National Training Aircraft Symposium (NTAS) 2018 - The Changing Role of the Pilot

Aug 15th, 9:30 AM - 10:45 AM

A Study of Human-Machine Interface (HMI) Learnability for Unmanned Aircraft Systems Command and Control

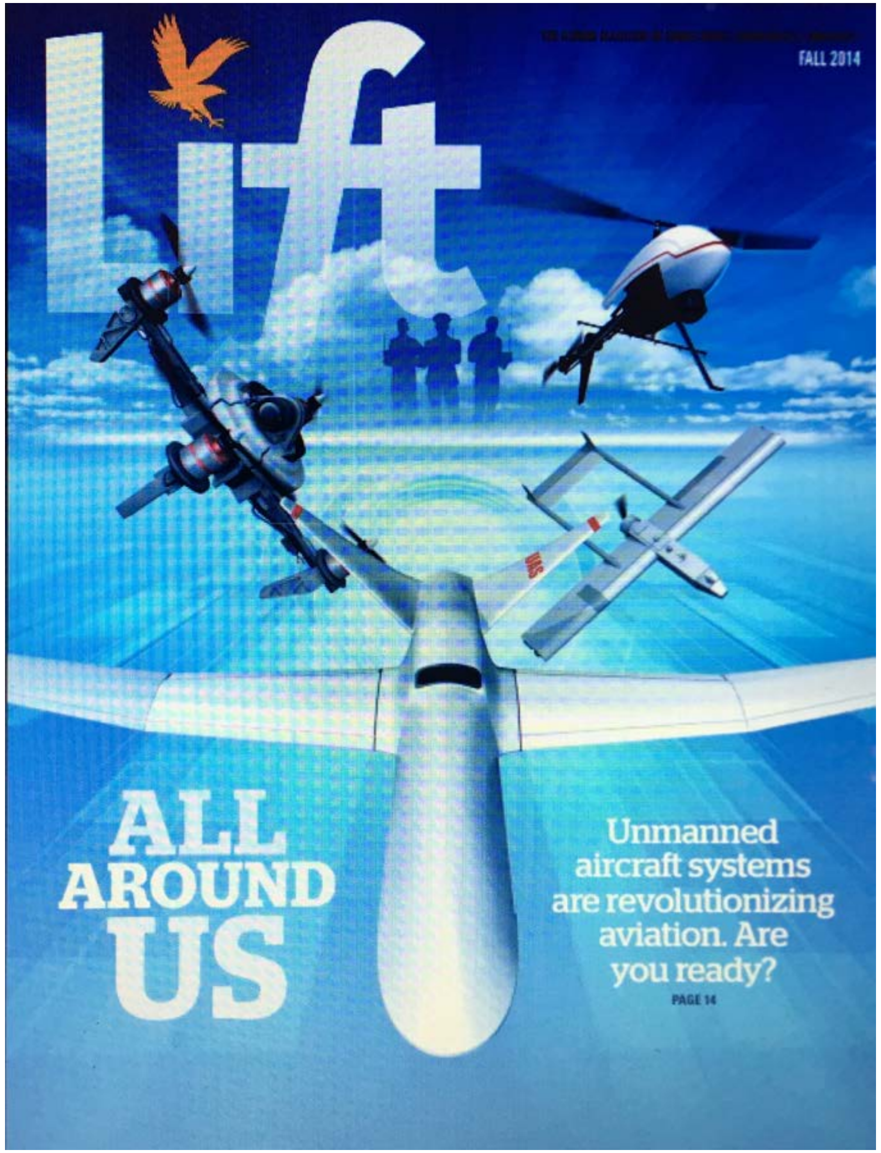
Tom Haritos Ph.D.
Assistant Professor of Aeronautical Science, Embry-Riddle Aeronautical University, tharitos@k-state.edu

A Study of Human-Machine Interface (HMI) Learnability for Unmanned Aircraft Systems Command and Control

Tom Haritos, Ph.D.
Department of Aeronautical Science
Embry-Riddle Aeronautical University

Agenda

- Background
- Motivation for the Study
- What is Learnability?
- Experimental Design
  - Research Questions
  - Methodology
  - Results
    - Effectiveness
    - Efficiency
    - Satisfaction
- Final Summary



Background

- Unmanned aircraft systems (UAS) have been instrumental for the DoD in the last two decades.
  - More recently, UAS have transitioned from *military* to *civilian*
  - A number of viable public and commercial applications have emerged.

- UAS operations is heavily restricted in the United States by the Federal Aviation Administration (FAA)

  - Many safety-related concerns of the public and users of the National Airspace System (Vincenzi, Terwilliger & Ison, 2015; GAO, 2012).

    - UAS industry is in a state of accommodation
    - Full-scale UAS operations may not be visible for a decade or longer
    - More recently, the FAA released 14 CFR Part 107 (sUAS)
      - For low-risk UAS operations
      - Reactive in nature and requires attention

Motivation for the Study

- Human-machine interaction (HMI) has been identified as a primary human factors concern.
  - Little emphasis has been placed on the design and development of display technologies for UAS command and control.

- Information presentation and information exchange has often been described as non-effective and non-efficient (Vincenzi, Terwilliger & Ison, 2015; Maybury, 2012; Cooke, 2008; Williams, 2004).

What is Learnability?

Learnability: Characterized and defined in many ways.

- The theory of learnability is a sub-principle of usability and relates to improving operator effectiveness and efficiency through human centered designs (Nielsen, 2012); Sauro & Lewis, 2012).
    - Some describe learnability as the ease of use on initial user performance and improvements in performance when interacting with a system over-time (Grossman et al., 2009; Chimbo et al., 2011).
    - Others suggest learnability is the capability of a software product to enable the user to learn how to use it effectively within a reasonable amount of time (Shamsuddin, Sulaiman, & Zamli, 2011).
    - Nielsen (1994) suggested a highly learnable system is one that allows users to reach a reasonable level of proficiency in a short span of time.
- For this study, learnability was defined as a user's initial performance with a system after instruction (i.e., initial learnability) and performance gains on a specific task after a user becomes familiar with the basic functionality of the system (i.e., extended learnability).

Experimental Design

- A causal-comparative or Ex Post Facto research design was established for this experiment.

- Causal-comparative research attempts to determine the cause for existing differences in the behavior or actions of individuals or groups (Gay, Mills & Airasian, 2012).

- The grouping variable for this research was experience:
  - Experience had three levels or factors:

    (1) no conventional flight experience

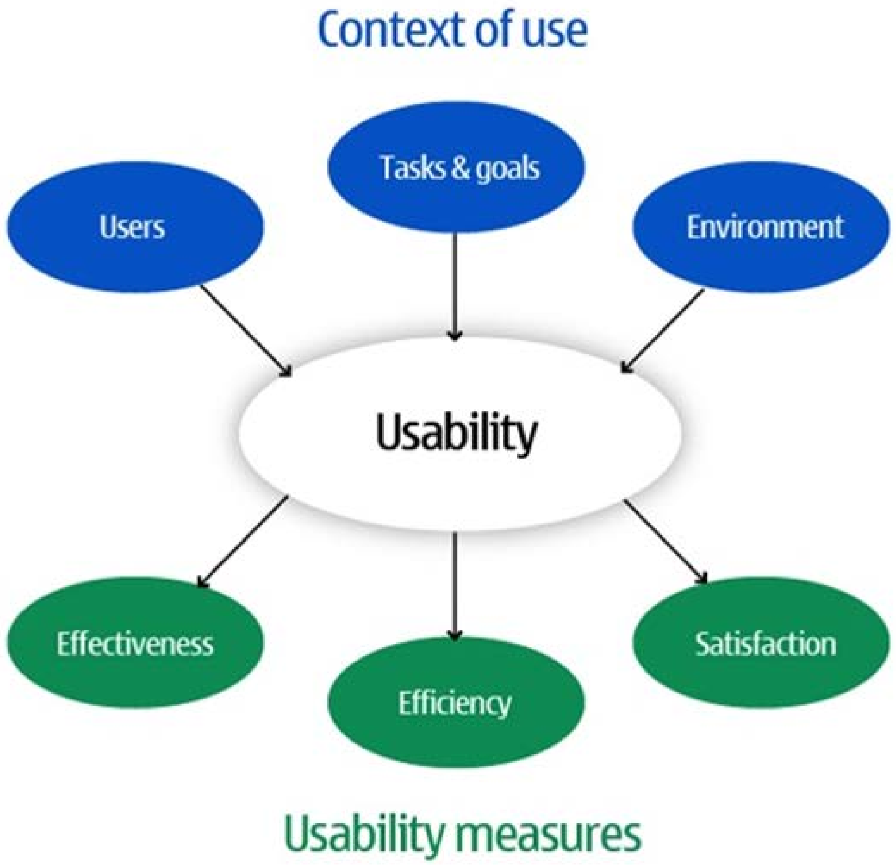
    (2) low conventional flight experience

    (3) high conventional flight experience

Causal-Comparative Research Design		
No Pilot Experience *n=15*	Low Pilot Experience *n=15*	High Pilot Experience *n=15*
Dependent Variables	**Task Success/Completion Rate**	
	Number of Errors	
	Failure Rate	
	Total Time on Task	
	Satisfaction	

Methodology: Dependent Variables

- **Effectiveness:**
  - Task Completion/Success Rate
  - Failure Rate
  - Errors
- **Efficiency:**
  - Total Time on Task
- **Satisfaction:**
  - System Usability Scale



Research Questions

1. How accurately did task completion rates such as task completion time, time until failure, total time on task, and errors (Sauro & Lewis, 2012) serve to measure the learnability of the UAS HMI representation?

2. Were participants satisfied with the level of interaction to perform the specific set of operational UAS tasks as regards the System Usability Scale (Brooks, 1996)?

3. Based on the System Usability Scale as scored by Sauro and Lewis (2012), did participants find the UAS HMI usable and learnable?

4. Was incremental learning exhibited as participants become more familiar with the HMI (i.e., reduction in terms of task completion rates and errors)?

5. To what degree did the level of conventional flight experience (i.e., subsequent learning) impact system learnability as regards the dependent variables and perceived satisfaction when compared to those without any conventional flight experience?

Methodology

Effectiveness: Task Completion Rate

- Data was handled in a binary manner and corresponded to task success or task failure
- Coding: (1) = success and (0) = Failure

Effectiveness: Errors

- An error was defined as any unintended actions, slips, mistakes, or omissions a user made while attempting the task.

- Errors were recorded as a count of the total number of errors committed by a participant per trial iteration.

Methodology

Satisfaction: was measured using the System Usability Scale (SUS) in the original format and without modification as defined by Brooks (1996).

- The SUS instrument provided measures for a composite SUS score and two sub-scale scores:
  - (1) usability and (2) learnability.

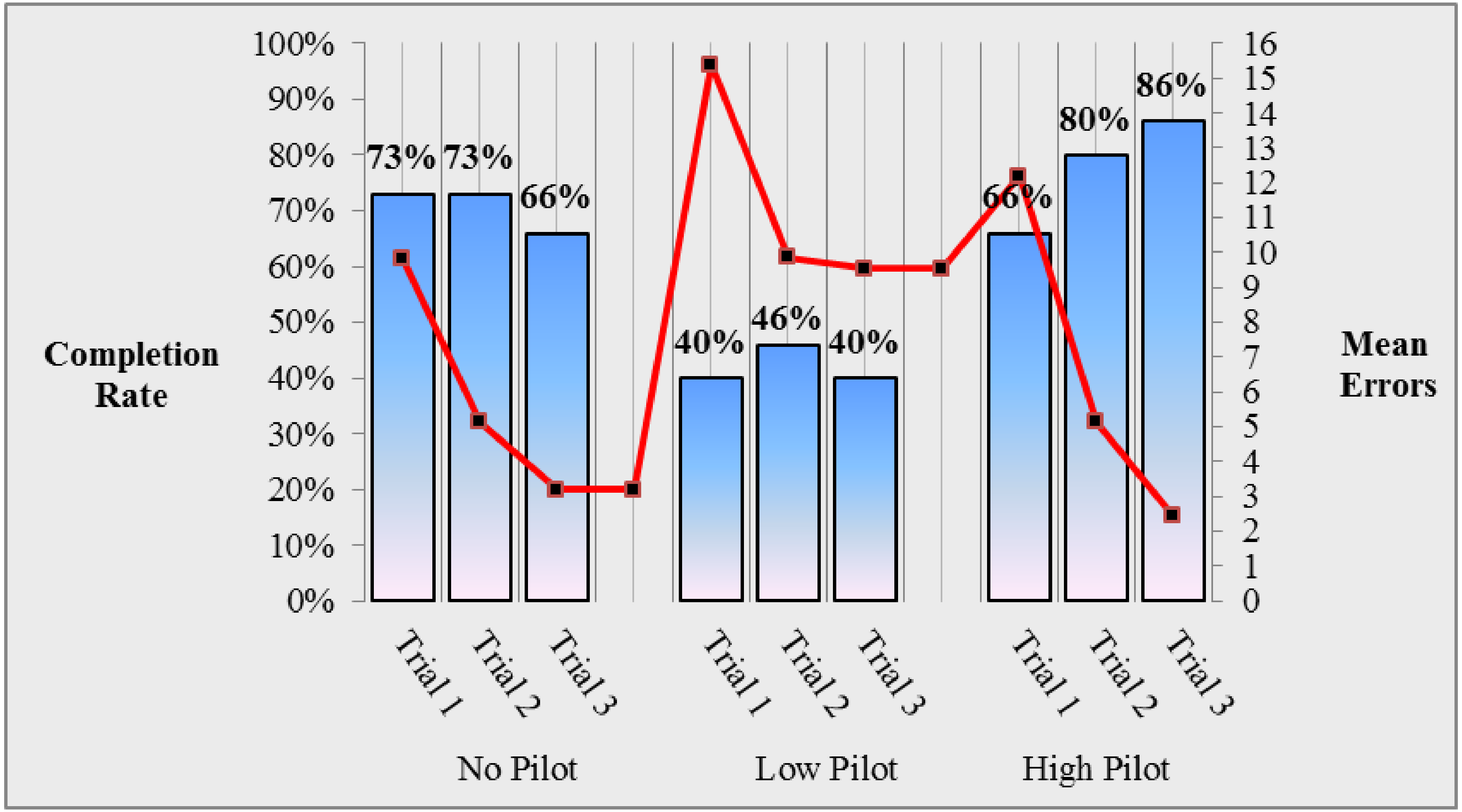
- The industry benchmark for an average SUS score is a 68.

System Usability Scale (Brooke, 1996)

	Strongly disagree 1	2	3	4	Strongly agree 5
1. I think that I would like to use this system frequently					
2. I found the system unnecessarily complex					
3. I thought the system was easy to use					
4. I think that I would need the support of a technical person to be able to use this system					
5. I found the various functions in this system were well integrated					
6. I thought there was too much inconsistency in this system					
7. I would imagine that most people would learn to use this system very quickly					
8. I found the system very cumbersome to use					
9. I felt very confident using the system					
10. I needed to learn a lot of things before I could get going with this system					

Results: Effectiveness

Task Completion Rate vs Mean Number of Errors



Results: Efficiency

Efficiency: Total Time on Task

- Efficiency was determined by calculating the total time on task for each participant's iteration
- Simulator start-up time and any observable simulator lag or latency was corrected

Example: Data extracted from simulator log file for a sample participant

Time	MagHeading	TrueHeading	Pitch	Roll	Speed	AOA	HAT	MSL	Lat
00:00.1	0.000004	0.000004	-0.12092	-0.07697	0.044452	-88.555405	4.125553	4603.863	31.58654
00:01.2	0	0	-0.1178	-0.0512	0.043598	-88.584694	4.126238	4603.863	31.58654
00:02.3	-0.000007	-0.000007	-0.11518	-0.05839	0.043805	-88.577995	4.126783	4603.864	31.58654
00:03.3	-0.000008	-0.000008	-0.11472	-0.0584	0.043781	-88.578773	4.126885	4603.864	31.58654
00:04.4	-0.000008	-0.000008	-0.11472	-0.0584	0.043781	-88.578773	4.126885	4603.864	31.58654
00:05.6	-0.000008	-0.000008	-0.11472	-0.0584	0.043781	-88.578773	4.126885	4603.864	31.58654
00:06.7	-0.000008	-0.000008	-0.11472	-0.0584	0.043781	-88.578773	4.126885	4603.864	31.58654
00:07.7	-0.000008	-0.000008	-0.11472	-0.0584	0.043781	-88.578773	4.126885	4603.864	31.58654
00:08.8	-0.000008	-0.000008	-0.11472	-0.0584	0.043781	-88.578773	4.126885	4603.864	31.58654
00:09.9	-12.000008	-0.000008	-0.11472	-0.0584	0.043781	-88.578773	4.126885	4603.864	31.58654
00:10.9	-12.000008	-0.000008	-0.11472	-0.0584	0.043781	-88.578773	4.126885	4603.864	31.58654
00:11.0	-12.000008	-0.000008	-0.11472	-0.0584	0.043781	-88.578773	4.126885	4603.864	31.58654
00:13.1	-12.000008	-0.000008	-0.11472	-0.0584	0.043781	-88.578773	4.126885	4603.864	31.58654
00:14.1	-12.000008	-0.000008	-0.11472	-0.0584	0.043781	-88.578773	4.126885	4603.864	31.58654
00:15.2	-12.000008	-0.000008	-0.11472	-0.0584	0.043781	-88.578773	4.126885	4603.864	31.58654
00:16.3	-12.000008	-0.000008	-0.11472	-0.0584	0.043781	-88.578773	4.126885	4603.864	31.58654

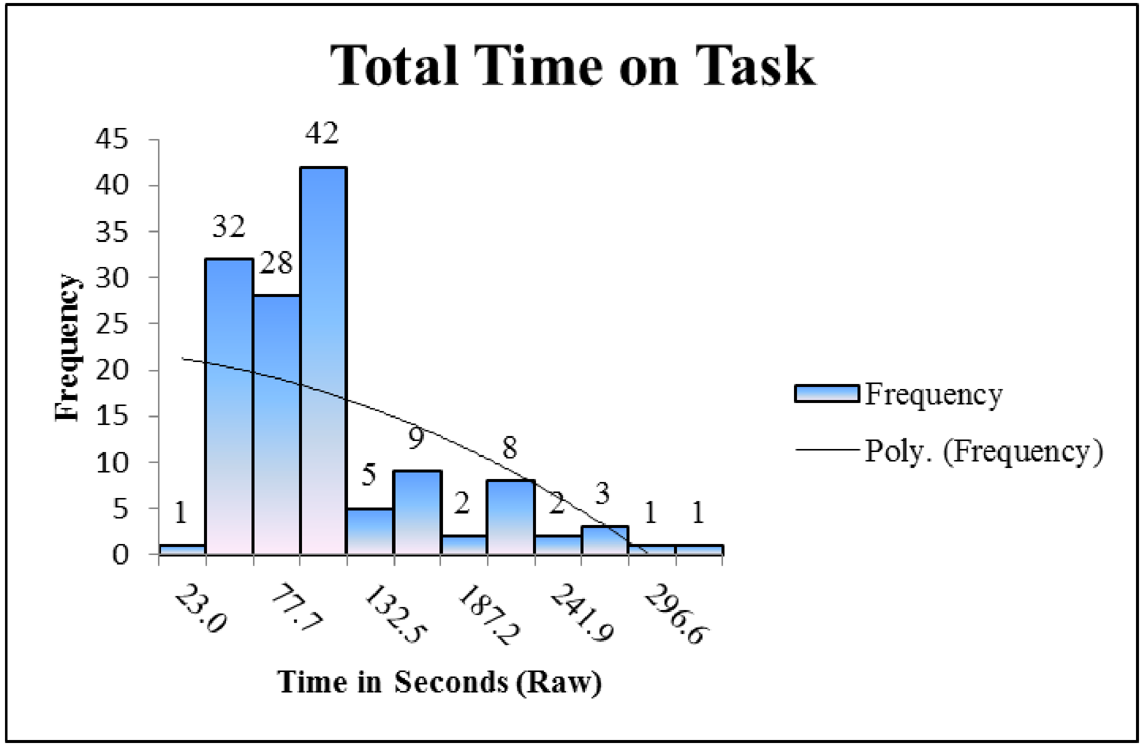
Efficiency: Total Time on Task

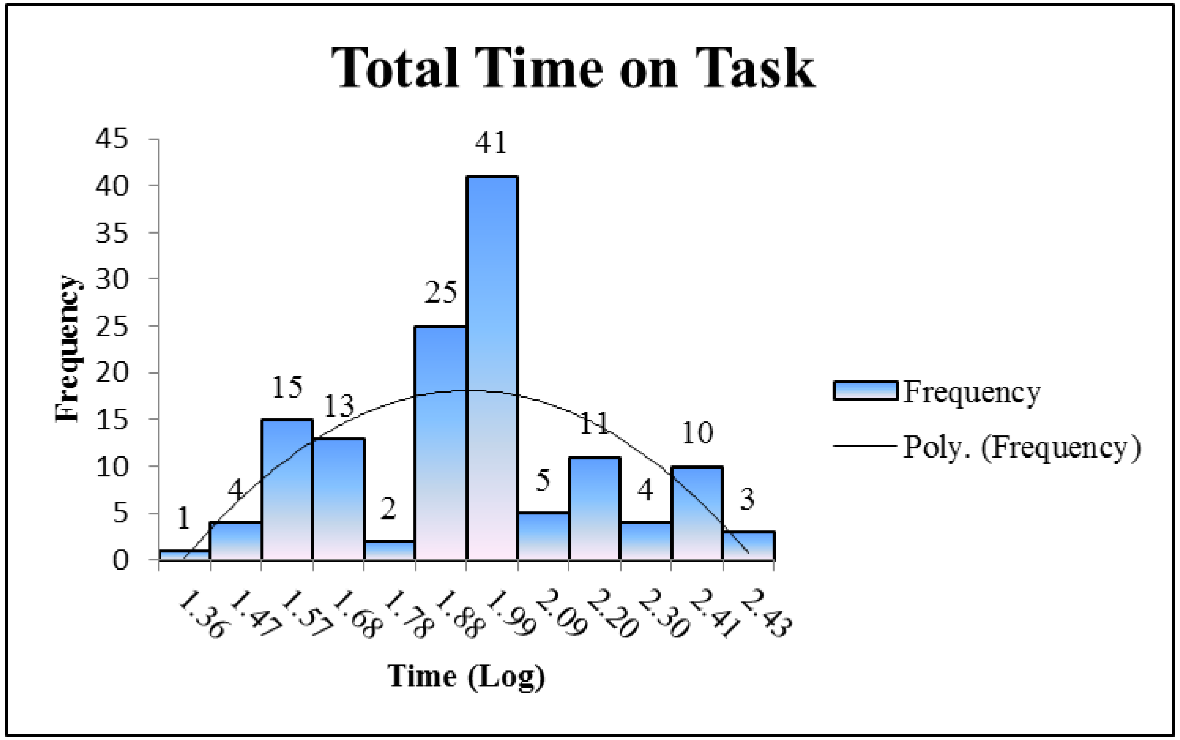
Descriptive Statistics for Total Time on Task (Raw)

	No Pilot/No UAS			Low Pilot/No UAS			High Pilot/No UAS		
	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 2
Mean	118.6	84.5	70.4	135.1	103.5	94.1	111.7	73.9	55.7
Standard Error	14.1	13.6	10.4	21.9	17.0	16.7	15.2	11.3	6.7
Median	103.0	64.0	68.0	94.0	90.0	77.0	91.0	76.0	42
Mode	66.0	#N/A	81.0	#NA	96.0	36.0	61.0	#NA	26
SD	54.4	52.8	40.2	84.6	65.9	64.5	58.8	43.9	674.5
Variance	2961.4	2791.0	1612.8	7162.9	4348.6	4160.7	3461.4	1929.3	-1.9
Kurtosis	-1.10	1.57	5.51	0.49	3.26	0.07	-0.92	5.70	-1.87
Skewness	0.56	1.43	1.99	1.23	1.91	1.13	0.40	1.91	0.25
Range	151.00	177.00	158.00	283.00	238.00	195.00	187.00	183.00	68.00

Results: Efficiency

Side-by-Side Comparison





Results: Efficiency

One-way **ANOVA** computations for ***Total Time on Task***

A One-way ANOVA indicated the effect level of experience on the dependent variable ***total time on task*** was significant within-subjects comparison for trial one and trial three: ***High pilot experience group*** [F (1, 28) = 10.9, p = 0.002].

ANOVA: Within-Subjects Comparison

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.609979	1	0.609778	10.95447	0.002575	4.195972
Within Groups	1.559125	28	0.055683			
Total	2.169103	29				

Results: Efficiency

One-way **ANOVA** computations for ***Total Time on Task***

Expert vs. Levels of Experience

- A One-way ANOVA indicated that the effect expert on the dependent variable total time on task was significant [$F(3, 56) = 21.0$, $p = 0.000000003$] for trial one and [$F(3, 56) = 7.44$, $p = 0.000277912$] and for trial three.
- Bonferroni corrected t-test revealed significant effects for all trials when compared to the benchmark total time on task.
  - All participant groups spent significantly longer as regards total time on task than an expert performing the same task.
- The comparison is important from a training perspective, as a significant amount of time and monetary resources is spent to train individuals to operate these types of UAS effectively and efficiently.

Results: Satisfaction

The descriptive statistics for the SUS scores across the independent grouping variables are presented

	No Pilot	Low Pilot	High Pilot
Mean	72.0	62.7	67.3
Standard Error	4.757	5.409	6.097
Median	77.5	55	75
Mode	82.5	75	75
Standard Deviation	18.42	20.95	23.61

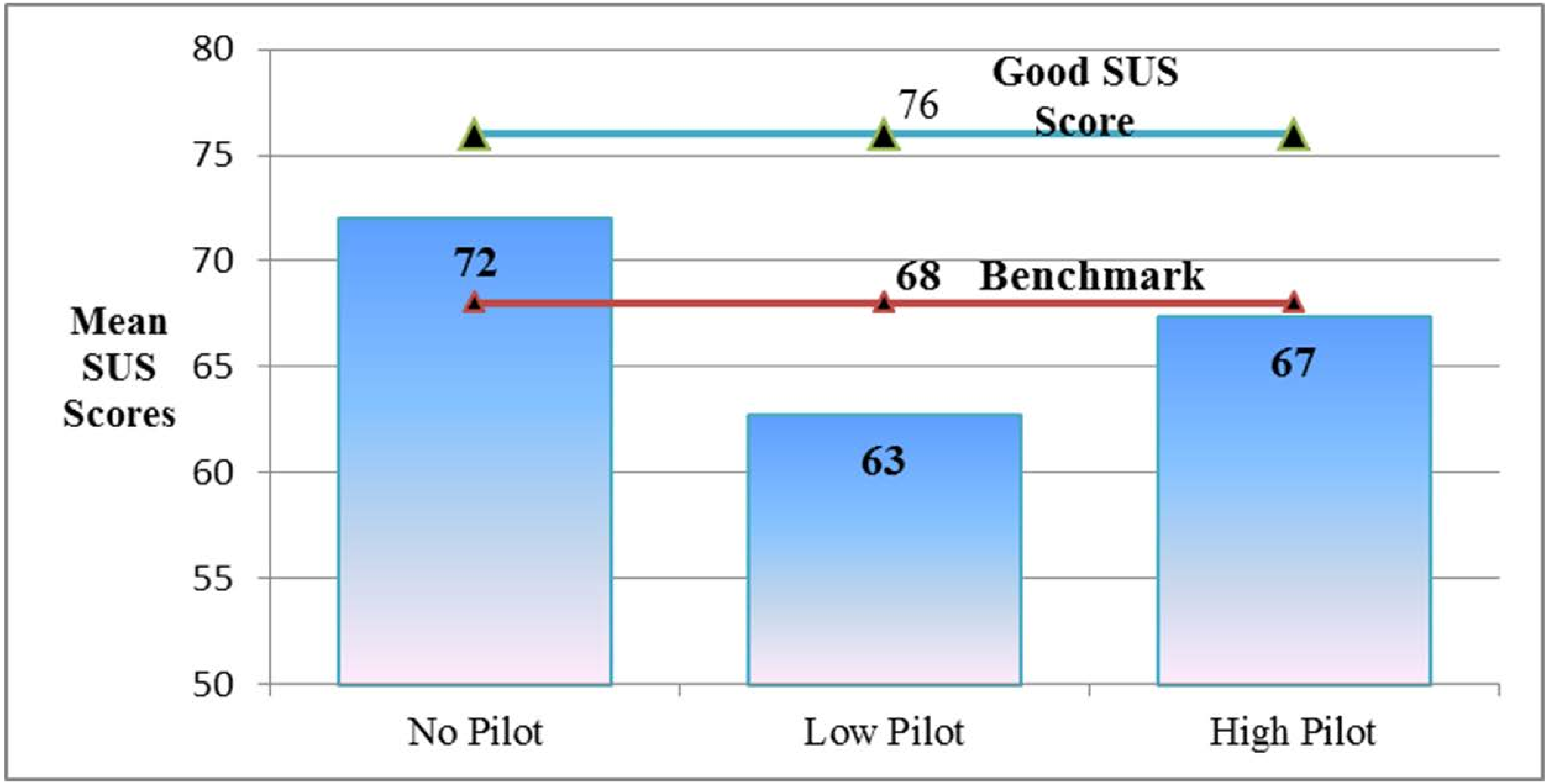
Average SUS Score = 68

- A one-way analysis of variance was executed to determine main effects on satisfaction across level of experience.

- The ANOVA output statistic for a between subjects comparison indicated no significant effects

Results: Satisfaction

Mean SUS Score Comparison by Participant Grouping Variable

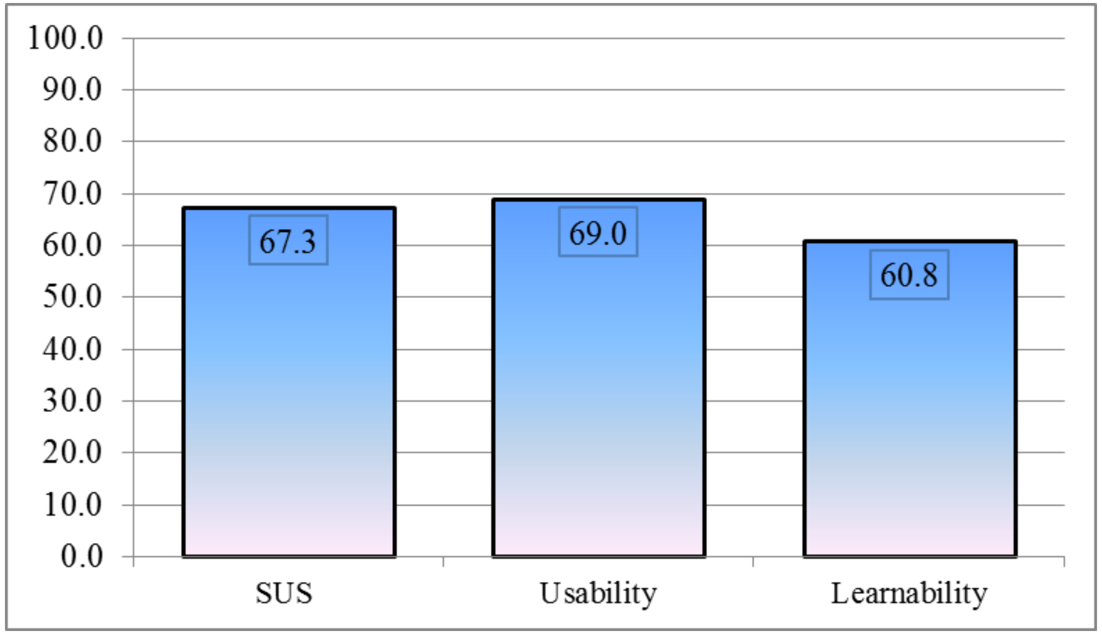


Results: Satisfaction

	No Experience	Low Experience	High Experience
SUS	72.0	62.7	67.3
Usability	74.8	63.8	68.3
Learnability	60.8	58.3	63.3

Mean Scores : SUS, Usability and Learnability

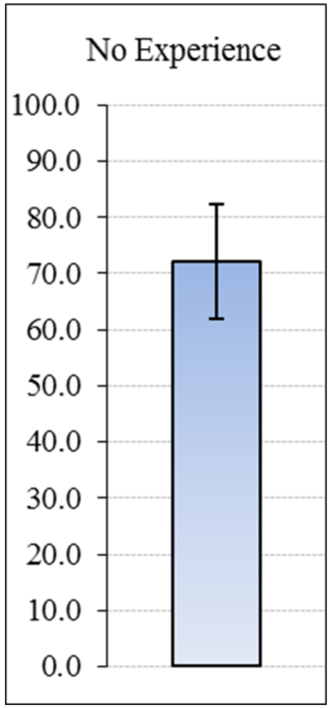
Mean System Usability Scale and Subscale Scores for all Participants

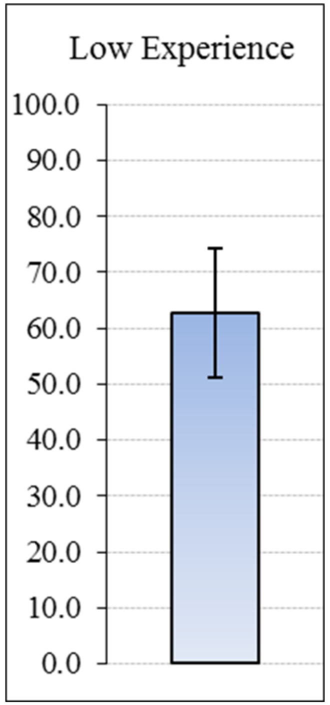


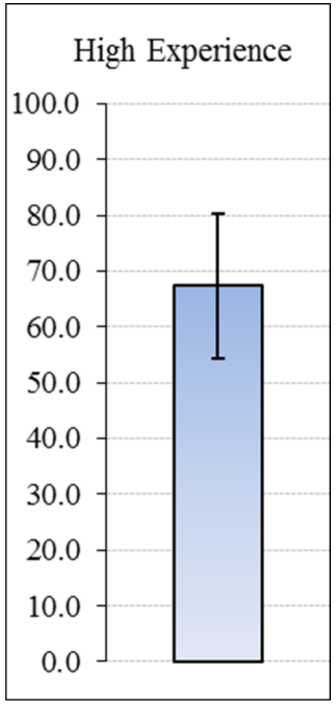
Results: Satisfaction

Confidence Intervals around the Mean SUS Score

CI around a SUS Score	
SUS Mean	67.3
SUS Standard Deviation	21
Sample Size	45
Low	60.9
High	73.6
Margin of Error	9%
Confidence Interval	95%







For this research, there was a 95% probability that the mean SUS scores could range between (61% -10.2 and 82% + 10.2) for the no experience group, (51% -11.06 and 74% +11.06) for the low experience group and (54% -13 and 84% +13) for the high experience group.

Final Summary

- This study ascertained a critical need for future research in the domain of unmanned aircraft systems designs and operator requirements as this industry is experiencing revolutionary change at a very rapid rate.

- The lack of legislation in the form of policy to guide the scientific paradigm of unmanned aircraft systems has generated significant discord within the UAS industry leaving many facets associated with the teleportation of UAS in dire need of research attention.

    - As regards the current state for user interface, practical HCI usability testing is obsolete from the industry (Maybury, 2012).

- The researcher believes this study furnished important information on the criticality for sound HCI principles in UAS applications and introduced the HCI community to a facet of usability testing related to complex UAS user interface as poor system usability has been identified as a leading cause for sub-optimal human performance in UAS operations.

- Last, future research should investigate procedural tasks on expert users in an effort to collect SUS data specific to the operation of medium altitude long endurance UAS as expert perceived satisfaction is desired as an initial construct to build a mental representation of user needs for future HMI design.


Questions