PERIOD 1 0800-0850	PERIOD 2 0855-0945	PERIOD 3 1000-1050	PERIOD 4 1055-1145	PERIOD 5 1245-1335	PERIOD 6 1340-1430	PERIOD 7 1435-1525	PERIOD 8 1545-1635	PERIOD 9 1640-1730
Mon			Intro + AD101-AD105 (125min)		ILS (AD01) (75min)		AD201-AD205 (150min)		
Tue	Refresh	ILS (AD02) (75min)	AD301-AD302 (100min)			ILS (AD03) (75min)		AD401-AD404 (100min)	
Wed									
Thu									
Fri	ILS (AD04A) (75min)	AD406-AD408 (100min)		ILS (AD04B) (75min)					
Mon							AD Revision		
Tue									
Wed	AD Exam								

Figure 3. New Aerodynamics Schedule for Blended Learning

The abbreviated lectures were designed to cover a collection of topics (e.g., AD101-AD105 lesson). The instructors used the traditional courseware as part of the instructor-led session portion, but focused more on the application of knowledge (Bloom Level 3) rather than recall or understanding of concepts (Bloom Levels 1 & 2) (Bloom, Krathwohl, & Masia, 1984).

RESULTS

Thirty-three of the students completed the demographic information questionnaire. Responses showed that there were four students with private pilot's licenses (PPL). Students rated their prior knowledge in Aerodynamics on a scale of 1 (low) to 5 (high) with an average response of $M = 2.33$. Interestingly, the PPL students' self-assessment was statistically indistinguishable from the non-PPL students. Table 2 depicts students' ratings of their previous knowledge and shows that nearly four out of every five students reported a 1, 2, or 3 whereas only 21% reported a 4 or 5. This suggests that, based on self-rating, a majority of students did not enter the training feeling confident with the material. Knowing that a Dunning-Kruger effect could be present where the non-PPL students rated their knowledge higher than would be warranted, performance across MCQ, FIB, essay, and total exam scores were also compared and are shown on Table 3 (Kruger & Dunning, 1999). An independent samples *t*-test revealed a significant group difference for total exam scores and a marginal difference for essay scores. As no pre-test was administered one cannot be certain that PPL students were more knowledgeable at the start of the topic, but their exam results at the end were superior.

Table 2. Self-rating of prior Aerodynamics knowledge

Rating	PPL (4)
1 (low)	33.3%
2	24.2%
3	21.2%
4	18.2%
5 (high)	3.0%

Table 3. PPL versus non-PPL exam performance

Score	PPL (4)	Non-PPL (29)	Difference
MCQ (40)	39.00	37.59	1.41
FIB (30)	25.50	24.00	1.50
Essay (30)	23.50	21.28	2.22^{+}
Total (100)	88.00	82.76	5.24^{*}

* $p = .05$
\+ $p < .10$

Objective 1: Determine the effect of blended learning on immediate knowledge acquisition

The Aerodynamics exam performance of the 34 students in the BL group was compared to the historical data of 202 students in the IL group. Total exam scores and the underlying MCQ, FIB, and essay sub-scores were compared. Table

4 shows the mean sub-scores and total for the two groups, and the score labels indicate the maximum number of points per score (e.g., 40 for MCQ). Inspection of the MCQ histogram (see Figure 4) illustrate the scores are skewed for the MCQ questions with a likely ceiling effect. Both the FIB (Figure 5) and essay (Figure 6) histograms also show a skewed distribution, although the ceiling effect is not quite as pronounced. It is noteworthy that there were no perfect essay scores, and there appears to be a ceiling of approximately 26 of the possible 30 points.

Table 4. Mean scores for total exam score and sub-scores

Score	Instructor-led	Blended Learning	Difference
MCQ (40)	37.74	37.76	0.02
FIB (30)	25.40	24.18	-1.22
Essay (30)	22.28	21.55	-0.73
Total (100)	85.42	83.39	-2.03

Differences between BL and IL groups were small and trended towards BL scores being lower than IL scores which ran counter to the hypothesis. An independent samples t-test showed that total exam scores, MCQ, and essay sub-scores were not significantly different. However, the FIB score difference was significant ($t(56) = 2.33$, $p < .05$), and the magnitude of this difference (1.22 points) was slightly less than half the value of a FIB question (3 points apiece). The interpretation of these results is not straightforward, and more participant data would be needed to draw firmer conclusions. However, given the observation of ceiling effects in the distribution coupled with only one sub-score being significantly different, one can conclude that the BL and IL groups performed similarly well on the exam with a small advantage to the IL group on FIB questions.

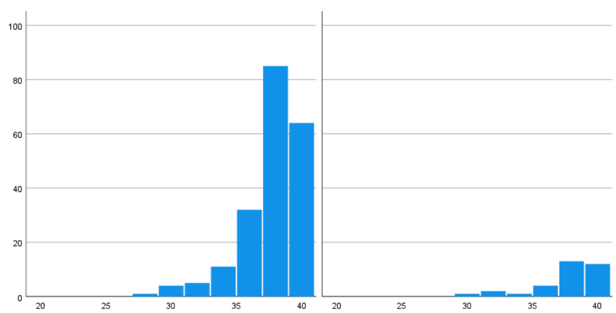

Figure 4. IL (left) and BL (right) MCQ histogram

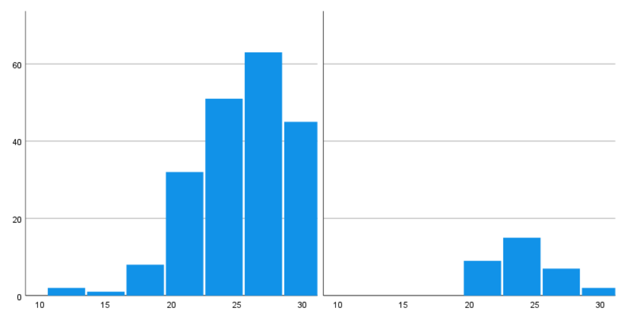

Figure 5. IL (left) and BL (right) FIB histogram

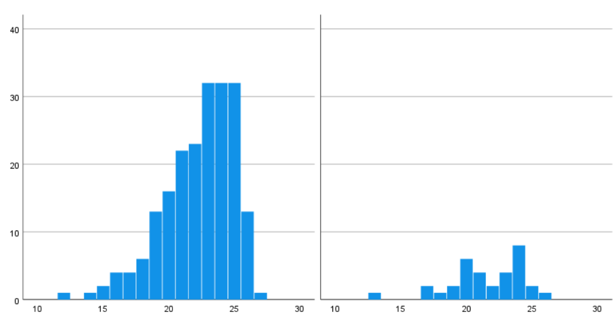

Figure 6. IL (left) and BL (right) Essay histogram

Objective 2: Determine the effect of blended learning on knowledge retention

An abbreviated surprise Aerodynamics exam was given 10 weeks later to both IL and BL students. Since UAV students had completed their portion of the training by that time, they did not participate in the surprise exam. Thus, 21 IL students (18 pilot trainees and three WSO) and 24 BL students (21 pilot trainees and three WSO) participated. On the original full exam (MCQ, FIB, and essay) with a cut score of 75, all 21 IL students passed and all but one (#24 in Figure 8) of the 24 BL students passed. As the surprise exam consisted of just the MCQ and FIB sections, there was a maximum score of 70 points with an adjusted cut score of 52.5. These surprise exam results were compared with only the MCQ and FIB portion of the original exam.

Figure 7 plots IL group's initial exam score (shown in yellow) compared to their surprise exam (shown in green). Although all of the students had a passing score on the original exam, only six of the 21 students (29%) achieved a passing score on the surprise exam 10 weeks later. Additionally, Student #20 and #15 were ½ or 1½ points, respectively, from the cut line and may not have passed if the essay portion had been included. As noted earlier, students typically lose about a point per essay question even if they scored high or perfectly on the MCQ and FIB sections. Given that the essay portion average score was 22.3 points and student #20 scored 51 points and student #15 scored 52 points, an average essay score would have had them below the cut line. If so, then the passing rate would fall to 19%. Regardless whether the passing rate was 29% or 19%, these results suggest that there is significant knowledge decay when taught with the IL method. Furthermore, the first exam performance was not significantly correlated with the second exam performance (r =.14), suggesting that knowledge decay was not uniform across students.

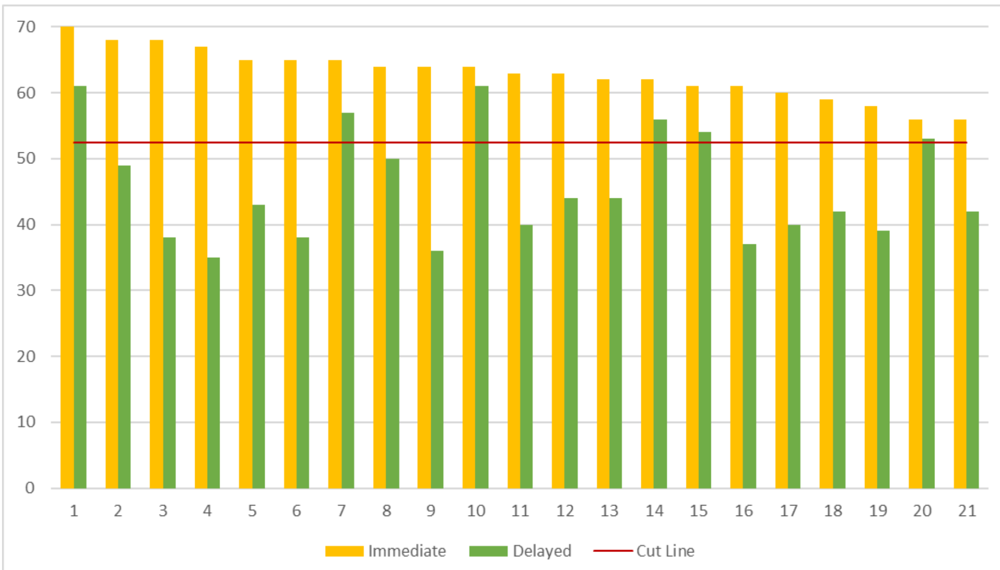

Figure 7. IL group MCQ & FIB immediate exam total (yellow) versus 10 weeks later (green)

Figure 8 plots the BL group's initial score (in yellow) against their later exam (in green). Like Figure 7, there are observed declines in performance, however, this decline appears less prominent. All but one of the students (#24) had a passing score on the original 100-point exam, and #24 would have passed if the essay was excluded. In the delayed exam, 15 of the 24 students (63%) would have passed. Only one student (#21) was within ½ and one can hypothesize that this student may not have passed if the essay portion had been included.

Table 5. IL and BL mean exam scores

	IL	**BL**
Immediate (0 weeks)	62.9	62.7
Delayed (10 weeks)	45.7	53.6
Knowledge Decay	17.2	9.1
% Knowledge Decay	27.3%	14.5%

Table 5 lists the mean scores for BL and IL groups on both the immediate and delayed exams as well as the calculation of knowledge decay as measured by decline in the mean exam performance. The IL group lost 17.2 points on a scale of 70, and the BL group lost 9.1 points. This 8.1 points is a significant difference ($t(43) = 3.51$, $p < .001$). This supports the notion that the BL group benefitted in some way that allowed them to retain that knowledge significantly better than the IL group.

Further analysis compared BL group PPL students (#1, 7, 16, and 20 in Figure 8) to non-PPL students. These PPL students showed less decay ($M = 3.3$) compared to that of the non-PPL students ($M = 10.3$), which is significant ($t(22) = 1.98$, $p < .05$). Thus, PPL student knowledge decay was less than their non-PPL classmates. As this IL group had only two PPL (# 9 and 13 in Figure 7) among the 21 students, no within group comparison could be made. The between-group knowledge decay comparison was reexamined without any PPL students, and the IL group (M = 16.6) still showed more decay than that BL group (M = 10.3), $t(37) = 2.54$, $p < .01$).

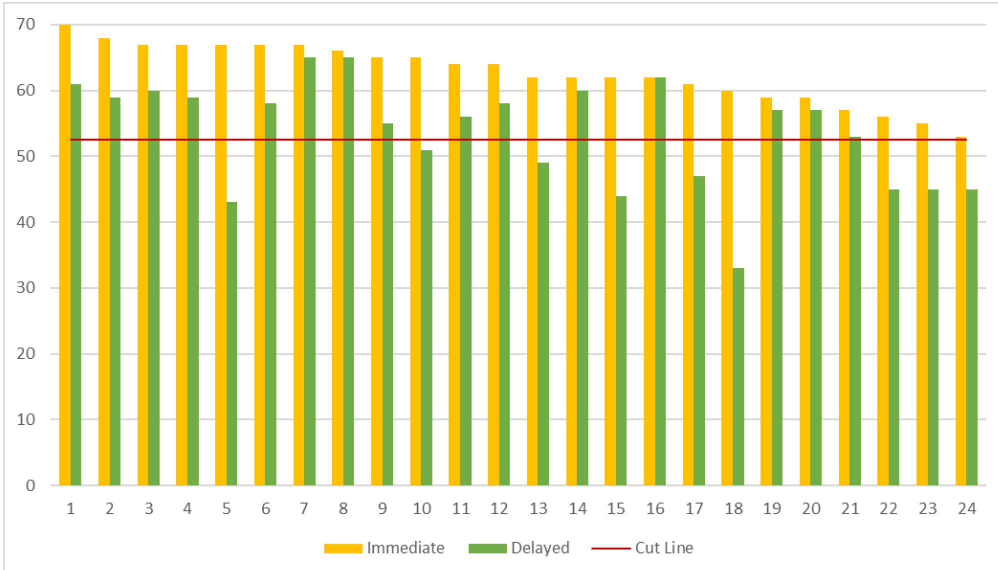

Figure 8. BL group MCQ & FIB immediate exam total (yellow) versus 10 weeks later (green)

It is important to add that there was considerable effort to ensure that the BL group's exam was as much of a surprise as it was for the IL group. First, the groups never directly interact with one another during their time at ground school. Next, the training leadership, instructors, and research staff never discussed this with the students. Finally, an after-exam set of random interviews of students and their leader was conducted where they were asked whether it had in fact been a surprise. Given the ramifications of deceiving superior officers in a highly competitive flight training program, these events give great confidence that the BL group's exam was a surprise for which they had not prepared.

Objective 3: Determine predictive power of adaptive e-learning process measures

Since the e-learning platform provides a rich set of information on the learner, those measures were investigated to determine whether they could predict learning outcomes. A regression model was created using these predictors with the total exam score as the outcome variable. The model with all six predictors (four metacognition scores, time spent learning, and number of incorrect probes) was significant ($F(6) = 6.50$, $p < .001$). Inspection of the partial correlations showed that the metacognitive percentages did not significantly contribute to the model. Thus, a model with only time spent learning and incorrect probe responses was created and this was significant ($F(2) = 13.42$, $p < .001$). The partial correlation for incorrect responses was $r = -.57$, $p <.01$, and total time was $r = -.32$, *ns*. Finally, a simple correlation examined the relationship between just incorrect probes and exam scores, and this was also significant ($r(33) = -.64$, $p < .001$). Figure 9 shows a scatterplot of this correlation with the failing exam scores below 75 marked as red. Thus, the number of students' incorrect probes were the primary predictor of their exam score performance. This is logical as the probes serve as a formative exam of sorts, so the frequency with which students don't know answers while working with the e-learning platform should be related to later exam performance.

Objective 4: Determine students' and instructors' perceptions of blended learning and e-learning usability

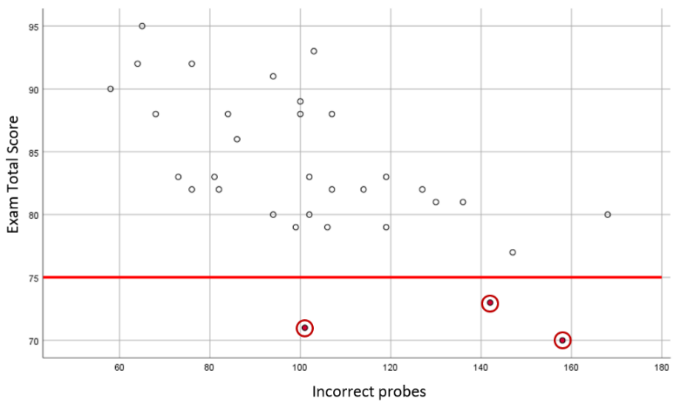

Figure 9. Scatterplot of the total exam score and number of incorrectly answered probes

Students' Perceptions

Of the 34 students in the class, 33 completed the System Usability Scale (SUS). The positive phrased questions (1,3,5,7,9) had an *M* = 4.16, and the negatively phrased questions (2,4,6,8,10) had an *M* = 2.31. The SUS score was calculated in the standard fashion (Brooke, 1986) revealing a score of 73.2. SUS scores above 68 are considered positive, and 73.2 is considered "good" or a "B" score (Will T, n.d.). Also, on the questionnaire, students were asked to elaborate on positives and negatives of their interaction with the system as well as the blended learning approach. These responses were qualitatively analyzed, and trends were noted:

- The software was easy to use and was accessible from a wide range of devices.
- Learning at their own pace was beneficial.
- Students had immediate feedback from probes when making a mistake, so doubts and misunderstandings could be addressed quickly.
- Students felt that they could focus for longer periods of time compared with instructor-led classes where they felt less attentive.

However, students also noted some challenges:

- The verbal 'coach' could be distracting, and some preferred reading the material themselves.
- Slides were not in order with the student study guide or the Flash courseware.
- Some topics (shockwaves, Dutch roll, gyroscopic effects, etc.) did not lend themselves to the e-learning platform and were hard to understand without an instructor.

Although students were asked to complete the adaptive learning modules, they also had access to the SSG. There was interest in learning more about how students split their time between the e-learning software and studying from the SSG, so a question was included on the questionnaire. Students reported that they spent a greater percentage of time with e-learning (*M* = 64.5%, *SD* = 17.8) compared with SSG (*M* = 35.5%, *SD* = 17.8). This suggests that they used the adaptive courseware as their primary means of self-paced Aerodynamics learning and is consistent with the notion that the adaptive courseware was seen as the primary learning tool.

Instructors' Perceptions

The ground school instructors are seasoned veterans who have successfully trained more than a thousand skilled aviators. Understandably, there was concern that if the approach proved unsuccessful it would be detrimental to the students and the students would likely need to repeat the Aerodynamics topic. Thus, recording the instructors' thoughts on the blended learning paradigm was important. In interviews, the instructors noted the following positive aspects:

- Students were more prepared for the instructor-led sessions.
- It generated discussion, talking, reading, and better interaction with the instructor compared with traditional lectures. One instructor noted that normally "when I stand up there, not many people pay attention to me. Really, do they understand? They are supposed to look at me, but they are so used to looking at SSG or courseware. I noticed the strong contrast with how they paid attention."
- All students benefited from the probe questions challenging their knowledge.
- Students interacted with one another well when working through the content in the self-paced portion.
- Students were more willing to engage with the instructors while learning the material.

- One noted that "the whole classroom seemed more relaxed."

However, they also noted some concerns:

- It was not obvious what the correct blend of instructors and adaptive e-learning should be.
- The adaptive system was not transparent, so it was difficult to explain its behavior to the students

DISCUSSION

The purpose of this study was to investigate the effect of blended learning with adaptive e-learning on student performance as measured by exam performance. BL and IL group differences were compared by analyzing the scores from the exam given promptly after the subject is taught. Although there was a difference between FIB scores with the IL group being slightly better, the general profile of those scores suggests that both groups performed quite well. If it had been an option, a novel performance measure with a better distribution of the data may have yielded greater insight into potential group differences. As it stands, the observed ceiling effect makes it difficult to distinguish these groups.

Next, differences on the surprise exam were examined. Since the exam took place 10 weeks after the initial content along with a lack of preparation by the students, scores were expected to decline for both groups. However, a robust advantage for the BL group was observed as they suffered about half the knowledge decay experienced by the IL group. This supports the notion that something positive took place in the blended learning environment that fostered improved retention of the knowledge. The PPL student performance was investigated, and not surprisingly PPL students in the BL group experienced less decay than non-PPL students. Given this and that there were twice as many PPL students in the BL group, the differences were examined without PPL students but the BL group advantage over the IL group remained. This important discovery points towards the need to further investigate the contributing factors that allowed the BL students to better retain the knowledge they initially acquired.

Given that the e-learning software uses a data-driven approach to model the learner, these measures were examined to see if they could predict exam performance. Of the six measures, only the number of incorrect probes predicted exam performance. This is a meaningful result as it suggests that instructors who use a system like this would benefit from knowing the students' probe performances. As a formative evaluation tool, it need not dictate continued progress in ground school but could serve as an early warning that a student needs help. In general, this study pointed out the necessity of aligning an e-learning interface with not only the students but also the instructors who need to monitor progress and intervene and assist when needed.

Finally, the perceptions of the blended learning experience were recorded to help understand what aspects could be improved in future work. Students and instructors noted positives such as the ability to work at one's own pace, improved attention and readiness during instructor-led sessions, better inter-student interaction, and the value of formative assessments via the probes. However, there were also negatives such as topics presented in a non-traditional order, a lack of transparency in how the adaptive engine works, and individual preferences that often tilted towards reading the content oneself versus listening to a text to speech voice. Additionally, comments on media type suggest future work exploring the impact of multimedia type on the learning of specific concepts.

A known limitation of this study was a focus on the adaptive e-learning portion of the blended learning approach. Given both instructors' lack of familiarity with this e-learning approach coupled with the investigator's inexperience with the context of this learning domain, there was not an opportunity to craft a genuinely active learning instructor-led session (Lombardi, Shipley, Astronomy Team, & Team, 2021). In future efforts it would be desirable to improve the instructor-led portion of blended learning so that it better complements the e-learning portion by expanding on the application of knowledge and skills. The goal would be to bring in active learning methods that counteract the passivity of a standard lecture approach. Additionally, skill components may be enhanced through deliberate practice techniques (Ericsson & Pool, 2016) that can improve skill-based learning exercises. While military flight training programs remain successful in producing skilled aviators, this study demonstrates that there is room for improvement. Specifically, blended learning may offer a path to help aviators retain valuable knowledge that could prove beneficial to subsequent skill development and may even raise the quality of graduating aviators.

ACKNOWLEDGEMENTS

We thank the students who shared their honest feedback on their experience, and the ground school instructors for their willingness to explore ways to make learning better. We also thank the Lockheed Martin and Area9 Lyceum collaborators who share a passion for providing excellent instruction and who helped make this effort a reality.

REFERENCES

Area9 Lyceum. (n.d.). Personalized adaptive learning in four dimensions. Retrieved from https://area9lyceum.com/

Barsuk, J. H., Cohen, E. R., Wayne, D. B., McGaghie, W. C., & Yudkowsky, R. (2018). A comparison of approaches for mastery learning standard setting. *Academic Medicine, 93*(7), 1079-1084.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational researcher, 13*(6), 4-16.

Bloom, B. S., Krathwohl, D. R., & Masia, B. B. (1984). Bloom taxonomy of educational objectives. In *Allyn and Bacon*: Pearson Education.

Bonk, C. J., Kim, K.-J., & Zeng, T. (2005). *Future directions of blended learning in higher education and workplace learning settings.* Paper presented at the EdMedia+ innovate learning.

Brooke, J. (1986). System usability scale. *Reading, England: Digital Equipment Corporation, 480*.

Chen, K. S., Monrouxe, L., Lu, Y. H., Jenq, C. C., Chang, Y. J., Chang, Y. C., & Chai, P. Y. C. (2018). Academic outcomes of flipped classroom learning: a meta-analysis. *Medical Education, 52*(9), 910-924.

DeLozier, S. J., & Rhodes, M. G. (2017). Flipped classrooms: a review of key ideas and recommendations for practice. *Educational psychology review, 29*(1), 141-151.

Durlach, P. J., & Spain, R. D. (2014). *Framework for instructional technology: Methods of implementing adaptive training and education*. Retrieved from

Ericsson, A., & Pool, R. (2016). *Peak: Secrets from the new science of expertise*: Houghton Mifflin Harcourt.

Finetti, B. d. (1965). Methods for discriminating levels of partial knowledge concerning a test item. *British Journal of Mathematical and Statistical Psychology, 18*(1), 87-123.

Furst, M., Hill, E., Krathwohl, W., Bloom, D., & Engelhart, B. (1956). Taxonomy of educational objectives: The classification of educational goals. *Handbook I: Cognitive Domain. New York: David McKay Company*.

Gardner-Medwin, A. R. (1995). Confidence assessment in the teaching of basic science. *Research in Learning Technology, 3*(1). doi:10.3402/rlt.v3i1.9597

Guskey, T. R. (2010). Lessons of mastery learning. *Educational leadership, 68*(2), 52.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology, 77*(6), 1121.

Lombardi, D., Shipley, T. F., Astronomy Team, B. T., Chemistry Team, Engineering Team, Geography Team, Geoscience Team,, & Team, P. (2021). The curious construct of active learning. *Psychological Science in the Public Interest, 22*(1), 8-43.

Manuela, A., Fernando, B., & Tiago, O. (2016). An e-Learning Theoretical Framework. *Journal of Educational Technology & Society, 19*(1), 292-307.

Risko, E. F., Anderson, N., Sarwal, A., Engelhardt, M., & Kingstone, A. (2012). Everyday attention: Variation in mind wandering and memory in a lecture. *Applied Cognitive Psychology, 26*(2), 234-242.

Vandewaetere, M., Desmet, P., & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior, 27*(1), 118-130.

Zimmerman, B. J., & Dibenedetto, M. K. (2008). Mastery learning and assessment: Implications for students and teachers in an era of high-stakes testing. *Psychology in the Schools, 45*(3), 206-216.