Human-in-the-Loop Method to Test the Effectiveness of Training Pilot Responses to Unexpected Events

Task 4: Training Development Plan

24 February 2022

Primary Authors; Jessica Cruit[1], Janeen Kochan[2], Mira Gruber[1], and P. A. Hancock[1]

Graduate Research Assistant; Yazmin Diaz[1]

Subject Matter Experts; Janeen Kochan[2], Shawn Pruchnicki[3], and Chris Reed[4]

University of Central Florida[1], Aviation Research, Training, and Services, Inc.[2], Human Factors Investigation and Education LLC.[3], Wings Sunward LLC.[4]

Submitted to: Bill Kaliardos, Federal Aviation Administration, Human Factors Integration Lead ANG-C1, NextGen Human Factors Division.

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Human-in-the-Loop Method to Test the Effectiveness of Training Pilot Responses to Unexpected Events. Task 4: Training Development Plan		5. Report Date February 2022 6. Performing Organization Code: N/A
7. Author(s) Jessica Cruit https://orcid.org/0000-0002-8436-8544 Janeen Kochan https://orcid.org/0000-0003-3783-9048 Mira Gruber https://orcid.org/0000-0001-8645-7906 Yazmin Diaz https://orcid.org/0000-0003-0991-5082 P.A. Hancock https://orcid.org/0000-0002-4936-066X Shawn Pruchnicki https://orcid.org/0000-0002-8758-2255 Chris Reed		8. Performing Organization Report No.
9. Performing Organization Name and Address University of Central Florida Department of Psychology 4111 Pictor Ln, Orlando, FL 32816		10. Work Unit No. N/A 11. Contract or Grant No. 692M151940001
12. Sponsoring Agency Name and Address Federal Aviation Administration (FAA) NextGen Human Factors Division 800 Independence Avenue, SW Washington DC, 20591		13. Type of Report and Period Technical report May 2021-February 2022 14. Sponsoring Agency Code ANG-C1
15. Supplementary Notes N/A		

16. Abstract

This report provides a primer as to how to conduct human-in-the-loop (HITL) research in which pilots' responses to unexpected events can be explored for potential mitigations through training. The results of this effort conclude the following findings and recommendations: Identification of key independent and dependent variables for assessing pilot responses during unexpected events; recommendations on types of training interventions and measurable skills and behaviors that pilots may exhibit during unexpected events; recommendations for HITL scenario selection criteria and candidate scenario examples; recommendations for conducting a HITL; and a recommended performance assessment method for measuring pilots' behaviors during unexpected events. Research results can ultimately lead to testing and validation to inform FAA personnel who develop evaluation criteria for pilot tasks, skills, knowledge, and proficiency and incorporate this information into human factors related documentation.

17. Key Words	18. Distribution Statement
Human factors; human performance; NextGen; resilience; automation; safety; unexpected events; surprise; startle; expertise; training; CRM; adaptability; adaptiveness; brittleness; decision-making; metacognition; robustness; training	No restrictions

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	99	N/A

Form DOT F 1700.7 (8-72)

Executive Summary

Background

Aircraft operations are highly proceduralized; however, existing, and even envisionable procedures do not completely cover all operational situations, especially during unexpected events. Evidence-based recommendations can inform pilot training on the topic of unexpected events, thereby maintaining acceptable levels of safety and response efficiency.

Scope of Report

The scope of our effort focuses on providing recommendations to test aircrew training interventions for those tasks that facilitate quantitative measurements of performance. Specifically, these recommendations are anticipated to be in increasing pilot performance and adaptation into future 14 CFR Part 121 Subparts M (Airman and Crewmember), N (Training Program), O (Crewmember Qualifications), and Y (Advanced Qualification Program [AQP]) conduits of improvement.

Purpose of Report

The purpose of this document is to articulate a detailed experimental protocol, a list of candidate scenarios, an assessment tool for measuring pilots' responses to unexpected events, and to provide recommendations for testing training interventions aimed at increasing pilot performance during unexpected events.

Organization of Report

The overarching aim of this project was to provide recommendations for researchers and training instructors to train air carrier pilots on how to manage unexpected aviation events. To this aim, four tasks were completed, and Tasks 2-4 resulted in technical reports. The tasks were as follows:

- *Task 1: Research Plan:* Research team met to create project plan of research.
- *Task 2: Relevant Research Assessment:* Synthesized the extensive literature on pilots' behaviors and responses to unexpected events.
- *Task 3: Pilot Needs Analysis:* Gathered expert input on current pilot performance and feedback to improve responses to unexpected events.
- *Task 4: Training Development Plan:* Recommend training interventions to increase pilot performance during unexpected events.

The present report presents the work of Task 4: Training Development Plan.

This report is organized into fourteen chapters with appendices. We begin with how our previous work (Tasks 1-3) associated with this effort drove the contents of this report. Chapters 2-8 discuss relevant terminology, selected candidate training methods, the suggested study variables, the recommended experimental hypotheses, the ideal study population, the study design, and the apparatus selection. Based heavily on subject matter expert (SME) input and knowledge elicitation interviews with pilots, Chapter 9 lays out how to design, develop, and select candidate scenarios for unexpected events. Chapter 10 provides a thorough example of a task analysis based on a candidate Weight-on-Wheels switch failure scenario while Chapter 11 walks the audience through an experimental protocol for the candidate scenario. Moreover, Chapter 11 provides a performance assessment for researchers or training instructors to use when assessing pilot performance during the experiment. Chapter 12 provides recommendations for analyzing the statistical data and Chapters 13 discusses lessons learned from those involved (i.e., researchers, SMEs, and stakeholders) with the project from start to finish. Chapter 14 summarizes the report and provides recommendations for future work. This report includes appendices that provide the audience with valuable references, forms, and measures for when/if they conduct the HITL.

Summary of Tasks 1-4

Task 1: Research Plan

The purpose of the research plan was for the entire research team to gather together to collaborate and develop a plan to determine how to provide recommendations for researchers and training instructors to train air carrier pilots on how to manage unexpected aviation events. The result derived from all subsequent meetings to date have established that the project's ultimate goal is to *provide the FAA with training information so that FAA can provide guidance through outlets such as Advisory Circulars* for the training requirements of 14 CFR 121 Subpart N (Training Program), Subpart O (Crewmember Qualifications), Subpart Y (Advanced Qualification Program), and 14 CFR 142 (Training Centers). Such information may have applicability and be useful for 14 CFR 61 (Certification: Pilots, Flight Instructors, and Ground Instructors), 14 CFR 141 (Pilot Schools), and 14 CFR 91 Subpart K (Fractional Ownership Operations).

Training pilot resiliency is not necessarily a goal of our project. However, it could be a potentially beneficial and secondary outcome of the knowledge developed. The current project was designed explicitly to build upon the report by Pruchnicki, Key, and Rao (2019) on training resiliency literature review. Specifically, we have looked to identify the resiliency body of knowledge with respect to pilot expertise in existing and newly forming teams. In addition, our project team leveraged insights derived from the observations of Holbrook et al. (2019), Chandra et al., (2020), and the EASA's mandatory training resiliency program (Flin, 2019). While we sought to emphasize a theory-driven approach to guide our methods, we also employed applied methods to deliver a final product that provides ecological validity as well as applicational utility.

Task 2: Relevant Research Assessment

The primary purpose of Task 2 was to understand and synthesize the extensive literature surrounding pilots' behaviors and responses to unexpected events in order to determine how to better prepare them to respond with a positive outcome to future aviation challenges. The results of the literature review from Task 2 revealed critical key points for moving forward with studying pilots' responses during unexpected events:

- First, there was no empirically validated training program aimed at increasing pilot resilience during unexpected events.

- Second, numerous definitions were associated with the term, *resilience*, we recommend focusing on the behaviors and skills associated with increasing positive outcomes during unexpected events.

- Third, there is evidence to suggest that behaviors and skills associated with positive outcomes and resilience can be observed, such as **domain expertise** (i.e., knowledge/skills/abilities pertaining to piloting aircraft and completing specific piloting tasks), **critical thinking** (i.e., using cognitive skills/strategies to procure desirable outcomes), **self-efficacy** (i.e., confidence in one's abilities to execute a particular task)**, metacognitive awareness** (i.e., recognition of one's current level of understanding), **cognitive flexibility** (i.e., restructuring one's current action script to adaptively respond to dynamic/unexpected situations), and **decision-making strategies** (i.e., techniques of seeking and evaluating information)**.** Moreover, while **domain expertise** is a necessary pre-requisite, it is insufficient at capturing the whole picture of a resilient system. More research is needed to determine how to improve these measures.

- Fourth, **scenario-based training**, and more specifically, **variable** and **unpredictable scenario-based training** (i.e., scenarios that deny information about upcoming events) have shown demonstrable effectiveness at increasing pilots' skills during unexpected aviation events (Landman et al., 2020).

The results of the literature review served to frame inquiries central to Task 3: The Knowledge Elicitation Interviews, whose goal was to capture both expert and novice narratives about how flightcrew have executed positive behaviors to unexpected events.

Task 3: Pilot Needs Analysis

The purpose of the Pilot Needs Analysis procedure was to evaluate the findings from the literature review and to gain additional insight prior to designing a human-in-the-loop (HITL) experiment. We aimed to identify what events pilots found surprising or unexpected, how these events were handled by aircrew, and how performance and positive outcomes during unexpected events can be facilitated. By conducting semi-structured interviews with US Air Carrier pilots, instructor pilots, and evaluators, we obtained expert input on current levels of pilot performance in line operations as well as gathered feedback on how to improve pilots' responses to these off-nominal incidents. The nature of the data collection (interviews) provided limited insight into actual pilot performance; however, the interviews did provide useful insights that were not necessarily evident from the literature review. The results of Task 3 revealed the following findings (bullets listed below) and HITL recommendations (sub-bullets listed below) for developing a HITL for unexpected aviation events (Please see Section 1.3.4 for a more comprehensive list and explanation of the findings from Task 3):

- The top three surprising or unexpected events were (1) aircraft systems events (e.g., engine, hydraulic, or gear failure); (2) environmental events (e.g., icing, wake turbulence, or weight and balance events); and (3) human events (e.g., events caused by a crew member and/or passenger behavior).
  - Given that pilots found these types of events to be the most unexpected or surprising, designing HITL scenarios with such events may provide the best path forward for evaluating the effectiveness of training interventions.

- When faced with unexpected events, pilots used the strategies of "aviate, navigate, communicate," following checklists and procedures, "winding the clock" (i.e., slowing down and taking time to think before acting), maintaining situation awareness, and utilizing their team both within and beyond the flight deck.
  - The identified strategies should be included as independent or dependent variables in the HITL. As independent variables, we may test their effectiveness at improving performance during unexpected events. As dependent variables, we may test *if* and *how* the training intervention bolstered the use of these beneficial strategies.

- When faced with unexpected events for which there is no known procedure (or the procedure/checklist is incomplete or incorrect), pilots reported adapting known procedures and using their system knowledge. Pilots also reported leveraging past experiences for use in the current unexpected situation. That is, if the pilot has experienced a similar situation in the past, they might apply their knowledge of the event and/or strategies that were helpful in overcoming that previous event to the problem at hand. For example, one pilot in the interviews reported diagnosing smoke as being caused by an electrical fire based on having experienced an electrical fire in the past.
  - These behaviors may be valuable to include as dependent variables in the HITL study.

- In general, pilots claim they handle unexpected events well, though some better than others.
  - The method of assessing performance in the HITL must be sensitive as to capture even subtle differences in performance of the unexpected event.

- Training on handling unexpected events included Crew Resource Management (CRM), Line Oriented Flight Training (LOFT), Advanced Qualification Program (AQP), expanded envelope training, and leadership training.
  - The HITL study ought to test the effectiveness of interventions *other* than what is currently being used.

- Crew procedures found useful when faced with unexpected or surprising events were maintaining flight deck discipline, employing active and open communication, and reaching consensus among the flightcrew.
  - Teamwork and communication are crucial for responding successfully to unexpected events, thus, they ought to be included as dependent variables in the HITL study.

- Overall, pilots believed that more simulator, ground, and LOFT training would help improve responses to unexpected events. Regarding simulator training, they emphasized that training on unexpected events and/or events that impart startle and surprise would be particularly beneficial. They also believed that more training on events found to be unexpected in the past would be beneficial for handling unexpected events in the future.

- Simulator training involving unexpected or surprising events should be considered as a training intervention for the HITL study.

The results of Task 3 suggest that more studies are still needed to determine how these constructs could be promoted during training in a way that elicits resilient behaviors during any unexpected events.

Task 4: Training Development Plan to Test the Training Effectiveness of Pilot Responses for Unexpected Events

The purpose of Task 4 is to provide recommendations for testing training interventions aimed at increasing pilot performance during unexpected events.
Tasks 1-3 provide the findings that allowed the researchers on this team to design an experimental protocol for testing training interventions for air carrier pilots during unexpected events. These findings were mentioned above. Therefore, Task 4 provides researchers and training instructors with a method (based on the findings of Tasks 1-3) for running a HITL for air carrier pilots during unexpected events including the following:

- A thorough summary of the findings from Tasks 1-3

- Proposed training methods
  - (i.e., variable training, training with mnemonics)

- List of the independent and dependent variables to be measured in the HITL:

  - *Metaprocedure (*An easily administered training technique or memory aid in the form of a high-level procedure that can be used to enhance pilot performance during an unexpected event (e.g., mnemonic)).

  - The proposed metaprocedure serves as a form of **independent variable.** This variable consists of two levels: metaprocedure training intervention and the allied control of no metaprocedure training intervention.

  - *Variable Training:* Variable training serves as a second **independent variable.** This manipulation has two levels: variable training intervention and no variable training intervention. Participants who receive this training intervention are presented with unpredictable and variable scenarios prior to the exposure of the test unexpected event scenario.

  - *Performance:* Performance on an unexpected event task serves as the **dependent (i.e., outcome) variable.** Performance is to be measured via a comprehensive evaluative tool. Participants will be evaluated based on their (a) decision-making ability during the event; (b) ability to recognize the unexpected event; (c) response to the unexpected event; and (d) efficacy in the category of teamwork/communication. Table 4 breaks down this measure.

  - *Domain Expertise:* Domain expertise is envisaged to serve as a covariate in the study. Domain expertise is to be measured by participant flight experience and knowledge assessments instrument. By including domain expertise as such a covariate, we can account for and control for, participant's pre-existing levels of domain expertise. The latter can clearly influence their performance on the unexpected event before any training interventions are implemented.

- Proposed hypotheses grounded in theory to test in the HITL:

  - *Metaprocedure (Treatment A)*
    - H_0: After controlling for domain expertise, we postulate that there will be no difference in performance on the unexpected event scenario between participants receiving the metaprocedure variable training intervention and participants in the comparable control group.
    - H_1: Controlling for domain expertise, participants receiving the metaprocedure training intervention will outperform contrasts on the unexpected event scenario.

  - *Variable Training (Treatment B)*
    - H_0: Controlling for domain expertise, there will be no difference in performance on the unexpected event scenario between participants receiving the variable training intervention and participants in the control group.
    - H_2: Controlling for domain expertise, participants receiving the variable training intervention will perform better on the unexpected event scenario compared participants in the control group.

  - *Interaction of Training Interventions (Treatment A & Treatment B)*
    - H_0: Controlling for domain expertise, there will be no difference in performance on the unexpected event scenario between participants receiving both training interventions and participants in the control group.
    - H_{3A}: Controlling for domain expertise, participants receiving both training interventions will perform better on the unexpected event scenario compared to participants receiving only one of the training interventions.
    - H_{3B}: Controlling for domain expertise, participants receiving both training interventions will perform better on the unexpected event scenario compared to participants in the control group.

- Recommended study population for the HITL.

- Proposed candidate scenarios for unexpected events:

  - Squat switch (Weight-on-Wheels, proximity, air/ground) switch failure on take-off

  - Cabin pressurization loss in climb (above 15,000 feet)

  - Runaway stabilizer trim in climb

- Severe turbulence in climb
- ATC erroneous vector
- Rejected take-off (RTO) with Brake Temperature Monitor System deferred per the MEL (See pp. 32-33 for example https://fsims.faa.gov/wdocs/mmel/b-737_rev_61.pdf) – this is a short scenario to observe the RTO and whether the procedures per the MEL guidance are followed.
- Glidepath fails to capture on RNAV approach (on an approach that needs LPV minimums).
- High engine vibration readings in moderate turbulence – with the appropriate system knowledge, the pilot would know that environmental perturbations should not increase engine vibrations (much) and it will be telling if the checklist is consulted or not.
- Slow degradation of engine performance (fuel control unit impending failure) – an insidious failure may not get the attention of the pilot until alerts occur but diagnosing and responding to the condition will show the process of responding to the event.
- Passenger illness requiring diversion (very common occurrence in commercial aviation) and unable to contact company medical service for advice– how will the pilot react?
- “Door Light” on climb-out – While there is a checklist for this, will it be appropriately accomplished?
- Engine Bleed Air Shutoff Valves (PRSOV) deferred and flight encounters unforecast icing conditions.
- Degraded Flight Control Law Mode – An unexpected degradation of to a degraded flight control law without evidence of the cause at high altitude.

- Proposed guidance and examples on how to develop scenarios:
  - Examples of how to elicit variable measurement in the candidate scenario.
  - An example task analysis performed on a candidate scenario.
  - An example performance assessment designed to measure pilot behaviors during unexpected events.
  - An example protocol and detailed script (Weight-on-Wheels Switch Failure Scenario) to use in the HITL.
  - A recommended HITL experimental design and method for analysis.
  - Appendices that provide corresponding survey instruments for measuring the study variables and forms typically used in conjunction with scenarios for unexpected events.

Summary and Conclusions

The present document provides a primer as to how it is possible to conduct a human-in-the-loop (HITL) research procedure. Outcome results will be important to current and prospective aviation challenges. The insights gained from this document are accrued from the previous tasks associated from this effort (see Section 1). The results of this effort conclude the following findings and recommendations:

- Identification of key independent (i.e., variable training, mnemonic applications) and dependent variables (i.e., decision-making ability, recognition, responding, and teamwork/communication) for assessing pilot responses during unexpected events, in addition to directly assessing pilot task performance for a given event response, these variables provide indirect measures of pilot responses or thinking that are correlated with "good" (e.g., resilient) responses.

- Recommendations from the literature and pilot elicitation interviews on types of training interventions (i.e., variable training, mnemonic applications) and measurable skills and behaviors that pilots may exhibit during unexpected events (e.g., crew procedures found useful when faced with unexpected or surprising events were maintaining flight deck discipline, employing active and open communication, and reaching consensus among the flightcrew).

- Recommendations for scenario selection criteria. For the purpose of this research, which is for filling in the gaps when procedures do not exist or are not completely applicable, it is important to select the appropriate unexpected event scenario for HITLs in order to evaluate training interventions. Furthermore, we have identified several candidate scenario examples and provided a detailed scenario description and application (i.e., Weight-on-Wheels Switch Failure scenario).

- Recommendation of framework for conducting a HITL that also incorporates best practices for aviation HITL scenarios and simulations.

- Recommendation of performance assessment method for measuring pilots' behaviors during unexpected events. (Although the example performance assessment offered in this document is specific to the Weight-on-Wheels Switch Failure scenario, it can easily be adapted to any unexpected event scenario after performing a tasks analysis on the chosen scenario prior to developing the performance assessment.)

Ultimately, we want to create safer aviation systems with the developing complexities that are interjected into the evolving airspace system. This document provides a method supported by empirical evidence to explore the effects of training interventions on pilot responses to unexpected events. Although candidate training interventions are presented, the focus is on recommendations for dependent measures and event scenarios. Research results can ultimately lead to testing and validation, to inform FAA personnel who develop evaluation criteria for pilot tasks, skills, knowledge, and proficiency and incorporate this information into human factors related documentation. FAA documentation can be used to evaluate training and qualification programs.

Table of Contents

Figures

Tables

1. Background and Overview

Aircraft operations, especially for air carrier flight, are highly proceduralized. Specifications of pilots' tasks, as well as procedures of others involved in airspace operations have been major contributors to improvements in overall safety and the evident performance efficiency of the National Airspace System (NAS). Such procedures are enabled and enacted through extensive pilot training for both normal as well as off-nominal operations. Within the context of the present project, we define pilot training as can be found in 14 CFR Part 121 Air Carrier airman certification. This includes academic pilot training, air carrier ground school, simulation training, and in-flight operational experience. To preserve safety within the National Airspace System, the FAA is directed by Congress to maintain security, sustainability, and pilot performance within the navigational infrastructure (FAA, 2016).

In the operational lifetime of complex interactive systems, unexpected events can be anticipated to occur. The aviation community cannot write checklists and procedures for all such eventualities since, by definition, many events are necessarily unpredictable. At such junctures what mitigates failure is the adaptive and resilient capacities of the system to respond. Determinative findings are required in order to develop recommendations to improve pilot behavior during these unexpected events; thereby maintaining acceptable levels of safety and response efficiency. Considering these premises, we propose a series of evaluations which are designed to understand and promote such increased pilot response by the air carrier flightcrew to unexpected events. The scope of our effort focuses on those tasks that facilitate quantitative measurements of performance. From this foundation can be established steps to identify recommend training interventions. Specifically, these are anticipated to be in increasing pilot performance and adaptation into future 14 CFR Part 121 Subparts M (Airman and Crewmember), N (Training Program), O (Crewmember Qualifications), and Y (Advanced Qualification Program [AQP]) conduits of improvement. Tasks completed toward this goal are:

- Task 1: Research Plan
- Task 2: Relevant Research Assessment
- Task 3: Pilot Needs Analysis
- Task 4: Training Development Plan to Test the Training Effectiveness of Pilot Responses for Unexpected Events

These tasks are described in more detail in the sections which follow. This document represents Task 4, the Training Development Plan to Test the Training Effectiveness of Pilot Responses for Unexpected events of our overall project. The purpose of this document is to articulate a detailed experimental protocol, a list of candidate scenarios, an assessment tool for measuring pilots' responses to unexpected events, and to provide recommendations for testing training interventions aimed at increasing pilot performance during unexpected events.

1.1 Task 1: Research Plan

The inaugural project meeting which occurred on December 13, 2019, provided our research team with the opportunity for mutual clarification and collaboration and an opportunity to discuss more specific directions and details for the tasks outlined in the proposed work. All members of the research team were included in this meeting. Dr. William Kaliardos and Dr. Kathy Abbott attended on behalf of the FAA. They served to identify and elaborate on current general concerns related to pilot responses to unexpected events in contemporary and prospective aviation systems. The result derived from all subsequent meetings to date have established that the project's ultimate goal is to *provide the FAA with training information so that FAA can provide guidance through outlets such as Advisory Circulars* for the training requirements of 14 CFR 121 Subpart N (Training Program), Subpart O (Crewmember Qualifications), Subpart Y (Advanced Qualification Program), and 14 CFR 142 (Training Centers). Such information may have applicability and be useful for 14 CFR 61 (Certification: Pilots, Flight Instructors, and Ground Instructors), 14 CFR 141 (Pilot Schools), and 14 CFR 91 Subpart K (Fractional Ownership Operations). However, since the research plan was completed, the project's research scope was adjusted from providing validated recommendations on training characteristics, to providing a framework to test and evaluate training interventions.

The information our team has provided was therefore directed to a focus on training to improve strategies to build expertise, cognitive flexibility, and performance adaptation during unexpected events. We have largely limited the focus of our work to air carrier operations. However, we wish to point out that results of our efforts can provide guidance and recommendations for other aviation operations and for other industries as well.

It is important to note that training pilot resiliency is not necessarily a goal of our project. However, it could be a potentially beneficial and secondary outcome of the knowledge developed. The current project was designed explicitly to build upon the report by Pruchnicki, Key, and Rao (2019) on training resiliency literature review. Specifically, we have looked to identify the resiliency body of knowledge with respect to pilot expertise in existing and newly forming teams. In addition, our project team leveraged insights derived from the observations of Holbrook et al. (2019), Chandra et al., (2020), and the EASA's mandatory training resiliency program (Flin, 2019). While we sought to emphasize a theory-driven approach to guide our methods, we also employed applied methods to deliver a final product that provides ecological validity as well as applicational utility.

1.2 Task 2: Relevant Research Assessment

The primary purpose of Task 2 was to understand and synthesize the extensive literature surrounding pilots' behaviors and responses to unexpected events in order to determine how to better prepare them to respond with a positive outcome to future aviation challenges. Table 1 serves as a general summary of these the findings from this comprehensive literature review (and see Hancock et al., 2021).

Table 1. Summary of Task 2 Relevant Research Assessment Findings

Literature Review Finding	Throughput from Task 2	Output to Following Study Tasks
Prevalence of the construct of resilience in different domains (business, finance, transportation, psychology, engineering, etc.)	Focused perspective on aviation context	Narrow the contextual scope of the construct
Identify the comprehensive range of definitions for concepts related and integral to adaptation and resilience	Determine study definitions of key terms	Provide consistent terminology for sequential study tasks
Concepts related to resilience and adaptation are intertwined	Address similarities and differences in terms susceptible to misinterpretation (e.g., surprise/startle; resilience/adaptability)	Operationalization of study variables specific to identified appropriate concept
Challenges of studying the construct of resilience due to the complexity of the impinging factors	Investigate ways to clarify factors involved in resilience such as concept mapping and/or Principal Components Analysis	Output of such activities shows how each factor contributes to overall individual (and system) resilience
System-level vs. individual-level resilience	Focus on individual resilience	Structure study questionnaires and scenarios to focus on pilot responses to the system, as opposed to system resilience per se
Few identified studies specifically use resilience as a variable	Identified variables used as indicators of resilience (i.e., metacognition, cognitive flexibility, decision-making strategies, adaptive expertise)	Consider previous research variables in analysis of interview and scenario data
Most useful studies focused on unexpected events (Pruchnicki et al., 2019; Field et al., 2018; Landman et al., 2018; Kochan, 2005)	Noted study designs, variables, scenarios, methods, apparatus, data collection, data analyses, and results of identified, related studies	Lessons learned from previous specific research studies (successes, challenges, and failures) which then contribute to our own task designs

The results of the literature review from Task 2 revealed numerous critical key points for moving forward with studying pilots' responses during unexpected events. First, since there was no empirically validated training program aimed at increasing pilot resilience during unexpected events and numerous definitions were associated with the term, *resilience*, we recommend focusing on the behaviors and skills associated with increasing positive outcomes during unexpected events. Second, it is evident from our literature assessment process and case studies of positive outcome events, that the behaviors and skills most associated with resilience are:

- *Cognitive flexibility* – The ability to quickly and accurately restructure the current action script to adaptively respond to dynamic situations and the inevitable unexpected event.

- *Critical thinking* – Using one's cognitive skills and strategies (e.g., assessing the validity of arguments, testing hypotheses, employing creative thinking, detecting bias) to increase the likelihood of a desirable outcome (Magno, 2010).

- *Decision-making strategies* – Demonstrating effective techniques of seeking and evaluating information to include the influence of biases and other cognitive factors on decision quality.

- *Domain expertise* – The knowledge, skills, abilities, and understanding of the essential aspects of piloting an aircraft and/or the knowledge, skills, abilities, and understanding needed to complete a specific task.

- *Metacognitive awareness – Recognizing and reflecting upon one's current level of understanding to decide when it is and is not adequate.*

- *Self-Efficacy* – Confidence in one's abilities to execute a particular task (Bandura, 2006).

Moreover, while domain expertise is a necessary pre-requisite, it is insufficient at capturing the whole picture of a resilient system. Third, scenario-based training, and more specifically, variable and unpredictable scenario-based training have shown demonstrable effectiveness at increasing pilots' skills during unexpected aviation events (Landman et al., 2018).

1.2.1 Positive Outcomes in Past Aviation Events

Through identifying the positive behaviors exhibited by aircrew that led to positive outcomes in off-nominal events, we can investigate what behaviors are historically used to maintain a safer flight. The current project can point to patterns of repeatability, trainability, and so ultimately maintain a safer aviation system. A Safety II approach serves to focus on all the positive behaviors pilots do to contribute to the successful outcome of any flight (see Hollnagel, 2006). This differs from an accident-driven focus of Safety I, which centers more around human error (and see Reason, 1990). Consider the following three events observed in relatively recent history (see Table 2). These provide examples of how pilots and crew exercised appropriate responses when faced with the unexpected. A team of three human factors experts independently analyzed each of the four events displayed in Table 1. They then coded all positive behaviors revealed by aviation personnel and passengers associated with that event. Then, the positive behaviors were categorized by Safety II constructs (i.e., decision-making strategies, metacognition, cognitive flexibility, and expertise). From the extracts derived from these examples, it is evident that crewmembers relied not only on standard flight training to critically maneuver through these events, but Crew Resource Management (CRM) training and lessons learned through previous live flights as well. Of course, we fully acknowledge the nature and limitations inherent in any such small sample extraction processes, nevertheless we believe the lessons are of value, especially as they pertain to the overall problems associated with 'rare-event' evaluations (and see Hancock, 2021). Looking at specific events, take for example, US Airways Flight 1549. There is a long list of positive behaviors associated with the captain, crew, and even passengers' decision-making. This observation demonstrates the need for a continuation and a stronger standardization of documenting positive actions in all aviation events. From these, instructors can maintain relevant and effective training to educate pilots as to how to exhibit and enact these positive behaviors during unexpected events.

The results of the literature review and the information encapsulated in Table 2 (Positive Behaviors Associated with Aviation Events) served to frame inquiries central to Task 3: The Knowledge Elicitation Interviews, whose goal was to capture both expert and novice narratives about how flightcrew have execute positive behaviors to unexpected events.

Table 2. Positive Behaviors Associated with Aviation Events

Event	Date	Safety II Constructs from Research Assessment	Positive Behaviors
United Airlines Flight 232 (Sioux Gateway Airport)	July 19, 1989	Cognitive Flexibility	The ability for aircrew to quickly learn to work with new teams such as the National Guard.
		Decision Making Strategies	Positive communication among the aircrew and the respect of each crewmember.
		Expertise	High level of domain and judgment expertise among the crew.
Aloha Airlines Flight 243	April 28, 1988	Expertise	High level of flight expertise between the captain and first officer.
		Cognitive Flexibility	The ability to flex communication styles from verbal to nonverbal due to ambient noise
		Decision Making Strategies	Thinking quickly under time pressure and then responding quickly.
US Airways Flight 1549	January 15, 2009	Decision Making Strategies	Effective communication and coordination between captain and first officer.
		Decision Making Strategies	The allocation of tasks to passengers and their openness to perform tasks.
		Cognitive Flexibility	Flight attendants were able to flex the evacuation commands by instructing those who were able to jump over seats.
		Cognitive Flexibility	The ability to mentally adjust checklists in order to perform under time pressure, applying successes from past experiences.
		Metacognition	Captain being aware that he had the confidence to land the plane in the Hudson River.
Southwest Flight 3472	August 27, 2016	Decision Making Strategies	Decision by the crew to complete the SWA Engine Fire or Engine Severe Damage or Separation Checklist; Decision by Captain to delegate tasks to First Officer and cabin crew members.
		Cognitive Flexibility	Crew coordination and the ability to change the mode of communication throughout the event (e.g., crew switched from verbal to non-verbal communication when the noise level made words inaudible.)

1.3 Task 3: Pilot Needs Analysis

1.3.1 Purpose

The purpose of the Knowledge Elicitation Interview procedure was to evaluate the findings from the literature review and to gain additional insight prior to designing a human-in-the-loop (HITL) experiment. We aimed to delve into such situations to identify what events pilots found surprising or unexpected, how these events were handled by aircrew, and how performance and positive outcomes during unexpected events can be facilitated. By conducting semi-structured interviews with US Air Carrier pilots, instructor pilots, and evaluators, we obtained expert input on current levels of pilot performance in line operations as well as gathered feedback on how to improve pilots' responses to these off-nominal incidents.

1.3.2 Method

Two human factors researchers conducted semi-structured interviews with fifty (*n*=50) airline pilots. Each session lasted approximately one hour and entailed 10 questions that had further prompts to elicit expert knowledge from these participants. These questions elicited information regarding (a) examples of events the participant found unexpected; (b) their current procedures regarding responding to unexpected events; (c) how desired behaviors of the flightcrew are compared with the current behaviors of the flightcrew; (d) where gaps in procedures or checklists may exist regarding unexpected or ill-defined events; and (e) what kind(s) of training might improve responses to unexpected events. In addition to the interview questions and prompts, we administered five judgment expertise surveys: self-efficacy and metacognition, generalized self-efficacy, trust in automation, cognitive flexibility, and aviation self-efficacy (see Appendix C.2). The purpose of these surveys was to obtain measures of four factors (human traits/states/behaviors) found in the literature (Hancock et al., 2021) that aid in responding to unexpected events.

The responses from the interviews were coded by two researchers using a grounded theory framework. The researchers reviewed the interviews/transcripts and made note of what factors were involved in (a) what was described to be surprising or unexpected; and (b) the circumstances surrounding the event and its resolution. After the initial review of relevant factors, the researchers created categories of variables based on how similar the described events and factors were. The categories that emerged from this process were then combined into themes associated with the resiliency needed to deal with unexpected events.

1.3.4 Results/Discussion

The results from the knowledge elicitation interviews provided (a) categories of events the participants found to be unexpected; (b) procedures regarding responding to unexpected events; (c) examples of behaviors associated with flightcrew response; (d) where gaps in procedures or checklists may exist regarding unexpected or ill-defined events; and (e) what kind(s) of training might improve responses to unexpected events. Table 3 describes these results in more detail.

Table 3. Summary of Task 3 Relevant Findings

Task 3 Finding	Output to Following Study Tasks
The top three surprising or unexpected events were (1) aircraft systems events (e.g., engine, hydraulic, or gear failure); (2) environmental events (e.g., icing, wake turbulence, or weight and balance events); and (3) human events (e.g., events caused by a crew member and/or passenger behavior).	Given that pilots found these types of events to be the most unexpected or surprising, designing HITL scenarios with such events may provide the best path forward for evaluating the effectiveness of training interventions.
Technology events (e.g., loss of flight management system (FMS)) were least frequently reported as surprising or unexpected.	The scenarios for the HITL study should not focus on this kind of event.
When faced with unexpected events, pilots used the strategies of "aviate, navigate, communicate," following checklists and procedures, "winding the clock" (i.e., slowing down and taking time to think before acting), maintaining situation awareness, and utilizing their team both within and beyond the flight deck.	The identified strategies should be included as independent or dependent variables in the HITL. As independent variables, we may test their effectiveness at improving performance during unexpected events. As dependent variables, we may test *if* and *how* the training intervention bolstered the use of these beneficial strategies.
When faced with unexpected events for which there is no known procedure (or the procedure/checklist is incomplete or incorrect), pilots reported adapting known procedures, using their system knowledge, and leveraging past experiences. Leveraging past experiences involves applying knowledge/strategies used to overcome past events to the respond to problem at hand. For example, one pilot reported diagnosing smoke as being caused by an electrical fire based on having experienced an electrical fire in the past.	These behaviors ought to be included as dependent variables in the HITL study.
In general, pilots claim they handle unexpected events well, though are some better than others.	The method of assessing performance in the HITL must be sensitive as to capture even subtle differences in performance of the unexpected event.
Training on handling unexpected events included Crew Resource Management (CRM), Line Oriented Flight Training (LOFT), Advanced Qualification Program (AQP), expanded envelope training, and leadership training.	The HITL study ought to test the effectiveness of interventions *other* than what is currently being used.
Crew procedures found useful when faced with unexpected or surprising events were maintaining flight deck discipline, employing active and open communication, and reaching consensus among the flightcrew.	Teamwork and communication are crucial for responding successfully to unexpected events, thus, they ought to be included as dependent variables in the HITL study.
Pilots believed that more simulator, ground, and LOFT training would help improve responses to unexpected events. Regarding simulator training, they emphasized that training on unexpected events and/or events that impart startle and surprise would be particularly beneficial. They also believed that more training on events found to be unexpected in the past would be beneficial for handling unexpected events in the future.	Simulator, ground, and LOFT training should be considered as training interventions for the HITL study.

The Pilot Needs Analysis served to identify constructs that proved to be most relevant to training as to how to respond to unexpected events. Participants reported that events related to the aircraft system, environment, and humans generated the most surprising and unexpected events, while technology was often not the cause of unexpected events. In addition to identifying the most common unexpected/surprising events, Task 3 generated specific examples of events that may be used when creating possible HITL study scenarios. For example, a gear failure event.

This task also served to identify the strategies and behaviors that are beneficial when faced with unexpected or surprising events. Participants reported relying on the maxim of "aviate, navigate, communicate" and reported "winding the clock," following procedures, maintaining situation awareness, and teamwork. In situations where the procedure is missing or ill-defined, participants reported adapting and leveraging their expertise. These constructs align with the elements associated with resilient behaviors that were identified within the literature review phase (i.e., decision making, biases, cognitive flexibility, metacognitive awareness, self-efficacy, and teamwork/communication. The identified strategies and behaviors will aid in the design of the HITL study in two ways. First, we might consider training interventions intended to improve these beneficial strategies. If participants reported relying on these strategies to overcome unexpected/surprising events in the past, it is possible that these strategies are trainable and may serve to improve performance in future and novel unexpected events. Second, these strategies and behaviors might serve as dependent measures to assess the effectiveness of the chosen training interventions in the HITL. If a given intervention is designed to teach or help the pilot to implement these strategies, the presence of these strategies during the HITL scenario would indicate that the training intervention was successful. For instance, if the training intervention prompts the pilot to utilize their team beyond the flight deck, and they communicate with ATC during the event, we might consider the intervention to be successful.

We also learned that most pilots believe that unexpected or surprising events are handled well, though some pilots handle them better than others. Experience was the most frequently cited reason for why a pilot might struggle in handling unexpected events. If most pilots handle unexpected events well, finding significant difference in performance might be challenging. In psychological research, this phenomenon is known as the "ceiling effect." Thus, any method of assessing performance in the HITL must be sensitive enough to capture even minute differences in performance. For this reason, we recommend a robust performance assessment that can capture a variety of variables that indicate successful performance.

Finally, we learned that most pilots have participated in training on unexpected events, including Crew Resource Management (CRM), Line Oriented Flight Training (LOFT), Advanced Qualification Program (AQP), expanded envelope training, and leadership training. Pilots believed that more simulator, ground, training on past events, and LOFT training would help improve responses to unexpected events. Thus, we might consider choosing interventions for the HITL that use the methods described as currently lacking in the pilots' training, such as simulator and ground training.

The results of Task 3 also suggest that more studies are still needed to determine how these constructs could be promoted during training in a way that elicits resilient behaviors during any unexpected events (i.e., variable and unpredictable scenario training and mnemonic training). For example, the results of the proposed HITL could establish that training cognitive anchoring can help mitigate decision-making during unexpected weather events, then training instructors would need to specify how to implement techniques for mitigating this cognitive bias into their training scenarios.

1.4 Implications for HITL Study for Unexpected Events

The results derived from Tasks 1-3 set the foundation for developing an experimental protocol for testing any proposed training interventions aimed at increasing pilots' responses during unexpected events. The remaining portions of the current work provide researchers, practitioners, trainers and readers with a detailed experimental protocol, a list of candidate scenarios based on a task analysis, and an associated cognitive walkthrough, an assessment tool for measuring pilots' responses to unexpected events, and recommendations for training interventions aimed at increasing pilot performance during unexpected events.

In preparation for a HITL, the following elements are identified to facilitate the experimental design's effectiveness aimed at studying the behaviors that best prepare pilots' responses to off-nominal events. These elements are, as noted, specifically supported by an assessment of the current literature, a needs analysis, a cognitive task analysis, and an expert informed cognitive walkthrough of the candidate scenario. Throughout this effort we explicitly consulted with subject matter experts (SMEs) who possessed extensive expertise in flight operations, flight instructing, designing pilot training, and behavioral research methods.

2. Terminology for HITL Study on Unexpected Events

The results of the literature review and knowledge elicitation interviews in Tasks 2 and 3 respectively, support the need to further specify the trainable characteristics pilots exhibit in their successful performance during unexpected events. The Relevant Research Assessment explored and critiqued terms (e.g., adaptivity, confidence, decision-making strategies, expertise, cognitive flexibility) that featured during successful pilot behaviors during unexpected events. Use of these concepts was further supported by the results of the Pilot Needs Analysis which analyzed unexpected event narratives from airline pilots, instructors, and evaluators. The results of both Tasks 2 and 3 conclude that there are certain skills/behaviors/strategies that may serve as predictors of successful context-driven pilot performance during unexpected events which therefore can be tested in a HITL. Below, we list additional terms and constructs that are central to the HITL study.

2.1 Additional Terms and Constructs

Adaptability – In a general sense, we can view adaptability as the search for stability when new conditions arise. Resilience, in contrast, is the achievement of a *new state of stability* after new conditions arise.

Adaptive expertise – The ability to use and apply facts, knowledge, procedures, and decision-making strategies and combine them in new ways to deal with unanticipated situations.

Availability heuristic – Making a judgment based on the ease with which instances can be brought to mind (Tversky & Kahneman, 1974).

Cognitive anchoring – Making a judgment based on the first piece of information offered (Furnham & Boo, 2011).

Communication (verbal and non-Verbal)/CRM/TEM – The extent to which crewmembers provide and receive necessary information at the appropriate time (for example, initiating checklists and alerting others to developing problems). Decisions are clearly communicated and acknowledged, and there is an atmosphere that invites open and free communications.

Decision-making – A systematic approach to the mental process used to consistently determine the best course of action in response to a given set of circumstances.

Judgment expertise – An umbrella term to describe the cognitive skills needed to respond to unexpected events, including decision-making skills, cognitive flexibility, metacognitive skills, and adaptive expertise (Kochan, 2005).

Metacognition – The ability to monitor one's current level of understanding and decide when it is and when it is not adequate.

Metaprocedure – An easily administered training technique or memory aid in the form of a high-level procedure that can be used to enhance pilot performance during an unexpected event (e.g., mnemonic).

Monitoring – "The observing of the aircraft's flight path and systems and actively cross-checking the actions of other crewmembers" (FAA, 2004).

Novelty – A property of a stimulus that has not been previously presented to or observed by and is thus unfamiliar to the subject (Gordon & Luo, 2011).

Recognition – Identifying something totally with sense, perception, awareness and/or behavior.

Resilience – With a focus on system states: the capacity of a system to exhibit a new state of operational stability when adaptation to recover the prior base state has failed. And, pertaining directly to the human in the system: the ability to adapt to changing circumstances by attaining a differing form of operational stability through situation assessment, self-review, decision making, and action.

Safety – The condition where risks are managed to acceptable levels (ICAO, 2018).

Startle – A physiological reflex reaction to a sudden, intense stimulus triggering an involuntary physiological response to include eye blink, increased heart rate, and increased tension of the muscles.

Stress – "A response to threatening situations that involves biological, cognitive, behavioral, and emotional components" (Dismukes et al., 2015).

Surprise – An emotional and cognitive response to unexpected and difficult-to-explain events (Landman et al., 2017).

Unexpected event – An unexpected event is (a) An event incongruent with expectations as determined by base rate probabilities (average probability of event occurring) and the contextual information available; may be normal, abnormal, or emergency in nature; it may also be frequent, infrequent, or novel, or (b) the absence of an expected event (Kochan et al., 2004).

3. Candidate Training Methods

Through an extensive literature review regarding methods that have shown to be promising toward increasing resilience and operator responses to off-nominal events, the following two candidate methods are recommended for HITL evaluations. Other training methods may also be considered; these do not affect the remaining HITL recommendations in this document.

3.1 Variable Training

One of the key questions present throughout all of the present program has been how to train pilots for the unexpected so that they can practice applying a select range of applicable skills and then transfer those skills to situations when they face a surprising or unexpected encounter in real operational circumstances. When training becomes standardized, it begins to exhibit a level of predictability and consequently, assessing pilots' skills becomes uniform and less comprehensive. While talking to airline pilots and training instructors, our own in-house SME's were often reminded how challenging it would be to (1) train pilots for every type of abnormal situation, and (2) keep these training scenarios confidential so that pilots who had not yet participated would remain surprised. Fortunately, there is a solution to both of those concerns detailed above. This is via variable and unpredictable training.
With both variable and unpredictable training, we are not seeking to find a training scenario that will train airline pilots how to master every single unexpected situation. Pilots possess the ability to deviate off script and make decisions based on how they understand the situation. Individuals draw from past experiences to help interpret novel situations. Individuals rely on heuristics to aid in timely decision-making but sometimes, unfortunately, they enact cognitive biases that could cause their performance during these unexpected events to be inadequate. Beyond how to respond to a specific set of behaviors for a specific event, we can train pilots on how to develop the underlying skills that equip them to make the most appropriate response for many forms of unexpected event.

The skills learned during predictable and unvaried training are not thought to transfer well to unexpected situations and may not be ideal for training pilots on how to respond to unexpected events (Landman et al., 2017). However, variable and unpredictable training present pilots with simulator scenarios and exercises with which that are unfamiliar and lack specific information that pilots would come to expect. Unpredictable scenario training may lack the sensory cues that pilots come to expect, whereas variable training presents different characteristics of problems in various ways. Pilots are able to practice different combinations of problems, therefore, learning the desired knowledge and skills for each problem, but not ingraining the routine into procedural memory. Landman and colleagues (2018) found that pilots performed better during unexpected events when presented with unpredictable and variable scenario training compared to pilots in a control group.

3.2 Metaprocedure through Mnemonic Application

A mnemonic is a decision-making aid that pilots memorize to enhance their ability to respond during an unexpected event. We consider this mnemonic to be a form of "metaprocedure." Such procedures are easily administered during training as a memory aid that can be used to enhance pilot performance during an unexpected event. The following mnemonics represent examples of procedures proposed and studied to improve pilot management of the unexpected:

3.2.1 COOL (Calmdown, Observe, Outline, Lead)

The COOL mnemonic is a decision-making aid that pilots memorize to enhance their ability to respond during an unexpected event (Landman et al., 2020). COOL stands for: C = *Calm down*, take a deep breath, relax your shoulders, and become aware of your control inputs. O = *Observe*, scan the appropriate flight parameters without focusing on the problem. O = *Outline*, focus on the problem, what doesn't make sense, what do you think is going on? L = *Lead*, make a plan and take action (Landman et al., 2020).

Landman et al. (2020) showed the effectiveness of pilots' performance when using the mnemonic (COOL Procedure) during an unexpected event. Limitations to their specific study procedure included pilots rushing through the steps of the COOL technique and not focusing on the problem per se. In cases where the mnemonic becomes a distraction, the authors suggest that the mnemonic is trained repeatedly before pilots are presented with the unexpected event so that utilizing the mnemonic during flight becomes as natural as possible (2020). There are several different mnemonics that have been studied for pilots facing unexpected events. These mnemonics are listed below. We have chosen to use the COOL mnemonic for the current study because of its effectiveness, simplicity of understanding, and ease of training.

3.2.2 Other Example Metaprocedures

OODA Loop (Observe-Orient-Decide-Act) (Boyd, 1996)

- "Observe"
    - Understanding circumstances
    - Outside information
    - Unfolding interaction with environment
- "Orient"
    - Cultural traditions
    - Genetic heritage
    - New information
    - Previous experiences
    - Analyses and synthesis
- "Decide"
    - Implicit guidance and control
    - Create hypotheses

- "Action"
  - Test hypotheses
  - Unfolding interaction with environment

ROC (Reset-Observe-Confirm) (Boland, 2016)

- "Relax" – Take physical distance, breathe, relax muscles, check colleague
- "Observe" – Call out observations, what do/did they see, hear, feel, and smell.
- "Confirm" – Current and future situation; Is that true? How sure are we about this? Can we back up this single source of information?

BAD (Breathe-Analyze-Decide) (Martin, 2017)

- "Breathe" – Take physical distance, breathe, relax muscles, check colleague
- "Analyze" – Call out observations, what do/did they see, hear, feel, and smell.
- "Decide" – Current and future situation; Is that true? How sure are we about this? Can we back up this single source of information?

URP (Unload-Roll-Power) (Field et al., 2018)

- "Unload" – A technique for relaxing after being startled such as creating mental distance by deliberately pushing back into the seat, deep breathing, and conscious relaxing of muscles.
- "Roll" – Means regaining one's situational awareness which can be facilitated by each pilot stating out loud, "what they see, hear, feel, and smell."
- "Power" – A technique to reinforce metacognitive skills by monitoring critical thinking and projecting the consequences of threats into the future by asking questions about the reliability and validity of the information sources and assumptions being used.

LCAP (Learn-Coordinate-Adapt-Plan) (Jefferies & AA personnel, 2020)

- "Learn" – Applies what was previously learned, demonstrates a positive interest in acquiring knowledge and improving
- "Coordinate" – Ask other crew member for input or assistance, delegate and divide tasks, monitor automation, etc.
- "Adapt" – Address unanticipated new pressure, adjust communication method or pause based on other pilot's workload, change automation level/mode/programming for changing condition
- "Plan" – Conduct a thorough briefing, develop "what if" scenarios and plans for contingencies

4. Study Variables

4.1 Overview of Study Variables

The identified study variables that we advise, are neutrally comprised of both independent and dependent variables. They are ones that emerged most prominently from both our literature review and the knowledge elicitation procedure. These variables are described below.

Metaprocedure (Treatment A). The proposed metaprocedure serves as a form of independent variable. This variable consists of two levels: metaprocedure training intervention and the allied control of no metaprocedure training intervention.

Variable Training (Treatment B). Variable training serves as a second independent variable. This manipulation has two levels: variable training intervention and no variable training intervention. Participants who receive this training intervention are presented with unpredictable and variable scenarios prior to the exposure of the test unexpected event scenario.

Performance. Performance on an unexpected event task serves as the dependent (i.e., outcome) variable. Performance is to be measured via a comprehensive evaluative tool. Participants will be evaluated based on their (a) decision-making ability during the event; (b) ability to recognize the unexpected event; (c) response to the unexpected event; and (d) efficacy in the category of teamwork/communication. Table 4 breaks down this measure.

Domain Expertise. Domain expertise is envisaged to serve as a covariate in the study. Domain expertise is to be measured by participant flight experience and knowledge assessments instrument. By including domain expertise as such a covariate, we can account for and control for, participant's pre-existing levels of domain expertise. The latter can clearly influence their performance on the unexpected event before any training interventions are implemented. Table 5 breaks down this measure.

4.2 Defining and Measuring the Indicators of Study Variables (Operationalization of Concepts)

This section shows how the specific variables in the study are reflected in tasks (indicators) that are measurable. First, we define the independent, covariate, and dependent variables and then we show the indicators and how they can be measured (see Tables 4 and 5 for a list of all the variables and how to measure those variables).

4.1.2 Independent Variables

Two independent variables are those which emerged from the literature review and the pilot needs analysis (i.e., metaprocedure and variable training). We hypothesize that each of these variables will exert effects on pilots' performance (dependent variable) during unexpected events. As a consequence of their importance, the variables are explained in more detail below.

Metaprocedure (Treatment A). The COOL mnemonic is a decision-making tool that pilots memorize to enhance their ability to respond during an unexpected event (Landman et al., 2020).
Variable Training (Treatment B). Variable and unpredictable scenario-based training will be presented to the pilot with different simulator scenarios and exercises that are unfamiliar and excise specific procedural information that pilots might then come to anticipate.

4.1.3 Dependent Variables

Table 4 describes each of the dependent variables of performance to be used in the articulated study. Table 4 also describes assessment types, how to measure such performance, and the methods of measurement themselves.

Table 4. Variables of Successful Pilot Performance during Unexpected Events

Dependent Variable of Interest	Assessment Type	Indicator Variable	Methods of Measurement
1. Decision Making			
1.1. General (Task Management)	Performance Assessment	• Tasks are appropriately prioritized to maximize efficiency and secondary operational tasks (e.g., dealing with communications to the company) are prioritized to allow sufficient resources for primary flight duties. • Potential distractions posed by automated systems are anticipated, and appropriate preventative actions are taken	• Verbal Protocol Analysis (Talks through decision making in the moment) • Post interview questions (Elicits possible reasons why pilots made certain decisions and how open they were to alternative decisions (cognitive flexibility) • View playback of recorded scenario with pilot (determines how or why pilots made certain decisions)
1.2. Biases	Performance Assessment	• Determine whether the pilot used any cognitive biases or heuristics when making decisions (e.g., Did the pilot make decisions based on information that came readily to mind?)	
1.3. Cognitive Flexibility	Performance Assessment	• Trigger pilots to make alternative decisions and determine their response (e.g., For encountered weather vs briefed weather, does the pilot deviate or stay the course?)	
2. Recognition of Unexpected Events (Metacognitive Awareness)	Performance Assessment	• Determine what the pilots were observing and concluding prior to their final determination of the event. • Determine if and how long does it take for pilots to recognize unexpected event. • Determine what means were used to recognize the event. • Determine how accurate was pilot's identification of the problem (Based on SMEs input to performance assessment).	• Observation and simulation data • Verbal Protocol Analysis (Determines when and how pilot recognizes event) • Post Interview Questions (Determine when and how pilot recognized event) • View playback of recorded scenario with pilot (Determine when pilot recognized event)
	Survey Measurement	• Determine general metacognitive awareness to correlate with metacognitive awareness measured in performance assessment	• Survey of questionnaires (given once at the start of experiment)
3. Pilot Response to Unexpected Event			

Dependent Variable of Interest	Assessment Type	Indicator Variable	Methods of Measurement
3.1. General Responding	Performance Assessment	• Determine how long it takes the pilot to react to the unexpected event once it was recognized. • Determine if the pilot freezes. • Determine how well pilots respond to the unexpected event (based on Task Analysis of event response). • Determine if all available resources are used to accomplish the task. • Determine if time available for the task is well-managed. • Determine if demands on resources posed by operation of automated systems and recognized and managed. • Determine if the pilot purposefully takes no action.	• Observation and simulation data • SME subjective assessment • Verbal Protocol Analysis (Determines when pilot recognizes and then reacts to event, what pilot was thinking, and whether they expressed self-efficacy) • Post Interview Questions (Determine when and how pilot reacted to event)
3.2. Metacognitive Awareness in Responding	Performance Assessment	• Determine if the pilot is aware of thinking or thought processes during the unexpected event.	
3.3. Self-Efficacy in Responding	Performance Assessment/Survey Measurement	• Determine the pilot's self-efficacy during the event; did they show confidence, composure, risk-taking, etc.?/Questionnaire aimed at measuring pilots' confidence in their own skills or abilities for a particular task.	
4. Teamwork/Communication	Performance Assessment	• Determine conformance to CRM policy. For example: • Determine if pilot delegates if needed • Determine if pilot communicates to appropriate and available resources • Determine if crewmembers speak up and state their information with appropriate persistence until there is some clear resolution. • Determine if operational decisions are clearly stated to other crewmembers and crewmembers acknowledge their understanding of decisions.	• Verbal Protocol Analysis (Explores team communication) • Recorded simulation data (Explores all communication among pilot, crew, and other resources (ATC))

4.1.4 Covariate

Domain Expertise. Domain Expertise will serve as a covariate in this study. Domain expertise will be measured by participant flight experience and knowledge assessments. By including domain expertise as a covariate, we can account and control for participant's pre-existing levels of domain expertise that might influence their performance on the unexpected event before any training interventions are implemented. See Table 5 to see a breakdown of this measure.

Table 5. Measures of Task Specific and General Domain Expertise

Covariate Variable of Interest	Assessment Type	Indicator Variable	Methods of Measurement
Task Specific Domain Expertise	Survey Measurement	Counterbalanced knowledge exam written from the scenario task analysis	Survey of questionnaires (given once at start of experiment)
General Domain Expertise	Survey Measurement	Flight hours, certificates, ratings	Survey of questionnaires (given once at start of experiment)

5. Hypotheses

The following hypotheses were developed as explicitly based upon our review of the literature (Task 2) and the results of the knowledge elicitation study (Task 3).

Metaprocedure (Treatment A)

H_0: After controlling for domain expertise, we postulate that there will be no difference in performance on the unexpected event scenario between participants receiving the metaprocedure variable training intervention and participants in the comparable control group.

H_1: Controlling for domain expertise, participants receiving the metaprocedure training intervention will outperform contrasts on the unexpected event scenario.

Variable Training (Treatment B)

H_0: Controlling for domain expertise, there will be no difference in performance on the unexpected event scenario between participants receiving the variable training intervention and participants in the control group.

H_2: Controlling for domain expertise, participants receiving the variable training intervention will perform better on the unexpected event scenario compared participants in the control group.

Interaction of Training Interventions (Treatment A & Treatment B)

H_0: Controlling for domain expertise, there will be no difference in performance on the unexpected event scenario between participants receiving both training interventions and participants in the control group.

H_{3A}: Controlling for domain expertise, participants receiving both training interventions will perform better on the unexpected event scenario compared to participants receiving only one of the training interventions.

H_{3B}: Controlling for domain expertise, participants receiving both training interventions will perform better on the unexpected event scenario compared to participants in the control group.

6. Study Population

Using a representative sample for the study population can serve to foster generalizability in order so that results can be applied to the population of affected airline pilots. Ideally, the study should include airline pilots who possess a variety of experience levels and who fly across a variety of geographic regions and terrain (e.g., tropical weather, icing and snow, mountainous terrain, and other conditions of flight).

Choosing an appropriate sample size is critical in order to identify significant differences across conditions if such differences exist. We have determined that sample for the specified study should include 56 (n=56) participants (airline pilots). This recommendation is derived from a G*power analysis (see Faul et al., 2007), in order to achieve a Cohen's f^2 effect size of .15 or greater. This is considered a substantive and impactful effect size (Cohen, 1988) (and also, see Dattalo's (2013) method for MANCOVA power analysis). However, it may not be feasible to achieve this desired sample size.

However, even if it is not possible to recruit this target number of participants, the study can still provide valuable information about such training interventions. We therefore offer four suggestions for conducting the proposed HITL with a sample size.

1. *Run the Proposed HITL as a Pilot Study.* This strategy limits the ability to extrapolate strong conclusions about the interventions' effectiveness. However, a pilot study may provide sufficient evidence to warrant into a full-scale investigation.

2. *Remove Interactive Effects.* That is, the interaction between the metaprocedure and variable training intervention should be excised from the analysis by examining singular effect in order to achieve a more aligned power (Hopkin et al., 2016). Though this strategy would limit conclusions about the combined effect of the two training interventions, results for the two interventions separately may be critically meaningful.

3. *Remove One Independent Variable.* Removing one of the training interventions from the HITL may well allow for a more standard power level using fewer participants. This strategy may be effective if one of the training interventions is better suited for one's goals and resources compared to the other. The excluded intervention can be tested subsequently.

4. *Focus on Data Visualization and Descriptive Statistics.* For small participant numbers in which hypothesis testing offers only restricted value, it may be useful and insightful to focus on data visualization and descriptive statistics. As with the pilot study, this method may provide "proof of concept" visualization and guide the investment of resources for elaborated studies.

6.1 Participant Selection

The sample selection process should ideally also include a stratified procedure in which participants are drawn from different sub-groups of the overall pilot population. The target population for a study would include an ATP rated line pilot employed by a U.S. Part 121/135 operator. It is also advisable to include First Officers and Captains in the selection process since they would be working together during the event. When dual pilot crews are needed for a study, it is important to pair pilots from the same operator. This will help manage differences in training/line operation policies and procedures. It is recommended to draw a sample of participants who fly various routes, have variable expertise and training levels, and draw also from a wide range of ages. Furthermore, it is important to identify experience requirements for pilot participation in a HITL. For example, minimum: Licensing/certificates, type rating, recent flight experience (number of hours within a specified time period), and current crewmember role/function in their company's operation (e.g., Captain, First Officer).

7. Study Design

The postulated study is designed to assess the individual and/or combined effectiveness of the two training interventions relative to a control group. Thus, this study is designed to employ a 2x2 factorial design. The ideal sample will be represented by a random one split into four different groups with 14 participants in each group. Table 6 provides a further breakdown of each of these group.

Table 6. Study Groups

Group	Condition	Treatment	Number of Participants
Group 1	(Treatment A) Metaprocedure (through mnemonic)	Mnemonic presented to participants prior to unexpected event scenario	14
Group 2	(Treatment B) Unpredictable/Variable training	Unpredictable and variable scenarios presented to the participants prior to the unexpected event scenario	14
Group 3	(Treatments A&B) Both metaprocedure (through mnemonic and unpredictable/variable training	Both procedures presented to the participants prior to the unexpected event scenario	14
Group 4	(No Treatment) Control group	Participants in this group only receive the unexpected event scenario	14

8. Apparatus Selection

We suggest that the experiment should be conducted in ground schools, training devices, with Full Flight Simulators (FFSs) that possess enhanced aerodynamic models representative of the aircraft. These should be applicable to the study population.

9. Scenario Development for Unexpected Events Research

General scenario development guidance is detailed in Advisory Circular 120-54A, Advanced Qualification Program (FAA, 2017) and in Advisory Circular 120-35, Flightcrew Member Line-Operational Simulations: Line-Oriented Flight Training, Special Purpose Operational Training, Line Operational Evaluation (FAA, 2015) where a structured methodology is offered. The scenarios are composed of "event sets" that are independent segments in which specified purposes, including an event trigger, possible distracters, and supporting events are presented. These design principles can also be followed in the construction of scenarios designed specifically for the present focus on unexpected events, though there are several caveats to be considered for this area of research.

Often aviation activities prove to be highly proceduralized and non-normal or emergency events can have procedures or checklist protocols to be followed to resolve the situation. However, there are circumstances in which standard operating procedures and checklists cannot, or do not, resolve the problem. Regardless of the origin of the event(s), the common denominator is that these events (or the combination of events) may not have been specifically trained, and/or do not have an associated procedure, and/or the associated procedure(s) does not completely apply to the situation. The purpose of this section is to discuss scenarios that can be created to specifically train, evaluate, and investigate pilots' performance in response to these unexpected events.

9.1. Defining an Unexpected Event

When considering the design of scenarios to study reactions to unexpected events, an understanding of the working definition is needed to set the stage. While events that "have never happened before" would certainly be unexpected and create surprise, there are other events and combinations of situations that can be equally surprising and unexpected. The definition that we have adopted for this research is that an unexpected event is: "*(a) An event incongruent with expectations as determined by base rate probabilities (average probability of event occurring) and the contextual information available; may be normal, abnormal, or emergency in nature; it may also be frequent, infrequent, or novel, or (b) the absence of an expected event.*"

A taxonomy of the terms used in various studies of unexpected events is presented in Figure 1. These observations are taken from the work of Kochan et al. (2005). Events here are divided in terms of severity, frequency, and the degree of unexpectancy involved. Events may be divided, as a first pass set of differentiations into normal, abnormal, or emergency. Similarly, we can then sub-divide them into categories that can be labelled common, unusual, or entirely novel; and, they may be expected or unexpected. This provides the tree structure as shown in Figure 1 below. Clearly, these forms of differentiation also provide an intrinsic hierarchy of threat and risk. Novel, non-normal, emergency, and unexpected conditions are those which will be considered in the generation of study scenarios.

Some situations that are novel may not have a procedure or checklist that addresses the situation. Combinations of situations (such as a mechanical failure of the engine anti-icing system, while flying in icing conditions, and an inoperative alerting function) could cause surprise and confusion even though a procedure and checklist are available. Furthermore, situations could arise where the alerting function or information provided concerning the situation is ambiguous so that it is not easily discernable which procedure(s) or checklist(s) should be followed (for example, a Weight-on-Wheels switch failure). A situation that is seemingly mundane, such as a loss of GPS integrity on an RNP approach, can cause confusion and delayed action in response to this unexpected event. Even if the signal loss (or degradation) of the GPS is identified in a timely manner, the procedure to follow may not be clear-cut, air traffic control may be overwhelmed, and airspace complexity can become untenable in short order. Thus, we first must realize that the unexpected comes in many forms and can be the result of infinite combinations of circumstances. Although the definitions in Table 7 are somewhat academic, they provide context for the study of responses to unexpected events.

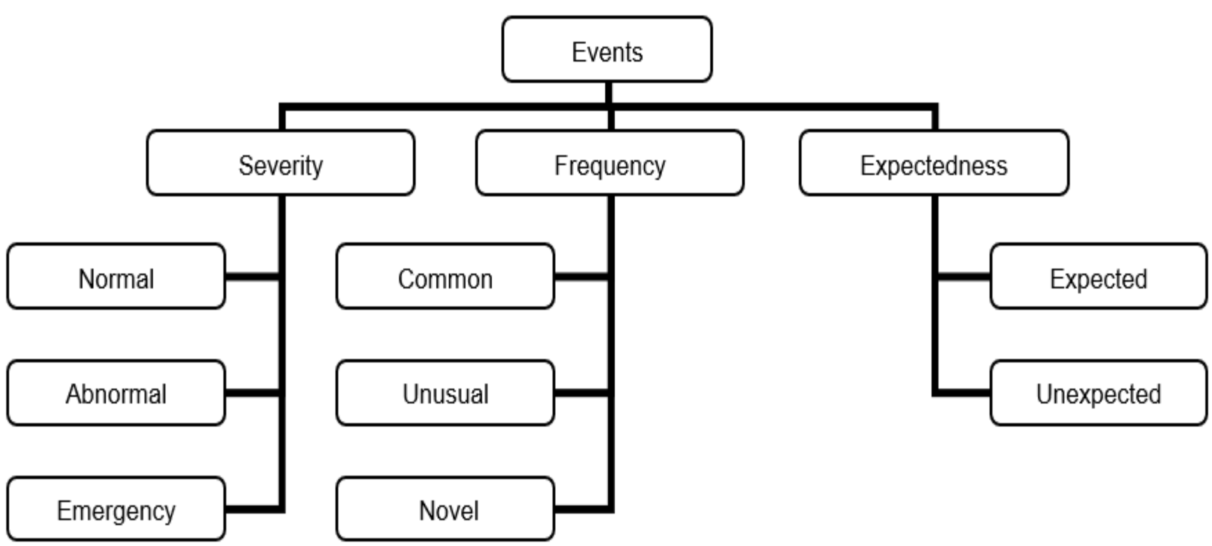

Figure 1. A Descriptive Taxonomy of Events in Terms of Severity, Frequency, and Expectedness

Table 7. Taxonomy of Events, Adapted from Kochan, Breiter, and Jentsch (2005)

Event Feature (Synonyms)	Working Definition
Severity of Event	
Normal (Nominal or Routine)	The status quo or homeostasis of the system consisting of a series of expected events leading to steady state operating conditions.
Abnormal (Off Nominal or Non-Normal)	An event requiring the use of additional equipment, standby systems, or resulting degraded system (human and/or machine) capabilities.
Emergency Event	A potentially hazardous event requiring timely resolution of the state; may evolve into a normal or abnormal state.
Frequency of Event	
Common Event (Frequent)	Events occurring with a high base rate; most often normal in nature, although may be abnormal or emergency events.
Uncommon Event (Unusual, Rare, Non-Frequent, or Non-Routine)	A seldom occurring event, varies with base rate of occurrences; may be normal, abnormal, or emergency in nature.
Novel Event	An event which has not been known to happen in the past; may be normal, abnormal, or emergency in nature.
Expectedness of Event	
Expected Event (Usual or Routine)	An event congruent with expectations of human and machine systems, and environment based on some prior knowledge, information, or preparatory information; may be normal, abnormal, or emergency in nature; it may also be frequent, infrequent, or novel.
Unexpected Event	(a) An event which occurs without the affordance of any prior information, cues, or warnings that could function to change (increase or decrease) the probability weighting of the event; may be normal, abnormal, or emergency in nature; it may also be frequent, infrequent, or novel (b) the absence of an expected event.

9.2 What Do Pilots Find Surprising or Unexpected?

When constructing a scenario designed to elicit surprise and present an unexpected event, it is important to understand what events pilots find surprising and unexpected. Research from pilot reports to the Aviation Safety Reporting System (ASRS) (Kochan et al., 2004) and from our pilot interviews (Hancock et al., 2021) shows us what types of events are unexpected and surprising to pilots. The database analysis showed the factors most frequently involved with surprising and unexpected events to be aircraft position/aircraft state, air traffic control, other crewmember actions, system status, automation, and environmental conditions (wind, turbulence, low visibility).

Similarly, our airline pilot SME interviews found that pilots found aircraft systems (e.g., engine, hydraulic, or gear failure events), environmental (e.g., icing, wake turbulence, or weight and balance events), and human events (e.g., events caused by a crewmember or passenger). Therefore, in general, all aspects of the *aviation system, human, machine, and environment* contribute to the generation of unexpected events. Again, for the purposes of creating study scenarios, the focus needs to be on those events or combination of events that may not have an associated procedure or checklist or there is one that in this case does not apply. For example, an aircraft may have a checklist for loss of hydraulic quantity, but there may be a case where the hydraulic system is still showing hydraulic pressure when it should not. Now the event is not clearly understandable as the conditions associated with the situation do not fit the procedure. Likewise, an extreme wake turbulence encounter may temporarily render flight instrument displays inoperative which combines an environmental event with a system anomaly. To evaluate how well pilots respond to unexpected events, it will be imperative that the test scenarios generate the requirements for the pilots to go beyond what is already trained and/or known or to be creative in resolving the situation presented to them.

9.3 Subject Matter Expert Input

Constructing scenarios to be used to train and evaluate pilots should adhere to the following guidelines to meet the goals of the training/evaluation or research sessions on unexpected events. SMEs should be solicited in order to provide content and context for the development of event sets to be used in scenario flights. For example, we have already used SMEs to help create the specified list of thirteen scenarios that would be ideal for training pilots on unexpected events. In addition, we have integrated components of real events that pilots described from knowledge elicitation narratives developed in Task 3. By obtaining expert accounts of events, the scenarios maintain high content validity.

9.4 Additional Characteristics of Study Scenarios

The following are additional characteristics that should be considered in the design of the study scenarios. When designing scenarios, one should consider relevance and realism, pertinence to the task specific learning objectives, whether or not they enable variable measurement, where to find scenario topics, and how many scenarios to use for a training session. In addition, this section explains the importance of conducting a proper task analysis as well as a list of candidate scenarios that meet all of these characteristics.

9.4.1 Relevant and Realistic

Scenarios or events that are embedded in a research study session must be (a) relevant, and (b) realistic. To be relevant, the events and the context presented in the scenario need to be applicable to the specific equipment and type of operation identified. For air carrier training/evaluation or research, in house data can be reviewed for trends and frequency of specific situations or combinations of situations found to be unexpected. These must still be translated into training and evaluation scenarios. When constructing the script for each scenario flight, the context (maintenance status, weather, deferred items, crew composition, etc.) should be as realistic as the training tools allow.

9.4.2 Pertinent to the Task-Specific Pilot Learning Objectives

Pilot learning objectives should be referenced in the creation of each scenario flight and embedded event. Scenarios must be sensitive to the stated learning or testing objective. An example of the learning objective for the task of understanding atmospheric conditions associated with mountain wave would be to understand the conditions leading to the existence of standing mountain waves and common indications and the effects of exposure to the atmospheric climb, descent, and roll rates which can be encountered in mountain wave and rotor conditions.

9.4.3 Enable Variable Elicitation and Measurement

For scenarios to be useful in training/evaluation, and particular research sessions, they need to possess elements that permit measurement of specific variables associated with the learning objectives. That is, the scenarios must contain specific events that will trigger behavioral indicators of the variables of interest. From Task 3 (Pilot Knowledge Elicitation Interviews), several variables of interest surfaced when analyzing the results of the narratives with the pilots (e.g., decision-making, cognitive flexibility, communication, and teamwork). The next step in the process would be to determine how one would elicit these variables in a scenario event.

There are many different applied human factors methods that can be used to determine how to elicit variables of interest in a scenario event. First, make a list of the variables of interest for the scenario. For example, we would list that we would like to elicit decision-making, cognitive flexibility, communication, and teamwork from an ideal scenario. Next, we would consult SMEs through structured interviews or focus groups to determine the knowledge, skills, and behaviors needed to perform each task during the scenario. This is also referred to as a task analysis. The task analysis will list all tasks within the chosen scenario and the corresponding knowledge, skills, and behaviors needed to perform each task successfully. At this point, the researchers determine if the variables of interest are included in the task analysis.

Some of the variables from the task analysis, like communication, will be easily recognizable (e.g., the crew communicates with ATC), but others may be more challenging to identify (i.e., cognitive flexibility). When this situation arises, there are different recommendations to effectively elicit a variable such as cognitive flexibility. One method is to determine how cognitive flexibility has been triggered in previous HITL experiments. Another method is to identify cases of cognitive flexibility being used in historical cases of aviation events. A third method is to return to your knowledge elicitation interviews and identify how cognitive flexibility was categorized from the pilot narratives. Once it is determined how you want to elicit cognitive flexibility, validate the approach with a group of SMEs.

After ensuring that the variables of interest can be elicited by the scenario, one must create a method of measuring the performance outcomes associated with those variables such as found in Table 4 (Variables of Successful Pilot Performance during Unexpected Events) and Appendix C (HITL Study Forms). This is done with a performance assessment. The assessment guides the record-keeping and scoring of variables associated with the pilot's performance during the unexpected event. It is essential that the performance assessment specifically relates to the knowledge, skills, and behaviors needed to perform successfully during the response to the events in the scenario. In addition, the performance assessment should be developed based on the task analysis and then assessed for content validity by multiple SMEs. Performance assessments are designed to measure the variables of interest in a scenario-based format. This begins in preflight and ends with the landing phase of flight. Assessing pilots through a scenario-based approach has shown to be more effective than a non-scenario-based approach since it allows pilots to sustain the mindset of their flight. It may also increase knowledge retention based on long-term memory (Cruit & Blickensderfer, 2017). We recommend that a scenario-based assessment be created specifically for the events embedded in a particular scenario.

9.5 Selecting Scenarios

When developing the experimental stimuli for the presently proposed human-in-the loop study, a first, major task is to generate a list of scenarios topics and then to develop the details for the specific scenarios. The first step is to identify situations that have been found to be or could in the future unexpected or surprising across a variety of variables. (e.g., industry, organization, aircraft fleet, etc.). Additionally, an in-depth literature review can provide the background for the situations and constructs of interest to aid in the development of scenarios (and see Hancock et al., 2021).

9.5.1 Subject Matter Expert (SME) Interviews

Initially, it is advisable to identify where training is specifically needed for air carrier pilots to better respond to unexpected events (e.g., surprise response, metacognition, etc.). Participant interviews should be structured to further elicit information regarding, (a) examples of events the SMEs find to be unexpected; (b) the current procedures regarding responding to such unexpected events; (c) how desired behaviors of the flightcrew could be compared with the current behaviors of the flightcrew; (d) where gaps in procedures or checklists may exist regarding unexpected or ill-defined events; and (e) what kind(s) of training might improve responses to unexpected events. Results from these interviews can inform what other training and response methods might be useful to aid in building a more resilient pilot (and see our own Task 3 report).

In addition, using SME expertise can also represent an avenue for the development and selection of scenarios applicable specifically to a training goal. Interviews can also be further conducted with SMEs both within and outside any specific organization. Cross reference institutional trends can be especially important for identification of elements to include in the scenarios.

9.6 How Many Scenarios?

We propose that multiple scenarios be developed to make a selection of events available depending on the purpose of the HITL study. The number of discrete scenarios used during the HITL will depend upon the type of training intervention(s) being tested. One consideration is whether some pilot participants have experience with a particular scenario while other participants do not. This asymmetric experience factor could readily skew results. In the best case, the scenario would be new to all participants. If not, the level of familiarization with the situation presented must be captured in the study debrief interview. This information must be accounted for in the analysis of the participants’ outcome performance.

9.7 Evaluate for Consensus and Rank Scenarios for Use in HITL

If a selection spectrum of scenarios is created, they can then be evaluated for the specific purpose desired. Here, they are able to be ranked on the basis of, for example, of (a) ease of deployment, (b) expected training value, and (c) applicability to the training objective. This exercise will serve to prioritize a hierarchy of scenarios so that development can proceed for those with the highest ranking.

9.8 Perform Task Analysis

Once a scenario is selected for testing, a task analysis should be performed to establish a baseline of pilot performance in responding to the unexpected event. The purpose of any task analysis is to understand the knowledge, skills, decisions, and behaviors that are needed for a pilot to best perform during candidate scenario(s). The results of the task analysis can be used to build the details of the scenario and evaluative performance measure for that scenario, as we have illustrated. The task analysis should be conducted by asking SMEs to walk through all the tasks needed to perform successfully during the chosen scenario. Once all the tasks for that scenario are listed, the SMEs should describe the knowledge, skills, decisions, and behaviors needed for the pilot to perform optimally for the given task. The result of the task analysis then will produce all the information indicated above. Researchers or training instructors can then access this information for further validation efforts or for selected training purposes. Refer to Section 10 for an example task analysis on a representative unexpected events scenario.

9.9 Candidate Scenarios

Appendix A provides a summary, detailed description, and historical examples of candidate scenarios. These scenarios allow for data collection and analysis on the constructs associated with successful performance in novel, ill-defined, or otherwise surprising situations (Hancock, 2021), and were chosen on the basis of the criteria identified in this section. These scenarios were also created through analysis of accident, incident, and event reporting, as well as our earlier and comprehensive literature reviews (Hancock et al., 2021). This process is further informed by our SME interview study.

10. Example Task Analysis of Example Scenario (Weight-on-Wheels Switch Failure)

The following section presents a represented scenario for illustrative purposes that can immediately be used to evaluate the efficacy of the training to be tested. Following the suggestions on selecting scenarios to study pilot performance in reacting to unexpected events, this particular event is ambiguous and can easily be misdiagnosed. Although a Weight-on-Wheels switch failure does not directly impose danger to the crew or aircraft, how pilots respond to the situation can provide ample opportunities to evaluate whether training methodologies are useful, and to what degree.

10.1 Technical Background for Example Scenario (Weight-on-Wheels Switch Failure)

Most aircraft use some form of Weight-on-Wheels (WoW) Sensor or Switch (also known as proximity switch, squat switch, or air/ground switch) that serves to activate when the aircraft is on the ground. These come in many different sizes shapes and technologies and can be in various locations in the aircraft most often in proximity to the landing gear assembly. One thing they all have in common is they complete circuitry which is required to do many additional things on the aircraft; that is, they are pre-potentiating of many other, often vital, systems. Multiple systems are affected by the WoW switch such as landing gear failure to retract, thrust reversers, nose wheel steering, pressurization system, engine and airframe anti/de-ice systems, stall warnings, and others depending on aircraft type and configuration. A faulty or incorrectly adjusted WoW switch/sensor may therefore cause vital systems not to function or problematically function intermittently.

10.2 Example Scenario Overview (Weight-on-Wheels Switch Failure)

Our specific scenario occurs with a departure from Las Vegas, NV (KLAS). This flight has an intended destination of Orlando, FL (KMCO). Current weather at KLAS is visual meteorological conditions, though thunderstorms are forecast for shortly after departure time. The flight proceeds nominally until just after lift-off, when the landing gear fails to retract (due to a failure of the Weight-on-Wheels switch). The actions of the crew are captured throughout all phases of the flight. It is expected that they will proceed to a holding pattern and/or return to the departure airport (KLAS). Section 11 provides a detailed description of this experimental protocol. The scenario script also indicates potential data collection points for the associated experimental session.

10.3 Task Analysis for Example Scenario (Weight-on-Wheels Switch Failure)

As discussed in Section 9.8, a task analysis is necessary to create the performance assessment. Here, we present an example task analysis for the Weight-on-Wheels switch failure (Table 8). Details of this representative scenario are found in Section 11 and expanded upon in Appendix B.

The task analysis for a scenario encompasses the entirety of the flight, not just a single event (for example, the WoW switch failure we present). The researchers interviewed SMEs within the context of the WoW scenario (i.e., Researchers asked SMEs, “what are the tasks needed to successfully perform a Weight-on-Wheels scenario and then what knowledge, skills, and attitudes are needed to perform each of those tasks successfully?”). Therefore, each task by phase of flight influences subsequent tasks and pilot decision making and is documented to design the entirety of the scenario. We have organized the task analysis and the experimental protocol script by phase of flight. Each phase of flight introduces factors that will later add to the operational complexity of the flight. For example, during the preflight, the preparation for the flight includes complexities such as terrain, unpredictable convective weather, fuel consumption based on the stage length (length of the flight) which would not allow for continued flight to the destination due to excessive fuel consumption if the gear were extended), a complex RNAV departure procedure and a NOTAM pertinent to the flight.

Table 8. Task Analysis of WoW Scenario

Phase of Flight	Task #	Task Description	Step #	Steps
Preflight	1	Check myself	1.1	"I'm safe" checklist; Have I been ill? Am I fatigued? Hydration? Have I had proper nutrition? Am I distracted?
			1.2	Think about external pressures, make sure I'm compartmentalized enough to stay focused on the task at hand
	2	Take note of any maintenance discrepancies; get feel for maintenance history	2.1	Review general maintenance status within compliance inspections
			2.2	Go back page by page in logs, see what kind of write-ups there have been
			2.3	Go back far enough to see if there are any noteworthy write-ups
			2.4	Get idea if the aircraft has been flying with deferred items, and if so, what those items are
			2.5	Brief crew on logbook review
	3	Check weather. Pay attention to discrepancies between forecast and radar	3.1	Look at departure and destination weather so when I look at the bigger picture, I can see why the weather at those locales is that way.
			3.2	Go through weather briefing
	4	Be aware of terrain in area knowing that I'm going to be returning to the departure airport or takeoff alternate depending on weather and terrain clearance I can maintain if I cannot retract my landing gear	4.1	General awareness using charting; see what terrain is like around the airport, especially if not familiar with the area
			4.2	Look at altitudes required for departure procedure; see if any of them have terrain constraints associated with them
	5	Crew briefing to discuss pre-flight maintenance status of aircraft; departure procedure, terrain, weather	5.1	Open door for communication; let them know their opinion is valued, considered, and equally important
			5.2	Letting the crew tell me what they think before I say what I think. "What do you think about this weather, departure procedure, and terrain?" Open door for communication; eliciting their viewpoint
			5.3	Come to a shared mental model on how we're going to proceed
	6	Note fuel state to determine if I would be below my maximum landing weight at takeoff	6.1	Being aware that takeoff weight is under maximum
Taxi	7	*All tasks normal/typical		
Takeoff/Climb	8		8.1	Verify gear is down
			8.2	Be mindful of gear speed

Phase of Flight	Task #	Task Description	Step #	Steps
		Ensure positive aircraft control and follow designated flight path for the procedure	8.3	Verify we can make climb gradient for departure
			8.4	Verify to stay on course
	9	Communicate that the landing gear appears to not be retracted; Indicate what appears to be the problem	9.1	Ask for confirmation that the gear is not up
			9.2	Communicate with other crewmembers
			9.3	Verify ATC is advised of landing gear anomaly
	10	Make sure we have a holding pattern and stay in the vicinity of the airport as we figure out what is going on (unsure of what the failure is, so we want to stay close)	10.1	Complete the applicable checklist(s)
			10.2	Use procedure that would put the landing gear back down and retract flaps
			10.3	Pull appropriate circuit breaker
			10.4	Land at nearest suitable airport
Descent	11	*All tasks normal/typical		
Approach and Landing	12	Request to land at nearest airport	12.1	It would all be normal. Anything abnormal would be seen from checklist
			12.2	Manual speed brakes on landing
	13	Write up discrepancy and have maintenance take care of it		

11. Example Unexpected Events Human-in-the-Loop Experimental Protocol with Performance Assessment for Weight-on-Wheels Switch Failure

This section provides an overview of the experimental protocol for a representative flight using a Weight-on-Wheels Switch Failure event in the scenario. Included in the protocol are boxed statements that indicate the experimental purpose or background for the study and/or scenario as well as the Performance Measures to be used for that phase of flight.

This is an example of the format of boxed statements that indicate the experimental purpose or background for the study and/or scenario. These are included in the protocol below.

The following is an outline of how the scenario could be used to drive the observations and data to be collected and address elements that can be trained in the future.

11.1 Scenario Overview

This scenario presents the pilots with an ill-defined mechanical event where multiple failures and ambiguous alerts could be presented. Furthermore, it encompasses a weather factor (convective activity) and added complexity due to the RNP departure procedure (with a NOTAM). The flight is planned from KLAS (Las Vegas Reid International Airport) to KDFW (Dallas Ft. Worth International) at FL 350. The clearance is the NIITZ 1 RNAV Departure as filed. The current and forecast weather is:
All operations are normal until just after lift-off, when the landing gear fails to retract. Annunciations associated with a failed WoW (or air/ground) switch are displayed (per the type of aircraft used). Note, that even a fairly basic aircraft type training device could accomplish this scenario. The performance criteria will be generated for each element to be observed and rated in much the same fashion as Cruit and Blickensderfer (2017) and Landman (2018).

11.2 Technical Background

Most aircraft utilize some type of WoW Sensor or Switch (also known as proximity switch, squat switch, or air/ground switch) that activates when the aircraft is on the ground. Example photographs of the WoW switch mechanism are shown in Figure 2. They come in many different sizes shapes and technologies and can be in various positions in the aircraft and landing gear. The one thing they all have in common is they complete the circuitry required to do many other things on the aircraft. Multiple systems can be affected by the WoW switch such as landing gear fails to retract, thrust reversers, nose wheel steering, pressurization system, engine and airframe anti/de-ice systems, stall warning, and others depending on aircraft type. A faulty or incorrectly adjusted switch/sensor may cause vital systems to not function or function intermittently.

There many reasons a landing gear may fail to retract which makes the trouble-shooting decision tree have many branches. Some of the more common reasons a landing gear may fail to retract and might need to be considered with this scenario are:

- Improper rigging
- Improper repairs or maintenance
- Parts worn beyond their allowable service limits
- Improper installation of parts
- Improperly secured parts
- Use of non-standard or unapproved parts
- Failure or fatigue of parts
- Rupture of hydraulic lines.
- Failure of electrical wire connections, relays, contactors, and/or actuators
- Malfunctions of warning systems
- Inoperative limit and safety switches,
- Uplocks failed to release
- Down locks failed to engage.
- Wheels jammed or hung up in wheel wells.
- Lack of lubrication
- Lack of hydraulic fluid
- Retraction of landing with tow bar still attached

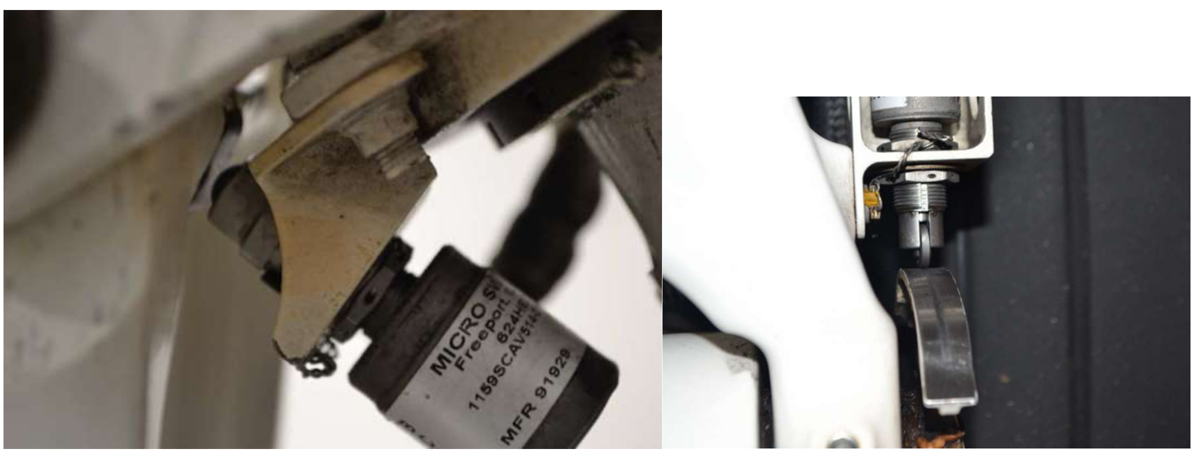

Figure 2. Example Weight-on-Wheels Switch Photos

11.3 Procedure Overview

The study will be conducted using a training device appropriate to the scenario of choice. As an example, we would choose a simulator for an aircraft type in use by the study population. The study "story" presented to the participants is that we are interested in observing airline pilots flying in a new airspace design (such as the new metroplex in Central Florida).

This HITL study is to evaluate the impact of training elements designed to improve pilot performance when faced with an unexpected, novel, ill-defined, or otherwise confusing event.
The participants in the study will receive training element "A", training element "B," both, or neither.

During the preflight process, participants will review a complete flight dispatch with route of flight, weather, NOTAMS, etc. and an aircraft maintenance logbook providing preparatory information for the flight. The flight is planned from KLAS (Las Vegas Reid International Airport) to KDFW (Dallas-Ft. Worth International) at FL 350. The clearance is the NIITZ 1 RNAV Departure as filed. The flight is planned for 2 hours and 20 minutes. The current and forecast weather at KLAS is:

KLAS 260356Z 00000KT 10SM SCT020 BKN065 BKN100 26/21 A3001 RMK AO2 SLP136 VCSH T02610211 52012
KLAS 262320Z 2618/2718 VRB05KT P6SM FEW040 BKN080 BKN150
FM270400 15012KT P6SM VCTS SCT080CB BKN120
FM270800 19008KT P6SM SCT100
FM272000 19007KT P6SM FEW100 SCT120
FM272300 VRB05KT P6SM FEW120
The forecast weather at KDFW is:
KDFW 262325Z 1500/1606 20007KT P6SM SCT040
FM270700 20007KT P6SM SCT012 BKN035
TEMPO 1510/1514 BKN010
FM27 1500 30015KT P6SM BKN025
FM27 1700 33018G28KT P6SM SCT080
FM27 2300 35013KT P6SM SKC

11.4 Flight Profile

The experimental flight profile is expected to last approximately 30-45 minutes depending on the actions of the pilots. Departure from KLAS is planned for 0400Z (night) for a planned revenue flight to KDFW. The flight profile would be standardized by utilizing scenarios stored in the computer database which include preprogrammed weather, winds, and the unexpected event malfunction.

11.5 Flight Profile Scenario Script

The flight profiles would be scripted to ensure consistency in the experimental protocol. In the event the script does not fit the current situation, standard ATC communications will be used until able to return to a portion of the standardized script. The study administrator will monitor the communication radio frequency selections for ATIS frequencies at which time the appropriate simulated recorded information will be delivered.

11.6 Study Introduction Transcript

Thank you for your participation in our study. As you know, we are interested in pilot's opinions on some of the new airspace redesign features in the National Airspace System. This will be a normal revenue flight. You will be provided adequate time to prepare for the flight. We will ask you to fill out a brief demographic questionnaire and complete a survey online prior to receiving your dispatch paperwork electronically. Do you have any questions?

The purpose of the contrived story regarding the purpose of the flight is to help mitigate the expectation of a malfunction or other event which would normally be expected in a simulator training, evaluation, or experimental session.

11.7 Simulator Session Overview by Phase of Flight

11.7.1 Preflight

- Preflight dispatch release and weather are reviewed
- Maintenance log does not indicate any recent maintenance
- Current Radar at KLAS and NOTAM for the NIITZ departure are shown in Figure 3

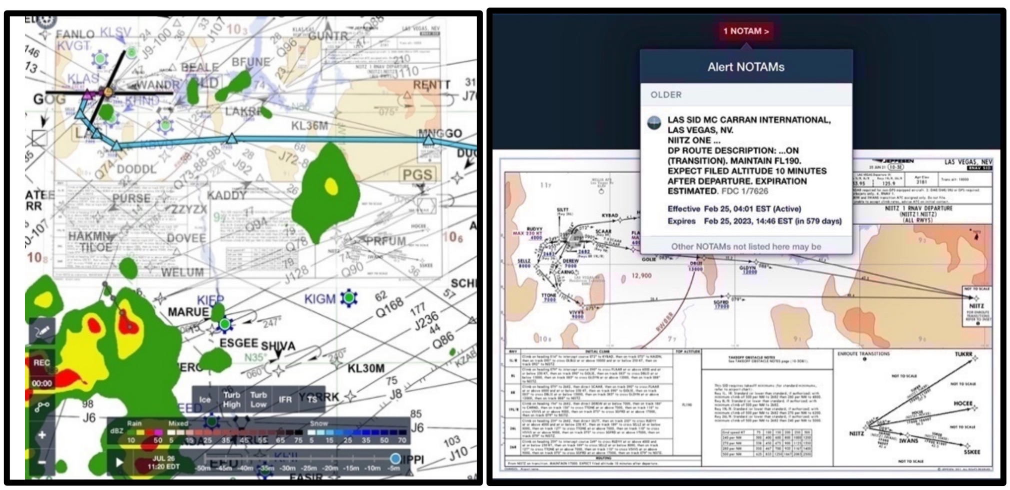

Figure 3. The Current Radar Depiction and NOTAMs Associated with the NIITZ 1 RNAV Departure

This preflight action has elements contrived to increase the complexity of the scenario. There is (a) high terrain both west and east of the departure airport, (b) convective weather moving toward the airport, (c) a complex RNAV departure procedure and a relevant NOTAM for the departure procedure. As the crew proceeds through the flight, we can identify specific features of their performance that we expect to show a difference based on the treatment (A, B, A&B, or none) that they received.

The performance assessment follows the scenario script by phase of flight and is used to record (real-time and post-flight review) the behavior, skills, and attitudes observed during the. Tables 9 through 15 display the segment of the performance assessment associated each phase of flight.

Table 9. Preflight Performance Assessment

<table>
<tr><th>Behavior, Skills, Attitudes</th><th colspan="4">Yes (Check if Yes)</th><th>Corresponding DV(s) of Interest</th></tr>
<tr><td>Does the crew make a decision about what to do with the terrain?</td><td colspan="4"></td><td>General Decision Making, Domain Expertise</td></tr>
<tr><td>Does the crew reference the chart to avoid terrain?</td><td colspan="4"></td><td>Domain Expertise</td></tr>
<tr><td>Was someone engaged at all times?</td><td colspan="4"></td><td>Domain Expertise, Task Management</td></tr>
<tr><td>Has the crew flown this path previously (consider expertise)?</td><td colspan="4"></td><td>Domain Expertise</td></tr>
<tr><td>Does the crew talk about the escape route (i.e., turn to head northeast)?</td><td colspan="4"></td><td>Teamwork, Communication</td></tr>
<tr><td>NOTAM:</td><td colspan="4"></td><td rowspan="3">General Decision-Making, Teamwork, Communication</td></tr>
<tr><td>Did the crew brief the NOTAM?</td><td colspan="4"></td></tr>
<tr><td>Did the crew make a change in the FMS if needed?</td><td colspan="4"></td></tr>
<tr><td>Convective Weather:</td><td>Yes</td><td>Low</td><td>Med</td><td>High</td><td rowspan="5">General Decision-Making, Task Management, Teamwork, Communication</td></tr>
<tr><td>Is there awareness about the weather?</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Is there management of the threat?</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Does the crew have a shared mental model about the threat?</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Is there a plan for mitigation?</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Communication (Notice how the crew establishes open communication):</td><td colspan="4"></td><td rowspan="5">Teamwork, Communication</td></tr>
<tr><td>Does everyone appear willing to speak up?</td><td colspan="4"></td></tr>
<tr><td>Is the Captain open and approachable?</td><td colspan="4"></td></tr>
<tr><td>Does the Captain ask questions?</td><td colspan="4"></td></tr>
<tr><td>Does the Captain introduce themselves?</td><td colspan="4"></td></tr>
<tr><td colspan="6">Notes:</td></tr>
</table>

11.7.2 Engine Start, Pushback, and Taxi

- Engine start, pushback, and taxi are nominal
- The flight is cleared to taxi to RWY 26R for the departure (see Figure 4)

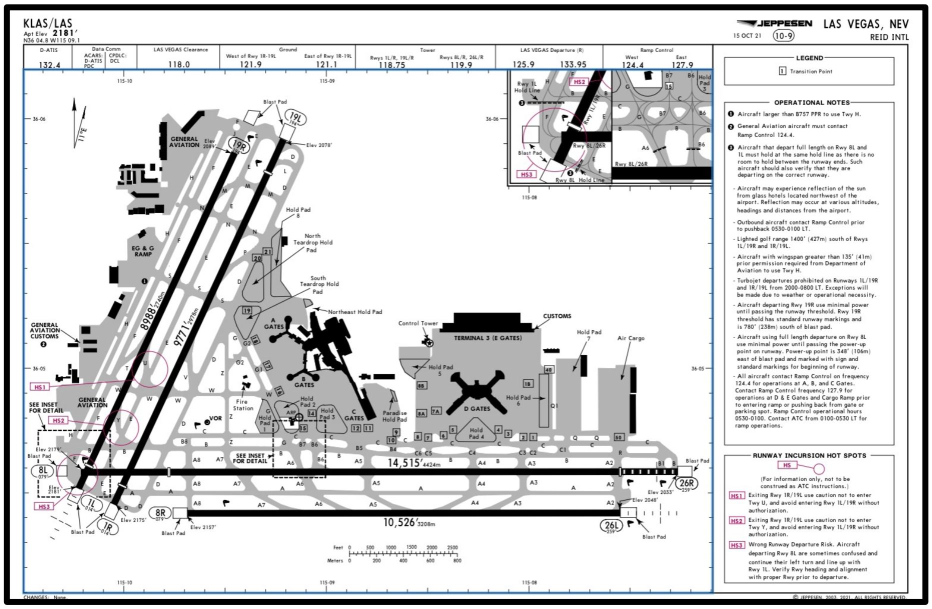

Figure 4. Taxi Chart for KLAS

Since the winds are currently calm (as reported in the ATIS), RWY 26R is in use. This will become a factor (bias) in the decision-making if the flight returns to KLAS for landing as the winds will have increased and RWY 19L will then be in use.

Table 10. Taxi Performance Assessment

Behavior, Skills, Attitudes	Yes (Check if Yes)	Corresponding DV(s) of Interest
Is there appropriate prioritization given to flight and control?		Domain Expertise, General Decision Making
Pitch		
Bank		
Power		
Speed		
Is the crew communicating back and forth?		Teamwork, Communication
Is the crew looking for threats along the way?		Cognitive Flexibility, Domain Expertise
Is the crew carefully managing time?		Time Management
Notes:		

11.7.3 Takeoff

- The aircraft is cleared for takeoff on RWY 26R
- The takeoff is nominal until after liftoff and the gear fails to retract (or the landing gear handle does not move)

At this point it is expected that several the performance measures can be accomplished. Specifically, the identification of surprise can be evaluated. In addition, the initial response to the malfunction will allow observation of whether the first two steps (calm down, observe) of the metaprocedure is utilized.

Table 11. Takeoff Performance Assessment

Behavior, Skills, Attitudes	Yes (Check if Yes)	Corresponding DV(s) of Interest
Is there appropriate prioritization given to flight and control?		Domain Expertise, General Decision Making
Pitch		
Bank		
Power		
Speed		
Did the crew show recognition to the malfunction?		Domain Expertise, Recognition
How long (in seconds) did it take the crew to recognize the malfunction. This should be measured by time of malfunction to time of recognition.	**Time:**	Domain Expertise, Recognition
Does the crew focus on the flight path as primary task?		Domain Expertise, Task Management
Does the crew communicate problem with each other?		Communication, Teamwork
Does the crew communicate problem to ATC?		Communication
Does the crew accurately correct altitude?		Domain Expertise
Does the crew focus on flying the plane instead of troubleshooting the gear after takeoff?		Cognitive Flexibility, General Decision-Making
Does the crew have a shared mental model (evident by discussion and communication)?		Teamwork, Communication
Is the crew communicating in a composed manner?		Self-efficacy, Communication, Assesses C in *Calm*
Does the crew assess their fuel?		Domain Expertise, General Decision-Making
Does crew identify appropriate checklists?		Domain Expertise
If emergency was declared, what was the reason why they declared an emergency?		General Decision Making
Notes:		

11.7.4 Climb

- The aircraft (is expected to) continue the departure procedure (see Figure 5)

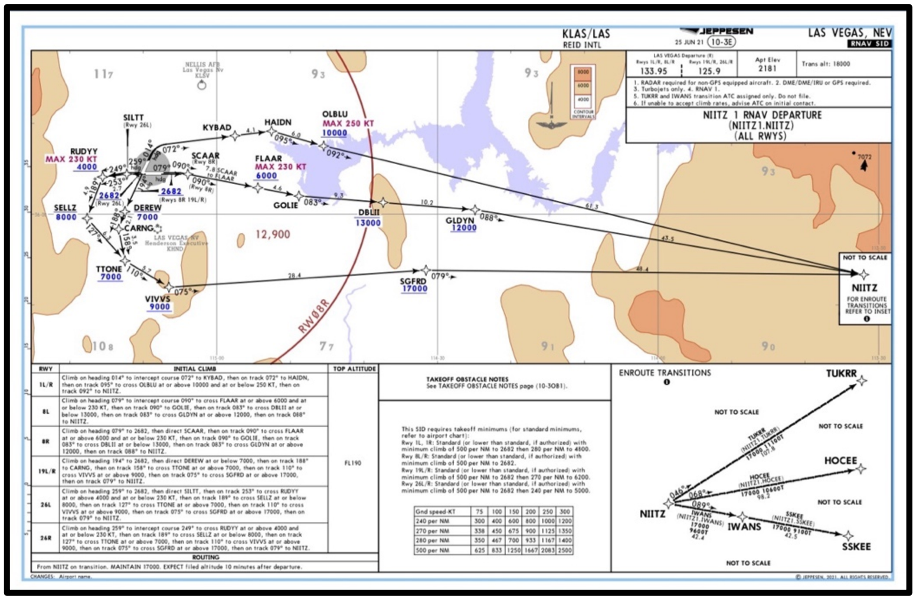

Figure 5. The NIITZ 1 RNAV Departure Procedure Chart

At this point it is expected that the flight will continue. Performance measures at this point include the "outline" step of the metaprocedure as well as the decision making, communication, and teamwork variables. Inter and intra-flight deck communications will be analyzed for indications of stress and self-efficacy. Throughout the flight, aircraft flight data recorder data and simulator track data will be measured as optimum, acceptable, or unacceptable. This will give an indication of attention to flight path management.

Table 12. Climb Performance Assessment

Behavior, Skills, Attitudes	Yes (Check if Yes)	Corresponding DV(s) of Interest
Is there appropriate prioritization given to flight and control?		Domain Expertise, General Decision Making
Pitch		
Bank		
Power		
Speed		
Does the crew focus on the flight path as primary task?		Domain Expertise, Task Management
Does the crew communicate what is going on with each other and take steps to correct any challenges?		Communication, Teamwork, *Outline*
Notes:		

11.7.5 Departure and Cruise (Holding)

- The aircraft continues per the decisions of the pilots.

Performance measures at this point include the "lead" step of the metaprocedure as well as the decision making, metacognitive, communication, and teamwork variables. Inter and intra-flight deck communications will be analyzed for indications of stress and self-efficacy. Throughout the flight, aircraft flight data recorder data and simulator track data will be measured as optimum, acceptable, or unacceptable. This will give an indication of attention to flight path management at this point. How the situation is handled will include measures of identification of the anomaly, use of checklists (see Figure 6), and how the situation proceeds. The WoW switch failure event in this scenario enables the evaluation of several skills. First, the identification of the correct checklist must be made given the indications, cues, and clues available. The sequence of events to handle the checklist must be carefully followed. The example checklist shown in Figure 6 shows the complexity of the checklist.

▼Landing Gear Lever Will Not Move Up After Takeoff continued▼

3 Choose one:

◆Intermittent cabin altitude/configuration warning horn **stays silent** and the TAKEOFF CONFIG lights do **not illuminate** after the flaps are fully retracted and the thrust levers are advanced beyond the vertical position:

Note: This indicates a failure of the landing gear lever lock solenoid.

►►**Go to step 4**

◆Intermittent cabin altitude/configuration warning horn **sounds** or the TAKEOFF CONFIG lights **illuminate** after the flaps are fully retracted and the thrust levers are advanced beyond the vertical position:

Note: This indicates a failure of the air/ground system.

Do **not** retract the gear.

►►**Go to step 7**

235K maximum.

4 LANDING GEAR lever LOCK OVRD switch . . . . Push and hold

5 LANDING GEAR lever . . . . UP

6 Continue normal operation.

■ ■ ■ ■

▼ Continued on next page ▼

▼Landing Gear Lever Will Not Move Up After Takeoff continued▼

7 Plan to land at the nearest suitable airport.

8 Do **not** arm the autobrake for landing. Use manual braking.

9 **Checklist Complete Except Deferred Items**

Deferred Items

Descent Checklist

Pressurization . . . . LAND ALT ____

Recall . . . . Checked

Autobrake **. . . . . .OFF**

Landing data . . . . VREF ____, Minimums ____

Approach briefing . . . . Completed

Approach Checklist

Altimeters . . . . ____

Gear Down Verification

LANDING GEAR lever. . . . . Verify DN

Landing Checklist

[Without automatic ignition]

ENGINE START switches . . . . CONT

Speedbrake . . . . ARMED

Landing gear **. . . . . . Down (previously verified)**

Flaps . . . . ____, Green light

■ ■ ■ ■

Figure 6. An Example Checklist for a WoW Switch Failure

Table 13. Departure and Cruise Performance Assessment

Behavior, Skills, Attitudes	Yes (Check if Yes)	Corresponding DV(s) of Interest
Is there appropriate prioritization given to flight and control?		Domain Expertise, General Decision Making
Pitch		
Bank		
Power		
Speed		
Does the crew focus on the flight path as primary task?		Domain Expertise, Task Management
Does the crew communicate problem with each other?		Communication, Teamwork
Does the crew accurately correct altitude?		Domain Expertise
Does the crew have a shared mental model (evident by discussion and communication)?		Teamwork, Communication
Is the crew communicating in a composed manner?		Self-efficacy, Communication
Does the crew assess their fuel?		Domain Expertise, General Decision-Making
Does crew identify appropriate checklists (example checklist above)?		Domain Expertise
Notes:		

11.7.6 Diversion and Descent

- The flight proceeds per the decisions of the pilots

As the flight continues, the pilots will be observed handling the situation and deciding on a plan of action. Choices include returning to the departure airport (KLAS), going to another diversion airport (there is no takeoff alternate listed in the dispatch), or continuing to the destination (KDFW). Decision triggers include (a) not enough fuel on board to continue to destination with the gear down, (b) the oncoming convective activity approaching KLAS, (c) mountainous terrain to the east and west of the departure path, and (d) new ATIS information at KLAS. It is expected that the pilots will return to KLAS. Performance measures at this point include effects of variable training as well as the decision making, metacognitive, communication, and teamwork variables. Inter and intra-flight deck communications will be analyzed for indications of stress and self-efficacy. Throughout the flight, aircraft flight data recorder data and simulator track data will be measured as optimum, acceptable, or unacceptable. This will give an indication of attention to flight path management. The diversion and descent decision-making could be influenced by the convective weather still in the area (see Figure 6).

Table 14. Diversion and Descent Performance Assessment

Behavior, Skills, Attitudes	Yes (Check if Yes)	Corresponding DV(s) of Interest
Is there appropriate prioritization given to flight and control?		Domain Expertise, General Decision Making
Pitch		
Bank		
Power		
Speed		
Does the crew communicate they have a plan?		Teamwork, Communication
Describe how the crew manages the plan:		
How did the crew manage any exceptional activities?		
Notes:		

11.7.7 Approach

- The flight proceeds per the decisions of the pilots
- KLAS ATIS is now reporting winds 23 at 10 gusts 20 and landing RWY 19L

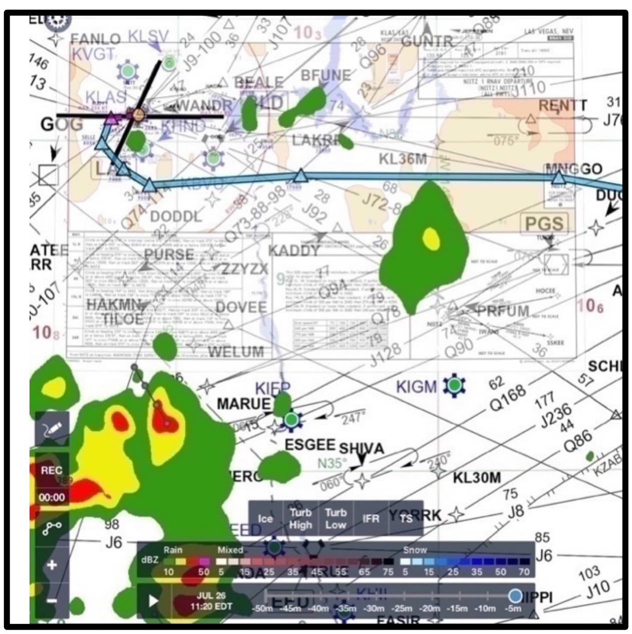

Figure 7. The Radar at KLAS Shown on High-Altitude Enroute Chart with Departure Procedure Overlay

At this point it is expected that the pilots are preparing to return to KLAS. The changing and impending weather introduces a decision point. Although RWY 19L is being advertised on the ATIS, RWY 26R would be an option (straight in approach, longer runway, closer runway, equal crosswind component, etc.) Performance measures at this point include effects of variable training as well as the decision making, metacognitive, communication, and teamwork variables. Inter and intra-flight deck communications will be analyzed for indications of stress and self-efficacy. Throughout the flight, aircraft flight data recorder data and simulator track data will be measured as optimum, acceptable, or unacceptable. This will give an indication of attention to flight path management.

11.7.8 Landing and Taxi

- The landing is nominal

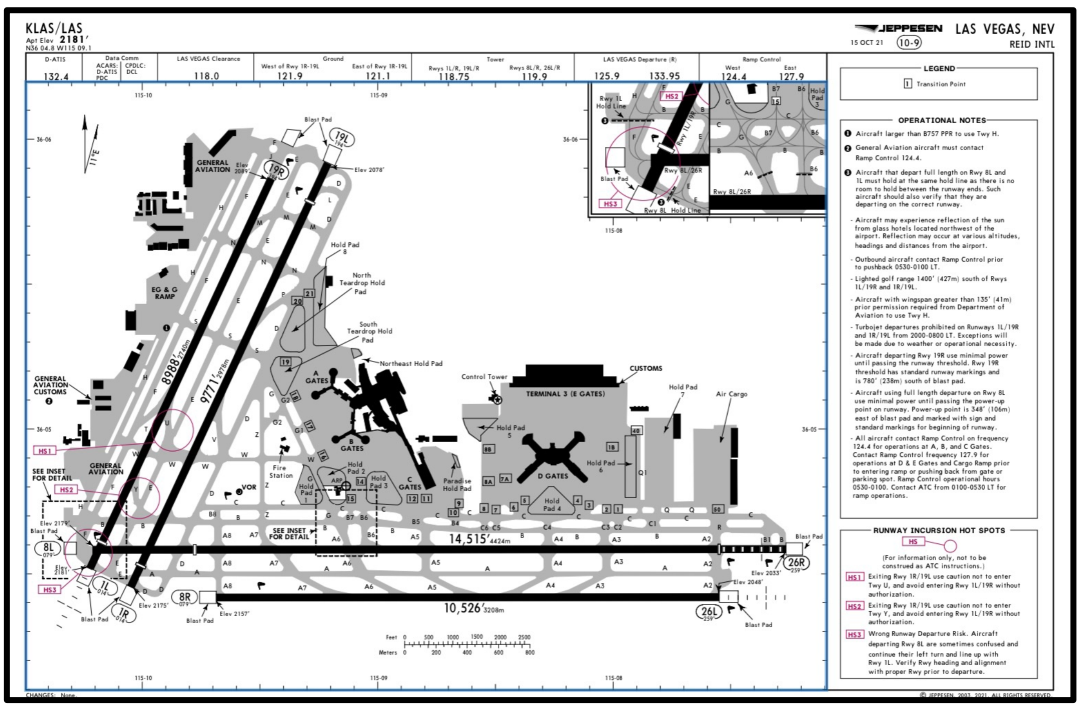

Figure 8. The Taxi Chart for KLAS

After landing, the decision point will include whether the pilots choose to taxi to the gate or stop and have maintenance check the gear and install gear pins. This will be a measurement point as to whether the crew had correctly identified the malfunction and/or the level of caution being exhibited.

Table 15. Approach and Landing Performance Assessment

Behavior, Skills, Attitudes	Yes (Check if Yes)	Corresponding DV of Interest
Is there appropriate prioritization given to flight and control?		Domain Expertise, General Decision Making
Pitch		
Bank		
Power		
Speed		
Completes end-of-flight tasks		Domain Expertise, Task Management
Localizer deviation		
Glideslope deviation		
Airspeed fluctuation		
Notes:		

12. Data Collection and Analysis

Data is to be collected via surveys, observational, and simulator track data. Examples of forms that can be used for data collection are found in Appendix C.

12.1 Statistical Analysis Methods

Collected data can be analyzed using standard statistical analysis packages, such as SPSS Version 28 (IBM Corporation, 2021). Descriptive statistics can be computed for demographic data and reduced to the first a second distributional moments for inferential interpretation. Performance scores, and domain expertise are also amenable to inferential statistical analyses. Assumptions of linearity, multivariate normality, outliers identified, homogeneity of regression slopes, and homogeneity of variances and covariances should be assured prior to further analysis (see Dattalo, 2013). A two-way Multivariate Analysis of Covariance (MANCOVA) with domain expertise as a covariate could then be computed to test the predicted main effects and interactions of the four outcome variables (i.e., decision-making performance, recognition of the event performance, response to the event performance, and teamwork/communication performance).

13. Additional Insights Gained for Conducting the HITL

Throughout the current project, our team of researchers and SMEs possessed the advantage of being immersed in the culture of resilience, with an awareness and familiarity of unexpected events, and air carrier operations across a number of years. This experience is evident in our assessment of the resilience and aviation safety literature, our data from the knowledge elicitation interviews with 50 air carrier pilots, choosing and writing a candidate scenario centered around an unexpected event, completing a task analysis on the identified scenario, and then developing an assessment to measure and evaluate pilot performance to validate proposed hypotheses. The collective work is directed to distilling what both the literature reports and what current operational air carrier pilots say about what behaviors and skills are needed to ensure flightcrew have the best outcome during unexpected events.

It is important to further note that all reported information is intended for researchers and experimentalists to test our hypotheses, constructs of interest, and performance profile expectations. The present report offers guidance which is based on the research and data that has been collected to conduct an experiment with air carrier pilots to test the propositions to distinguish paths of recovery during an unexpected event. We do not advocate such predictions be implemented in the absence of the testing identified.

One challenge that is identified in the literature on expertise is to transfer our knowledge and insights to those who follow so that the experimental design is executed as precisely as possible. Further, that results are interpreted appropriately, and the implications for training are considered so that ultimately, pilots receive superior training. Below, we have generated a list of recommendations that we believe useful in conducting the experimental protocol. The list of recommendations applies to both researchers and training instructors for creating the most effective and controlled experimental design and then recommendations for interpreting outcome results.

13.1 Hypotheses

We here identify potential data patterns for each of the offered hypotheses and how garnered results can be used to inform relevant and effective training regimens. Given there are multiple performance measures, organizational needs, and requirements that have to be considered when interpreting results determining how to incorporate the experimental results into a training program is a contingent and summary process. If results suggest overwhelming support (i.e., a consistent pattern of data for all performance measures) for one training intervention, the decision as to which intervention to implement may be clear. However, it is possible that performance profiles vary across measures for each training treatment intervention. For instance, what can be concluded when Treatment A (metaprocedure) leads to improved pilot responding and recognition, while Treatment B (variable scenario training) leads to improved decision-making and recognition? In these cases, we recommend considering additional decision-making criteria such as compliance with SMS and risk-assessment analysis. This issue of measure diversion Is one which pervades behavioral testing and evaluation (see Hancock & Matthews, 2021).

13.1.1 Metaprocedure (Treatment A)

H_1: Controlling for domain expertise, participants receiving the metaprocedure training intervention will perform better on the unexpected event scenario compared to participants in the control group.

If this proposition is supported, there would be a significant main effect of metaprocedure training evident on one or more of the performance measures (decision-making, recognition, responding, teamwork, or communication). Participants in the metaprocedure group would exhibit higher performance scores compared to their control group peers.

If unsupported, the outcome is either (a) no significant main effect of metaprocedure training on one or more of the performance conditions; or (b) a significant main effect, but in the opposing direction, i.e., a contradictory pattern. Performance scores in this eventuality, would be lower in the metaprocedure group compared to the control. If this occurred, it will be appropriate to consider the following:

- Did the metaprocedure create more distraction? It is important to present mnemonics in a way that the flightcrew does not lose relevant situation awareness, which enables them to land the plane safely.

- An additional issue concerns who evaluates performance. It is best to have both an experienced training instructor and a researcher evaluate performance. If an untrained evaluator assesses performance, they could omit important non-verbal behavior that the flightcrew demonstrate for example, when they develop a shared mental model.

Implications for Training. If it is determined that the metaprocedure does exert a significant main effect, then recommendations for training that mnemonic need to be advanced, such as practicality, time management, and costs associated with implementing the mnemonic. Alternatively, if there is no significant main effect on performance, researchers and training instructors should assess how well the experiment was conducted and whether sample of pilots used in the study in order to assess its value.

13.1.2 Variable Training (Treatment B)

H_2: Controlling for domain expertise, participants receiving the variable training intervention will perform better on the unexpected event scenario compared to participants in the control group.

If supported, a significant main effect of variable training on one or more of the performance measures (decision-making, recognition, responding, teamwork, or communication) will be evident. Participants in the variable training group would exhibit higher performance scores compared to controls.

If unsupported, results will indicate (a) no significant main effect of variable training on one or more of the performance conditions exist; or (b) a significant main effect, but in the opposite direction has accrued. Performance scores may be lower in the variable training group compared to the control.
Implications for Training

If a significant main effect is found on performance through the addition of variable scenario training, researchers and training instructors can consider implementing such variable scenario training into curricula. If time management, or other logistical issues arise during the consideration of implementing variable scenario training, one might think about the possibility of no jeopardy or computer-based training that allowed pilots to practice on their own with lower assessment pressures. This type of training has been recommended by pilots during knowledge elicitation interviews as reported in Task 3. In addition, various forms of this latter training are already being implemented in military operations (US

Navy) as a way to enhance and modernize training so that the same form of intensive, personalized training is also being employed by the U.S. Air Force.

13.1.3 Both Training Interventions (Treatment A and Treatment B)

H_{3A}: Controlling for domain expertise, participants receiving both training interventions will perform better on the unexpected event scenario compared to participants receiving only one of the training interventions.

If supported, we would witness a significant interaction between the metaprocedure intervention and the variable training intervention. Participants receiving both interventions would show higher performance scores (on one or more of the performance metrics) than participants in (a) the metaprocedure-only group; and (b) the variable training-only group, as well as the baseline control group.

If unsupported, there would be either (a) no significant interaction between the metaprocedure and variable training interventions; or (b) there would be a significant interaction, but in the contra-indicatory directions. That is, performance scores would be lower for the metaprocedure, and for the variable training group compared to either (a) the metaprocedure-only group; or (b) the variable training-only group or the control.

H_{3B}: After controlling for domain expertise, participants receiving both training interventions will perform better on the unexpected event scenario compared to participants in the control group.

If supported, results would show a significant interaction between the metaprocedure intervention and the variable training intervention. Participants receiving both interventions would show higher performance scores (on one or more of the performance metrics) than participants in the control group. This finding would suggest that the combination of both interventions (variable training and the metaprocedure combined) leads to improved performance above and beyond that of just one intervention.

If unsupported, there would be either (a) no significant interaction between the metaprocedure and variable training interventions; or (b) there will be a significant interaction, but in the opposite direction. That is, performance scores will be lower for the metaprocedure, and variable training group compared to the control.

13.2 Performance Evaluation

Performance evaluations are designed to measure air carrier pilot performance during a specific event (e.g., Weight-on-Wheels Switch Failure scenario). This panoply of evaluations was developed after performing a detailed task analysis and then validating both that task analysis and performance evaluation for content validity via SMEs. The constructs in the performance evaluation are intended to measure optimal behavior, skills, and attitudes during an unexpected event. Although the content of the performance evaluation can generalize to many different, unexpected aviation event scenarios, we recommend that a task analysis be performed on each event and then be validated for content. Furthermore, we recommend that multiple trained (a minimum of (*n*=3) three) raters assess such pilot performance using the performance evaluation techniques recommended and then measures of inter-rater reliability should be calculated. Raters should include both trained researchers and training instructors who hold a high level of expertise in the domain.

14. Summary and Conclusions

The present document provides a primer as to how to conduct human-in-the-loop (HITL) research in which pilots' responses to unexpected events can be explored for potential mitigations through training. Although candidate training interventions are presented, the focus is on recommendations for dependent measures and event scenarios.

This report does not include each and every possible variation in the manipulation of independent variables and the selection and interpretation of dependent variable outcomes. However, it does provide a step-by-step sequence through which a specific, and problematic flight scenario can be enacted and explored. Outcome results will be important to current and prospective aviation challenges, most especially as great degrees of automation and autonomy are injected into everyday operations in all forms of aviation and space activities.

The results of this effort conclude the following findings and recommendations:

- Identification of key independent (i.e., variable training, mnemonic applications) and dependent variables (i.e., decision-making ability, self-efficacy, metacognitive awareness, cognitive flexibility, critical thinking, recognition, responding, and teamwork/communication) assessing pilot responses during unexpected events. In addition to directly assessing pilot task performance for a given event response, these dependent variables provide indirect measures of pilot responses or thinking that are correlated with "good" (e.g., resilient) responses.

- Recommendations from the literature and pilot elicitation interviews on types of training interventions (i.e., variable training, mnemonic applications) and measurable skills and behaviors that pilots may exhibit during unexpected events (e.g., crew procedures found useful when faced with unexpected or surprising events were maintaining flight deck discipline, employing active and open communication, and reaching consensus among the flightcrew).

- Recommendations for scenario selection criteria. For the purpose of this research, which is for filling in the gaps when procedures do not exist or are not completely applicable, it is important to select the appropriate unexpected event scenario for HITLs in order to evaluate training interventions. Furthermore, we have identified several candidate scenario examples and provided a detailed scenario description and application (i.e., Weight-on-Wheels Switch Failure scenario).

- An overall recommended framework for conducting a HITL that also incorporates best practices for aviation HITL scenarios and simulations.

- A performance assessment for training instructors and future researchers for measuring pilots' behaviors during unexpected events. Although the example performance assessment offered in this document is specific to the WoW Switch Failure scenario, it can easily be adapted to any unexpected event scenario after performing a tasks analysis on the chosen scenario prior to developing the performance assessment.

The insights gained from this document are accrued from the previous tasks associated from this effort (see Section 1). We recommend referencing those documents for a comprehensive understanding of the material presented in the current document. Furthermore, the appendices of this document provide details that are essential for running the type of HITL we recommend for testing training interventions during unexpected aviation events (e.g., candidate scenarios, performance measures, survey instruments, WoW Switch Failure experimental script and procedure). It is prudent to reference both the previous tasks mentioned in Section 1 and the appendices of the current document.

Ultimately, we want to create safer aviation systems with the developing complexities that are interjected into the evolving airspace system. This document provides a method supported by empirical evidence to validate a training intervention and an assessment to train and assess pilots' behaviors and skills during unexpected events. The advocated HITL procedure cannot be simply conducted without useful interpretations of the complicated pattern of outcome findings. It is this *required* step in the research process that will produce recommendations useful to the aviation community. This analysis becomes more complex as technology continues to evolve and the context of aviation operations change rapidly (Hancock & Hoffman, 2015). Research results can ultimately lead to testing and validation, to inform FAA guidance on pilot training characteristics.

References

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science*, *1*(2), 164–180. https://doi.org/10.1111/j.1745-6916.2006.00011.x

Boland, E. (2016). *Managing startle & surprise* [Conference presentation]. PACDEFF. https://pacdeff.com/wp-content/uploads/2016/11/PACDEFF-Startle-Surprise-Management.pdf

Boyd, J. (1996). *The essence of winning and losing* [Unpublished briefing]. https://fasttransients.files.wordpress.com/2010/03/essence_of_winning_losing.pdf

Cohen J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Academic Press.

Cruit J., Frederick C., Blickensderfer B., Keebler J. R, Guinn T. (2017). Predicting general aviation pilots' weather-related performance through a scenario-based assessment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 2000–2004. https://doi.org/10.1177/1541931213601995

Dattalo, P. (2013). Analysis of multiple dependent variables. *Oxford Scholarship*. https://doi.org/10.1093/acprof:oso/9780199773596.001.0001

Dismukes, R. K., Goldsmith, T. E., & Kochan, J. A. (2015*). Effects of acute stress on aircrew performance: Literature review and analysis of operational aspects* (NASA TM-2015-218930). NASA Ames Research Center. https://human-factors.arc.nasa.gov/flightcognition/Publications/NASA_TM_2015_218930-2.pdf

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Federal Aviation Administration. (2017). *Advanced qualification program.* (Advisory Circular 120-54A). https://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentID/1028646

Federal Aviation Administration. (2004). *Crew resource management training* (Advisory Circular 120-51E). https://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentid/22879

Federal Aviation Administration. (2015). *Flightcrew member line-operational simulations: Line-oriented flight training, special purpose operational training, line operational evaluation*. (Advisory Circular 120-35D). https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_120-35D.pdf

Field, J. N., Boland, E. J., van Rooj, J. M., Mohrmann, J. F. W., Smeltink, J. W. (2018). *Startle effect management* (NLR-CR-2018-242). European Aviation Safety Agency. https://www.easa.europa.eu/document-library/research-reports/easarepresea20153

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, *40*(1), 35–42. https://doi.org/10.1016/j.socec.2010.10.008

Gordon, C. L., & Luo, S. (2011). The Personal Expansion Questionnaire: Measuring one's tendency to expand through novelty and augmentation. *Personality and Individual Differences, 51*(2), 89–94. https://doi.org/10.1016/j.paid.2011.03.015

Hancock, P. A. (2014). Automation: How much is too much? *Ergonomics, 57*(3), 449–454. https://doi.org/10.1080/00140139.2013.816375

Hancock, P. A. (2017). Imposing limits on autonomous systems. *Ergonomics*, *60*(2), 284–291. https://doi.org/10.1080/00140139.2016.1190035

Hancock, P. A. (2021). *Reacting and responding to rare, uncertain, and unprecedented events.* Revision in Submission.

Hancock, P. A., & Cruit, J. (2020). *The route to resilience.* Manuscript in Submission.

Hancock, P. A., Cruit, J., Kaplan, A. D., Riaz, Y., Kochan, J. A., & Pruchnicki, S. (2021). *Pilots' responses to unexpected events: Concepts and constructs*. Revision Submitted.

Hancock, P. A., Cruit, J., Kochan, J., Gruber, M., Diaz, Y. (2021). *Air carrier training recommendations to address limitations of pilot procedures during unexpected events in NextGen operations*. Interim report Technical reported submitted to the Federal Aviation Administration, Washington D.C.

Hancock, P. A., & de Ridder, S. (2003). Behavioral response in accident-likely situations. *Ergonomics*, *46*(12), 1111–1135. https://doi.org/10.1080/0014013031000136386

Hancock, P. A., & R. R. Hoffman. (2015). Keeping Up with Intelligent Technology. *IEEE Intelligent Systems*, *30*(1), 62–65. https://doi.org/10.1109/MIS.2015.13

Hancock, P.A., & Volante, W. (2020). Quantifying the qualities of language. *PLoS ONE*, *15*(5): e0232198.

Hopkin, C. R., Hoyle, R. H., & Gottfredson, N. C. (2015). Maximizing the yield of small samples in prevention research: A review of general strategies and best practices. *Prevention Science, 16*(7), 950–955. https://doi.org/10.1007/s11121-014-0542-7O

IBM Corp. (2021). *IBM SPSS Statistics for Macintosh* (Version 28.0) [Computer software]. https://www.ibm.com/products/spss-statistics

Jefferies, C., & AA personnel (2020). *Trailblazers into Safety-II: American Airlines' learning and improvement team.* SKYbrary. https://www.skybrary.aero/index.php/Trailblazers_into_Safety-II:_American_Airlines%E2%80%99_Learning_and_Improvement_Team

Kochan, J. (2005)*. The role of domain expertise and judgment in dealing with unexpected events.* [Unpublished doctoral dissertation]. University of Central Florida. https://stars.library.ucf.edu/etd/456/

Kochan, J. A., Breiter, E. G., & Jentsch, F. (2004). Surprise and unexpectedness in flying: Database reviews and analyses. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(3), 335–339. https://doi.org/10.1177/154193120404800313

Kochan, J. A., Breiter, E. G., & Jentsch, F. (2005). Surprise and Unexpectedness in Flying: Factors and Features. *Proceedings of the 13th International Symposium on Aviation Psychology*. Oklahoma City: Wright State University.

Landman, A., Groen, E. L., van Paassen, M. M., Bronkhorst, A. W., & Mulder, M. (2017). Dealing with unexpected events on the flight deck: A conceptual model of startle and surprise. *Human Factors*, *59*(8), 1161–1172. https://doi.org/10.1177/0018720817723428

Landman, A., van Middelaar, S. H., Groen, E. L., van Paassen, M. M, Bronkhorst, A. W., & Mulder, M. (2020). The effectiveness of a mnemonic-type startle and surprise management procedure for pilots. *The International Journal of Aerospace Psychology*, *30*(3–4), 104–118. https://doi.org/10.1080/24721840.2020.1763798

Landman, A., van Oorschot, P., van Paassen, M. M. R., Groen, E. L., Bronkhorst, A. W., & Mulder, M. (2018). Training pilots for unexpected events: A simulator study on the advantage of unpredictable and variable scenarios. *Human Factors*, *60*(6), 793–805. https://doi.org/10.1177/0018720818779928

Magno, C. (2010). The role of metacognitive skills in developing critical thinking. *Metacognition and Learning, 5*(2), 137–156. https://doi.org/10.1007/s11409-010-9054-4

Martin, W. L. (2017). *Developing startle and surprise training interventions for airline training programs* [Conference Presentation]. PACDEFF. http://pacdeff.com/wp-content/uploads/2017/08/PACDEFF- FC- Forum-Presentation- on- Startle.pdf

Pruchnicki, S., Key, K., & Rao, A. (2019). *Problem solving/decision making and procedures for unexpected events: A literature review* (DOT/FAA/AM-20/01). Office of Aerospace Medicine. https://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/2020s/media/202001.pdf

Spiro, R. J., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1995). Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. In L. P. Steffe & J. E. Gale (Eds.), *Constructivism in education* (pp. 85–107). Lawrence Erlbaum Associates, Inc.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Bibliography

Burke, C., Wilson, K. A., & Salas, E. (2005). The use of a team-based strategy for organizational transformation: Guidance for moving toward a high reliability organization. *Theoretical Issues in Ergonomics Science*, *6*(6), 509–530. https://doi.org/10.1080/24639220500078682

Chandra, D. C., Sparko, A., Kendra, A., Kochan, J. (2020). *Operational complexity in performance-based navigation (PBN) arrival and approach instrument flight procedures (IFPs)* (DOT-VNTSC-FAA-20-02). U.S. DOT Volpe National Transportation Systems Center. https://rosap.ntl.bts.gov/view/dot/43835

Damm, G. F. (2017). *Is resilience the most valuable skill for a pilot?* Aerotime Hub. https://www.aerotime.aero/aerotime.team/19764-is-resilience-the-most-valued-skill-for-a-pilot?page=2

Dekker, S. W. A. (2003). Illusions of explanation: A critical essay on error classification. *The International Journal of Aviation Psychology, 13*(2), 95–106. https://doi.org/10.1207/S15327108IJAP1302_01

Federal Aviation Administration. (2016). *Pilot training and checking*. (Advisory Circular 120-114). https://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentid/1030049

Flin, R. (2019). CRM (nontechnical) skills: A European perspective. In B. G. Kanki, J. Anca, & T. R. Chidester (Eds.), *Crew resource management* (pp. 185–206). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-812995-1.00006-3

Hoffman, R. R., & Hancock, P. A. (2017). Measuring resilience. *Human Factors, 59*(4), 564–581. https://doi.org/10.1177/0018720816686248

Holbrook, J. B., Stewart, M. J., Smith, B. E., Prinzel, L. J., Matthews, B. L., Avrekh, I., Cardoza, C. T., Ammann, O. C., Adduru, V., and Hull, C. H. (2019). *Human performance contributions to safety in commercial aviation* (NASA/TM-2019-220417). NASA Ames Research Center. https://hsi.arc.nasa.gov/publications/NASA-TM-2019-220417-comp.pdf

Hollnagel, E. (2006). Achieving system safety by resilience engineering. *2006 1st IET International Conference on System Safety*, 184–195. https://ieeexplore.ieee.org/document/4123704

Hunter, R. & Kochan, J. A. (2016). *System Safety Management Transformation 2014 Table Top Exercises Report.* Volpe National Transportation System Center.

Industrial Economics, Incorporated (2018). *Enhancing the use of expert judgment at the U.S. Federal Aviation Administration* (Final Report June 15, 2018).

International Civil Aviation Organization (ICAO) (2018). *Safety Management Manual (Doc 9859) 4th Edition.*

Kahneman, D. (2011). *Thinking, fast and slow.* Farrar, Straus and Giroux.

Kochan, J. A., Breiter, E.G., & Jentsch, F. (2005). Surprise and unexpectedness in flying: Factors and features. *Proceedings of the 13th International Symposium on Aviation Psychology*. Oklahoma City: Wright State University. https://corescholar.libraries.wright.edu/isap_2005/59

Landman, A., van Middellaar, S. H., Groen, E. L., Van Paassen, M. M., Bronkhorst, A. W., & Mulder, M. (2019). Testing the applicability of a checklist-based startle management method in the simulator. *Proceedings of the 20th International Symposium on Aviation Psychology*, 325–330. https://corescholar.libraries.wright.edu/isap_2019/55/

Landman, A., van Middellaar, S. H., Groen, E. L., Van Paassen, M. M., Bronkhorst, A. W., & Mulder, M. (2019). Pilot evaluations of a non-verbal startle and surprise management method, tested during airline recurrent simulator training. *Proceedings of the 20th International Symposium on Aviation Psychology*, 337–342. https://corescholar.libraries.wright.edu/isap_2019/57

Morgan, M. G. (2014). Use (and abuse) of expert elicitation in support of decision making for public policy. *Proceeding of the National Academy of Sciences of the United States of America*, *111*(20), 7176–7184. https://doi.org/10.1073/pnas.1319946111

National Association of Flight Instructors (NAFI) (2021). *Creating simulation scenarios.* NAFI MentorLIVE! https://www.mentorlive.site/program/48.html

National Transportation Safety Board. (1989). *Aircraft Accident Report Aloha Airlines Flight 243 Boeing 737-200, N73711, near Maui, Hawaii, April 28, 1988* (NTSB/AAR-89/03). https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR8903.pdf

National Transportation Safety Board. (1990). *Aircraft Accident Report United Airline Flight 232 McDonnell Douglas DC-10-10, Sioux Gateway Airport, Sioux City, Iowa, July 19,1989 (*NTSB/AAR-90/06). https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR-90-06.pdf

National Transportation Safety Board. (2010). *Loss of Thrust in Both Engines After Encountering a Flock of Birds and Subsequent Ditching on the Hudson River, US Airways Flight 1549 Airbus A320-214, N106US, Weehawken, New Jersey, January 15, 2009* (NTSB/AAR-10/03). https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR1003.pdf

National Transportation Safety Board. (2016). *Aviation Accident Final Report Southwest Airlines Flight 3472 Boeing 737-7H4, N766SW, Pensacola, Florida, August 27, 2016* (DCA16FA217). https://www.skybrary.aero/bookshelf/books/5822.pdf

Oser, R. L., Gualtieri, J. W., & Cannon-Bowers, J. A. (1998). Enhancing training systems design: Implementing an event-based approach. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*(18), 1309–1313. https://doi.org/10.1177/154193129804201807

Pilots: Extended Envelop Training, 14 C.F.R. § 121.423. (2020). https://www.ecfr.gov/current/title-14/chapter-I/subchapter-G/part-121/subpart-N/section-121.423

Rivera, J., Talone, A. B., Boesser, C. T., Jentsch, F., & Yeh, M. (2014). Startle and surprise on the flight deck: Similarities, differences, and prevalence. *Proceedings of the Human Factors and*

Ergonomics Society Annual Meeting, *58*(1), 1047–1051. https://doi.org/10.1177/1541931214581219

Saab Sensis Corporation. (2013). *Integrated safety assessment model design* (No. 840-035452).

Woods, D. D. (2015). Four concepts for resilience and their implications for systems safety in the face of complexity. *Reliability Engineering and System Safety, 141*, 5–9. https://doi.org/10.1016/j.ress.2015.03.018

Appendix A: Example Candidate Scenarios in Training for the Unexpected

Appendix A.1 summarizes example candidate scenarios in training for the unexpected. Appendix A.2 expands on these scenarios by providing the type of scenario, synopsis of the scenario, and real-world examples of the scenarios.

A.1 Summary of Example Candidate Scenarios

1. Squat switch (Weight-on-Wheels, proximity, air/ground) switch failure on take-off
2. Cabin pressurization loss in climb (above 15,000 feet)
3. Runaway stabilizer trim in climb
4. Severe turbulence in climb
5. ATC erroneous vector
6. Rejected take-off (RTO) with Brake Temperature Monitor System deferred per the MEL (See pp. 32-33 for example https://fsims.faa.gov/wdocs/mmel/b-737_rev_61.pdf) – this is a short scenario to observe the RTO and whether the procedures per the MEL guidance are followed.
7. Glidepath fails to capture on RNAV approach (on an approach that needs LPV minimums) – common problem as reported in ASRS reports; pilot will need to perceive the situation, process the information, then perform the appropriate response (go missed)
8. High engine vibration readings in moderate turbulence – with the appropriate system knowledge, the pilot would know that environmental perturbations should not increase engine vibrations (much) and it will be telling if the checklist is consulted or not.
9. Slow degradation of engine performance (fuel control unit impending failure) – an insidious failure may not get the attention of the pilot until alerts occur but diagnosing and responding to the condition will show the process of responding to the event.
10. Passenger illness requiring diversion (very common occurrence in commercial aviation) and unable to contact company medical service for advice– how will the pilot react?
11. “Door Light” on climb-out – While there is a checklist for this, will it be appropriately accomplished?
12. Engine Bleed Air Shutoff Valves (PRSOV) deferred and flight encounters unforecast icing conditions.
13. Degraded Flight Control Law Mode – An unexpected degradation of to a degraded flight control law without evidence of the cause at high altitude.

A.2 Descriptions and Examples of Candidate Scenarios with Historical Events

#	Scenario Category/Categories	Scenario Name	Synopsis	Historical Events
1.	System Malfunction/Weather Event	Squat switch (weight-on-wheels, proximity, air/ground) failure on take-off	This scenario presents the pilots with an ill-defined event where multiple failures and ambiguous alerts could be presented. The flight is planned from KMCO to KCVG at FL 350. The clearance is the McCoy 2 Departure as filed. The weather is KMCO 250153Z 13009KT 1/2SM BR 23/22 BKN 13 OVC 4000 A2985 RMK AO2 SLP219 T02280178 and the flight is scheduled to depart at 0100Z (night). All operations are normal until just after lift-off, when the landing gear fails to retract. Annunciations associated with a failed WoW (or air/ground) switch are displayed (per the type of aircraft used). Note, that even a fairly basic aircraft type training device could accomplish this scenario.	Jul 18, 2021. A Korean Air Cargo Boeing 747-400 freighter, registration HL7601 performing flight KE-351 from Ho Chi Minh City (Vietnam) to Bangkok (Thailand), was climbing out of Ho Chi Minh's runway 25L when the crew could not retract the landing gear. The aircraft stopped the climb at 5000 feet and returned to Ho Chi Minh City for a safe landing on runway 25L about 20 minutes after departure. The aircraft remained on the ground in Ho Chi Minh City for 25.5 hours, then departed for the flight to Bangkok. (Hyperlink)
2.	System Malfunction/Weather Event	Cabin pressurization loss in climb (above 15,000 feet) in icing conditions	Cabin pressurization loss in climb (above 15,000 feet) in icing conditions	Sep 17, 2021. A KLM Boeing 737-900, registration PH-BXO performing flight KL-1387 from Amsterdam (Netherlands) to Kiev (Ukraine), was climbing out of Amsterdam's runway 18L when the crew stopped the climb at FL100 to solve an issue. The crew subsequently reported they had a pressurization problem, needed to burn off fuel and would extend gear and flaps early which would produce quite some noise. The aircraft was vectored out to the North Sea and subsequently landed safely on Amsterdam's runway 18R about 65 minutes after departure. A replacement Boeing 737-900 registration PH-BXP reached Kiev with a delay of about 3 hours. (Hyperlink)
				Jun 26, 2020. A Skywest Canadair CRJ-200 operating as United Airlines flight number UA-5071 lost cabin pressure on route to Prescott, Arizona. The 50-seat regional jet, registration number N431SW, was flying from Denver International Airport (DEN) to Prescott Regional Airport (PRC) when the incident occurred. (Hyperlink)

#	Scenario Category/Categories	Scenario Name	Synopsis	Example
				Aug 14, 2005. Helios Airways Flight 522 (HCY 522 or ZU522) was a Helios Airways Boeing 737-300 flight that crashed into a mountain at 12:04 EEST, north of Marathon and Varnavas, Greece. An engineer had carried out a pressurization leak check on the aircraft earlier that day. However, this had involved the engineer setting the pressurization system to manual to avoid running the engines for the check. Upon completing the check, they forgot to reset the pressurization system to auto, which ultimately led to the loss of pressure that incapacitated the flight's crew. (Hyperlink)
3.	System Malfunction/Weather Event	Runaway stabilizer trim in climb	Runaway stabilizer trim in climb in low IFR conditions	Nov 6, 2019. A Republic Airways Embraer ERJ-175, registration N117HQ performing AA-4439 from Atlanta, GA to New York La Guardia, NY (USA) with 6 passengers and 3 crew, was climbing out of Atlanta's runway 09L when the crew declared emergency reporting they had a trim runaway, the crew stopped the climb at about 14,000 feet and positioned for a return to Atlanta's runway 10. The crew subsequently reported, while cleared for a right downwind to runway 10, they were in a stalling situation and subsequently added they couldn't get their pitch down, they were trying to descend nonetheless. ATC offered runway 08L, 10 or 09R, ATC could clear anyone out of the way. The crew advised they were able to take a turn and received vectors to runway 10. Instead of descending the aircraft began to climb again, then descended, the crew advised they got a system warning to cut out, got the problem under control and were now okay, they had been fighting with the aircraft for a while. The aircraft joined the final for runway 10, ATC again offered runway 10 or 09R, "your call", and cleared the aircraft to land either runway. The aircraft landed safely on runway 10 about 19 minutes after departure and about 15 minutes after the emergency call. (Hyperlink)

#	Scenario Category/Categories	Scenario Name	Synopsis	Example
4.	System Malfunction/Weather Event	Severe turbulence in climb	Severe turbulence in climb results in high engine vibration readings in moderate turbulence – with the appropriate system knowledge, the pilot would know that environmental perturbations should not increase engine vibrations (much) and it will be telling if the checklist is consulted or not.	Aug 7, 2021. A KLM Cityhopper Embraer ERJ-175, registration PH-EXU performing flight KL-1251 from Amsterdam (Netherlands) to Nice (France), was climbing out of Amsterdam when the crew requested to stop the climb at FL290 to work a problem and consult maintenance. The crew subsequently indicated they needed to return to Amsterdam, an engine (CF34) showed high vibrations but didn't need to be shut down. The aircraft returned to Amsterdam for a safe landing on runway 27 about 40 minutes after departure. A replacement Embraer ERJ-175 registration PH-EXR reached Nice with a delay of 2.5 hours. The occurrence aircraft is still on the ground in Amsterdam about 14 hours after landing back.(Hyperlink)
5.	Crew Resource Management Event	ATC erroneous vector	ATC erroneous vector for RNP approach	
6.	System Malfunction Event	Rejected take-off with brake temperature monitor system deferred	Rejected take-off (RTO) with Brake Temperature Monitor System deferred per the MEL	See pp. 32-33 for example- this is a short scenario to observe the RTO and whether the procedures per the MEL guidance are followed. (Hyperlink)
7.	System Malfunction Event	Glidepath fails to capture on RNAV approach	Glidepath fails to capture on RNAV approach (on an approach that needs LPV minimums) – common problem as reported in ASRS reports; pilot will need to perceive the situation, process the information, then perform the appropriate response (go missed)	
8.	System Malfunction/Weather Event	High engine vibration readings in moderate turbulence	High engine vibration readings in moderate turbulence – with the appropriate system knowledge, the pilot would know that environmental perturbations should not increase engine vibrations (much) and it will be telling if the checklist is consulted or not.	

#	Scenario Category/Categories	Scenario Name	Synopsis	Example
9.	Mechanical Event	Slow degradation of engine performance (fuel control unit impending failure)	Slow degradation of engine performance (fuel control unit impending failure) – an insidious failure may not get the attention of the pilot until alerts occur but diagnosing and responding to the condition will show the process of responding to the event.	Aug 8, 2021. A Delta Airlines Airbus A330-300, registration N808NW performing flight DL-304 from Honolulu, HI to Atlanta, GA (USA), was climbing through FL260 out of Honolulu about 80nm northeast of Kahului, HI (USA) when the crew declared Mayday reporting the failure of the right-hand engine (PW4168) and decided to return to Honolulu. Descending towards FL160 the crew advised they had reduced the engine to idle due to a compressor stall followed by overtemperature and vibrations. The aircraft landed safely back on Honolulu's runway 08L about one hour after departure. The flight was cancelled. The aircraft is still on the ground in Honolulu awaiting a new engine. (Hyperlink)
10.	Passenger Event	Passenger illness requiring diversion	Passenger illness requiring diversion (very common occurrence in commercial aviation) and unable to contact company medical service for advice.	
11.	Alerting Event	Door light illuminates on take-off	"Cargo Door" light illuminates on take-off roll. If flight is continued, the aircraft does not pressurize. Log-book write-up showing previous instance of event with "adjustment to the door switch" as corrective action.	Oct 17, 2016. A United Boeing 767-300, registration N657UA performing flight UA-934 from Newark, NJ (USA) to London Heathrow, EN (UK), was accelerating for takeoff from Newark's runway 22R when the crew rejected takeoff at low speed reporting an open cargo door indication and returned to the apron. The aircraft departed about 2 hours later and reached London with a delay of just under two hours. A passenger reported the captain announced on the PA that a cargo door light had illuminated prompting the reject. Following maintenance and refueling the aircraft departed about two hours later. (Hyperlink)
12.	System Malfunction Environmental Event	Engine Bleed Air Shutoff Valves deferred	Engine Bleed Air Shutoff Valves (PRSOV) deferred and flight encounters unforecasted icing conditions.	

#	Scenario Category/Categories	Scenario Name	Synopsis	Example
13.	System Malfunction Event/Flight Envelope Protection	Degraded Flight Control Law Mode	An unexpected degradation of to a degraded flight control law without evidence of the cause at high altitude.	July 27, 2016. EasyJet flight U25462 was in cruise at FL340 over northern France when it experienced an oscillation in pitch and in normal acceleration during which a fault was detected in the stabilizer system. The flight crew were alerted to the fault when the autopilot disengaged, and the Master Caution annunciated with ECAM message stab jam. The co-pilot took control of the aircraft while the commander carried out the ECAM actions. The checklist required the flight crew to check that the manual trim was available and to move the stabilizer trim until the elevator was in the neutral position. The commander stated that they moved the stabilizer trim wheel a little but the co-pilot stated that he felt that the aircraft was largely in trim and so they decided not to move the stabilizer significantly after that. As a result of the fault the control law degraded into 'Alternate Law', which provided reduced levels of protection and the use of the autopilot was lost. However, 'load factor demand law' was maintained as were load factor protection and low/high speed stability functions. During the time when the flight crew were performing the checklist items, the aircraft started a gradual climb 100 feet from its assigned altitude; however, the flight crew were able to bring the aircraft back to the assigned altitude with minor control stick inputs. The flight crew descended below RVSM airspace and continued the flight to Gatwick with the autopilot disengaged. When the landing gear was lowered during the approach the control law changed to 'Direct Law', as designed, and an uneventful landing was carried out. (Hyperlink)

Appendix B: Task Analysis for Weight-on-Wheels Switch Failure Scenario

This table is an expansion of Table 8, the task analysis for the WoW scenario. In addition to providing information on the breakdown of tasks and steps associated with the WoW scenario, the task analysis elicited information on the knowledge, skills, decisions, and behaviors needed for the pilot to perform optimally for the given task. The SME provided information on which tasks/steps required difficult cognitive skills and when and what kind of expertise is critical for each task/step. The SME also provided information on strategies for dealing with the event and why certain tasks/step might be more challenging or susceptible to potential errors.

Stages of flight	What tasks would be needed to have a successful outcome in the scenario?			Break down each task into steps	Knowledge audit			
I. Preflight	**Task #**	**Task Description**	**Step #**	**Description**	**Which steps require difficult cognitive skills?**	**When is expertise used? What kind of expertise?**	**Strategies for dealing with event**	**Why difficult?/Potential Errors**
	1	Checked myself; daytime flight	1.1	"I'm safe" checklist; have I been ill? Am I fatigued? Hydration? Have I had proper nutrition? Am I distracted?	Decision-making to be aware of biases that can come about; I'm not feeling well but "it's not that bad." If I'm telling myself "It's not that bad," that's a red flag	Each task requires general expertise (judgment and domain technical expertise). A baseline level of expertise is necessary. Working knowledge of aircraft systems, of yourself, of crew, of environmental factors, of the interaction between these factors	• Not rush; verbalize and acknowledge rushing; then stop it. Rushing does not allow a thorough preflight assessment • Use metacognition to see if you are rushing and slow down • Keep track of how things are interrelated • Someone might not miss a particular item, but they might not combine it with a different anomaly. Seeing the relationship between different items is a method of risk assessment • Working smarter not harder is important for working with crew. Don't try to do it all yourself. Allocate tasks to others as necessary.	• Might be difficult for novices if not aware of consequences, feel time pressure, not want to come early • Novices could struggle with the big picture. It takes practice reflecting back on the outcome of the flight based on what you did in the preflight • Rushing; doing too much in too little time • Difficult during fatigue (difficult to assess capabilities when fatigued) • Difficult when there is a lot of external pressure (e.g., wanting to go home) • Inattention and not paying attention to what is being done preflight for the equipment
			1.2	Think about external pressures, make sure I'm compartmentalized enough to stay focused on the task at hand	Not just going through the checklist but listening to my responses. If I am tired, inform crew; tell them to not hesitate to speak to speak up if they think I'm not doing well			
	2	Take note of maintenance discrepancy for previous hydraulic problem and any other previous write-ups on airplanes; get feel for maintenance history	2.1	General maintenance status within compliance inspections	If there is a trend of related malfunctions, keep that in mind during your flight so if one of those occurs again it won't be surprising	Baseline level of domain and judgment expertise		
			2.2	Go back page by page in logs, see what kind of write-ups there have been	Synthesizing maintenance history of airplane; creating mental model of areas where we might see something happen	Knowledge of landing gear system; Beyond baseline understanding		
			2.3	Go back far enough to see if there are any additional		Recognizing anomalies (useful in all of the tests); using cues and		

				malfunctions in the hydraulic systems (check for any hydraulic write-ups)		clues to make predictions	• Get someone else's opinion • Adding to the flight plan • Sometimes when you have a new crewmember you have to mentor them and make sure they know what they are doing and help them grow. • Be aware of my own performance • Check for bias; step back and look at the situation. If I say "it's not that bad" that's a red flag that I'm biased • Asking other for input; try to go back to quantitate measurements • Maintenance log with previous hydraulic problem or other landing gear write-ups	
			2.4	Get idea if the aircraft has been flying with deferred items, and if so, what those items are		Having sufficient expertise to be able to recognize what things should be like		
			2.5	Brief crew on logbook review				
	3	Check weather; some discrepancies between forecast and radar; pay attention to that	3.1	Look at departure and destination weather so when I look at the bigger picture, I can see why the weather at those locales is the way it is	Is the weather particularly better or worse than what was forecasted? Be aware and keep this in back of your mind			
			3.2	Go through weather briefing (this is about 30 steps)				
	4	Environment: Be aware of terrain in area knowing that I'm going to be returning to the departure airport or takeoff alternate depending on weather and terrain clearance I can maintain if I cannot retract my landing gear	4.1	General awareness using charting; see what terrain is like around the airport, especially if not familiar with the area	Being able to correlate the capabilities of the aircraft, the terrain, the elevation needed.			
			4.2	Look at altitudes required for departure procedure; see if any of them have terrain constraints associated with them	Realizing that if we get off the departure procedure, we may find ourselves closer to terrain than we want to be			
	5	Crew briefing to discuss pre-flight maintenance status of aircraft; departure procedure, terrain, weather	5.1	Open door for communication; let them know their opinion is valued, considered, and equally important	Listening skills (being able to listen to different positions)			
			5.2	Letting the crew tell me what they think before I say what I	Synthesizing what the other crew members are saying			

				think. "What do you think about this weather, departure procedure, and terrain?" Open door for communication; eliciting their viewpoint				
			5.3	Come to a shared mental model on how we're going to proceed				
	6	Note fuel state to determine if I would be below my maximum landing weight at takeoff	6.1	Being aware that takeoff weight is under maximum				
II. Taxi	**Task #**	**Task Description**	**Step**	**Step Description**	**Description**	**Which steps require difficult cognitive skills?**	**When is expertise used? What kind of expertise?**	**Strategies for dealing with event**
	7	****All general/typical until event happens**	7.1					
III. Takeoff/climb	**Task #**	**Task Description**	**Step**	**Step Description**	**Description**	**Which steps require difficult cognitive skills?**	**When is expertise used? What kind of expertise?**	**Strategies for dealing with event**
	8	Ensure positive aircraft control and follow designated flight path for the procedure	8.1	Make sure gear is down	Manage workload (accomplishing tasks properly and in the right order) to avoid distraction from managing aircraft and flight path	Know the systems but use good judgment	• Maintaining aircraft control is important • Failed squat switches with subtle indications resulting in failure on a flight test. Not noticed due to inattentiveness • Decide if there is a reason to declare an emergency to give us priority to land • Improvisation would be kept to a minimum. Might be inclined to try a few things to get the gear up, but it is not a good idea of conditions do not indicate to forgo that procedure	• Having an event that disrupts the normal flow of the flight can have a more profound effect on novice pilots • Most pilots will have a difficult time because they do not want to declare an emergency
			8.2	Be mindful of gear speed	Stress management (normal part of flying): "Breathe and focus"	Using resources to best accomplish the resolution of the situation		
			8.3	Make sure we can make climb gradient for departure	Focus on flying the airplane	Metacognition, make sure you are thinking about the appropriate thing		
			8.4	Make sure to stay on course				
	9	Communicate that the landing gear appears to not be off the	9.1	Ask for confirmation that the gear is not up	Clear, concise, continuous, and appropriate use of terminology. No ambiguous language.	Workload management	Will have to change course of action as decisions change	Novice pilots don't want disruptions and don't want to declare an emergency

		ground; Indicate what appears to be the problem			Essentially think out loud to allow crewmembers to fill in any gaps			
			9.2	Communicate with other crewmembers	Judgment: make sure everyone is aware	Getting to destination vs suitable alternative destination		
			9.3	Make sure ATC is advised of landing gear anomaly	Problem: There may be steps not on the checklist	Problem solving: evaluating situation and determining what can be done and what is most likely based on evaluation and use of checklist		
	10	Make sure we have a holding pattern and stay in the vicinity of the airport as we figure out what is going on (unsure of what the failure is, so we want to stay close)	10.1	Complete the checklist	Make sure weather allows this			
			10.2	Use procedure that would put the landing gear back down and retract flaps	Breathing and focus to maintain awareness of big picture			
			10.3	Pull appropriate circuit breaker				
			10.4	Land at nearest suitable airport				
IV. Descent	**Task #**	**Task Description**	**Step**	**Step Description**				
	11	****All normal**	11.1					
V. Approach and Landing	**Task #**	**Task Description**	**Step**	**Step Description**				
	12	Request to land at nearest airport	12.2	It would all be normal. Anything abnormal would be seen from checklist				
			12.2	Manual speed brakes on landing				
	13	Write up discrepancy and have maintenance take care of it						

Appendix C: Human-in-the-Loop Study Forms

This appendix contains the following forms used for data collection. A brief description of the purpose and use of the form, questionnaire, or survey follows.

C.1 Participant Demographic Questionnaire Development

This is an outline of the demographic data that can be capture electronically on a platform such as Qualtrics. The choice of variables in this questionnaire facilitates controlling for confounding factors in data analysis as well as for categorizing participants.

Background Information
Age in Years__________
Highest Educational Level Obtained______________________________
Current Pilot Position/Title_____________________________________

Airman Certificates, Ratings, and Qualifications
Private Pilot (ratings) ____
Commercial Pilot (ratings)____
Airline Transport Pilot (ratings)____
Certified Flight Instructor (ratings)____
Company Instructor (aircraft types)____
Pilot Examiner (authorizations)____
Check Airman (authorizations)____
Years flying__________Total flight hours____________Flight Hours Last 12 Months________________
Types of aircraft flown____
Current airplane type(s) flying now?   ____
Were you or are you in the military?__________Which Branch?_______Flight Status?_________
Are you familiar with the Interviewer?________

C.2 Judgment Expertise Surveys

These surveys were included in the experimental protocol for the Pilot Needs Analysis (Phase I-Task 3). They will be administered electronically via Qualtrics (a data collection tool) as part of the Human-in-the-Loop study.

The purpose of the surveys is to obtain measures of four factors (human traits/states/behaviors) found in the literature (Hancock et al., in preparation) that contribute to being able to best respond to unexpected events. The Technical Expertise survey will give a score to be used in the determination of participants' domain expertise. The data from these surveys will be included in the analysis of the HITL data.

Participants will be instructed to complete the following instruments that will be posted as a single survey. The surveys (measurement instruments) and primary references are:

Survey #1 – Self-Efficacy and Metacognition

Pintrich, P. R., Smith, D. A., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, *53*(3), 801–813. https://doi.org/10.1177/0013164493053003024

Schmidt, A. M., Ford, K. J. (2003). Learning within a learner control training environment: the interactive effects of goal orientation and metacognitive instruction on learning outcomes. *Personnel Psychology 56,* 405–429. https://doi.org/10.1111/j.1744-6570.2003.tb00156.x

Wells, A., & Cartwright-Hatton, S. (2004). A short form of the metacognition questionnaire: properties of the MCQ-30. *Behaviour Research and Therapy*, *42(*4), 385–396. https://doi.org/10.1016/S0005-7967(03)00147-5

Survey #2 – Generalized Self-Efficacy

Scholz, U., Doña, B. G., Sud, S., & Schwarzer, R. (2002). Is general self-efficacy a universal construct? Psychometric findings from 25 countries. *European Journal of Psychological Assessment*, *18*(3), 242–251. https://doi.org/10.1027/1015-5759.18.3.242

Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale*.* In J. Weinman, S. Wright, & M. Johnston (Eds). *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35–37). NFER-NELSON. http://userpage.fu-berlin.de/~health/engscal.htm

Survey #3 – Trust in Automation

Wojton, H. M., Porter, D., Stephanie T. Lane, S. T., Bieber, C., & Madhavan, P. (2020). Initial validation of the trust of automated systems test (TOAST), *The Journal of Social Psychology*, *160*(6), 735–750. https://doi.org/10.1080/00224545.2020.1749020

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Survey #4 – Cognitive Flexibility

Dennis, J. P., & Vander Wal, J. S. (2010). The Cognitive Flexibility Inventory: Instrument development and estimates of reliability and validity. *Cognitive Therapy and Research*, *34*(3), 241–253. https://doi.org/10.1007/s10608-009-9276-4

Survey #5 – Aviation Self-Efficacy

Cruit. J. (2016). *Predicting general aviation pilots' weather-related performance through a scenario-based written assessment* [Doctoral dissertation]. Embry-Riddle Aeronautical University. https://commons.erau.edu/edt/198

Domain Expertise – Technical Expertise Assessment

This survey is a selection from the FAA Airline Transport Pilot Multiengine knowledge test (https://www.faa.gov/training_testing/testing/test_questions/media/atm_questions.pdf).

C.2.1 SME Survey #1 (Self-Efficacy and Metacognition)

Participant ID #_______

DIRECTIONS: Please answer the following questions by circling the appropriate rating.

1. I believe I will receive excellent ratings for my performance on this interview.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
a. 1	2	3	4	5

2. I'm certain I can handle the most difficult situations that arise in flying.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
a. 1	2	3	4	5

3. I memorize key words to remind me of the important concepts when studying.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
a. 1	2	3	4	5

4. I practice material mentally while "chair flying."

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
a. 1	2	3	4	5

5. Considering the difficulty of the flying task and my skills, I think I do well.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
a. 1	2	3	4	5

6. I believe that I perform within the top 10% of all participants on the flying task.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
a. 1	2	3	4	5

7. I read over my notes and the course materials often when I am in training.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
a. 1	2	3	4	5

8. I expect to do well in my flying.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
a. 1	2	3	4	5

9. I am confident I can do an excellent job on my flying tasks.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
a. 1	2	3	4	5

10. I make lists of important terms and memorize the lists.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
a. 1	2	3	4	5

C.2.2 SME Survey #2 (Generalized Self-Efficacy)

Participant ID #_______

DIRECTIONS: Indicate for each statement below how true it is for you.

1. I can always manage to solve difficult problems if I try hard enough.

Not at all True	Hardly True	Moderately True	Exactly True
1	2	3	4

2. If someone opposes me, I can find the means and ways to get what I want.

Not at all True	Hardly True	Moderately True	Exactly True
1	2	3	4

3. I am certain that I can accomplish my goals.

Not at all True	Hardly True	Moderately True	Exactly True
1	2	3	4

4. I am confident that I could deal efficiently with unexpected events.

Not at all True	Hardly True	Moderately True	Exactly True
1	2	3	4

5. Thanks to my resourcefulness, I can handle unforeseen situations.

Not at all True	Hardly True	Moderately True	Exactly True
1	2	3	4

6. I can solve most problems if I invest the necessary effort.

Not at all True	Hardly True	Moderately True	Exactly True
1	2	3	4

7. I can remain calm when facing difficulties because I can rely on my coping abilities.

Not at all True	Hardly True	Moderately True	Exactly True
1	2	3	4

8. When I am confronted with a problem, I can find several solutions.

Not at all True	Hardly True	Moderately True	Exactly True
1	2	3	4

9. If I am in trouble, I can think of a good solution.

Not at all True	Hardly True	Moderately True	Exactly True
1	2	3	4

10. I can handle whatever comes my way.

Not at all True	Hardly True	Moderately True	Exactly True
1	2	3	4

C.2.3 SME Survey #3 (Trust in Automation)

Participant ID #_______

DIRECTIONS: Please answer the following questions regarding aircraft flight management systems by circling the appropriate rating.

1. I understand what the system should do.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1	2	3	4	5

2. I understand the limitations of the system.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1	2	3	4	5

3. I understand the capabilities of the system.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1	2	3	4	5

4. I understand how the system executes tasks.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1	2	3	4	5

5. The system helps me achieve my goals.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1	2	3	4	5

6. The system performs consistently.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1	2	3	4	5

7. The system performs the way it should.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1	2	3	4	5

8. I am rarely surprised by how the system responds.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1	2	3	4	5

9. I feel comfortable relying on the information provided by the system.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1	2	3	4	5

C.2.4 SME Survey #4 (Cognitive Flexibility)

Participant ID #_______

DIRECTIONS: Please use the scale below to indicate the extent to which you agree or disagree with the following statements when dealing with an unexpected event in aviation.

1. I am good at "sizing up" situations.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1	2	3	4	5

2. I have a hard time making decisions when faced with difficult situations.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1	2	3	4	5

3. I consider multiple options before making a decision.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1	2	3	4	5

4. When I encounter difficult situations, I feel like I am losing control.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1	2	3	4	5

5. I like to look at difficult situations from many different angles.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1	2	3	4	5

6. I seek additional information not immediately available before attributing causes to behavior.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1	2	3	4	5

7. When encountering difficult situations, I become so stressed that I cannot think of a way to resolve the situation.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1	2	3	4	5

8. I try to think about things from another person's point of view.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1	2	3	4	5

9. I find it troublesome that there are so many different ways to deal with difficult situations.

   Strongly Disagree Somewhat Disagree Neutral Somewhat Agree Strongly Agree
   1 2 3 4 5

10. I am good at putting myself in others' shoes.

    Strongly Disagree Somewhat Disagree Neutral Somewhat Agree Strongly Agree
    1 2 3 4 5

11. When I encounter difficult situations, I just don't know what to do.

    Strongly Disagree Somewhat Disagree Neutral Somewhat Agree Strongly Agree
    1 2 3 4 5

12. It is important to look at difficult situations from many angles.

    Strongly Disagree Somewhat Disagree Neutral Somewhat Agree Strongly Agree
    1 2 3 4 5

13. When in difficult situations, I consider multiple options before deciding how to behave.

    Strongly Disagree Somewhat Disagree Neutral Somewhat Agree Strongly Agree
    1 2 3 4 5

14. I often look at a situation from different viewpoints.

    Strongly Disagree Somewhat Disagree Neutral Somewhat Agree Strongly Agree
    1 2 3 4 5

15. I am capable of overcoming the difficulties in life that I face.

    Strongly Disagree Somewhat Disagree Neutral Somewhat Agree Strongly Agree
    1 2 3 4 5

16. I consider all the available facts and information when attributing causes to behavior.

    Strongly Disagree Somewhat Disagree Neutral Somewhat Agree Strongly Agree
    1 2 3 4 5

17. I feel I have no power to change things in difficult situations.

    Strongly Disagree Somewhat Disagree Neutral Somewhat Agree Strongly Agree
    1 2 3 4 5

18. When I encounter difficult situations, I stop and try to think of several ways to resolve it.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1	2	3	4	5

19. I can think of more than one way to resolve a difficult situation I am confronted with.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1	2	3	4	5

20. I consider multiple options before responding to difficult situations.

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1	2	3	4	5

C.2.5 SME Survey #5 (Aviation Self-Efficacy)

Participant ID #_______

Directions: A number of situations are described below that pertain to situations during flying in which you encounter something surprising or unexpected. Please rate in the blanks below how confident you are that your decisions for the given situations will result in a positive outcome.

Rate your degree of confidence of a positive outcome to the situation by entering a number from 0 to 100 using the scale given below:

0	10	20	30	40	50	60	70	80	90	100
Cannot do at all					**Moderately can do**					**Highly certain can do**

Situation	Confidence of Positive Outcome (0-100)
Contradictory Resolution Advisories from TCAS	_____
An event for which you were not trained	_____
Severe and sudden weather phenomena	_____
Incapacitated crewmember	_____
Loss of Situation Awareness	_____
An event that has no prescribed procedure	_____
Loss-of-communication with ATC	_____
Engine failure or power loss on takeoff	_____
Jet upset	_____
Loss of reliable airspeed	_____
Last minute Instrument Approach Procedure change	_____
Last minute runway change	_____

C.2.6 Domain Expertise (Technical Expertise Assessment)

Participant ID #_______

DIRECTIONS: Please answer the questions below.

1. As required by Part 121, an airport may be listed as an alternate in the flight release only if the weather forecast indicates that conditions will be at or above the
   A. Alternate weather minima specified in the operation specifications at the time of arrival.
   B. Lowest available IAP minima at the time of arrival.
   C. Lowest available IAP minima for 1 hour before to 1 hour after the time of arrival.

2. What effect does extending leading edge slats have on an airplane's wing?
   A. Increases the pitch up moment of an airfoil.
   B. Increases the camber and CL-MAX.
   C. Allows for earlier airflow separation.

3. Under what conditions might a pilot expect the possibility of hydroplaning?
   A. When landing on a wet runway that is covered in rubber from previous landings.
   B. When departing a grooved runway with less than a thousandth of an inch of water.
   C. When the adiabatic lapse rate is high, and steam is rising from the landing surface.

4. Which is a common symptom of hyperventilation?
   A. Visual acuity.
   B. Decreased breathing rate.
   C. Tingling sensations.

5. Altitude-induced hypoxia is caused by what atmospheric condition?
   A. Significantly less oxygen molecules at high altitude.
   B. Insufficient partial pressure of the inhaled oxygen.
   C. Incorrect balance of oxygen and carbon dioxide.

6. When using a flight director system, what rate of turn or bank angle should a pilot observe during turns in a holding pattern?

   A. 3° per second or 25° bank, whichever is less.
   B. 1-1/2° per second or 25° bank, whichever is less.
   C. 3° per second or 30° bank, whichever is less.

7. How does an increase in an aircraft's weight affect its climb performance?
   A. The aircraft will climb at a lower angle of attack, which allows for a higher TAS and higher rate of climb.
   B. Both parasite and induced drag are increased, which will lower the reserve thrust available to climb.

C. A higher aircraft weight requires that the aircraft is configured for climb earlier in the departure which allows a greater climb gradient.

8. What is the absolute ceiling of an airplane?
   A. The point where the minimum rate of climb becomes lower than the optimum L/DMAX speed.
   B. The altitude at which the aircraft is unable to climb at more than 100 feet per minute.
   C. When the maximum rate of climb and the maximum angle of climb speeds converge.

9. In a turbojet aircraft, when is braking performance optimized during landing?
   A. Before the nose wheel touches down.
   B. Wheel spin-up at touchdown.
   C. Maximum weight on main wheels.

10. To conduct an RNAV (GPS) approach to LPV minimums, the aircraft must be furnished with
    A. A GPS/WAAS receiver approved for an LPV approach by the AFM.
    B. A GPS (TSO-C129) receiver certified for IFR operations.
    C. An IFR approach-certified system with required navigation performance (RNP) of 0.5.

11. How does the stall speed (KCAS) vary as you climb from sea level to 33,000 feet?
    A. It varies directly with a change in altitude.
    B. It remains relatively unchanged throughout the climb.
    C. It varies indirectly with a change in altitude.

12. While on an ILS approach, what is the proper way to recover from an impending stall?
    A. Engage the autopilot.
    B. Changing flap settings.
    C. Reducing the angle of attack.

13. The crew monitoring function is essential,
    A. Particularly during high altitude cruise flight modes to prevent CAT issues.
    B. Particularly during approach and landing to prevent CFIT.
    C. During RNAV departures in class B airspace

14. One purpose of Crew Resource Management (CRM) is to give crews tools to
    A. Recognize and mitigate hazards.
    B. Maintain currency with regulations.
    C. Reduce the need for outside resources.

15. When piloting a turbojet transport airplane, what is a possible result when operating at speeds 5-10 percent above the critical Mach number?
    A. Increased aerodynamic efficiency.
    B. Decreased control surface effectiveness.
    C. Occasional low speed Mach buffet warnings

C.3 Performance Assessment for an Unexpected Event Scenario

C.3.1 Preflight

Participant ID #_______

1. **Low:** No indication that a behavior is exhibited
2. **Medium:** It is clear that a behavior is exhibited but there is no discussion
3. **High:** Both pilots are clearly on the same page as evident by their communication

Behavior, Skills, Attitudes	Yes (Check if Yes)				Corresponding DV(s) of Interest
Does the crew make a decision about what to do with the terrain?					General Decision Making, Domain Expertise
Does the crew reference the chart to avoid terrain?					Domain Expertise
Was someone engaged at all times?					Domain Expertise, Task Management
Has the crew flown this path previously (consider expertise)?					Domain Expertise
Does the crew talk about the escape route (i.e., turn to head northeast)?					Teamwork, Communication
NOTAM:					General Decision-Making, Teamwork, Communication
Did the crew brief the NOTAM?					
Did the crew make a change in the FMS if needed?					
Convective Weather:	Yes	Low	Med	High	General Decision-Making, Task Management, Teamwork, Communication
Is there awareness about the weather?					
Is there management of the threat?					
Does the crew have a shared mental model about the threat?					
Is there a plan for mitigation?					
Communication (Notice how the crew establishes open communication):					Teamwork, Communication
Does everyone appear willing to speak up?					
Is the Captain open and approachable?					
Does the Captain ask questions?					
Does the Captain introduce themselves?					
Notes:					

C.3.2 Taxi

Behavior, Skills, Attitudes	Yes (Check if Yes)	Corresponding DV(s) of Interest
Is there appropriate prioritization given to flight and control?		Domain Expertise, General Decision Making
Pitch		
Bank		
Power		
Speed		
Is the crew communicating back and forth?		Teamwork, Communication
Is the crew looking for threats along the way?		Cognitive Flexibility, Domain Expertise
Is the crew carefully managing time?		Time Management
Notes:		

C.3.3 Takeoff

Behavior, Skills, Attitudes	Yes (Check if Yes)	Corresponding DV(s) of Interest
Is there appropriate prioritization given to flight and control?		Domain Expertise, General Decision Making
Pitch		
Bank		
Power		
Speed		
Did the crew show recognition to the malfunction?		Domain Expertise, Recognition
How long (in seconds) did it take the crew to recognize the malfunction. This should be measured by time of malfunction to time of recognition.	**Time:**	Domain Expertise, Recognition
Does the crew focus on the flight path as primary task?		Domain Expertise, Task Management
Does the crew communicate problem with each other?		Communication, Teamwork
Does the crew communicate problem to ATC?		Communication
Does the crew accurately correct altitude?		Domain Expertise
Does the crew focus on flying the plane instead of troubleshooting the gear after takeoff?		Cognitive Flexibility, General Decision-Making
Does the crew have a shared mental model (evident by discussion and communication)?		Teamwork, Communication
Is the crew communicating in a composed manner?		Self-efficacy, Communication, Assesses C in *Calm*
Does the crew assess their fuel?		Domain Expertise, General Decision-Making
Does crew identify appropriate checklists?		Domain Expertise
If emergency was declared, what was the reason why they declared an emergency?		General Decision Making
Notes:		

C.3.4 Climb

Behavior, Skills, Attitudes	Yes (Check if Yes)	Corresponding DV(s) of Interest
Is there appropriate prioritization given to flight and control?		Domain Expertise, General Decision Making
Pitch		
Bank		
Power		
Speed		
Does the crew focus on the flight path as primary task?		Domain Expertise, Task Management
Does the crew communicate what is going on with each other and take steps to correct any challenges?		Communication, Teamwork, *Outline*
Notes:		

C.3.5 Departure and Cruise

Behavior, Skills, Attitudes	Yes (Check if Yes)	Corresponding DV(s) of Interest
Is there appropriate prioritization given to flight and control?		Domain Expertise, General Decision Making
Pitch		
Bank		
Power		
Speed		
Does the crew focus on the flight path as primary task?		Domain Expertise, Task Management
Does the crew communicate problem with each other?		Communication, Teamwork
Does the crew accurately correct altitude?		Domain Expertise
Does the crew have a shared mental model (evident by discussion and communication)?		Teamwork, Communication
Is the crew communicating in a composed manner?		Self-efficacy, Communication
Does the crew assess their fuel?		Domain Expertise, General Decision-Making
Does crew identify appropriate checklists (example checklist above)?		Domain Expertise
Notes:		

C.3.6 Diversion and Descent

Behavior, Skills, Attitudes	Yes (Check if Yes)	Corresponding DV(s) of Interest
Is there appropriate prioritization given to flight and control?		Domain Expertise, General Decision Making
Pitch		
Bank		
Power		
Speed		
Does the crew communicate they have a plan?		Teamwork, Communication
Describe how the crew manages the plan:		
How did the crew manage any exceptional activities?		
Notes:		

C.3.5 Approach and Landing

Behavior, Skills, Attitudes	Yes (Check if Yes)	Corresponding DV of Interest
Is there appropriate prioritization given to flight and control?		Domain Expertise, General Decision Making
Pitch		
Bank		
Power		
Speed		
Completes end-of-flight tasks -Localizer deviation -Glideslope deviation -Airspeed fluctuation		Domain Expertise, Task Management
Notes:		

Instructor Assessment

Please rate on a scale of 1-100 how well you think each pilot performed overall (1 being the lowest score and 100 being the highest).

__________%

C.3.6 Post-Flight Debriefing Questions

To be completed by evaluator after simulation		Comments
Overall Assessment	• On a scale of 1-100, how would you rate your overall performance during this flight?	
Preflight	• Did you have experience with this same flight path? • How did decide how to prioritize tasks? • Talk about why you decided to choose that escape route. • What were you thinking when you identified the terrain?	
Taxi	• During taxi, did you feel like you and your crew were developing rapport and a shared mental model? • If not, how could this be improved?	
Cruise	• When did you notice the threat? • How did you decide to handle the threat? • What else were you thinking during this time?	
Descent	• How did you decide to handle any exceptional activities happening during the descent? • How did you and the crew decide to prioritize tasks and manage time?	
Approach and Landing	• Is there anything you would have done differently if you could go back and do the flight again? • What do you think made the flight successful?	

C.4 Behavioral Rating Scales

(This would be completed with reference to the task analysis.)

Event	Behaviors	Measure	Training Intervention Aspect being Measured
Landing Gear Fails to Retract	2a Recognizes anomaly	Reaction Time – in seconds Video – facial expression of surprise (1-5) Audio – verbalization of surprise (1-5)	
	2a Determines cause of anomaly	Video – searches for information (yes/no) Audio - verbalizes situation (yes/no)	
	2a Identifies anomaly	Video – identifies landing gear fails to retrace (yes/no) Calls out correct situation (yes/no)	
	2a Continues to monitor flightpath	Video – (yes/no)	
Multiple Alerts in View (for example, landing gear lights, antiskid, cabin altitude, configuration warning)	1. Recognizes alerts	Reaction Time – in seconds Video – facial expression of surprise (1- 5) Audio – verbalization of surprise (1-5)	
	2. Calls for checklist	Calls for correct checklist (yes/no)	
	3. Communicates situation to ATC	Number of phrases to ATC it takes to communicate the failure (1 – x) Number of separate transmissions (1 – x)	
	4. Requests holding to troubleshoot	(yes/no)	
	5. Completes appropriate checklist	(yes/no)	
	6. Makes decision to divert (Return to KLAS or alternate airport)	Requests landing at nearest suitable airport (KLAS) (yes/no/other airport) (time to decide)	
	7. Returns for landing	Completes flight to end of task Performance measures – localizer deviation, glideslope deviation, airspeed fluctuation	

C.5 Debriefing Questions

1) What happened on this flight?
2) What would you expect when the landing gear fails to retract?
3) What would you expect from multiple alerts (annunciations)?
4) Do you recall any write-ups in the maintenance logbook?
5) Did any of the events of the flight surprise you?
    a. Which ones?
6) How do you think you performed on this flight?
7) Do you remember the weather forecast for this flight?
8) If so, was anything surprising with the weather during the flight?
9) Looking back, would you have done anything differently?
    a. What?

C.6 Experimenters' Checklist for Real-Time Behavioral Markers

****Sim Unfreeze Time:_________________

1. The apparent anxiety level of the participant prior to flight:

Very Relaxed	Relaxed		Anxious	Very Anxious	N/A
1	2	3	4	5	6

Comments: ____

****Unexpected Event Time:____________________

2. Did participant recognize unexpected event? _____Yes _____No

Comments: ____

3. Did participant actively respond to failure? _____Yes _____No

Comments: ____

4. The apparent anxiety level of the participant:

Very Relaxed	Relaxed		Anxious	Very Anxious	N/A
1	2	3	4	5	6

Comments: ____

5. Facial expression of surprise was:

No Surprise	Somewhat	Very	Startled	N/A	
1	2	3	4	5	6

Comments: ____

6. Verbal expression of surprise was:

No Surprise	Somewhat	Very		Startled	N/A
1	2	3	4	5	6

Understatement?________________
Overstatement?_________________
Irony?__________________________

Comments: ____

7. Task (flying) interruption indication of surprise was:

None	Minor Deviations		Gross Deviations	Loss of Control	N/A
1	2	3	4	5	6

Comments: ____

8. Was ATC contacted regarding the unexpected event? _____Yes _____No

Comments: ____

9. Was an emergency declared? _____Yes _____No

Comments: ____

10. Was a diversion requested? _____Yes _____No

Comments: ____

11. ATC communications were:

Ambiguous	Broken		Adequate	Very Clear	N/A
1	2	3	4	5	6

Comments: ____

12. Flight outcome was:

Disaster	Near Disaster		Safe	Impeccable	N/A
1	2	3	4	5	6

Comments: ____

****Experiment End Time:______________

C.7 Post Analysis Behavioral Markers

1. The apparent anxiety level of the participant prior to flight:

Very Relaxed	Relaxed		Anxious	Very Anxious		N/A
1	2	3	4	5	6	

2. Did participant recognize the unexpected event? ______Yes ______No

Comments:____

(Reaction Time:____________)

a. Reaction to the gear failed to retract was:

Subtle					Overt	N/A
1	2	3	4	5	6	

b. Reaction to the alerts was:

Delayed					Timely	N/A
1	2	3	4	5	6	

Comments:____

3. Did participant actively respond to failure? ______Yes ______No

(Reaction Time:____________)

a. Response to landing gear failed to retract was:

Subtle					Overt	N/A
1	2	3	4	5	6	

b. Response to multiple alerts was:

Delayed					Timely	N/A
1	2	3	4	5	6	

Comments:____

4. The apparent anxiety level of the participant after event was:

Very Relaxed	Relaxed		Anxious	Very Anxious	N/A
1	2	3	4	5	6

Comments:____

5. Did participant recognize the unexpected event? ______Yes ______No

(Reaction Time:___________)

a. UEE recognition was:

Subtle					Overt	N/A
1	2	3	4	5	6	

b. UEE recognition was:

Delayed					Timely	N/A
1	2	3	4	5	6	

(Reaction Time:___________)

c. UEE expression of recognition was:

No surprise					Startled	N/A
1	2	3	4	5	6	

No fear					Frightful	N/A
1	2	3	4	5	6	

No frustration					Frustration	N/A
1	2	3	4	5	6	

Comments:____

6. Facial expression of surprise was (see examples below):

No Surprise	Somewhat		Very	Startled	N/A
1	2	3	4	5	6

Comments:____

7. Verbal expression of surprise was:

No Surprise	Somewhat		Very	Startled	N/A
1	2	3	4	5	6

a. Understatement of surprise was:

Subtle ______________________________ Overt N/A

1 2 3 4 5 6

b. Overstatement of surprise was:

Subtle ______________________________ Overt N/A

1 2 3 4 5 6

c. Irony of surprise was:

Subtle ______________________________ Overt N/A

1 2 3 4 5 6

d. Participant's actions were:

Haphazard ______________________________ Purposeful N/A

1 2 3 4 5 6

Comments:____

8. Task (flying) interruption indication of surprise was:

None ______ Minor Deviations ______ Gross Deviations ______ Loss of Control N/A

1 2 3 4 5 6

Comments:____

9. Was ATC contacted regarding UEE? ______Yes ______No

Comments:____

10. Was an emergency declared? ______Yes ______No

Comments:____

11. Was a diversion requested? ______Yes ______No

Comments:____

12. ATC communications were:

Ambiguous Broken ______________ Adequate Very Clear N/A

1 2 3 4 5 6

Comments:____

13. Flight outcome was:

Disaster		Near Disaster		Safe	Impeccable	N/A
1	2	3	4	5	6	

Comments:____

C.7.1 Examples of Faces of Surprise

(a) Anger (b) Disgust (c) Fear (d) Happiness (e) Sadness (f) Surprise

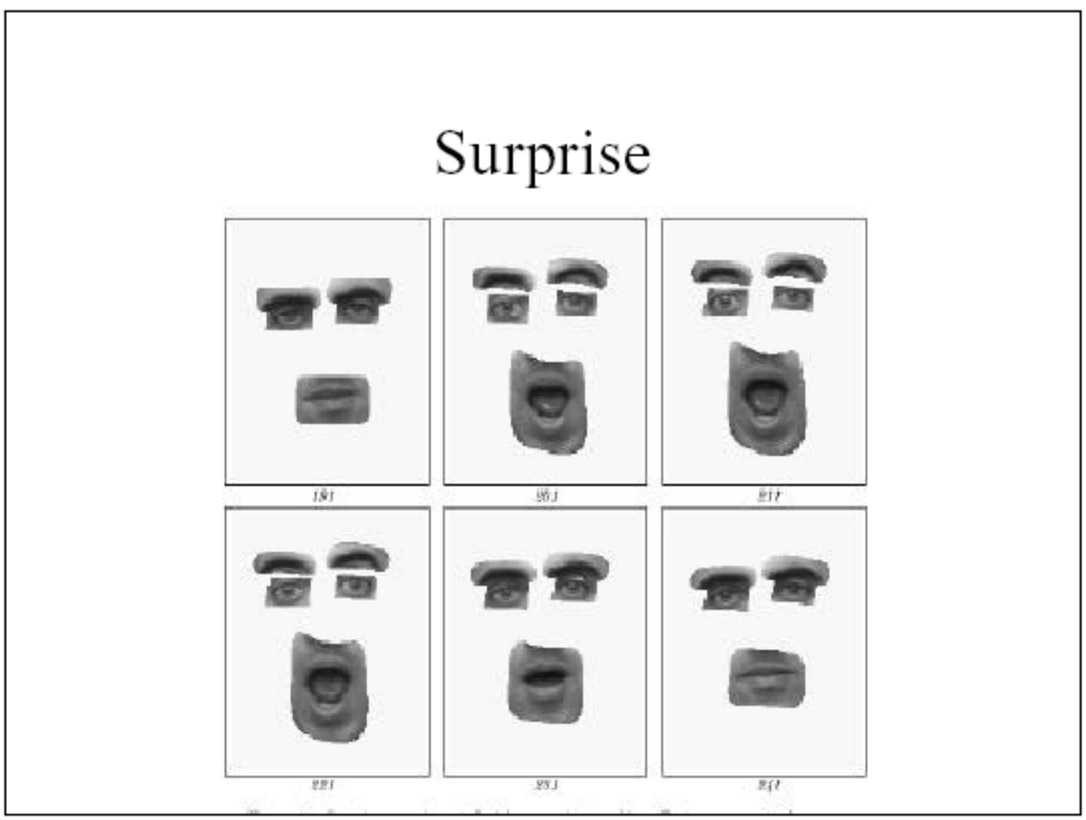

Figure 1: (a) Example prototypical expressions of six basic emotions and a neutral face for actor "JJ" in Ekman and Friesen's Pictures of Facial Affect (POFA) (Ekman and Friesen, 1976). (b) Morphs from fear to sadness and happiness to disgust, generated from the corresponding prototypes.