

Building User Proficiency in Piloting Small Unmanned Aerial Vehicles (sUAV)

Siya Kunde and Brittany Duncan

Abstract—Assessing proficiency in small unmanned aerial vehicles (sUAVs) pilots is complex and not well understood, but increasingly important to employ these vehicles in serious jobs such as wildland firefighting and infrastructure inspection. The limited prior work with UAVs has focused on user training using modalities like simulators and VR and no performance assessments with line-of-sight UAVs. This paper presents a training methodology for novice pilots of sUAVs. We presented two studies: the *Baseline* study (21 participants) and the *Training* study (16 participants). Our work is of interest to sUAV operators, regulators, and companies developing this technologies to produce a more capable workforce capable of consistent, safe operations. We successfully utilized the method developed in [1] to assess user proficiency in flying UAVs. We presented a UAV pilot training schedule for novice users (in the *Training* study), and were able to determine the minimum training time necessary to observe performance gains and mitigate damage. Results indicate that task completions noticeably improved and crashes minimized by day 10 of training, with a training plateau observed by day 15.

I. INTRODUCTION

Building user proficiency with small unmanned aerial vehicles (sUAVs) through quick and effective training programs can prepare users to be integrated into human-robot teams. Currently the Federal Aviation Administration (FAA) only relies on a knowledge-based assessment (Part 107 [2]) in order to issue a certificate for commercial operation. However, as mentioned by Schneider in [3], this can lead to the fallacy where trainers assume that the learner’s conceptual understanding of the system will lead to proficiency in the operational setting. In reality, users often need a great deal of experience operating the system even after they have learned to conceptualize it accurately. For example, a car driver may find it relatively easy to understand the manual transmission system, but for the gear shifting to become proficient the user will have to spend substantial time driving the car.

In this paper, we focus on understanding how novice pilots can be trained, and how long they must be trained to achieve a level of proficiency (and observable training plateau).

Our research questions include:

- (R1):** *Based on the method developed in [1], can we assess user proficiency for flight skills?* We believe our observations across two additional studies support this classification technique.

Siya Kunde and Brittany Duncan are with School of Computing, University of Nebraska-Lincoln, USA, siya.kunde@huskers.unl.edu, and bduncan@unl.edu

This work was supported by grants from National Science Foundation (IIS-1750750, IIS-1925368, and IIS-2221648).

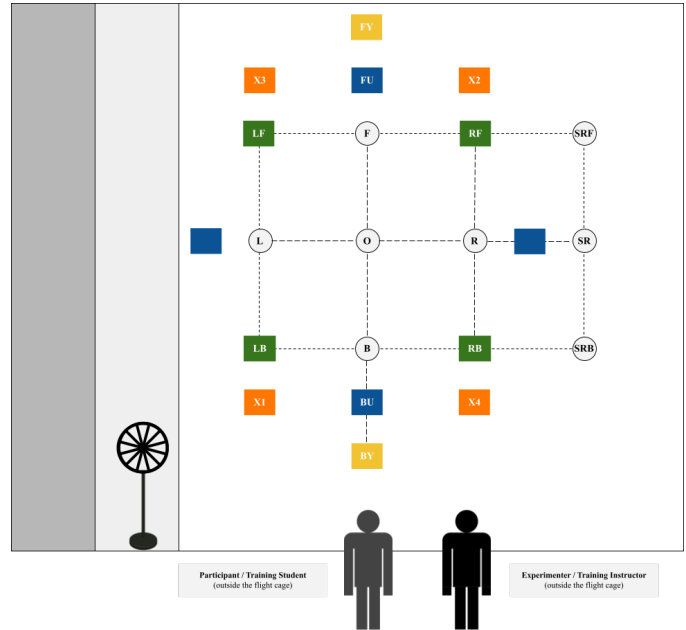


Fig. 1: Top-view visualization of the flight cage area. The grey areas on the left are stairs. The fan in the lower left side was used only during the windy condition tasks described in section III-B.2 (under “All sessions after the first”) in *Training* study.

- (R2):** *Can we determine the minimum training time necessary to observe performance gains and mitigate damage?* We present a plateau after 3 weeks of consistent training.

II. RELATED WORK

In this section we examine literature on training methods, modalities and focused skills from general aviation and UAV domains.

A. Training Methods

In general aviation, daily flying practice has been found to be the frequency of choice for trainees [4]. Hays [5] found that total higher training hours produced improved training outcomes. We incorporate these findings from general aviation domain into our training schedule to be able to create an effective training program.

Training programs should be personalized and adaptive to the individual to assure a trainee has reached proficiency before moving on [6], [7]. Such programs have been found to be more effective than the ones where all trainees proceed at the same pace [5]. Proficiency based advancement can improve motivation for the individual to continue the training [5], reduce burn-out and increase productivity in the training

program [8]. Hence during the training phase of our study, we will employ a self paced learning strategy.

In an interesting study, Lindo [9] assessed performance changes by testing in a cockpit that the user had not trained on. Results indicate that training users solely with autonomous aid can hinder training adaptation to a more manual approach later. We would however like to be able to use partially assistive autonomous modes (like Altitude Hold mode) in the training program, to facilitate targeted manual practice of other controls (roll, pitch and yaw).

B. Training Modality

In general aviation, simulators have been an essential tool in assisting pilot training [6] and have been widely used for training and assessment of pilot skills [10], [11], [12], [13], [9]. Simulators have been enhanced with motion cuing platforms [11] and mixed reality [12], enabling improved flight performance and training. They have been found to be effective in training not just experienced pilots but also new trainees, where there was a 90% reduction in aircraft time required to attain course objectives [14]. Simulator plus aircraft training is found to produce training effects rapidly compared to aircraft training alone [5].

[15] conducted a 3 day UAV training study, using virtual reality, with various levels of training including familiarization, simple and precise waypoint navigation, obstacle avoidance, and skill transfer. Differences were found between the high and low performers based on VR-based performance metrics, workload and stress. Hing [13] conducted a simulator study with drones focused on improving training outcomes by increasing user situation awareness.

Simulators and VR environments offer a clear advantage in terms of reducing training cost, preventing crashes of real vehicles, by providing a safe environment to practice potentially dangerous procedures, and allowing rapid and multiple repetition of events. These modalities however cannot replace the experience of training with a line of sight UAV, with appropriate precautions in place.

C. Skill Training

A few different types of specific skill training are found in literature. One study by Hing [13] found that operator performance could be improved to minimize accidents by adding emphasis on operator situation awareness skills during training. This was done by testing the utility of chase viewpoints. Hazard, instrument, and time-sharing skills learned during training can lead to more successful and prepared pilots. For example, Molesworth [16] found that interactions with hazards during training can help pilots learn skills for future flights to be able to recognize and assess dangerous situations. In our research we hope to train users in dealing with unexpected situations.

III. EXPERIMENT

This section presents the experimental details of two studies (participant overview in Table I): the *Baseline* study

TABLE I: Overview of participants in the *Baseline* and *Training* studies.

	Study	
	<i>Baseline</i>	<i>Training</i>
Sample Size	21	16
Gender: Female, Male, Unstated	5, 15, 1	8, 8, 0
Prior Flight Experience	All	2

used to understand user proficiency, and the longitudinal *Training* study designed to build user proficiency with UAVs.

Participants in these (IRB approved) studies were recruited at University of Nebraska-Lincoln (UNL) and surrounding UAV clubs, using mailing lists and flyers.

NIST [17] created test methods for small unmanned aircraft systems that can be used to quantitatively evaluate various system capabilities and remote pilot proficiency, and were primarily intended for systems with an onboard camera and remote pilot display. Our method developed in [1] focuses primarily on automatically detecting proficiency levels of users based on features computed from user flight paths. In the current paper we extend this idea and along with a training plan to demonstrate changes in pilot proficiency over time. While the NIST system provides a more comprehensive testing and assessment guide, we simplified the tasks, setup, and assessments for the purpose of our studies.

A. Baseline Study

We conducted this study with users who reported to have at least some UAV piloting experience.

1) *Participants*: We were able to recruit 22 participants, from which 1 dropped out mid-way. While we would have preferred to balance the population for gender, it is difficult to find female UAV pilots based on this requirement and our experience running previous studies. Henceforth we will present statistics and results for the 21 participants (5 female, 15 male, and 1 unstated) who completed the study.

The amount of prior flight experience varied across users with flights of 116.38 hours on average (std. dev. = 320.14, min = 1 hr and max = 1500 hrs). Flight frequency differed as well where users reported to practice yearly (1 user), monthly (8), weekly (11), and daily (1) respectively. 12 users reported to possess the FAA Part 107 Remote Pilot License.

In contrast, the flight experience in the last 6 months was 26.37 hours on average (std. dev. = 35.12, min = 1 hr and max = 140 hrs). Users flew the UAV either monthly (5 users), weekly (13), or daily (1) respectively.

2) *Protocol*: In the *Baseline* Study users were asked to report prior experience at the time of signing up for the study. This study was conducted across two sessions, lasting two hours each. Participants were compensated \$20 after completing the first session and \$10 for the second. Once they arrived at the study location, we recorded their demographic data and they then completed three tests: the *Prediction Test*, the *Sound Test*, and *Control Test* with Xbox controller. During the second session, participants first completed the *Control Test* with the Futaba controller and then the *Baseline Flight Test* with the Futaba controller (described in section III-A.3).

The *Prediction Test* involved user observing a series of controller states and given a start pose for a toy UAV, predicting the flight movements. The *Sound Test* involved user predicting UAV movements based on sounds recorded of a UAV. Both the *Control Tests* involved the user applying control inputs using the two controllers mentioned, while looking at a video of a flying UAV. These four knowledge tests will not be covered in this paper. In this paper we will focus solely on the *Baseline Flight Test*.

3) *Baseline Flight Test*: In this test the user was first allowed up to 5 minutes to familiarize themselves with the Flamewheel F450 UAV and Futaba RC (See Fig. 2 for both), and then asked to complete a series of UAV flying tasks. The tasks in this section utilize the flight cage map from Fig. 1. The tasks in our studies were adopted from the sub-sequences in NIST tests [17], which asked users to fly the UAV from single control stick inputs to multiple controls simultaneously across both sticks. The following tasks varied in difficulty based on UAV orientation (nose-out (NO) or nose-in (NI)) and complexity of the control sequence/combination.

- T1** Hover(NO) : Take-off from O to a height of 1m, hover in place for 30 seconds, and then land back at the O.
- T2** Hover(NI) : Take-off from O to a height of 1m, hover in place for 30 seconds, and then land back at the O.
- T3** Long-Hover(NO) : Long hover (3 minutes) over O. Don't inform the user how long they need to hover.
- T4** Land-Near(NO) : Take-off from O to a height of 1m, approach BU and land in the center of the BU.
- T5** Land-Near(NI) : Take-off from O to a height of 1m, approach BU and land in the center of the BU.
- T6** Land-Far(NO) : Take-off from O to a height of 1m, go to FU and land in the center of the FU.
- T7** Land-Far(NI) : Take-off from O to a height of 1m, go to FU and land in the center of the FU.
- T8** Square(NO) : Take-off from RB to a height of 1m, complete a square along the green tiles clockwise, come back to the same location and land.
- T9** Diamond(NO) : Take-off from BU to a height of 1m, complete a diamond along the blue tiles clockwise, come back to the same location and land.
- T10** Alpha(NO) : Take-off from X1 to 0.5, go to X2 diagonally upwards to 1.5m, go horizontally sideways to X3 by maintaining the altitude at 1.5m, go to x4 diagonally downwards to 0.5m, and land at X4.
- T11** Eight(NO) : Takeoff from BY to a height of 1m, complete a figure 8 starting clockwise that passes through O, then FY, then back through O, back to BY and lands.
- T12** Hover(NO) : This last task is the same as the first.

B. Training Study (TS)

We conducted this study with users who reported to have zero to minimal UAV piloting experience.

1) *Participants*: We were able to recruit 19 participants (9 female and 10 male), of which 3 dropped out early in the process. We present statistics and results for the 16 participants (8 female and 8 male) who completed the study.

Apart from two users (1 male and 1 female), all reported to have never flown a UAV. One of them, who had prior experience, was recruited from the *Baseline* Study after testing under the lowest proficiency group (N), as they expressed interest in this study as well. The other participant reported to have interacted with a UAV for a couple of hours.

2) *Protocol*: The *Training* Study was divided into 3 phases: pre-training evaluation (phase 1), training (phase 2), and post-training evaluation (phase 3). The activities of phases 1 and 3 were same, and similar to the tests conducted in the *Baseline* Study. The evaluations in phases 1 and 3 were performed to assess the overall improvement in proficiency of the user. Additionally, similar to the *Baseline* Study, we will only discuss the process and outcomes of the *Baseline Flight Test* (from phases 1 and 3) and the training in phase 2 in this paper. For both the phases 1 and 3, participants were compensated \$30 after completing the entire phase.

Phase 2 of the *Training* Study lasted 4 consecutive weeks, with participants coming in for a 30 minutes training session on 5 days per week (Monday through Friday). Participants were compensated \$20 after completing every 5th session. Due to longitudinal nature of the study, it was expected that participants may need to reschedule few sessions in case of emergencies. We asked them to come in for make-up sessions on a different day (during the weekday or weekend) to ensure all participants were trained for same amount of days and time. The total cash compensation over the course of the longitudinal study added up to \$140, and users were further incentivized to continue participation by offering them a mini drone on completion of the study.

First training session: Participants were provided with materials that include a pre and post flight checklist they were asked to follow every day, a basic UAV and remote controller operations manual, training guide with a set of tasks to be practiced, instructions on charging batteries used for the UAV etc. The instructor walked the user through all the materials and provide a demonstration wherever necessary (e.g. the instructor demonstrated how to fly the UAV and how to charge the batteries). Apart from the first session, some of the instruction steps were repeated as necessary whenever a new UAV was used (start of week 2 and week 3).

All sessions after the first: Once a participant arrived, they first answered the pre-flight survey, performed all the pre-flight checks, and then started training by following the training guide. Once 20 minutes of the training session were completed, the last 10 minutes were utilized to conduct the *Training Flight Test* (section III-B.3). Testing time was gradually increased to up to 15 minutes, so towards the end participants could practice for up to 15 minutes before that. In the fourth week however, there was no deliberate practice, as we had designed various flight *experiences* for the participants. These included flying in nose-in orientation, in the presence of wind, and with different flight modes.

During the training time of the session, participants were asked to first work on the lower level tasks and then progress to the next task. For tasks 1 to 6, users could proceed to a higher numbered task, only after completing

a lower numbered task. Once they reached the tasks with the higher level of difficulty (tasks under 7 and 8), they could practice whichever task they wanted, out of order. This was enforced to ensure that users would master the basic controls before trying out complicated maneuvers. We did not seek perfection or assess the quality of the practice flight ourselves, but users were asked to self-assess and pace themselves through the training tasks as necessary.

We progressively increased the testing tasks by week so that users would have enough time to train beforehand. The number of tasks to be used for testing increased on the fifth session of the week. The *Training Flight Test* conducted for the first four sessions of the week used the F450 only, but the one conducted on the fifth session of the week first used the UAV used for training and then the F450. This schedule is described in Table II.

Lastly before leaving for the day, participants performed all post-flight checks, and answered a post-flight survey.

3) *Training Flight Test*: The tasks in this section utilize the flight cage map from Fig. 1. All tasks were typically completed in nose-out orientation, except for session 1 of week 4 when users completed some tasks nose-in as well.

- T1** : Takeoff-Land : Take-off from O, stabilize, land at O.
- T2** : Hover 30s : Flight test task **T1**.
- T3** : Line horizontal : Take-off from O to a height of 1m, go to SR and land in the center of the SR.
- T4** : Line vertical : Take-off from B to a height of 1m, go to F and land in the center of the F.
- T5** : Square : Flight test task **T8**
- T6** : Diagonal : Take-off from B to a height of 1m, go to SRF and land in the center of the SRF.
- T7** : Diamond : Flight test task **T9**
- T8** : Up-diagonal : Take-off from RB to 0.5m, go to LF while thrusting up to a height of 1.5m, and land at LF.
- T9** : Down-diagonal : Take-off from SR to 1.5m, go to BY while descending to 0.5m, and land at BY.
- T10** : Alpha : Flight test task **T10**
- T11** : Circle : Take-off to a height of 1m from the right blue tile, complete a circle along the blue tiles clockwise, come back to the same location and land.
- T12** : Eight : Flight test task **T11**

C. Materials and Data Collection

The UAVs and RCs used in the study are shown in Fig. 2. The DJI Flamewheel 450 sUAV was used in *Manual / Stabilize* mode for the *Baseline Flight Test* (Section III-A.3) and the *Training Flight Test* (Section III-B.3). The Holystone HS170 mini UAV and the Flamewheel F330 (with prop-guards) were used in the *Training Study*.

In both the studies, the user control inputs were captured as RC_OUT data from *mavros* Robot Operating System (ROS) package at 50 Hz. This is a higher data capture rate compared to the one mentioned in [1], and was an additional reason for collecting the flight data in *Baseline Study* again. The sUAV pose data was recorded using the Vicon motion capture system set to 200 Hz.



Fig. 2: Holystone HS170 mini UAV was used with the white RC, and Flamewheel F330 (with prop-guards) and Flamewheel F450 were used with the Futaba controller.

IV. METHOD

In the *Baseline* and *Training* studies, we were able to record the RC data at a higher resolution so we introduced power spectral density controls frequency mean and standard deviation as features. Additionally, we included two more features: ratios of participant and ideal trajectory's distances and duration. We compared the user trajectory against the expected trajectory of the UAV without actually flying the UAV. Following are all the features computed, for each task:

- 1) Dynamic time warping distance similarity measures of the flight paths in X (DTW_X), Y (DTW_Y), and Z (DTW_Z) axes considered individually; XY (2D points) together (DTW_{XY}); and XYZ (3D points) together (DTW_{XYZ}).
- 2) Fréchet distance (FR) measures of the flight paths.
- 3) Mean and standard deviation of the first derivative of the flight paths in X (V_X), Y (V_Y), and Z (V_Z) axes considered individually; XY (2D points) together (V_{XY}); and XYZ (3D points) together (V_{XYZ}).
- 4) Power spectral density features: Mean and standard deviation of the power of the control inputs (roll (P_R), pitch (P_P), yaw (P_Y) and throttle (P_T)), and frequency of the control inputs (PF).
- 5) $PathLengthRatio = (l_{u/a}) \triangleq \frac{ParticipantTrajectoryLength}{IdealTrajectoryLength}$
- 6) $PathDurationRatio = (t_{u/a}) \triangleq \frac{ParticipantTrajectoryDuration}{IdealTrajectoryDuration}$

We utilized the methodology defined in [1] (with features computed above and rater rankings of the trajectories from best to worst) for proficiency labeling in the current two studies. The data from the *Baseline* study was clustered to produce labels and then used as a reference for classifying the data from the *Training* study.

We did not use all the tasks from the *Baseline Flight Test* (in section III-A.3) of *Baseline* study for the clustering process. Since it was observed in the results of [1] that landing tasks are not very discriminative, they were left out. We only used the 30 second hover (T1) and the square, diamond, alpha, and eight pattern tasks (T8, T9, T10 and T11) for the clustering. Consequently the same tasks were picked from the data of the *Training* study for the classification process. As mentioned in the phase 2 schedule (Table II), all tasks mentioned in the *Training Flight Test* (in

TABLE II: Phase 2 schedule. Tasks are referenced from section III-B.3

Day(s)	Week	Session	Testing UAV(s)	Flight Type	Test Tasks
1	1	1	-	Training	No Test
2-4	1	2-4	F450	Training	1, 2
5	1	5	Holystone, F450	Training	1,2,3,4,5
6-9	2	1-4	F450	Training	1,2,3,4,5
10	2	5	F330, F450	Training	1,2,3,4,5,6,7
11-14	3	1-4	F450	Training	1,2,3,4,5,6,7
15	3	5	F450	Training	1,2,3,4,5,6,7,8,9,10,11,12
16	4	1	F450	Nose-in flights	NO (1,2,3,4,5,6,7,8,9,10,11,12), and NI (1,2,3,4,5,6)
17-18	4	2-3	F450	Flights with and without wind	0,0,1,2,3,4,5,6,7,8,9,10,11,12
19	4	4	F450	Flights with and without wind in Manual and AltHld modes	0,2,3,4,6,8
20	4	5	F450	Switching flight mode mid-flight and autonomous flights	No Test

TABLE III: Leave-one-out cross-validation accuracy for clustering methods for number of clusters = 2,3,4,5.

Method	Dunn Index				Task Completion Acc.			
	No. of Clusters							
	2	3	4	5	2	3	4	5
Agglomerative	0.47	0.43	0.45	0.40	0.90	0.92	0.94	0.96
KMeans	0.49	0.38	0.38	0.39	0.92	0.93	0.95	0.99
Spectral	0.49	0.29	0.27	0.28	0.94	0.96	0.92	0.94

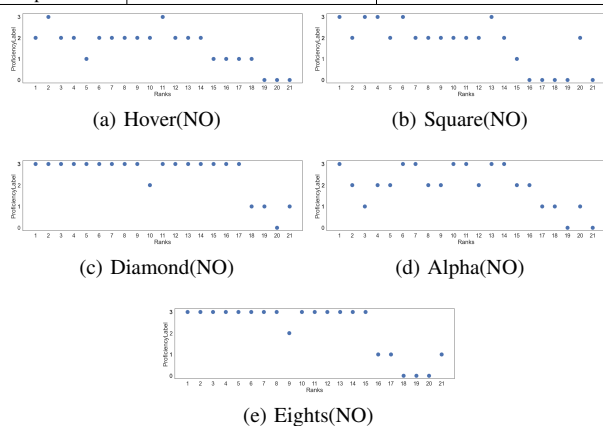


Fig. 3: Visualization of proficiency labels (y axis) (Top to bottom:- 3 (E: Expert), 2 (I2: Intermediate 2), 1 (I1: Intermediate 1), and 0 (N: Novice)) produced by the clustering method (described in [1]) versus ranks provided by rater (x axis) for the data from *Baseline* study.

section III-B.3) were not conducted every day. For each day, we used whatever tasks data was available (datasets available publicly [18]).

V. RESULTS AND DISCUSSION

A. Baseline Study User Proficiency Groups

Baseline Flight Test data from the *Baseline* study was clustered to produce baseline proficiency groups (with the method described in section IV) that could be further used to classify the *Training* study data.

After performing leave-one-out cross validation experiments across methods for different clusters (results displayed in Table III), we chose to perform the final clustering with 4 clusters with the Agglomerative method. The clustering for each task used is shown in Fig. 3, and the final user proficiency was computed by averaging across all tasks (black circles visualized in Fig. 4).

B. Novice User Piloting Proficiency Development

Fig. 4 presents the results of clustering the *Baseline Flight Test* data from the *Baseline* study (indicated with filled black

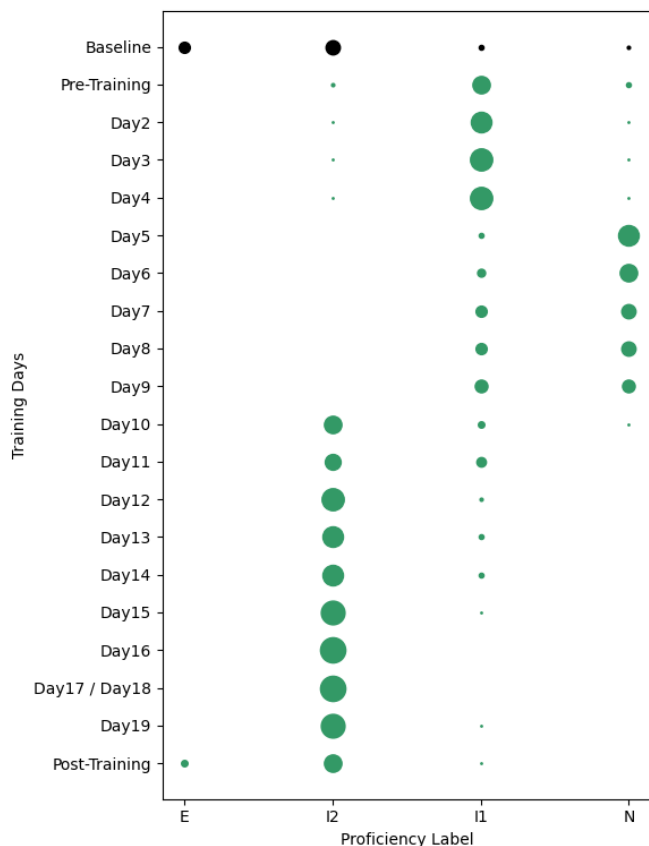


Fig. 4: User proficiency classification. Labels on x-axis: E: Expert, I2: Intermediate 2, I1: Intermediate 1, N: Novice.

circles), and classifying the pre-training (*Baseline Flight Test* data), training (*Training Flight Test* data) and post-training (*Baseline Flight Test* data) data from the *Training* study (indicated with green circles with black border).

Before beginning training, when novice participants completed the flight test, they were classified into one of the lower proficiency groups (I2, I1 and N), with most users classified under I1 group. In the first week (days 2-4) we do not observe much difference in the user proficiency. There are a number of potential reasons for this outcome. Use of one task for the classification (since only 30s hover was completed during the test) is one such possibility. Alternatively the reason could be because it was too early in the process where users still needed to practice more to make a noticeable difference in the performance.

From day 5 to 9, we have data from the two tasks (30s hover and square), and a gradual shift of users from group N to group I1, confirming the effectiveness of the training they

VI. RECOMMENDATIONS AND FUTURE WORK

All flight tests were conducted indoors while pilot strategies can change outdoors. Due to the limited number of participants we chose to conduct these tests in a more controlled (indoor) setting. We tried to recreate some outdoor experiences (e.g. using a fan to create wind) during the final week of training. Now with initial results from a controlled setting, future work can be conducted where users are trained in outdoor environments, with our work as a baseline.

Task difficulty becomes important in use when the pilot needs to utilize their skills to complete a difficult task, complete a simple task with time limits, or while distracted with a secondary task. We included tasks of varying difficulty but did not impose any time limits or introduce secondary tasks. We did however utilize task completion ratio and velocity metrics which account for the ideal task completion time and the mean velocity maintained to achieve this.

We saw many users in the *Training* study plateau into the I2 proficiency group. While this is a comparative rating (to data from the *Baseline* study), it would be interesting to explore how to improve flight skills further. One way to explore this would be through testing of various training strategies, with the current one as a control. One aggressive training strategy could train users only in *manual* mode through out the training (swap out the *altitude-hold* mode used in week 2 with *manual* mode). Another gentler training strategy could test the impact of the “buddy” system (as described in [19]) by having the instructor or the computer involved as a flight buddy. The “buddy” system would be similar to how we used the *altitude-hold* in week 2, except we could tailor the training to not just help with maintaining the altitude, but to practice flying with assistance provided for each control (or combination of controls). This system can potentially replace the use of the mini UAV from week 1 of training, by having the buddy control roll, pitch, and yaw, while the user focuses on how to safely takeoff and land the UAV gently (using throttle). We specifically found in this work that UAV orientation can impact flight, and giving users a chance to gather more experience with that control will help in overall flight outcomes.

VII. CONCLUSION

Understanding how to train novice users to build flight skills can help make drones safer for everyone involved. In this work, we presented two studies: the *Baseline* study (21 participants) and the *Training* study (16 participants). We successfully utilized the method developed in [1] to assess user proficiency in flying UAVs (answering *R1*) and were able to observe the following results. We presented a UAV pilot training schedule for novice users (in the *Training* study), and were able to determine the minimum training time necessary to observe performance gains and mitigate damage. Related to *R2* results indicate that task completions noticeably improved and crashes minimized by day 10 of training, with a training plateau observed by day 15.

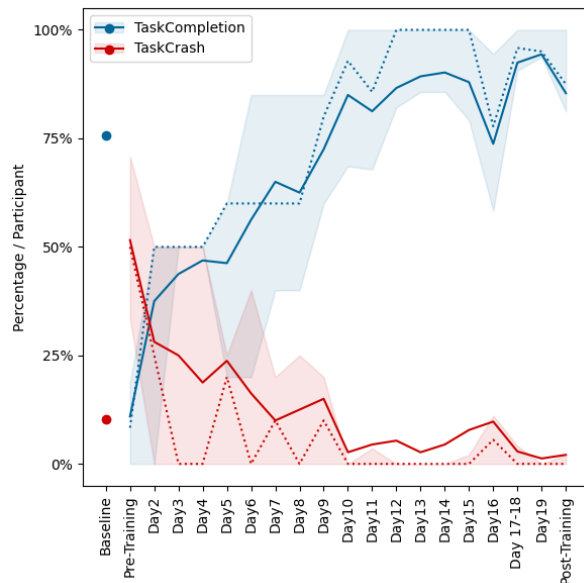


Fig. 5: Percentage of task completions and UAV crashes. *Baseline* data is indicated with solid circles. For *Training* data, solid line indicates mean, dotted line indicates median.

were receiving. From day 6 users had started practicing on the F330 in altitude hold mode, which potentially contributed to better development of roll and pitch controls.

Unlike day 5 where new tasks had worsened the proficiency levels, on day 10 we see a jump when many users are bumped up to group I2. From day 10 to 14, we have data from three tasks (30s hover, square, and diamond), and we see a slight change in the cluster sizes as users shifted between groups I1 and I2.

On day 15, 16, and 17/18, we have data from all tasks that were used in the clustering process of *Baseline* data, while day 19 had just the one task (30s hover). After day 15 we do not see a lot of change in user performance. This is also the point where majority of the users are able to complete all of the tasks and we see minimal crashes. We conclude this to be the training plateau.

Finally, we can examine the outcome of the post-training evaluation, which was conducted as early as the day after training concluded and as late as 2 weeks after, pending user availability. We see a highly variable performance where some users performed better than they did on day 19, while one user’s performance degraded. This leads us to believe that recency of experience may impact a user’s skills of piloting the UAV.

C. Task completions and crashes

During the training phase (see Fig. 5), we see an increase in the percentage of tasks completed, and a decrease in the number of crashes. Within 10 days of training, participants were able to complete on average 85% of tasks, and crashed the UAV only 2.6% of the times. With 20 days of training, this performance was superior even to that of the *Baseline* data group. The task completions and crashes deteriorated slightly on *Day 16* when the participants completed NI tasks.

REFERENCES

- [1] S. Kunde, E. Palmer, and B. Duncan, "Recognizing user proficiency in piloting small unmanned aerial vehicles (suav)," *IEEE Robotics and Automation Letters*, vol. 7, pp. 2345–2352, 2022.
- [2] "Certificated remote pilots including commercial operators," 2023. [Online]. Available: https://www.faa.gov/uas/commercial_operators
- [3] W. Schneider, "Training high-performance skills: Fallacies and guidelines," *Human factors*, vol. 27, no. 3, pp. 285–300, 1985.
- [4] R. E. King, T. R. Carretta, P. Retzlaff, E. Barto, M. J. Ree, and M. S. Teachout, "Standard cognitive psychological tests predict military pilot training outcomes," *Aviation Psychology and Applied Human Factors*, 2013.
- [5] R. T. Hays, J. W. Jacobs, C. Prince, and E. Salas, "Flight simulator training effectiveness: A meta-analysis," *Military psychology*, vol. 4, no. 2, pp. 63–74, 1992.
- [6] P. W. Caro, "Aircraft simulators and pilot training," *Human Factors*, vol. 15, no. 6, pp. 502–509, 1973.
- [7] —, "An innovative instrument flight training program." *Paper No. 710480, Society of Automotive Engineers*, 1971.
- [8] D. Gopher and R. A. North, "Manipulating the conditions of training in time-sharing performance," *Human Factors*, vol. 19, no. 6, pp. 583–593, 1977.
- [9] R. S. Lindo, J. E. Deaton, J. H. Cain, and C. Lang, "Methods of instrument training and effects on pilots' performance with different types of flight instrument displays," *Aviation Psychology and Applied Human Factors*, 2013.
- [10] M. Friedrich, S. Y. Lee, P. Bates, W. Martin, and A. K. Faulhaber, "The influence of training level on manual flight in connection to performance, scan pattern, and task load," *Cognition, Technology & Work*, pp. 1–16, 2021.
- [11] J. T. Hing and P. Y. Oh, "Development of an unmanned aerial vehicle piloting system with integrated motion cueing for training and pilot evaluation," *Journal of Intelligent and Robotic Systems*, vol. 54, no. 1, pp. 3–19, 2009.
- [12] —, "Mixed reality for unmanned aerial vehicle operations in near earth environments," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2010, pp. 2509–2510.
- [13] J. T. Hing, K. W. Sevcik, and P. Y. Oh, "Improving unmanned aerial vehicle pilot training and operation for flying in cluttered environments," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 5641–5646.
- [14] P. W. Caro, "Transfer of instrument training and the synthetic flight training system." HUMAN RESOURCES RESEARCH ORGANIZATION ALEXANDRIA VA, Tech. Rep., 1972.
- [15] J. Hayes, S. Dwivedi, R. Karthikeyan, M. Abujelala, J. Kang, Y. Ye, E. Du, and R. K. Mehta, "Identifying early predictors of learning in vr-based drone training," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 66, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2022, pp. 1872–1876.
- [16] B. Molesworth, M. W. Wiggins, and D. O'Hare, "Improving pilots' risk assessment skills in low-flying operations: The role of feedback and experience," *Accident Analysis & Prevention*, vol. 38, no. 5, pp. 954–960, 2006.
- [17] "National institute of standards and technology uas test methods," 2020. [Online]. Available: <https://www.nist.gov/ctl/pscr/uas-test-methods>
- [18] S. Kunde and B. Duncan, "Uav pilot flight proficiency progression dataset and materials," 2023. [Online]. Available: <https://github.com/siyakunde/UAV-Pilot-Flight-Progression>
- [19] J. R. Rogers, "Low-cost teleoperable robotic arm," *Mechatronics*, vol. 19, no. 5, pp. 774–779, 2009.