DOT/FAA/AM- 26/07

Aviation Safety

Office of Aerospace Medicine

Washington, D.C. 20591

Prioritizing Safety-Critical Information in the National Airspace System: A Four-Phased Human Factors Methodology and Its Future Applications

Author

Chanda S. Sanders

Civil Aerospace Medical Institute (CAMI)

Federal Aviation Administration

Oklahoma City, OK 73125

March 2026

NOTICE

This publication and all Office of Aerospace Medicine technical reports are available in full text from the Civil Aerospace Medical Institute's publications website and at the National Transportation Library's Repository & Open Science Access Portal.

Technical Report Documentation Page

<table>
<tr><td>1. Report No.

DOT/FAA/AM-26/07</td><td colspan="3">2. Title & Subtitle

Prioritizing Safety-Critical Information in the National Airspace System: A Four-Phased Human Factors Methodology and Its Future Applications</td></tr>
<tr><td colspan="2">3. Report Date

March 2026</td><td colspan="2">4. Performing Organization Code

AAM-520</td></tr>
<tr><td colspan="2">5. Author(s)

C. S. Sanders</td><td colspan="2">6. Performing Org Report Number

N/A</td></tr>
<tr><td colspan="2">7. Performing Organization Name & Address

Civil Aerospace Medical Institute (CAMI), FAA Oklahoma City, OK 73169</td><td colspan="2">8. Contract or Grant Number

N/A</td></tr>
<tr><td colspan="2">9. Sponsoring Agency Name & Address

NextGen Human Factors Division (ANG-C1)</td><td colspan="2">10. Type of Report & Period Covered

Technical Report</td></tr>
<tr><td colspan="4">11. Supplementary Notes

N/A</td></tr>
<tr><td colspan="4">12. Abstract

This report describes a methodology for sorting and prioritizing safety-critical aeronautical information, developed in response to a National Transportation Safety Board recommendation. The methodology integrates human factors and risk assessment principles to identify systemic vulnerabilities in information management and to align information delivery with operational, cognitive, and contextual demands. The approach applies the bow-tie risk model to represent human factors constructs as threats that weaken preventive barriers. Information characteristics, including volume, relevance, timeliness, and modality, are modeled as drivers of these threats and are explicitly linked to preventive and mitigative controls. The resulting framework supports operationally realistic filtering, sequencing, and delivery strategies. The methodology is executed in phased activities, including expert knowledge elicitation, scenario-based simulation, and development and validation of a decision support tool. Certified professional controllers and other operational roles complete realistic scenarios varying in complexity, traffic load, and environmental conditions. Data include event-linked performance metrics, post-scenario interviews, and standardized measures of workload, Situation Awareness, and trust. Integrated quantitative and qualitative analyses identify patterns in information use, decision making, and operational outcomes. Outputs include evidence-based recommendations for training, interface design, and policy and procedural improvements to reduce operational risk and support resilient operations. The scope is limited to the contiguous United States, with future research recommended for non-contiguous regions.</td></tr>
<tr><td colspan="2">13. Key Words

safety-critical information, contextual demands, standardized measures</td><td colspan="2">14. Distribution Statement

Document is available to the public through the National Transportation Library: http://www.faa.gov/go/oamtechreports/</td></tr>
<tr><td>Security Classification (of this report)
Unclassified</td><td>Security Classification (of this page)
Unclassified</td><td colspan="2">15. No. of Pages

122</td></tr>
</table>

Acknowledgments

This research was completed with funding from the Federal Aviation Administration (FAA) NextGen Human Factors Division (ANG-C1) in support of ATO Safety and Technical Training (AJI-152). The author wishes to express appreciation to numerous individuals for their contributions throughout the development and execution of this work. From AJI-152: Phillip Russ, whose steady engagement and patience sustained the effort despite a number of constraints, including the extended time required to address this complex topic. His willingness to provide information, share insight and resources, and consistently champion the project ensured that it achieved its full potential. From the NAS Human Factors Safety Research Laboratory (AAM-520): Dr. Jerry Crutchfield, Dr. Carla Hackworth, Tracy Streagle, MSA, and Dr. Braden Tanner. From the FAA Safe Operations in Aviation Research (SOAR) Lab (AAM-510): Dr. Katrina Avers, for guidance and support in recognizing the way an initial workbook of variables could be developed into a data collection instrument, serving as the foundation for a structured, four-phased methodology. Special appreciation is extended to Dr. Anthony Tvaryanas from the Aerospace Medical Research Division (AAM-600), whose expert review and feedback on the risk framework provided essential grounding for the methodology.

This methodology was reviewed and refined with the support of generative AI tools, including OpenAI’s ChatGPT, which was used to assist with reviewing, editing, and synthesizing technical content. All content was reviewed, validated, and refined by the lead researcher to ensure accuracy and alignment with project objectives.

Table of Contents

1. Executive Summary 1
2. Introduction 1
3. Purpose and Relevance 2
    3.1. Objectives 3
    3.2. Anticipated Benefits 3
    3.3. Strategic Relevance 3
4. Scientific and Critical Foundations Informing Methodology 3
    4.1. Safety-Critical Information 3
    4.2. Framing Human-Centered Risk Regarding Aeronautical Information 5
        4.2.1 Volume 5
        4.2.2 Relevance 6
        4.2.3 Timeliness 6
    4.3. The Modality Challenge in Safety-Critical Environments 7
    4.4. Human Factors and Aeronautical Information 8
        4.4.1 Cognitive Integration Complexity 8
        4.4.2 Modal Load Compatibility 9
        4.4.3 Attentional Demand Type 9
        4.4.4 Actionability and Coordination Burden 10
    4.5. Situation Awareness: A Foundational Performance Construct 10
    4.6. Shared Situation Awareness 12
    4.7. The Role of Trust in Interpreting Safety-Critical Information 13
5. Risk Perception and Air Traffic Control 15
    5.1. The Role of Risk Perception in Operational Decision-Making 15
        5.1.1 Interaction with Information Characteristics 15
    5.2. Trust and Risk: A Critical Interaction 16
        5.2.1 Implications for Safety-Critical Information Design 16
    5.3. The Swiss Cheese Model and Human Factors in ATC 17
        5.3.1 Latent and Active Failures 18
        5.3.2 The Swiss Cheese Model and the Analysis of Risk 19
        5.3.3 From Swiss Cheese to Bow-Tie: Structuring Vulnerabilities and Controls 20
6. The Bow-Tie Model 20
    6.1. Human Factors in the Bow-Tie 21
        6.1.1 A Human Factors Framework 22
        6.1.2 Cognitive Workload in the Bow-Tie Pathway 23
        6.1.3 Situation Awareness in the Bow-Tie Pathway 24

6.1.4 Shared Situation Awareness in the Bow-Tie Pathway 25
6.1.5 Trust in Information Source in the Bow-Tie Pathway 27
6.2. Contextual Complexity in the Bow-Tie Model 28
6.2.1 Factors Driving Contextual Complexity in the NAS 29
7. Methodology 32
7.1. Sampling Philosophy 34
7.2. Phase I: Establishing the Empirical Foundation 35
7.2.1 Research Questions 36
7.2.2 Participants 36
7.2.3 Data Collection Instruments 36
7.2.4 Data Analysis 38
7.3. Phase II: Simulation Testing and Assessment of Human Factors Threats 39
7.3.1 Research Questions 39
7.3.2 Hypotheses 39
7.3.3 Participants 40
7.3.4 Independent Variables 40
7.3.5 Measures 42
7.4. Phase III: Development and Validation of the Decision-Support Framework 57
7.4.1 Research Questions 57
7.4.2 Hypotheses 58
7.4.3 Participants 58
7.4.4 Independent Variables 58
7.4.5 Measures 59
7.4.6 Data Analysis 59
7.5. Phase IV: Implementation Planning and System Integration 61
7.5.1 Recommendations for Phase IV Integration 62
8. References 66
9. Appendix A. NTSB Recommendation A-18-024 76
9.1. FAA Response to NTSB Safety Recommendation A-18-24 77
9.2. FAA Further Response to NTSB Safety Recommendation A-18-24 79
10. Appendix B: Example Types of Aeronautical Information 81
11. Appendix C: Background Resources Reviewed 82
12. Appendix D: Contextual Complexity Dimensions 88
12.1. Information Characteristics 88
12.2. Contextual Complexity 90
13. Appendix E: Sample Size Approach 99

13.1. Sample Size Formulas....99
13.2. Sample Size Calculations....100
14. Appendix F: Phase I Data Collection Instruments....101
14.1. Decision Support Variables....101
14.1.1 Criticality of Information 101
14.1.2 Clarity/Interpretability of Information 101
14.1.3 Preventive Value of Information 102
14.1.4 Scope of Operational Impact 103
14.1.5 Severity of Consequences 103
14.1.6 Magnitude of Change Needed 104
14.1.7 Systemic Scope of Impact 104
14.1.8 Time Sensitivity 105
14.2. Structured Interview Protocol Example....106
15. Appendix G: Measures....107
15.1. Cognitive Workload....107
15.2. Situation Awareness....110
15.3. Shared Situation Awareness....111
15.4. Trust in Information Source....113
15.5. Perceived Usefulness of Information....114
15.6. Performance....115
15.7. Mental Models & Decision Strategies....116
15.8. Risk Perception....117

List of Figures

Figure 5-1 The "Swiss Cheese" Model of Human Error Causation 18
Figure 6-1 Complete Bow-Tie Model 21
Figure 6-2 Risk Pathway for Cognitive Workload 24
Figure 6-3 Risk Pathway for Situation Awareness 25
Figure 6-4 Risk Pathway For Shared Situation Awareness 27
Figure 6-5 Risk Pathway for Trust in Information Source 28
Figure 7-1 Proposed Study Phases 33

List of Abbreviations

Acronym	Abbreviation Explained
ARTCC	Air Route Traffic Control Centers
ATC	Air Traffic Control
CERAP	Combined Center Radar Approach
CPDLC	Controller Pilot Data Link Communication
FAA	Federal Aviation Administration
ICAO	International Civil Aviation Organization
NAS	National Airspace System
NOTAM	Notice to Airmen
NTSB	National Transportation Safety Board
PIREP	Pilot Report
PSSUQ	Post-Study System Usability Questionnaire
SA	Situation Awareness
SCI	Safety-Critical Information
SCM	Swiss Cheese Model
SIGMET	Significant Meteorological Information
SME	Subject Matter Expert
SMS	Safety Management Systems
SSA	Shared Situation Awareness
TFR	Temporary Flight Restriction
TRACON	Terminal Radar Approach Control

1. Executive Summary

This paper presents a methodology designed to develop an evidence-based framework for the sorting, prioritizing, and delivery of Safety-Critical Information (SCI) to operational personnel in the National Airspace System (NAS). Prompted by a recommendation from the National Transportation Safety Board (NTSB, 2018a, 2018b) to gain insight into human factors considerations related to the management of aeronautical information with regard to volume, relevance, and timing after a near-miss by Air Canada 759 in 2017, the proposed methodology integrates human factors principles and risk assessment concepts to identify systemic threats related to the handling of aeronautical information.

This proposed work acknowledges that excessive, ambiguous, or poorly timed information can contribute to operator workload and increase the potential for errors. To gain insight into these potential errors, this methodology proposes leveraging Reason's Swiss Cheese Model (1990) and an adapted version of the bow-tie risk framework (CCPS, 2020a; 2020b) to map failure pathways in the air traffic control (ATC) operating environment and utilize the sorting and prioritization of SCI as a key safety barrier within that framework.

To that end, this methodology is intentionally structured as a multi-phase effort within the ATC workforce that begins with domain knowledge elicitation, moves through simulation-based testing, and culminates in the development and validation of a decision-support framework. Although this document proposes research that includes positions related to ATCs, it is recommended to conduct the research using both pilot and ATCs to ensure that the resulting framework is operationally relevant, human-centered, and scalable across FAA lines of business. The recommended approach not only supports the development of technically sound solutions but also advances a model for collaborative safety research within government that recognizes the complexity of shared responsibility and the value of integrated operational insight. It is important to note that this proposal has been scoped to the contiguous United States (lower 48 states). While the findings offer actionable insights into the delivery of SCI within this domain, additional research is necessary to account for operational conditions unique to other regions, such as Alaska, Hawaii, and U.S. territories. These areas present distinct geographic, environmental, and airspace characteristics that may affect SCI communication practices.

Ultimately, this research aims to develop and propose a validation framework for sorting, prioritizing, and delivering safety-critical information for ATCs that accounts for cognitive and contextual demands, strengthens preventive barriers, and supports safer, more resilient NAS operations.

2. Introduction

Following a near-miss involving Air Canada Flight 759 at San Francisco International Airport on July 7, 2017, an investigation from the NTSB revealed that the Notice to Airmen (NOTAM) indicating the closure of runway 28L was embedded within a large volume of non-critical NOTAMs, which was either overlooked or misinterpreted, contributing to confusion about the active runway. Following the investigation, the NTSB issued Recommendation A-18-24, urging the FAA to manage persistent deficiencies in the ways safety-critical aeronautical information is communicated to pilots. The FAA has remained diligent in working to address this recommendation and engaged in multiple communications regarding elements of the ongoing research plans (See Appendix A). This incident was neither unprecedented nor isolated, and similar events have occurred in which crucial flight information, such as runway closures, hazardous weather, and other safety-relevant NOTAM content, was overlooked, misunderstood, or received too late to support safe operational decisions (NTSB, 2001a; 2001b; 2013; 2016; 2018c).

These recurring issues underscore a systemic need to improve the management, prioritization, and delivery of safety-critical aeronautical information to operational personnel.

Despite substantial efforts by the FAA and many collaborators to improve the NOTAM system, aeronautical information management more broadly has continued to present challenges, with multiple high-profile events underscoring ongoing risks. The proposed research methodology aims to address these challenges through the development of an evidence-based framework designed to more efficiently prioritize relevant information for the controller to present at a time and in a format that is operationally usable. The approach draws from human factors principles and aviation safety research and seeks to build an information decision-support framework that aligns with modern aviation operations. Ultimately, the goal is to ensure the right information reaches the right personnel, at the right time, and in the right format, supporting timely action and advancing the safety and resilience of the NAS.

Although this study is intentionally scoped to exclusively focus on how air traffic controllers interact with safety-critical information, the broader operational reality is that this information is jointly processed, prioritized, and acted upon across multiple NAS roles, including pilots, dispatchers, and flight service specialists. The ATC-focused scope presented here reflects project funding, access, and research design constraints, not an assumption of sole responsibility. This methodology is intentionally designed to be modular and extensible, supporting future adaptation for cross-role applications. This foundational work establishes the structure needed for later phases or complementary studies to examine and validate cross-role interoperability, shared information risk pathways, and domain-wide relevance across the NAS.

3. Purpose and Relevance

The safe and effective operation of the NAS depends on the timely delivery and accurate interpretation of safety-critical aeronautical information by ATCs, pilots, and other operational personnel. Current processes for prioritizing, presenting, and communicating such information can be challenged by workload, operational tempo, and environmental complexity. These challenges can lead to delays or misinterpretations, reducing the ability of distributed teams to maintain a shared and accurate understanding of the situation. This proposed methodology aims to develop and validate a framework for evaluating how safety-critical information is prioritized and delivered under realistic operational conditions, with the goal of informing policies, training, and a decision-making framework that improves safety. For the purposes of this research, the operational environment is limited to the contiguous United States. While SCI delivery processes are critical across the entire NAS, this study does not include analysis of regions outside the lower 48 states. In particular, Alaska and other non-contiguous areas present unique logistical and operational challenges that warrant dedicated research initiatives.

3.1. Objectives

1. Identify and evaluate key factors that influence the timely and accurate delivery of safety-critical information in the NAS.
2. Examine the impact of human factors elements on operator decision-making with regard to aeronautical information.
3. Develop an evidence-based framework to support FAA in sorting, prioritizing, and delivering aeronautical information.

3.2. Anticipated Benefits

Research using the proposed methodology will produce an evidence-based, replicable framework for more effective handling of safety-critical information across the NAS. This work is expected to generate actionable insights for identifying and mitigating communication-related factors that contribute to delays, overload, or missed cues, which impact operator decision-making. Together, these outcomes will strengthen communication reliability, enhance coordination across roles, and inform strategies to mitigate systemic risk while informing future policy, training, and system design.

3.3. Strategic Relevance

This research supports FAA strategic priorities by directly addressing human factors risks that can compromise safety in the NAS. The outcomes will help align operational practices with evolving system capabilities, regulatory requirements, and safety mandates, while enabling more resilient responses to rapidly changing operational conditions.

4. Scientific and Critical Foundations Informing Methodology

This methodology proposes a four-phased, scenario-based experimental design with knowledge elicitation as the foundational step. Input from certified professional controllers will provide insight into critical safety-related elements and guide the development of realistic scenarios, which will then be tested in human-in-the-loop simulations with varying levels of complexity, traffic load, and environmental conditions to address key research objectives. Data collected through objective performance measures, structured post-scenario interviews, and standardized instruments assessing workload, situation awareness, and trust will be analyzed to reveal patterns in information use, decision-making, and performance outcomes.

4.1. Safety-Critical Information

Safety-critical information refers to any aeronautical, environmental, operational, or procedural information that, if not accurately received, clearly understood, and appropriately acted upon in a timely manner, has the potential to contribute to an unsafe condition, degrade operational safety margins, or directly result in harm to aircraft, crew, passengers, or persons and property on the ground (Rausand, 2013; CCPS, 2020a; 2020b). Because SCI is highly dependent on operational context, it is difficult to establish strict rules for why one piece of information may be more important than another at a given time. For example, dissemination and distribution of hazardous weather information may only be deemed "safety-critical" to those NAS users who are currently operating in airspace where hazardous weather is occurring.

While such information is used by a variety of aviation personnel, including pilots, dispatchers, and safety analysts, the current research focuses exclusively on the ATC's role in receiving, interpreting, and acting on safety-critical information.

For the purposes of this research, safety-critical information is defined as an operational human factors construct, rather than a formally codified FAA regulatory term for ATC communication. Within the ATC operational context, safety-critical information includes any data that (1) reflects current or emerging hazards (e.g., closed or contaminated runways, hazardous weather, or inoperative safety-related equipment); (2) is intended support the prevention, mitigation, or recovery from operational risk events (e.g., loss of separation, runway incursions); and (3) requires timely adaptation or coordination by controllers to maintain operational safety. Examples relevant to this study include NOTAMs and advisories indicating runway closures, unserviceable navigation aids, construction activity, temporary procedural changes, or airspace restrictions; weather alerts or Pilot Reports (PIREPs) describing hazardous meteorological conditions such as severe turbulence or low visibility.

Information that is purely administrative, outdated, redundant, or unrelated to active operational hazards is explicitly excluded from this operational definition. For example, scheduled maintenance notices for non-operational systems or equipment (e.g., an inactive VOR or backup lighting systems) that do not impact a current flight or ATC procedures are not considered safety-critical, even when formally disseminated through official channels.

This definition is grounded in FAA safety guidance and established risk and human factors frameworks and emphasizes both the content of the information and its timeliness, clarity, and interpretability as key factors in determining safety criticality (FAA, June 9, 2025; FAA, February 11, 2025; Endsley, 1995; Wickens et al., 2015). In high-stakes, time-constrained environments like air traffic control, effective delivery and prioritization of safety-critical information are essential for supporting safe and efficient decision-making. Aeronautical information spans a wide range of formats, such as NOTAMs, PIREPs, SIGMETs, and TFRs, and varies in urgency, specificity, and operational impact. Its relevance is heavily dependent on the controller's role, airspace sector, traffic configuration, and temporal demands. See Appendix B for an example list of aeronautical information.

Aeronautical information supports decision-making across several operational roles, including pilots, dispatchers, and safety analysts, but ATCs are uniquely responsible for synthesizing and delivering this information under high cognitive demand. Controllers must continuously review, interpret, and act on incoming data streams during dynamic, high-pressure situations. This information must often be retained and recalled under significant time constraints and workload. Although challenges such as dense, poorly prioritized information affect many user groups, this research focuses exclusively on the ATC's role in processing, prioritizing, and applying safety-critical information. Without effective methods to highlight the most relevant data, controller workload increases, situation awareness can degrade, and operational safety may be compromised. (Bainbridge & Dorneich, 2016; Loft et al., 2007; Roth et al., 2021).

To better develop an understanding of the breadth of constraints related to the complex contextual environment, and to gain a greater understanding of some of the complexities related to users involved, background resources across a number of domains were compiled and reviewed (as seen in Appendix C) to ground the methodology.

4.2. Framing Human-Centered Risk Regarding Aeronautical Information

Before introducing the formal risk assessment framework, it is important to examine key characteristics of aeronautical information – specifically, volume, relevance, timeliness, and modality. These upstream factors shape how information is perceived, processed, and used in operational settings. Their influence manifests through human performance constructs such as cognitive workload, situation awareness, shared situation awareness, and trust in information sources (Endsley, 1995; Wickens et al., 2015; Lee & See, 2004).

Rather than describe the effects of these factors in isolation, the following sections trace how they interact with core constructs that mediate risk perception and decision quality. Volume, relevance, timeliness, and modality are treated as input features whose impact depends on task demands, operator goals, and cognitive constraints. Where relevant, examples of their effects are integrated within each construct's discussion, laying a foundation for the risk framework presented in the subsequent section.

4.2.1 Volume

ATCs must continuously manage a high volume of incoming information from multiple, often competing, sources – including pilot communications, weather advisories, radar data, flight plans, and system-generated alerts. While expanded access to information can enhance decision-making, excessive or poorly structured inputs may overwhelm cognitive resources and impair performance, especially during time-sensitive or high-tempo traffic conditions (Endsley, 1995; Wickens et al., 2015). Under such constraints, controllers may rely on heuristics or attend selectively to only a subset of cues (Sweller, 1988; Sweller et al., 2011; Todd & Gigerenzer, 1999), increasing the risk of overlooking or misprioritizing safety-critical elements.

Task complexity and time pressure further strain working memory and increase the chances that relevant information will be forgotten, missed, or misapplied (Baddeley, 2012; Loft et al., 2007). Even when operational data are technically accurate and available, their utility may be reduced if embedded in lengthy, non-prioritized content streams. These conditions not only degrade the accuracy of the individual controller's mental model of the traffic situation but may also disrupt coordination across controller teams or shifts, where shared understanding is essential to maintaining continuity of operations.

Persistent exposure to overwhelming or poorly filtered information can also erode trust in the reliability of information systems or human sources. When controllers experience repeated difficulty discerning relevance or clarity in critical updates, they may begin to disregard or discount important inputs (Lee & See, 2004). This undermines both confidence in information sources and the operational effectiveness of communication tools designed to support real-time decision-making.

4.2.2 Relevance

The relevance of aeronautical information strongly influences whether ATCs notice, process, and act upon it. Controllers must continuously filter incoming data, ranging from radar displays and flight progress strips to weather advisories and pilot communications, based on their current sector responsibilities, operational priorities, and phase of traffic flow (Endsley, 2015; Endsley & Jones, 1997). Information that is tangential or poorly aligned with immediate operational demands increases cognitive burden, making it more difficult for controllers to detect and prioritize safety-critical cues, such as active runway closures, urgent weather alerts, or other relevant and potentially critical advisories. Within high-tempo conditions, irrelevant updates may clutter working memory or lead to inattentional blindness, in which subtle but critical threats are overlooked in favor of more salient yet less relevant data (Wickens et al., 2015).

Relevance also directly influences coordination across ATC teams. When controllers working adjacent sectors or facilities receive information subsets that are mismatched in scope or operational utility, shared situation awareness can deteriorate. Misalignments in how relevance is interpreted across roles, such as when one controller emphasizes sector-specific traffic while another prioritizes broader flow management, can lead to inconsistent responses and reduced efficiency in collaborative decision-making during time-sensitive operations.

Over time, repeated exposure to low-relevance or off-target updates may erode trust in information sources. If controllers come to view updates as excessive, tangential, or misaligned with their operational context, they may begin discounting subsequent inputs, even those that are urgent or critical (Lee & See, 2004). This erosion of credibility poses significant risks in safety-critical environments, where timely use of relevant information is essential to preventing errors and maintaining operational safety.

4.2.3 Timeliness

In air traffic control operations, timeliness refers to how well the delivery of information aligns with the moment it is most needed. This temporal alignment is critical for effective information uptake, especially in the high-tempo, time-compressed environments of the NAS where hazards evolve quickly and priorities can shift without warning (Klein, 1993). When safety-critical updates arrive too late in the decision cycle, they may be obsolete or overtaken by new developments. Conversely, information delivered too early – before it can be acted upon – risks being forgotten, deprioritized, or overshadowed by more immediate stimuli (Wickens et al., 2015).

Controllers managing multiple data streams under high task load may not have sufficient attentional resources to process even relevant updates if they arrive during peak workload intervals (Endsley, 2017). Information introduced at the wrong time can result in delayed action, decision inertia, or misinterpretation – particularly when the controller has already shifted attention to another task or traffic flow issue (Endsley et al., 2003; Endsley & Conners, 2008; Endsley & Garland, 2000). Even technically accurate content becomes functionally ineffective if delivered outside of the controller’s optimal processing window.

Poor timing can also disrupt the controller’s ability to form and update an accurate mental model of the traffic situation. If critical information is delivered too late, it may fail to be integrated into the controller’s working understanding of aircraft trajectories, sector status, or coordination points. This limits the ability to anticipate system changes and increases reliance on reactive decision-making, which is more error-prone under pressure (Endsley, 1995).

At the team level, the effects of mistimed communication compound. Shared situation awareness across ATC positions – within and between facilities – depends on the near-simultaneous delivery and interpretation of information. Asynchrony caused by delays, workload interference, or inconsistent update mechanisms can result in divergent mental models, leading to communication breakdowns or coordination errors during high-tempo operations (Orasanu et al., 2011).

Timeliness also plays a significant role in shaping trust. When controllers repeatedly receive updates that are poorly timed – too late, too early, or misaligned with their operational rhythm – the perceived reliability of the information source can erode (Lee & See, 2004). Even when future messages are well-timed and accurate, prior experiences with untimely delivery can lead to them being discounted, ultimately undermining the safety benefit of the information delivery system.

4.3. The Modality Challenge in Safety-Critical Environments

While volume, relevance, and timeliness determine whether SCI is noticed, prioritized, and retained, modality shapes how efficiently that information is perceived and acted upon. Together, these properties form the foundation for managing incoming data in dynamic air traffic control environments. When mismatches occur across these characteristics – such as excessive volume, low relevance, poor timing, or inappropriate modality – critical cues may be overlooked or misinterpreted, undermining both individual performance and inter-controller coordination (Endsley, 1995; Wickens et al., 2015).

Controllers frequently operate across multiple sensory channels, particularly during high-tempo or off-nominal events. For example, a weather reroute may appear as a visual update on a scope, be accompanied by an auditory alert, and require verbal coordination with adjacent sectors – all within moments. If these streams are poorly synchronized or differ in priority framing, the controller must resolve conflicting inputs under time pressure, which can shape how risk is perceived, communicated, and managed. Misalignment across modality and timing may contribute to information asymmetry and decision-making delays.

The modality used to deliver SCI – whether visual, auditory, tactile, or multimodal – directly impacts how well controllers perceive, integrate, and respond to information in operationally constrained settings. Each channel presents unique strengths and limitations:

- **Visual** channels are effective for displaying persistent or structured data (e.g., radar displays, data blocks, status panels) but can quickly become saturated. During peak visual demand, adding more visual input may delay cue recognition or degrade responsiveness (Wickens, 2008).

- **Auditory** cues, such as radio calls, sector coordination, or aural alerts, are often better suited for urgent or time-sensitive updates when visual channels are heavily engaged. However, high-radio-traffic conditions can mask auditory cues or cause delays in message parsing (Endsley, 2021; Patterson, 1982).
- **Tactile** modalities, while less commonly employed in current ATC systems, have potential for rapid, intuitive alerts that bypass visual and auditory load. Examples might include seat or armrest vibration to indicate system failures or urgency conditions. Their success depends on controller familiarity, standardization, and minimal latency (Van Erp & Van Veen, 2004; Jones & Sarter, 2008).
- **Multimodal** delivery involves presenting information simultaneously across two or more channels (e.g., combining visual display changes with auditory tones), and can enhance salience, redundancy, and detection, especially when one channel is overloaded or compromised. However, for controllers under time constraints, multimodal inputs must be tightly aligned in priority and timing to prevent confusion or cognitive overload (Wickens, 2008; Drijvers & Holler, 2023).

For ATCs, the delivery modality of SCI interacts with task load, attention, and operational context. In this research framework, modality is treated as a critical factor influencing information usability and decision effectiveness under real-world ATC conditions.

4.4. Human Factors and Aeronautical Information

The interpretation and handling of safety-critical aeronautical information places considerable cognitive demands on ATCs, especially during high-tempo operations where safety margins are narrow and real-time decision-making is essential. In such environments, human factors play a critical role in shaping how information characteristics – such as volume, relevance, timing, and modality – interact with cognitive constraints. These interactions influence how information is perceived, prioritized, and acted upon under operational pressure.

To systematically examine these dynamics, this study applies an adapted framework derived from Wickens' (2008) four workload dimensions. The framework has been recast into constructs tailored for the air traffic control domain: Cognitive Integration Complexity, Modal Load Compatibility, Attentional Demand Type, and Actionability & Coordination Burden. These dimensions serve as the foundation for evaluating how different attributes of safety-critical information affect controller performance, particularly in relation to information sorting, shared understanding, and timely action.

4.4.1 Cognitive Integration Complexity

Adapted from Wickens' Perceptual-Cognitive Demands dimension, Cognitive Integration Complexity captures the mental effort required for ATCs to interpret, synthesize, and act on SCI in real time.

High-volume or marginally relevant SCI elevates cognitive demand, particularly under time pressure. Even accurate messages may lose operational value if they require excessive mental effort to reconcile with existing expectations or procedural knowledge. This is especially problematic when messages include ambiguous language, nested conditions (e.g., "Only applies when Runway 10L is active"), or require cross-referencing with charts or flight strips – issues often found in legacy NOTAM formats.

Controllers are especially vulnerable to these burdens during high-stakes operations such as weather diversions, reroutes, or emergency management. In these moments, information must be both relevant and immediately interpretable. Pre-filtered summaries or simplified contextual framing can mitigate risks associated with delayed or missed comprehension (Durso et al., 2017).

4.4.2 Modal Load Compatibility

Adapted from Wickens' Sensory Modalities dimension, Modal Load Compatibility examines whether the delivery mode of SCI – visual, auditory, tactile – is appropriate for the controller's current perceptual availability.

Controllers rely heavily on visual channels (e.g., radar scopes, flight plans), which can become saturated, especially in high-density traffic sectors. Simultaneously, auditory inputs such as radio transmissions or aural alarms are susceptible to congestion or masking, particularly during peak traffic periods. Tactile cues – though less common in current ATC systems – offer promise for bypassing overloaded channels but require standardization and controller training to be operationally viable (Van Erp & Van Veen, 2004; Jones & Sarter, 2008).

Delivery methods must match real-time workload. For example, low-priority updates might be deferred or sent through less-demanded modalities, while time-critical alerts may benefit from redundant multimodal delivery – e.g., simultaneous visual and auditory alerts – to ensure uptake without introducing new cognitive burdens.

4.4.3 Attentional Demand Type

Based on Wickens' Processing Channels dimension, Attentional Demand Type categorizes SCI by the level of attention it demands – passive monitoring, active review, or immediate interruption.

This construct is especially relevant for timing: effective SCI delivery requires not only relevant content but also awareness of the controller's current attentional bandwidth. For example:

- Low-urgency data (e.g., sector weather outlooks) may only require ambient awareness.
- Moderate-priority information (e.g., coordination requests or holding instructions) demands focused review.
- High-priority alerts (e.g., urgent reroutes, in-flight emergencies) require immediate attentional redirection.

SCI that arrives misaligned with task demands can disrupt attentional flow, increase reaction time, or result in overlooked cues. Proper classification and scheduling of alerts – based on urgency, operational phase, and sector complexity – supports more precise attention management and reduces unnecessary cognitive switching (Endsley, 2017; Klein, 1993; Gutzwiller & Clegg, 2013).

4.4.4 Actionability and Coordination Burden

Adapted from Wickens' Response Types dimension, Actionability and Coordination Burden addresses the extent of effort and collaboration required to respond to a piece of SCI.

Some SCI items are simple and actionable by a single controller. Others, however, may involve broader coordination across multiple facilities or teams. For instance, an alert about an airspace flow program (AFP) or large-scale reroute may require coordination across adjacent sectors, flow control units, and traffic management. Even brief messages can impose a high operational burden if they trigger procedural handoffs, inter-facility negotiations, or traffic sequencing adjustments (Orasanu et al., 2011; Endsley & Garland, 2000).

Messages with high coordination demands should be delivered earlier, clearly flagged, and designed for shared visibility. Conversely, low-burden SCI (e.g., weather updates, single-aircraft NOTAMs) may be scheduled closer to time-of-use. Prioritizing based on coordination demand supports smoother task flow and helps prevent bottlenecks during time-sensitive operations.

Together, these four adapted dimensions – Cognitive Integration Complexity, Modal Load Compatibility, Attentional Demand Type, and Actionability and Coordination Burden – provide an operationally grounded framework for examining how safety-critical aeronautical information imposes cognitive demands on ATCs. This structure supports targeted analysis of how information characteristics interact with real-world task constraints, helping to identify when and how SCI delivery may either support or hinder performance. However, workload alone does not fully explain how controllers interpret and act upon information in fast-moving, safety-critical contexts.

To understand how information impacts operational effectiveness in addition to mental effort, it is essential to consider situation awareness – the controller's evolving understanding of their environment and ability to anticipate future states. The following section explores SA to illustrate how information properties such as volume, relevance, timing, and modality also influence perception, comprehension, and projection in the dynamic decision cycles of ATC operations.

4.5. Situation Awareness: A Foundational Performance Construct

Situation Awareness (SA) refers to an ATC's ability to perceive elements in the operational environment, comprehend their significance, and project their likely evolution (Endsley, 1995). It is a dynamic, cognitively mediated process essential to maintaining safety and efficiency, particularly under high-tempo, high-stakes conditions. For controllers, SA is not just about having access to information – it involves filtering, integrating, and interpreting multiple data streams in real time while managing task loads, traffic complexity, and shifting priorities.

Volume and Organization of Information

Controllers are continuously required to process large volumes of safety-critical information delivered through voice communications, radar displays, flight progress strips, weather systems, and other sources. When information volume exceeds working memory capacity or is poorly structured, controllers may miss relevant cues or fail to perceive them in time (Level 1 SA: Perception). This is especially likely when critical data is embedded within lengthy, non-prioritized messages or arrives during periods of heightened operational demand, such as handoffs, reroutes, or emergency response.

Perceived Relevance and Filtering

Controllers naturally prioritize information that aligns with their current traffic picture, sector responsibilities, and operational phase. Safety-critical information that lacks contextual clarity – or appears disconnected from current controller goals – may be subconsciously filtered out or dismissed, even if technically accurate. Over time, repeated exposure to irrelevant or marginally relevant messages can lead to cognitive fatigue and reduced sensitivity to important updates. This degradation in perceived utility can weaken Level 2 SA (Comprehension), particularly when trust in the reliability of information sources is compromised (Lee & See, 2004).

Timeliness and Decision Windows

Timely delivery of safety-critical information is crucial to maintaining Level 3 SA (Projection) – the ability to anticipate future events and proactively manage system state. In ATC operations, even accurate information becomes less effective when it arrives outside the optimal decision window. A weather alert received too late to affect routing decisions, or a traffic advisory delivered during a handoff, may go unused due to timing misalignment. Conversely, information introduced too early, without reinforcement, may be forgotten or deprioritized as new traffic demands emerge.

Modality Compatibility and Information Uptake

The modality through which information is presented – visual, auditory, or tactile – shapes how effectively it is perceived and acted upon. In air traffic control, voice communication remains the dominant auditory channel, often saturated during peak traffic periods. Visual displays such as radar scopes, flight strips, or decision-support tools may also be overloaded. When safety-critical information is introduced via an already saturated modality, it risks being overlooked or misinterpreted. For example, auditory instructions may be masked in congested radio environments, while visual alerts may compete with high-priority traffic monitoring tasks. While alternative modalities such as haptics show promise, they require standardization and training to ensure they enhance, rather than complicate, information processing.

Integrated View: SA as a Mediator of Information Effectiveness

Ultimately, the effectiveness of safety-critical information in ATC depends on how well its characteristics – volume, relevance, timing, and modality – align with both cognitive constraints and real-time operational context. These variables do not function in isolation; they interact to shape the controller's evolving mental model of the traffic environment. If that model becomes inaccurate, incomplete, or fragmented, the quality of decision-making can degrade – leading to delayed interventions, missed handoffs, or inadequate responses to developing risks.

This research treats situation awareness as a core construct for understanding performance in information-rich, safety-critical ATC environments. It provides an empirical basis for assessing how specific properties of aeronautical information either support or hinder the controller's ability to detect, interpret, and act on critical events in real time.

4.6. Shared Situation Awareness

While individual SA enables ATCs to detect, interpret, and anticipate events within their own sector or area of responsibility, Shared Situation Awareness (SSA) refers to the degree to which controllers across positions, facilities, or sectors maintain a coordinated and consistent understanding of the operational environment and its likely evolution (Endsley & Connors, 2008; Endsley & Jones, 1997). SSA is especially vital in the distributed, multi-role structure of the NAS, where safe and efficient operations depend on seamless coordination across en route, terminal, and tower environments.

SSA extends Endsley's three levels of SA – 1) perception, 2) comprehension, and 3) projection – to the team level, where each controller's awareness is shaped by, and contributes to, a collective mental model. This shared model is critical for effective handoffs, coordinated reroutes, collaborative traffic management initiatives, and real-time responses to evolving weather or equipment issues.

Volume and Processing Capacity

In high-volume traffic environments, ATCs may prioritize immediate, sector-specific tasks, unintentionally filtering out updates or advisories that are relevant to coordination with adjacent positions. When safety-critical information is not clearly filtered, structured, or prioritized, it can strain working memory, reduce collaboration bandwidth, and weaken team-wide alignment. Even when all parties have access to the same data, variation in interpretation – driven by differences in workload, local traffic picture, or procedural constraints – can lead to SSA breakdowns.

Relevance and Role Alignment

Controllers are more likely to integrate information into their shared mental model when it is clearly relevant to their operational responsibilities and supports their coordination goals. When incoming information is ambiguous or appears tangential to current tasks, it may be dismissed, even if technically accurate or time-sensitive. Over time, inconsistent relevance or excessive peripheral updates can erode trust in information sources and foster divergent filtering strategies among team members – introducing variability in perception and decision-making across roles.

Timeliness and Synchronization

Timing is critical for maintaining SSA during collaborative decision-making and real-time traffic management. If one controller receives key information earlier or later than others – for example, a weather deviation advisory or holding pattern instruction – misaligned mental models may emerge. These discrepancies can compromise coordination during high-tempo phases such as sequencing, rerouting, or traffic flow adjustments. Without mechanisms to reinforce or confirm shared updates, the risk of coordination delays or miscommunication increases, particularly during time-sensitive operations.

Modality and Information Flow

SSA also depends on the modality of information delivery, especially in environments with overlapping visual and auditory demands. For example, verbal updates via landline or voice communications may compete with ongoing coordination tasks, while visual notifications (e.g., through decision-support tools or shared displays) may be missed if not seamlessly integrated into the controller's visual scan. When modality mismatches occur – such as auditory updates during radio congestion or visual alerts on peripheral displays – they can degrade not only individual awareness but the shared recognition of evolving events across the team.

Cumulative Impact on Operational Risk

SSA degrades when information is delivered through overloaded modalities, when its relevance to the role is unclear, when volume exceeds processing capacity, or when updates are not synchronized across positions. Because these dimensions often interact – especially under conditions of elevated workload – minor mismatches in perception can compound across controllers, escalating operational risk and increasing the potential for errors, delays, or redundant actions.

Maintaining shared situation awareness across distributed air traffic teams hinges not only on access to timely, relevant, and well-formatted information but also on the degree of trust each operator places in the sources providing that information. In high-tempo, safety-critical environments like the NAS, even accurate and well-timed updates can fail to support coordination if they are not perceived as credible or reliable. This trust – whether in human teammates, automation, or algorithmically prioritized data feeds – serves as a gatekeeper for integration into the shared mental model. Without sufficient trust, operators may dismiss, delay, or deprioritize critical updates, fracturing alignment across roles and increasing the risk of miscommunication, redundant action, or degraded performance. Thus, to understand how information characteristics shape team performance and safety outcomes, it is essential to explore the construct of trust in information sources and how it interacts with volume, relevance, timeliness, and modality across operational contexts.

4.7. The Role of Trust in Interpreting Safety-Critical Information

In air traffic operations, trust in the source of SCI – whether human (e.g., a fellow controller), automated (e.g., a decision-support tool), or algorithmically derived (e.g., AI-based alerts) – is a critical mediator of how that information is interpreted, prioritized, and acted upon. Especially under high workload and time pressure, operators may not have the cognitive bandwidth to thoroughly evaluate every incoming message. Instead, they often rely on trust-based heuristics to determine whether to attend to or dismiss the information (Lee & Moray, 1992; Rasmussen et al., 1994).

Trust is shaped dynamically over time based on the perceived reliability, consistency, and contextual relevance of prior information received from a source (Lee & See, 2004). In high-stakes environments like the NAS, where missed or misjudged cues can have cascading safety consequences, this trust calibration becomes essential to both individual and team performance.

Impact of Information Characteristics on Trust

Each of the four core information characteristics – volume, relevance, timeliness, and modality – can build or erode trust depending on how well they align with operational expectations and task context.

- Volume: When information is delivered in excessive or poorly filtered amounts, trust in the system's prioritization logic can degrade. ATCs may begin to selectively ignore messages, assume the system lacks intelligence, or rely more heavily on personal judgment, increasing the risk of inconsistency (Wickens et al., 2015). On the other hand, the absence of expected information, especially during dynamic traffic or weather events, can also undermine trust – leading controllers to question whether a system is functioning correctly or up to date (Gutzwiller & Clegg, 2013).
- Relevance: Trust depends not just on accuracy, but on operational usefulness. Repeated exposure to marginally relevant or tangential updates can lead to message fatigue and a generalized loss of confidence in the source. Over time, even high-priority messages may be discounted if users have learned that most outputs from a system are unhelpful or non-actionable (Lee & See, 2004).
- Timeliness: When information arrives too late to inform planning or too early to be actionable, it is often perceived as poorly aligned with task demands – even if technically accurate. Such mistiming signals to the operator that the source cannot be relied upon to support real-time decision-making. Timeliness is particularly influential during time-critical events such as reroutes, handoffs, or weather avoidance, where any lag can shift perception from "supportive" to "burdensome" (Endsley, 2015; Klein, 1993).
- Modality: If messages are delivered through incompatible or overloaded channels (e.g., auditory updates during radio congestion, visual alerts during heads-down coordination), they may be missed, misunderstood, or perceived as disruptive. Over time, operators may begin to deprioritize messages from specific modalities – not based on their content but due to sensory mismatch, further fracturing trust (Wickens, 2008).

Cognitive Load and Heuristic Trust Decisions

Under conditions of elevated mental workload, ATCs have limited capacity to critically evaluate each message. Instead, they default to trust heuristics: favoring sources that have previously aligned with their expectations and deprioritizing those that have not (Rasmussen et al., 1994; Hollnagel & Bye, 2000). These mental shortcuts are efficient but not infallible. When miscalibrated – either through over-trust or under-trust – they can result in the premature dismissal of valid SCI or the over-reliance on flawed or outdated systems.

Trust Is Not Enough: The Role of Risk Perception

Even when trust in a system is high, behavioral response is not guaranteed. Operators may still disregard a message if they underestimate the likelihood or severity of the hazard it references. Conversely, low-trust sources may occasionally produce credible warnings that are dismissed due to the operator's history with that system. Therefore, trust interacts with risk perception in determining whether SCI prompts action, hesitation, or inaction.

With a foundational understanding of how information characteristics influence workload, situation awareness, shared awareness, and trust, the next section introduces the concept of risk perception and how it mediates operator response to SCI. By applying the Bow-Tie Framework, we model the interplay between information, trust, and operational risk. This approach provides a structured, systems-level lens through which to visualize and mitigate how misaligned or degraded information flows can evolve into safety events when preventive and mitigative barriers fail.

5. Risk Perception and Air Traffic Control

In dynamic aviation environments, the effectiveness of decision-making hinges not only on the availability and accuracy of SCI, but also on how that information is perceived, interpreted, and trusted. Previous sections have outlined how information characteristics, such as volume, relevance, timeliness, and modality, interact with cognitive workload, shape individual and shared situation awareness, and influence trust in the information source. These elements all converge in the operator's risk perception, which plays a pivotal role in determining how information is prioritized and acted upon under time pressure and uncertainty. Even highly accurate, well-timed information may go unused if the perceived risk does not align with the operator's mental model or workload capacity (Endsley, 1995; Lee & See, 2004; Wickens et al., 2015). Because risk perception governs attention allocation and response strategy selection, it serves as a critical link between human cognition and operational safety outcomes.

5.1. The Role of Risk Perception in Operational Decision-Making

Risk perception refers to how individuals interpret the likelihood and severity of potential hazards and how those interpretations influence decision-making under uncertainty (Slovic et al., 1986; Fischhoff et al., 2013). In ATC environments, where decisions must often be made in seconds, perceived risk often overrides objective risk assessments. That is, even highly accurate, well-timed information may be underutilized or ignored if the controller does not perceive an immediate or meaningful threat (Wickens et al., 2015; Gutzwiller & Clegg, 2013).

Controllers must constantly prioritize among competing demands, and in doing so, often rely on heuristic-based filtering that favors familiarity, visual salience, recency, and perceived task relevance (Klein, 1993; Todd & Gigerenzer, 1999). These mental shortcuts, while adaptive under cognitive load, can lead to underweighting rare but high-impact hazards, especially when those hazards are embedded in lengthy, ambiguous, or low-priority information streams.

5.1.1 Interaction with Information Characteristics

Each core attribute of SCI – volume, relevance, timeliness, and modality – interacts with risk perception:

- Volume: Excessive data can overwhelm working memory and obscure signal amid noise, reducing a controller's ability to distinguish between low-risk routine updates and high-risk anomalies (Baddeley, 2012). When information density is high, perceived urgency can be dulled by sheer cognitive load.

- Relevance: Perceived risk increases when information is clearly and immediately applicable to a controller’s area of responsibility. Conversely, updates perceived as tangential – such as advisories for adjacent sectors – may be dismissed despite containing cues relevant to emerging hazards (Endsley & Jones, 1997).
- Timeliness: Risk is highly time-sensitive. SCI that arrives after a critical decision point may be interpreted as less useful, even if the underlying threat remains. This mismatch in timing can reduce perceived utility and trust in the system (Endsley, 2017).
- Modality: Risk perception can be degraded when information is delivered through overloaded or mismatched sensory channels. For example, if a visual cue about convective weather is delivered during a period of high visual demand, the controller may not register it in time to respond appropriately (Wickens, 2008; Jones & Sarter, 2008).

These characteristics also influence the emotional salience of a risk. Research has shown that operators are more likely to respond when hazards are framed in ways that highlight near-term operational impacts or are delivered through more "urgent" modalities, such as auditory alerts (Drijvers & Holler, 2023).

5.2. Trust and Risk: A Critical Interaction

Trust and risk perception are tightly coupled in ATC operations. Even when a system or information source is trusted, risk perception determines whether that trust translates into timely action. Conversely, a breakdown in trust – due to poor filtering, mistimed alerts, or modality mismatches – can skew the operator’s risk model, leading to discounting of even critical alerts (Lee & See, 2004).

Research in supervisory control and human-automation interaction confirms that miscalibrated trust – either overtrust or distrust – can lead to inappropriate risk responses (Parasuraman & Riley, 1997). For instance, controllers may underreact to real threats if the alerting system has previously issued irrelevant or false alarms, leading to habituation or alert fatigue.

5.2.1 Implications for Safety-Critical Information Design

Given the centrality of risk perception in information processing, SCI systems and procedures should be designed not only for accuracy and timeliness, but also for perceptual and contextual alignment with operator expectations. Effective risk communication in ATC requires:

- Cue congruence: Clear links between data and operational relevance.
- Temporal alignment: Information that maps to the correct phase of decision-making.
- Modality optimization: Delivery through underloaded channels to maximize salience.
- Trust calibration: Reinforcing the reliability of alerts through consistent, context-aware filtering.

Controllers benefit from systems that explicitly surface the risk context – for example, visually flagging information with “elevated,” “moderate,” or “urgent” risk markers, or using consistent auditory tones to signal alert tiers. These affordances help reduce reliance on ambiguous heuristics and support faster, more accurate prioritization.

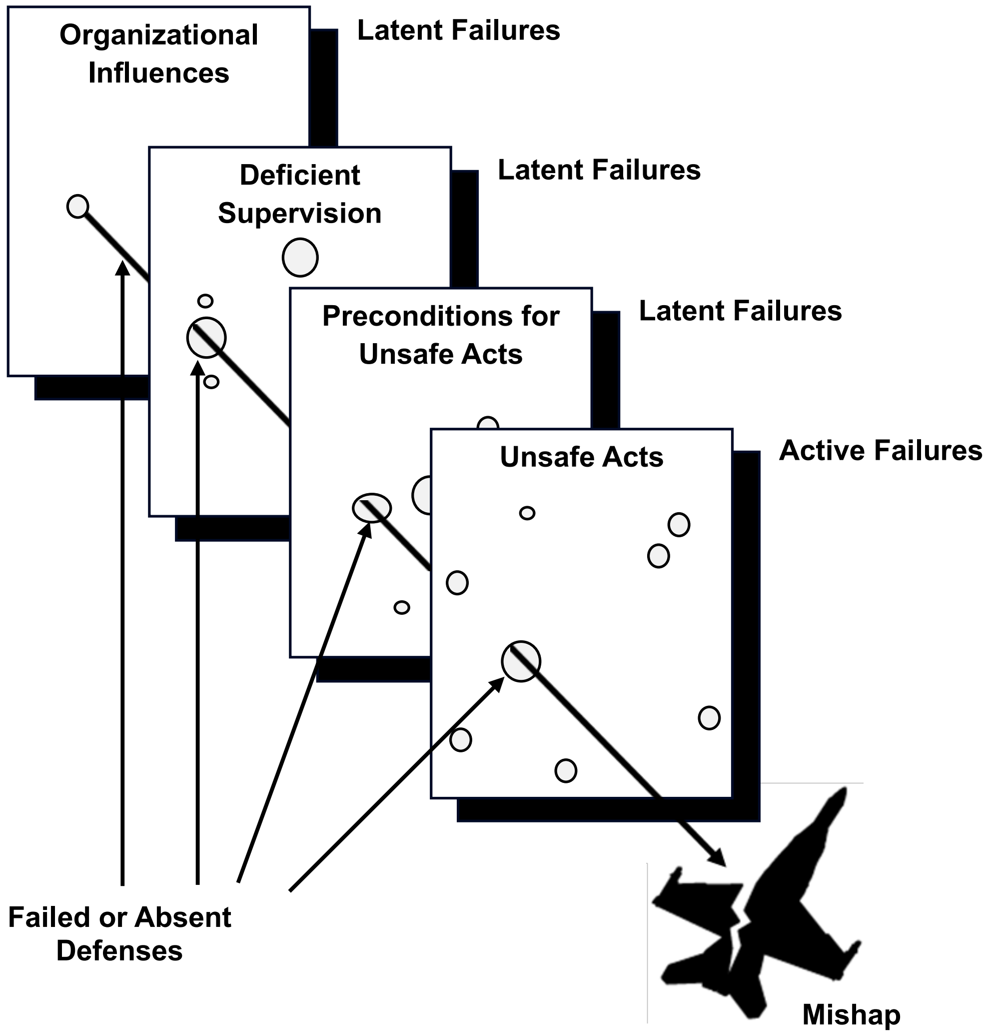
5.3. The Swiss Cheese Model and Human Factors in ATC

Accidents in complex systems rarely result from a single point of failure. Instead, they emerge through the accumulation and alignment of multiple factors across multiple domains (Levenson, 2011). An adapted version of James Reason's Swiss Cheese Model (SCM) (Reason, 1990) provides ATC-specific foundational lens for conceptualizing weaknesses in the pathway of communicating aeronautical information. The model depicts organizational, supervisory, operational, and human factors barriers designed by organizations to prevent adverse events as slices of Swiss cheese, each slice including inherent weaknesses or holes, representing latent or active vulnerabilities in the system (See Figure 5-1). When these vulnerabilities align, they create a pathway through which hazards can bypass defenses, leading to an incident.

In the context of ATC, these vulnerabilities may arise from multiple interacting factors, complex airspace structure, dynamic weather, procedural variability across facilities, or human performance factors such as workload and fatigue. These variables complicate efforts to address risk related to aeronautical information through any single barrier (e.g., procedures, training, technology). A structured sorting and prioritizing framework plays an essential role in reducing the size and number of holes in each layer of defense, thereby strengthening systemic resilience in the area of communication SCI.

Figure 5-1

The "Swiss Cheese" Model of Human Error Causation



5.3.1 Latent and Active Failures

1. Organizational Influences

The first layer of cheese in the model is organizational influence, and refers to latent conditions embedded in policies, resource allocation, and cultural norms that shape the operating environment long before a front-line decision is made. In ATC, this includes how resources are allocated for technology modernization, staffing, training, and infrastructure. Outdated systems, inconsistent SOPs, and organizational messaging about safety priorities can introduce vulnerabilities that impact operational decision-making (Levenson, 2011). For example, ambiguous or outdated procedures created without consideration of a dynamic NAS can make it difficult to prioritize critical information during dynamic operations. This type of latent organizational factor forms the first layer of defense and, if weakened, can be exploited by downstream failures.

2. Unsafe Supervision

The second layer reflects supervisory practices that shape day-to-day operations. Supervisory performance deficiencies are defined as deviations from established best practices, organizational standards, or training protocols that adversely influence workload distribution, decision-making consistency, or overall system reliability. Rather than implying direct “unsafe supervision”, these deficiencies are more appropriately characterized as inadequate, ineffective, or deficient supervisory controls (Reason, 1997; Shappell & Weigmann, 2000). Examples include insufficient oversight of training progression, staffing decisions that disregard fatigue risk or sector complexity, and inadequate reinforcement of safety-critical procedures. These supervisory conditions can degrade operational resilience by increasing front-line cognitive workload, narrowing decision bandwidth, and amplifying susceptibility to performance variability or error.

3. Preconditions for Hazards

The third layer includes performance shaping conditions that increase the likelihood of operational hazards. These preconditions – such as cognitive overload, loss of situation awareness, and coordination breakdowns – arise from factors like fragmented information, adverse mental states (e.g., fatigue, distraction), or environmental complexity. In the ATC context, such conditions are not failures themselves but elevate the potential for safety-relevant deviations when left unmitigated (Wickens, 2002; Reason, 1990). Their interaction with contextual factors like traffic density, airspace configuration, and weather complexity contributes to an increased risk of human error setting the stage for downstream safety events (Durso & Manning, 2008; Histon & Hansman, 2008).

4. Operational Hazards

The fourth layer represents frontline operational hazards – observable actions or omissions that pose elevated safety risk. These may include decision-making lapses, procedural deviations, or misprioritization of safety-critical information. For example, failing to disseminate an urgent weather update, overlooking a runway closure NOTAM, or misjudging separation distances in high workload scenarios can all constitute safety-relevant hazards (Wiegmann & Shappell, 2003; Eurocontrol, 2013). While such actions may not directly cause incidents on their own, they can represent the final active link in a chain of latent conditions and systemic vulnerabilities. As described in Reason’s Swiss Cheese Model (SCM) and reflected in FAA’s hazard-based SMS structure, these hazards are often shaped by upstream supervisory, organizational, or information environment factors (Reason, 1997; FAA, 2023b). Therefore, operational hazard analysis must be contextualized within the broader risk environment rather than focusing solely on individual behavior.

5.3.2 The Swiss Cheese Model and the Analysis of Risk

The SCM provides a conceptual foundation for risk analysis, drawing attention to the failures across multiple layers that must align for an adverse event to occur. Contextual variables such as weather, traffic density, class of airspace, and operational tempo may increase system complexity and make it difficult to fully address risk within any single barrier category.

Therefore, strengthening contextual variables' defenses requires a full understanding of those factors with regard to safety-critical information. A framework for sorting and prioritizing SCI directly targets the human and technical layers for ATCs, to aid more effective identification and handling of the most critical data during complex and dynamic operations. This may then reduce the likelihood that multiple vulnerabilities will align.

5.3.3 From Swiss Cheese to Bow-Tie: Structuring Vulnerabilities and Controls

While the SCM illustrates the way vulnerabilities align, it does not provide a structured way to map specific hazards, threats, barriers, and potential hazard-related consequences. It provides the framework for understanding how layered defenses can fail, but it does not specify how to operationalize those failures into structured risk controls. The bow-tie framework complements this by mapping vulnerabilities identified through Swiss Cheese into explicit threat pathways and barrier strategies.

For example, latent organizational vulnerabilities, such as incomplete integration of weather data or outdated procedures, may be identified as upstream holes in the SCM. These same vulnerabilities can then be translated into hazards and threats within the bow-tie diagram, with associated preventive and mitigative barriers and potential operational outcomes to move from a conceptual understanding of risk to a structured, barrier-based analytical framework, enabling more precise targeting of safety interventions.

6. The Bow-Tie Model

The bow-tie model is a risk diagram showing how various threats can lead to a loss of control of a hazard and allow the unsafe condition to develop into a number of undesired consequences. Figure 6-1 shows all elements mapped in a pathway that details the progression of identified hazards and threats to their resulting consequence. It was originally developed in engineering domains to analyze complex industrial hazards and can be easily used to provide a structured framework for understanding risk in the aviation context (CCPS, 2001, 2020a). Below is a list of terms seen in the bow-tie diagram and a definition for each.

Hazard. An operation, activity, or material with the potential to cause harm to people, property, the environment, or business; or simply, a potential source of harm. In aviation operations with regard to communication, this may be seen as aviation operations with incomplete information.

Top Event. In bow-tie risk analysis, a central event lying between a threat and a consequence corresponding to the moment when there is a loss of control or loss of containment of the hazard.

Consequence. The undesirable result of a loss event, usually measured in health and safety effects, environmental impacts, loss of property, and business interruption costs. This may also be referred to as an Outcome.

Threat. A possible initiating event that can result in a loss of control or containment of a hazard (i.e., the top event). This may also be referred to as the Cause or Initiating Event.

Prevention Barrier. A barrier located on the left-hand side of the bow-tie diagram that lies between a threat and a top event. It must have the capability on its own to completely terminate a threat sequence. This may also be referred to as Proactive Barrier.

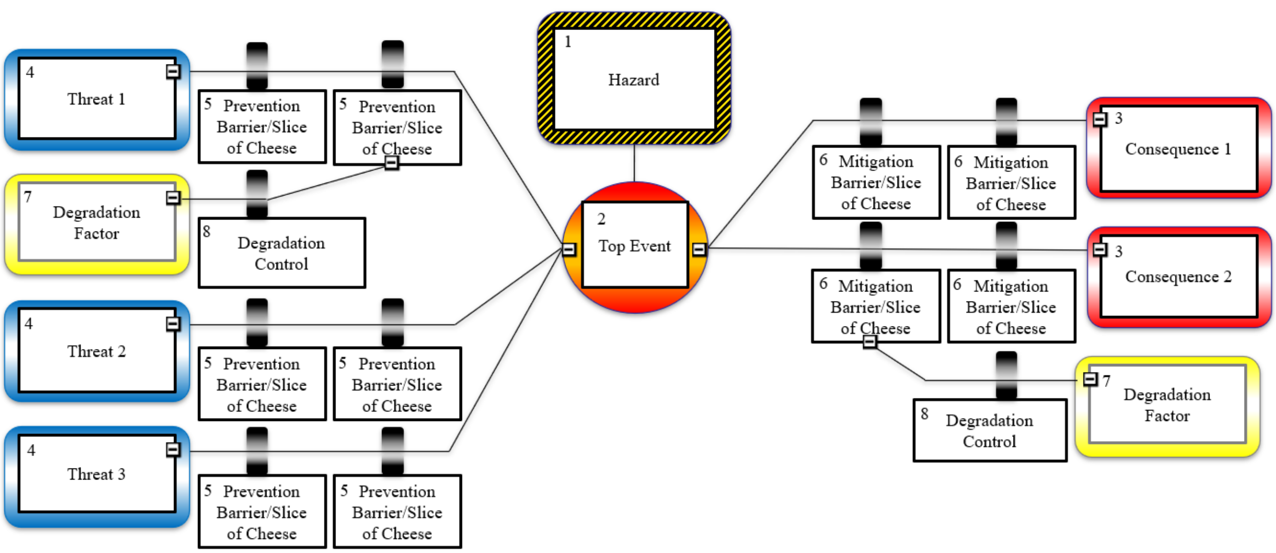
Mitigation Barrier. A barrier located on the right-hand side of the bow-tie diagram lying between the top event and a consequence. It might only reduce a consequence, not necessarily terminate the sequence before the consequence occurs. This may also be referred to as Reactive Barrier or Recovery Measure.

Degradation Factor. A situation, condition, defect, or error that compromises the function of a main pathway barrier, through either defeating it or reducing its effectiveness. If a barrier degrades, then the risks from the pathway on which it lies increase or escalate. This may also be referred to as Barrier Decay Mechanism, Escalation Factor, or Defeating Factor.

Degradation Control. Measures that help prevent the degradation factor from impairing the barrier. They lie on the pathway connecting the degradation threat to the main pathway barrier. This may also be referred to as Degradation Safeguard, Defeating Factor Control, Escalation Factor Control, or Escalation Factor Barrier.

The utility of this model lies in its ability to represent complex, information-centric safety threats, particularly those shaped by decision-making under incomplete or delayed information. The proposed methodology proposes using the model to illustrate how aviation operations with incomplete information can lead to a decision error and possible consequences.

Figure 6-1
Complete Bow-Tie Model



6.1. Human Factors in the Bow-Tie

Embedding human factors constructs into the bow-tie framework creates a direct link between related hazards, measurable degradation factors, and potential consequences for each. This integration allows scenario conditions to be developed and systematically manipulated in experimental phases to address the NTSB recommendation to research this critical area of interest.

Drawing on Cognitive Load Theory (Sweller, 1988), Multiple Resource Theory (Wickens, 2002), and Endsley's SA framework (1995), the proposed research examines how cognitive and perceptual limitations interact with operational complexity to shape risk outcomes. Cognitive workload is treated not only as a consequence of poor information design but also as a direct threat that increases the likelihood and severity of adverse events.

Similarly, breakdowns in SA or SSA are conceptualized as threats that propagate risk, especially in distributed operations where coordination is critical.

Trust in information sources further shapes operator behavior, whether cues are acted upon or dismissed, thereby amplifying or mitigating risk depending on alignment with system reliability. By embedding these constructs into the bow-tie framework, the methodology captures both systemic and human contributions to hazard progression.

6.1.1 A Human Factors Framework

The bow-tie model is used in this methodology to structure safety-critical information hazards, associated controls, and potential operational consequences in a visual and scenario-based format. While it provides a valuable framework for tracing latent conditions and points of control failure, the diagram alone does not determine risk in the formal sense outlined in FAA Order 8040.4B, which defines risk as a function of likelihood and severity (Reason, 1997; FAA, 2023e; CCPS, 2020a). Consistent with FAA Order 8040.4B, a hazard is defined as "a condition that could foreseeably lead to or contribute to an unplanned or undesired event (FAA, 2023e). In the SCI context, this includes situations where delayed, incomplete, or misprioritized information increases the likelihood of operational disruption or safety margin erosion. Thus, while the bow-tie model captures structural hazard-consequence pathways, it must be supplemented with empirical analyses to assess operational risk.

This methodology addresses that gap by layering in human performance theory to explain how and why these pathways are traversed in real-world contexts (De Dianous & Fiévez, 2006; CCPS, 2020b). Risk prioritization in this study is not assigned within the bow-tie structure itself but is derived through integrated methods, including real-time cognitive workload assessment (e.g., ATWIT), expert judgment panels, and scenario-based outcome severity evaluations.

This framework focuses on four interrelated human factors constructs critical to SCI processing and decision quality: cognitive workload, SA, SSA, and trust in information source. These constructs form the foundation for understanding human reliability and performance variability in safety-critical environments.

Each of these constructs is embedded in a shared risk pathway, characterized by (1) a hazard: aviation operations with incomplete information, (2) a top event: operator decision error, and (3) a consequence: operational safety event. A safety event in this context is defined as any operational occurrence in which safety margins are reduced or compromised due to delayed, missed, or misprioritized safety-critical information, resulting in degraded separation assurance, trajectory deviation, or similar adverse operational impact (Loft et al., 2007; Histon & Hansman, 2008; NTSB, 2016; Wiegmann & Shappell, 2017).

The following sections present construct-specific bow-tie diagrams aligned in this pathway in ATC operational settings. Each diagram is structured not as a conceptual illustration but as an operational tool supporting empirical scoring, simulation, and integration into the study's Phase II and III scenario-based validation efforts. All diagrams include hazard sources, human and technical barriers, degradation factors, and potential consequences grounded in validated human factors principles (De Dianous & Fiévez, 2006; CCPS, 2020a; 2020b).

6.1.2 Cognitive Workload in the Bow-Tie Pathway

As illustrated in Figure 6-2, the bow-tie pathway for cognitive workload begins with the (1) hazard: in this case, aviation operations with incomplete information, resulting in decision-making by ATCs with an incomplete picture of the NAS. This may lead to an occurrence of the (2) top event: a decision error, where inaccurate or delayed judgments occur due to strained cognitive resources.

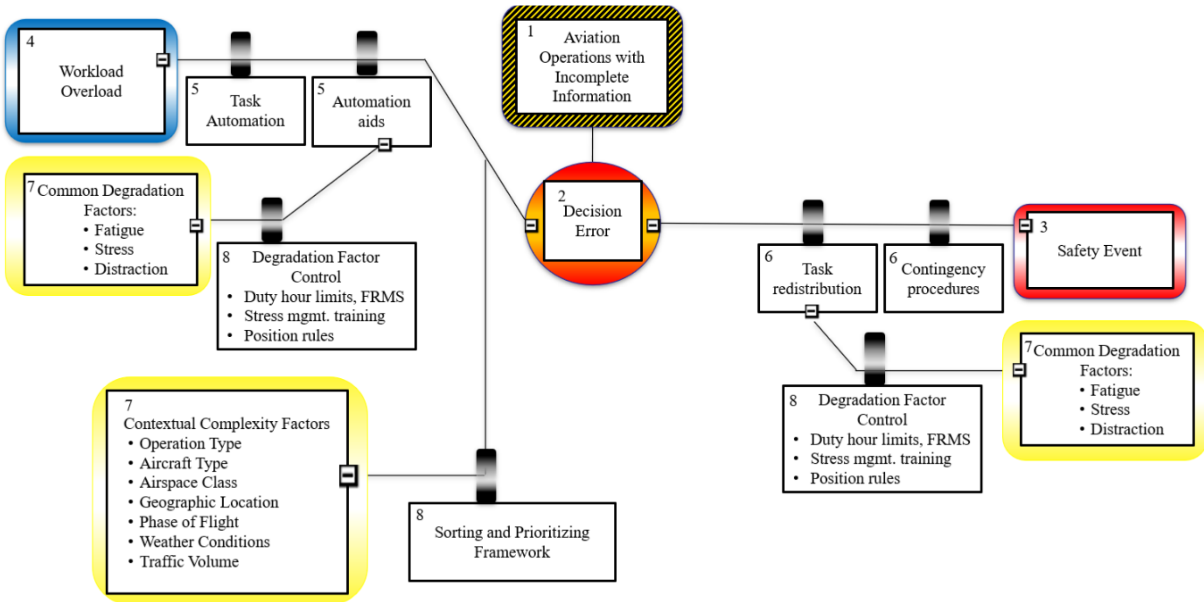
For the purpose of this research, cognitive workload is defined as the mental effort and attentional resources required for an ATC to perceive, process, and respond to operational demands while maintaining safe and orderly traffic flow (Wickens, 2002; Young et al., 2015). Within the ATC context, workload is impacted by multiple sources, including traffic volume, weather disruptions, coordination demands, and procedural variability (Damos, 1997; Loft et al., 2007; Histon & Hansman, 2008). Elevated workload occurs when these demands exceed an ATC's available cognitive capacity, impairing performance and increasing the risk of error (Wickens, 2002; Young et al., 2015; Loft et al., 2007).

Cognitive workload is inherently dynamic and context-dependent, fluctuating across operational tasks, phases of flight, and traffic conditions (Damos, 1997; Loft et al., 2007). In this study, workload will be examined relative to the timing and characteristics of contextual complexity as well as the timing and characteristics of SCI delivery. Example consequences from elevated or degraded workload states may include delayed handoffs during peak traffic or failure to integrate weather deviations into routing decisions under time pressure (Bainebridge & Dorneich, 1983, 2016; Loft et al., 2007).

The key (4) threat in this pathway is workload overload, which reduces attentional capacity, narrows situational focus, and limits cognitive flexibility for integrating new information (Wickens, 2002; Bainbridge & Dorneich, 2016; Young et al., 2015). To address these vulnerabilities, (5) prevention barriers such as task automation or automation aids may be utilized to counter any negative effects of overload (Wickens, 2002; Hilburn, 2004; Mosier et al., 2017a). If preventions fail, (6) mitigation barriers, including task redistribution and other contingency procedures, serve to buffer the severity of the (3) consequence or safety event, such as delayed clearances, separation losses, or reduced capacity to respond to emerging hazards (Vu et al., 2010). This figure shows common (7) degradation factors seen in aviation, such as fatigue, multitasking under pressure, or other resilience-reducing conditions which may erode the effectiveness of prevention barriers and may increase the risk of a consequence, as well as degradation factors related to the previously discussed elements of contextual complexity, along with (8) degradation controls for each, in the case of common degradation factors, we see controls such as duty hour limits and stress management training among others, and a sorting and prioritization framework for handling aeronautical information as a control for degradation factors related to contextual complexity.

Figure 6-2

Risk Pathway for Cognitive Workload



6.1.3 Situation Awareness in the Bow-Tie Pathway

As illustrated in Figure 6-3, the bow-tie pathway for situation awareness begins with the common (1) hazard, which creates conditions where operators may be forced to rely on partial or unclear inputs. This increases the likelihood of the (2) top event: a decision error, where judgments are based on inaccurate or incomplete mental models.

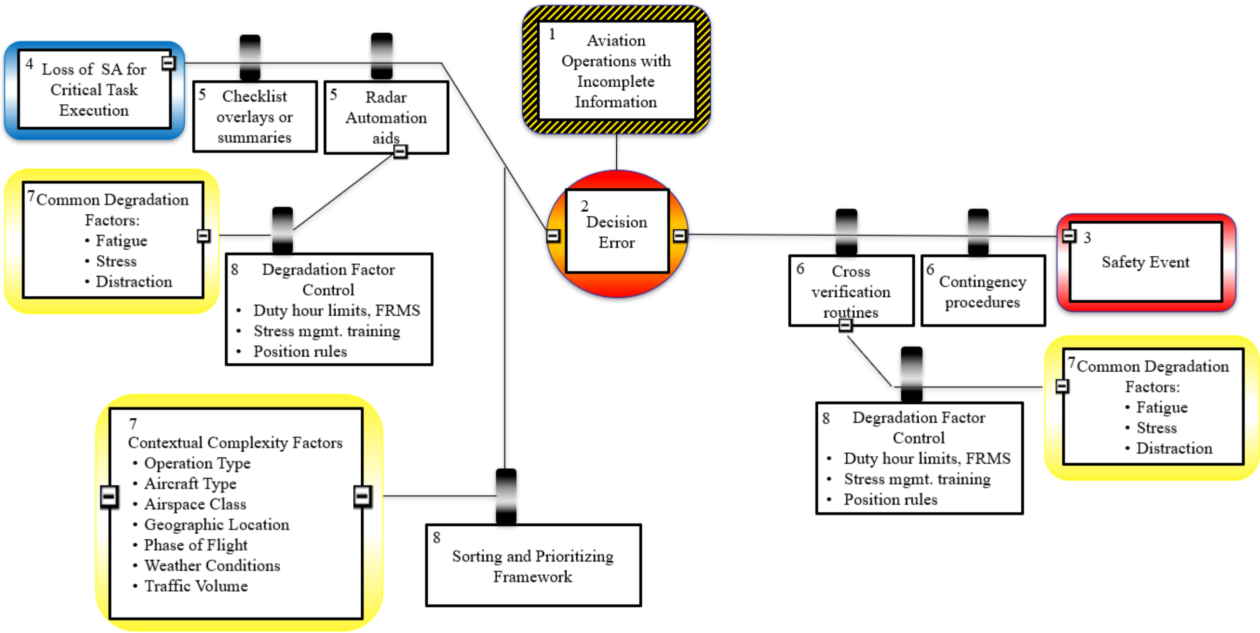
For the purpose of this research, SA is defined as an ATC's ability to perceive, comprehend, and project relevant elements of the operational environment necessary to ensure the safe and orderly flow of traffic (Endsley, 1995; Endsley & Jones, 1997). In the ATC context, this includes awareness of aircraft position and intent, traffic sequencing, terrain, weather, sector status, and other dynamic operational factors. Degraded SA is operationally defined as a loss, delay, or distortion in one or more of the three SA levels – perception, comprehension, or projection – which affects the controller's ability to prioritize and act on safety-critical information.

Because SA is inherently context-sensitive, its adequacy varies across tasks, phases of flight, operational environments, and tactical objectives (Endsley, 1995; Endsley & Jones, 1997; Histon & Hansman, 2008). In this study, SA will be evaluated in relation to the operational context (e.g., traffic complexity, weather conditions, sector boundaries, and the timing of SCI delivery) rather than as a static construct. This approach reflects how controllers dynamically adjust attention and interpretation based on changing operational demands (Wickens, 2002; Damos, 1997). Examples of degraded SA include missed or delayed detection of an aircraft entering controlled airspace, incomplete or inaccurate mental models of traffic flow, or failure to anticipate conflicts or weather impacts due to incomplete or poorly timed SCI (Endsley, 1997; Endsley & Jones, 1997; Roth et al., 2021).

The key (4) threat in this pathway is degraded situation awareness, which occurs when information is ambiguous, poorly timed, or cognitively misaligned, such as delayed PIREPs, vague or outdated NOTAMs, or cluttered visual interfaces (Endsley, 1997; Endsley & Jones, 1997). To address these vulnerabilities, (5) prevention barriers may be deployed, including the use of checklist overlays or summaries, as well as radar automation aids to increase the likelihood that ATCs maintain access to relevant cues (Wickens, 2002; Hilburn, 2004; Mosier et al., 2017a; 2017b). If these measures are insufficient, (6) mitigation barriers, such as cross-verification routines (e.g., verbal and visual confirmation), may help reduce the severity of a resulting (3) consequence or safety event, which may include tactical errors such as incorrect vectoring or misjudged spacing (Vu et al., 2010). As previously noted in the risk pathway for cognitive workload, this model also details common (7) degradation factors seen in aviation, as well as degradation factors represented as contextual complexity factors, explicitly discussed for this research, along with (8) degradation controls.

Figure 6-3

Risk Pathway for Situation Awareness



6.1.4 Shared Situation Awareness in the Bow-Tie Pathway

As illustrated in Figure 6-4, the bow-tie pathway for SSA begins with the (1) hazard, which creates conditions where not all controllers have shared access to the same flight-related information. This can increase the likelihood of the (2) top event: a decision error, where actions become misaligned across controllers and contradictory actions may occur.

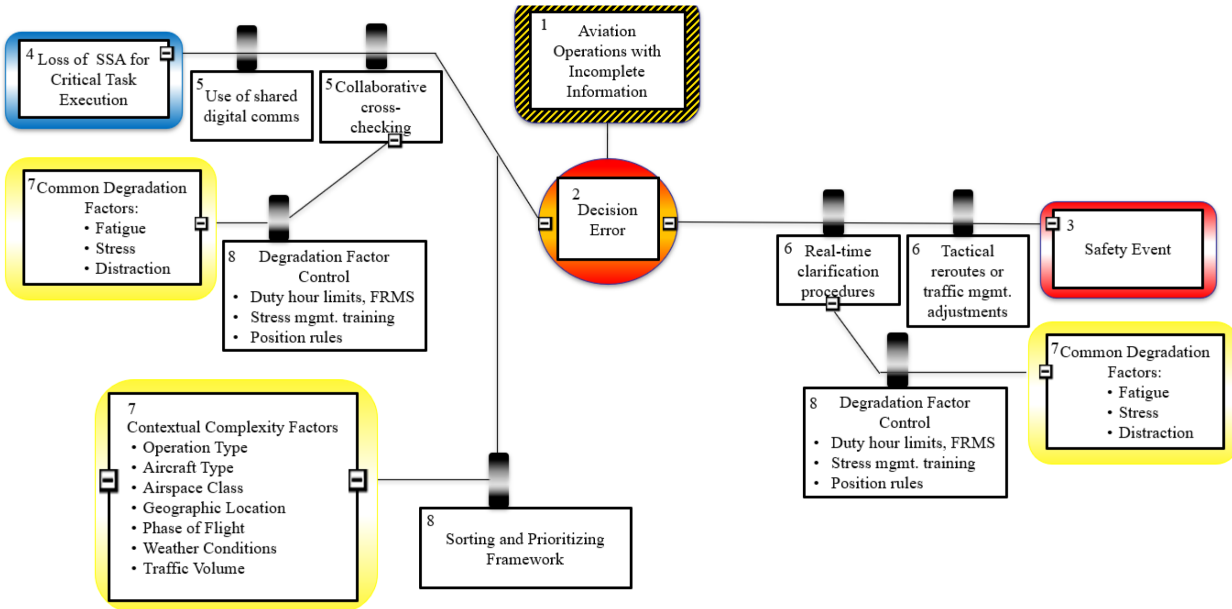
For the purpose of this research, shared situation awareness is defined as a common, accurate, and up-to-date understanding of the operational environment across multiple roles engaged in managing air traffic (Endsley & Jones, 1997; Salas et al., 1995. In the ATC context, SSA involves shared understanding of aircraft trajectories, weather hazards, sector conditions, reroutes, and other operational elements that influence coordinated decision-making (Endsley & Jones, 1997; Salas et al, 1995; Orasanu et al., 2001). Degraded SSA occurs when asynchronous information flow, inconsistent updates, or communication gaps cause controllers along the pathway of the flight to form divergent mental models (Lai et al., 2019; Hansman & Davison, 2000).

Because SSA is fundamentally co-constructed and time-sensitive, its adequacy depends on both the reliability of the information systems and the effectiveness of the communication processes under varying operational conditions (Salas et al., 1995; Gorman et al., 2006; Lai et al., 2019). In this methodology, SAA will be assessed in relation to how SCI is distributed over the duration of the flight and the impact of contextual complexity on timeliness and alignment of shared information and mental model development (Mosier & Fischer, 2017; Lai et al., 2019; Hansman & Davison, 2000). Examples of degraded SSA include inconsistent handoff information, outdated reroute messages, or a lack of synchronization between controllers' awareness of hazards (Loft et al., 2007; Histon & Hansman, 2008).

The key (4) threat in this pathway is misaligned team mental models. To address these vulnerabilities, (5) prevention barriers may be deployed and include the use of shared digital communication channels such as CPDLC and SWIM feeds, and collaborative cross-checking (Cardosi, 1993; Barshi & Farris, 2016; Cardosi, 2017; Cardosi et al., 1996; Cardosi et al., 1998). If these measures are insufficient, (6) mitigation barriers, such as real-time clarification procedures and tactical reroutes or traffic management adjustments, become critical for recalibrating team awareness to reduce the severity of a resulting consequence (Cushing, 1994; Histon & Hansman, 2008; Cardosi, 2017). If barriers fail, the pathway culminates in the (3) consequence or safety event, which may include conflicting clearances or degraded conflict detection. As with other pathways, this model also details common (7) degradation factors seen in aviation, as well as degradation factors represented as contextual complexity factors, explicitly discussed for this research, along with (8) degradation controls.

Figure 6-4

Risk Pathway For Shared Situation Awareness



6.1.5 Trust in Information Source in the Bow-Tie Pathway

As illustrated in Figure 6-5, the bow-tie pathway for trust in information source (trust) begins with the common (1) hazard, where ATCs are making decisions under uncertainty, which may increase the likelihood of the (2) top event: a decision error, where actions may be either based on unreliable information or valid information disregarded due to insufficient trust.

For the purpose of this research, trust in information source is defined as the ATC's willingness to rely on and act upon information provided by human or automated systems (Lee & See, 2004; Madhavan & Wiegmann, 2007). In the ATC context, this includes trust in other controllers, pilots, meteorological data, automated tools, and system displays. Degraded trust is operationally defined as either over-trust (inaccurate reliance on flawed or outdated data) or under-trust (unwarranted rejection or discounting of accurate, safety-critical information).

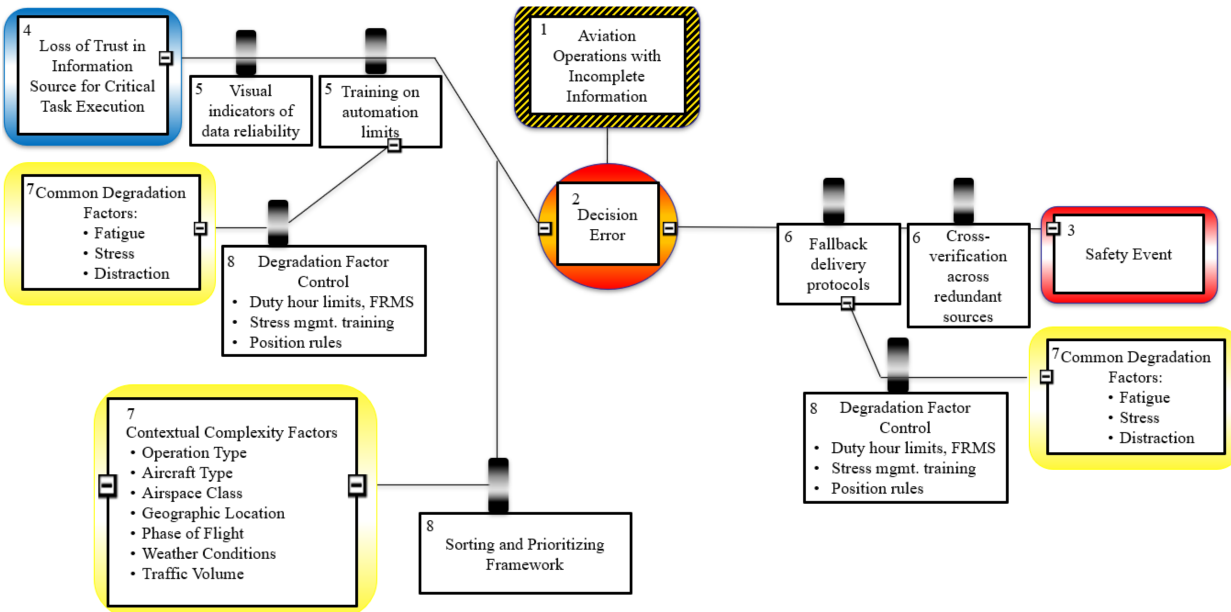
Trust calibration is highly sensitive to context and influenced by system reliability, prior experience, and operational tempo (Lee & Moray, 1992; Hoff & Bashir, 2015). In this methodology, trust will be assessed in relation to how SCI is presented (i.e., timing, modality, validation cues) and how contextual complexity shapes reliance behavior. Examples of degraded trust include ignoring valid weather alerts due to prior false alarms or following outdated reroute advisories based on overreliance on automation (Parasuraman et al., 2000; Madhavan & Wieigmann, 2007; Parasuraman & Riley, 1997).

The key (4) threat in this pathway is miscalibrated trust, a dynamic that can significantly impair decision-making and overall system safety, which may amplify or mitigate operational risk depending on its alignment with information validity (Lee & Moray, 1994; Lee & See, 2004). To address this vulnerability, (5) prevention barriers may be deployed, including visual indicators of data reliability or transparent system logic with explainable AI outputs (Hernandez et al., 2021). If these measures are insufficient, (6) mitigation barriers, such as fallback delivery protocols or cross-verification procedures, can help reduce the severity of a resulting (3) consequence (Parasuraman & Riley, 1997).

Should these barriers fail, the pathway culminates in the (3) consequence: a safety event, where reliance on outdated or flawed data, or disregard of valid alerts due to distrust in the system, degrades safety margins and system resilience. As previously noted in the risk pathway for cognitive workload, this model also details common (7) degradation factors seen in aviation, as well as degradation factors represented as contextual complexity factors, explicitly discussed for this research, along with (8) degradation controls.

Figure 6-5

Risk Pathway for Trust in Information Source



6.2. Contextual Complexity in the Bow-Tie Model

Even when prevention and mitigation barriers are well-designed, their effectiveness is shaped by the broader operational environment in which they are applied. Maintaining manageable levels of cognitive workload becomes increasingly difficult when the operational context itself imposes high and variable demands in a constantly shifting interplay of personnel, procedures, and interconnected technologies that together form a vast, distributed ecosystem (Wickens et al., 2015; FAA, 2016). These interacting operational conditions that shape how ATCs perceive, prioritize, and act SCI are referred to as contextual complexity, a multidimensional construct, which includes variables such as airspace class, geographic location, phase of flight, weather conditions, and traffic density. For en route controllers in particular, complexity is often compounded by handoff dynamics, where an aircraft entering a new sector may or may not have received relevant SCI from the previous controller or may be transferred without full situational context due to uncertainty, time pressure, or information ambiguity.

These dynamics are not merely background noise; they actively influence how SCI is framed, its urgency, and how cognitively salient it is to the receiving controller. For example, a reroute notification or PIREP may carry immediate relevance during peak traffic but appear low-priority when workload is low or weather conditions are benign. Operational context may determine the way controllers process, manage, and prioritize information.

Operational conditions within the NAS can change abruptly in response to weather disruptions, equipment outages, security events, or rapid fluctuations in traffic demand, introducing contextual complexity, which can independently and interactively alter both the cognitive and coordination burdens on operators (Damos, 1997; Loft et al., 2007; Salas et al., 2008; Histon & Hansman, 2008).

Operators in these settings work under very different procedures, use varied equipment, and rely on distinct communication methods. Aircraft capabilities also differ widely, from highly automated, data-rich commercial airliners to slower general aviation planes with basic, analog instruments (Dornheim, 2000; Hansman & Davison, 2000). Such differences can impact an ATC's ability to communicate information to all users and may contribute to varied interpretations of the same data, making it more difficult for pilots, controllers, dispatchers, and analysts to coordinate effectively. For example, a reroute issued via data link to a highly automated commercial aircraft may not be received by a nearby general aviation aircraft without that capability, requiring the controller to relay the same SCI verbally. In rapidly evolving weather or traffic situations, this mismatch in information pathways can create temporal delays, uneven situation awareness across users, inconsistent mental models of the situation, and coordination breakdowns, which can hinder decision-making and elevate operational risk (Endsley & Jones, 1997; Mosier et al., 2017a, 2017b).

To address this complexity, the proposed methodology proposes explicit attention to systematically researching contextual demands in ways that reflect real-world operational conditions. This approach enables assessment of how increasing complexity interacts with cognitive workload, SA, and SSA to influence the likelihood that SCI is noticed, accurately interpreted, and effectively acted upon. The resulting data will highlight the impact of complexity on both individual cognition and shared mental models, even when information is technically accurate and delivered on time (Orasanu et al., 2011; Wiegmann & Shappell, 2017).

6.2.1 Factors Driving Contextual Complexity in the NAS

In the bow-tie risk assessment framework, contextual complexity serves as a set of threat-shaping conditions that influence both the likelihood and potential severity of a decision error related to SCI effectiveness. Even well-timed and accurate safety-critical messages can be ineffective or even counterproductive if they are not delivered during the operational phase in which they are most relevant to the pilot's decision-making. In accordance with FAA Order 7110.65, ATCs are responsible for issuing weather and other essential operational information to flight crews. Delivering this information during phases of flight when it can be readily acted upon, such as approach and landing, when pilot workload is often high, ensures that the message is both operationally useful and cognitively accessible (Brimmer et al., 2023; Beringer & Schvaneveldt, 2002). Conversely, when information is delivered too early, too late, or during periods of high task saturation, its utility can degrade, increasing the likelihood that critical cues are missed or misinterpreted (Bailey et al., 2001; Schvaneveldt et al., 2000). Traffic complexity – which encompasses variability in pilot experience, procedural proficiency, and aircraft performance – interacts closely with equipage differences and routing patterns. These dynamic elements are captured across multiple dimensions of contextual complexity and are further defined in Appendix D.

Understanding these contextual influences is essential for designing prevention and mitigation barriers that ensure SCI is delivered in a way that supports ATC decision-making. This includes tailoring delivery based on expected task load, phase of control (e.g., initial contact, handoff, conflict resolution), and communication infrastructure. Such alignment is critical not only for individual situation awareness but also for maintaining shared mental models across sectors and minimizing coordination gaps during controller transitions.

The proposed study will account for key contextual complexity factors that influence how ATCs perceive, interpret, and respond to SCI. Contextual complexity factors influence workload, timing, and coordination demands, and must be accounted for when evaluating and delivering SCI. These dimensions will be used to design realistic scenarios, stress-test communication pathways, and evaluate SCI prioritization under varied operational conditions. Within the bow-tie risk framework, each factor functions as a threat-shaping condition that can degrade preventive barriers or exacerbate coordination breakdowns if not appropriately managed. Additionally, they will be integrated into variable coding and analytical frameworks to ensure ecological validity and operational relevance. The elements below represent specific contextual dimensions that will be integrated into the proposed study design.

Operation Type

While pilots operate under distinct regulatory categories (e.g., commercial air transport under Part 121, commuter and on-demand operations under Part 135, and general aviation operations under Part 91), ATCs must seamlessly manage traffic across these operational boundaries within shared airspace. Each operation type is associated with different equipage levels, performance capabilities, communication methods, and procedural expectations, which can shape how SCI is communicated and acted upon. For example, controllers may need to account for differences between highly automated, data-linked commercial aircraft and less-equipped general aviation traffic relying solely on voice communications (Hansman & Davison, 2000; FAA, 2025). These operational differences also create variability in urgency, timing, and format of SCI delivery, particularly in dynamic conditions such as convective weather or reroutes. For controllers, especially in en route sectors, this requires rapid assessment of whether critical information has been transmitted during previous sector handoffs and whether outgoing coordination will be constrained by operational uncertainty related to the type of flight operation (Histon & Hansman, 2008; Wickens, 2002).

Aircraft Type

Controllers regularly work with a diverse mix of aircraft types that vary in speed, climb/descent performance, avionics capabilities, and automation. These differences shape the nature and frequency of SCI needed (e.g., advisories for less-equipped aircraft), and the communication strategies required to maintain safety. For example, handling a low-performance aircraft with limited avionics or data link capabilities alongside a high-performance jet requires distinct SCI communication strategies. Differences in flight deck system capabilities and aircraft performance characteristics (e.g., speed, climb/descent profiles, and altitude ranges) can affect how information is transmitted, prioritized, and acted upon. When SCI is not appropriately adapted to these operational differences, controller workload may increase, and communication continuity across sectors may degrade (Hansman & Davison, 2000; Loft et al., 2007; Histon & Hansman, 2008)

Airspace Class

Airspace class defines the level of control and communication protocols in a given sector. In Class B or TRACON environments, dense traffic and layered restrictions require precise, time-sensitive SCI delivery to support separation and sequencing. In contrast, Class E airspace is controlled but not designated as Class A, B, C, or D, and IFR operations are under ATC authority. However, VFR aircraft are not required to establish communication with ATC. Class G airspace, by definition, is uncontrolled, and pilots are not required to communicate with ATC at all. These differences can lead to lower communication density, but greater ambiguity regarding pilot situation awareness, receipt of SCI, and controller expectations for information sharing. For controllers, especially those managing sector boundaries, mismatches in information expectations across airspace classes can result in degraded coordination and missed opportunities for risk mitigation (FAA, 2023a; ICAO, 2018; Wiegmann & Shappell, 2017)

Geographic Location

Controllers working in regions with complex terrain, infrastructure limitations, or constrained airspace (e.g., mountainous sectors or dense terminal areas) face increased contextual demands. These conditions can amplify SCI frequency and urgency, especially related to weather, terrain clearance, or congestion, and influence how controllers prioritize and distribute cognitive effort. Geographic complexity may also increase reliance on visual scanning tools or collaborative decision-making with adjacent facilities.

Phase of Flight

ATC workload and information demands shift with aircraft phase of flight. Approach and landing phases often compress decision timelines and increase SCI density, while climb or cruise phases may afford greater flexibility but require sustained monitoring. For controllers, phase of flight influences how SCI is sequenced and relayed, not just in terms of timing, but also in relation to traffic complexity, reroute constraints, and facility coordination. In high-density environments, the ability to intercept threats through early, well-framed SCI is especially critical during descent and approach.

Weather Conditions

Weather introduces dynamic variables that affect visibility, turbulence, icing, and operational risk. Controllers must process and transmit SCI related to these changes, often under significant time constraints. The rapid onset of convective activity or low ceilings can generate spikes in coordination workload and elevate the consequences of delayed or misunderstood SCI. Poor timing or unclear phrasing of weather-related SCI increases the likelihood that preventive barriers will be bypassed, especially during critical flight phases.

Traffic Density and Volume

Increased traffic density – defined as the concentration of aircraft within a specific sector, fix, or airspace segment – may elevate cognitive load and coordination requirements, particularly in en route sectors where controllers must manage converging flows and multiple handoffs in close proximity. High density can intensify the need to filter, sequence, and prioritize SCI to maintain safe separation and effective tactical decision-making.

In contrast, traffic volume refers to the overall number of aircraft transitioning through a larger airspace or facility over a given time period. High traffic volume can place sustained demands on resources and staffing, increasing the likelihood that small coordination breakdowns or missed SCI transmissions accumulate over time.

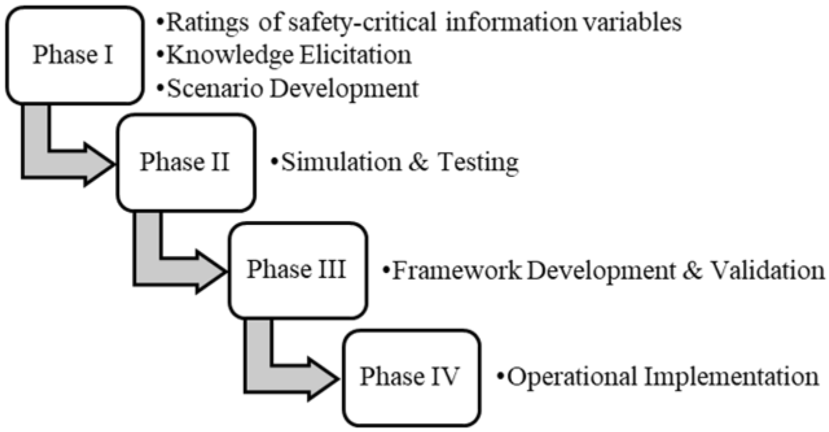
Both density and volume contribute to controller workload, but density tends to drive immediate tactical decision pressures, whereas volume shapes longer-term operational strain and the probability of error under time-sensitive conditions such as weather deviations, reroutes, or emergencies (Damos, 1997; Loft et al., 2007; Histon & Hansman, 2008; Wickens, 2002).

7. Methodology

Given the complex and dynamic nature of ATC operations, this methodology proposes a four-phased, controller-centered design that reflects the operational realities of the NAS and the cognitive demands placed on ATCs. It recognizes the interconnected nature of SCI transmission – including inter-sector coordination, handoff dynamics, and intra-facility communication – and emphasizes the need to define which types of information are most critical at a given time, and how they should be delivered to support accurate, timely, and coordinated decision-making.

Structured into four sequential phases (see Figure 7-1), the approach integrates human factors principles – such as workload, modality compatibility, and situation awareness – with formal risk assessment modeling (e.g., Bow-Tie analysis). This approach ensures that any research outcomes are empirically grounded, operationally relevant, and aligned with the priorities of ATC stakeholders, including facility leadership, training developers, and system designers. The design supports actionable insights into how SCI characteristics affect controller performance and safety outcomes under varying conditions of traffic density, weather, phase of flight, and contextual complexity.

Figure 7-1
Proposed Study Phases



The dynamic, high-consequence communication environment of air traffic control requires more than a one-time analysis or isolated research efforts conducted in organizational silos. Within the NAS, communication is interactive, distributed, and mediated by both human and technical systems, demanding a methodology that can evolve in response to operational variability, cognitive workload, and emerging risks. This proposed approach follows an iterative cycle of elicitation, testing, and refinement, progressively strengthening the preventive and mitigative controls identified through a bow-tie risk model focused on SCI effectiveness.

Phase I centers on knowledge elicitation with certified ATCs and supervisory subject matter experts (SMEs) to identify context-specific hazards, refine critical human factors constructs (e.g., workload, situation awareness, trust in information sources), and rate variables that shape SCI delivery and use. These SME ratings and qualitative insights guide the development of operationally relevant scenarios and serve as foundational inputs for SCI sorting and prioritization frameworks.

Phases II and III incorporate these elicitation results into structured human-in-the-loop simulations, which model realistic operational complexity, including sector traffic load, airspace transitions, environmental conditions, and system constraints. Phase II emphasizes scenario refinement and preliminary evaluation, while Phase III expands simulation fidelity, validates the SCI framework under elevated workload conditions, and captures objective controller performance metrics alongside subjective assessments (e.g., trust, usability, perceived risk).

Phase IV transitions into validation within operational or high-fidelity training environments, confirming that SCI delivery strategies are scalable, contextually appropriate, and applicable to FAA operational needs. Data and insights from all phases support continuous framework refinement, ensuring that the proposed methodology remains evidence-based, controller-centered, and responsive to the unique demands of en route, terminal, and TRACON facilities.

This methodology is intentionally interdisciplinary, integrating the expertise of ATCs, human factors researchers, cognitive scientists, and systems engineers. This collaboration ensures that SCI identification, prioritization, and communication strategies are examined from both the human performance and operational integration perspectives, reducing the risk of overlooking facility-level constraints, workload trade-offs, or inter-sector coordination challenges.

The sampling strategy reflects the distributed yet role-specific nature of SCI within ATC. A stratified design will ensure balanced representation across facility types (e.g., ARTCC, TRACON, Tower), geographic regions, and contextual complexity factors (e.g., weather intensity, traffic volume, phase of flight). All participants will be currently certified or active-duty ATCs, ensuring findings directly reflect the operational environments, procedural requirements, and human-system dynamics that shape SCI performance in the NAS.

7.1. Sampling Philosophy

Each phase of this research engages operational ATCs who are directly involved in the delivery, interpretation, and use of SCI. To ensure representation across the full range of ATC operational contexts and SCI risk levels, a stratified sampling strategy is proposed (see Appendix E).

Stratification Dimensions

Stratification is based on three primary dimensions:

1. Facility Type: Includes En Route (ARTCC), Terminal Radar Approach Control (TRACON), Tower, and CERAP. FAA Contract Towers are treated as a distinct stratum due to differences in equipment, procedures, and oversight.
2. SCI Risk Level: Classified as Very High, High, Medium, or Low, based on the likelihood and consequence of failures in SCI delivery, timing, or interpretation (CCPS, 2020; Rasmussen, 1997; Endsley, 1995). SCI risk levels are informed by FAA documentation on timeliness, operational relevance, and filtering complexity (FAA, 2023a). While the FAA Pilot's Handbook of Aeronautical Knowledge does not define SCI, its guidance on risk management and operational decision-making provides contextual framing for the way information may influence flight safety outcomes.
3. Operational Environment Factors: Includes traffic density, sector complexity, access to automation, airspace class, and frequency of high-risk coordination events, such as weather reroutes or aircraft handoffs across sectors.

This approach aligns with established safety risk assessment frameworks, which define risk as a function of probability and severity (CCPS, 2020) and is particularly applicable to ATC environments where delayed or misaligned information can degrade SA and SSA (Endsley, 1995).

Sampling Parameters

Population size estimates for each stratum are derived from internal FAA records and facility staffing data. Sample sizes are calculated using a 95% confidence level with a ±5% margin of error, applying finite population correction as appropriate. High-population or high-risk strata (e.g., ARTCC or TRACON facilities in major metro areas) will be sampled at standard levels (~370–384 participants), while smaller or higher-risk groups (e.g., CERAP or contract towers) may be oversampled to retain statistical power for subgroup analyses.

In resource-constrained scenarios, a ±7% margin of error (sample sizes ~175–200 per stratum) may be acceptable while still yielding actionable and generalizable results. Sampling will also account for regional diversity, sectoral complexity, and variability in technology use to ensure findings are grounded in the full range of NAS operational conditions.

Participants must meet all inclusion criteria to be eligible for this study. Specifically, they must be currently certified and actively serving as an ATC within the FAA or an FAA Contract Tower. All participants must also be at least 18 years of age. In addition, their current job responsibilities must include direct involvement in the interpretation, coordination, or delivery of SCI as part of their routine operational duties. These criteria ensure that all participants have relevant, up-to-date experience and are positioned to provide accurate insights into SCI communication and decision-making within real-world air traffic control environments.

Exclusion criteria include retirees, trainees, and non-certified support staff. This ensures that responses reflect current operational experience and directly inform SCI system design, prioritization models, and interface usability.

With the participant framework established, Phase I of the research serves as the empirical foundation for the full methodology. This initial phase is designed to ground all subsequent phases in operational realities and validated human factors concepts specific to air traffic control. As introduced in Figure 1, the bow-tie model serves as a conceptual scaffold to guide hazard identification during SME elicitation, helping structure the relationships among SCI threats, workload, and SA degradation, and potential operational consequences. Although it will be used to guide the initial hazard identification and construct development, this methodology is not proposing a validation for the bow-tie model itself. Rather, the constructs identified through it – such as workload, SA degradation, and information characteristics – are evaluated through scenario testing to inform practical decision-support tools.

7.2. Phase I: Establishing the Empirical Foundation

Phase I introduces the research through qualitative and quantitative elicitation methods to empirically ground the operational definition of SCI and the contextual variables that shape how it is perceived, prioritized, and operationally acted upon. This phase focuses on capturing the lived operational experience of ATCs and related SCI-interfacing roles to identify the decision factors, heuristics, and environmental conditions that most influence how information is sorted, filtered, and communicated under real-world constraints.

Structured interviews and a decision-support survey are proposed to systematically explore how operational users distinguish between safety-critical and non-critical information, how they perceive and respond to contextual cues such as workload, traffic complexity, and airspace class, and how these factors influence prioritization and coordination behaviors.

By drawing from a broad and representative sample of ATCs (Tower, TRACON, ARTCC) and other relevant users of SCI, Phase I ensures coverage of the full communication and decision-support chain that interacts with ATCs. Findings from this phase provide the empirical foundation for scenario design in subsequent simulation phases, particularly by clarifying which information types and contextual conditions most reliably affect perceived criticality, workload, and decision quality.

The goal of Phase I is to produce a validated and context-sensitive decision-support framework that serves as the basis for empirical testing in Phase II.

7.2.1 Research Questions

1. How do operational users distinguish between safety-critical and non-safety-critical aeronautical information in real-world contexts?
2. Which factors (e.g., consequence magnitude, time sensitivity, traffic complexity, procedural risk) most influence SCI prioritization decisions?
3. How do contextual variables – such as airspace class, sector type, operational complexity, and workload – shape perception and subsequent action?
4. What communication patterns, prioritization heuristics, and information filtering strategies are used across facility types (e.g., TRACON, Tower, En Route), and how do they influence SCI effectiveness?

7.2.2 Participants

A broad and representative sample will be drawn from key operational roles directly involved in the communication, interpretation, coordination, and analysis of SCI. These roles include FAA and FAA Contract Tower ATCs (across Tower, TRACON, ARTCC, and CERAP facilities). Together, these participants represent the full spectrum of ATC perspectives and analytical functions responsible for handling, prioritizing, and disseminating aeronautical information across the NAS.

While this research is specifically scoped to ATC roles, it is important to acknowledge that safety-critical aeronautical information is managed across multiple operational domains, including pilots, dispatchers, flight service specialists, and military ATC personnel at joint-use facilities. Due to funding constraints, institutional agreements, and the primary mission of this research, the participant sample is intentionally limited to FAA-certified and FAA Contract Tower controllers. This focused scope ensures methodological alignment with sponsor priorities and supports analytic clarity. Future extensions of this methodology may incorporate additional user groups to evaluate cross-role applicability and identify interface-level challenges.

Sample size targets for each group are calculated to achieve a 95% confidence level, with a ±5% or ±7% margin of error, depending on subgroup size and criticality. Finite population correction is applied where appropriate (Cochran, 1977; Israel, 1992).

7.2.3 Data Collection Instruments

Two instruments detailed in Appendix F will be used:

- A structured interview protocol designed to elicit mental models, communication strategies, heuristics, and perceptions of information salience under varying conditions.

- A decision-support survey, developed from validated variables, to capture ATC judgments about the relevance and safety-criticality of various aeronautical information types across flight phases and operational scenarios.

These tools will provide both qualitative and quantitative insights into controller decision-making and SCI prioritization in dynamic environments.

1. Structured Interviews

All interviews will be conducted by trained researchers using a standardized protocol to ensure consistency and reduce interviewer bias. Participants will review and refine a set of decision-support variables and independent contextual factors previously identified through literature review. Interviews will focus on:

- How SCI is defined and differentiated from non-critical information across roles;
- The contextual variables (e.g., traffic load, phase of flight, weather, workload) that influence SCI salience;
- The challenges faced in interpreting or communicating SCI under time pressure;
- Operational workarounds, heuristics, and decision-making strategies used during routine and non-routine operations.

Each participant will also review a pre-developed set of eight decision-support variables and eleven contextual variables related to SCI. Their feedback ensures these constructs are clearly defined, practically applicable, and aligned with the real-world conditions experienced across NAS positions. Responses will be audio recorded, transcribed, and thematically analyzed. These findings will shape scenario design for simulation and refine the operational definitions of SCI variables.

2. Decision-Support Survey

Following SME validation, the decision-support survey will be distributed to a broader sample of ATCs through a secure online platform. The survey will gather controller ratings on:

1. The perceived safety-criticality of various information types;
2. The operational phase(s) during which each information type is most critical;
3. Judgments on the usefulness, urgency, and optimal delivery method for SCI under different conditions.

After variable validation, a decision-support survey is deployed to a broader population sample across the same operational roles. The survey captures how each participant rates the safety-criticality of different aeronautical information types across operational conditions and flight phases. Participants also indicate during which operational phases (e.g., preflight planning, en route, coordination handoff, adverse weather response) they believe each information type is most critical. Controllers do not directly participate in formal preflight planning, but they may play a supporting role by providing or confirming time-sensitive aeronautical information prior to departure, especially when operational changes arise or pilots request clarification.

The survey includes Likert-type scales and open-ended questions to capture both quantitative patterns and contextual nuance. Results are used to validate and refine the criticality assessment framework and ensure phase-of-flight alignment for each SCI type across user roles. All data will be collected anonymously.

7.2.4 Data Analysis

SME Validation Analysis

Responses from SMEs will be analyzed to evaluate the clarity, consistency, and operational relevance of the survey items. Descriptive statistics, such as means and standard deviations, will be calculated for each decision variable across all aeronautical information types. Inter-rater agreement will also be examined to assess consistency within and across roles, using metrics such as intraclass correlation coefficients or Kendall's W to gauge consensus on variable importance. Phase-of-flight selections for each information type will be aggregated to identify patterns and areas of cross-role agreement, with items showing low agreement or inconsistent interpretation flagged for revision. In addition, qualitative feedback from open-ended SME responses will be thematically coded to refine item wording, improve construct clarity, and ensure comprehensive content coverage. Findings from this validation phase will inform revisions to the survey instrument prior to deployment to the broader operational population.

Findings from the SME validation phase will guide any necessary revisions to item structure before deploying the instrument to the broader operational population.

Full Sample Analysis

Once the finalized survey is administered to the full participant sample, the data will be analyzed to identify patterns in perceived safety-criticality, decision-support factors, and contextual relevance. Descriptive statistics will be calculated for each decision variable across information types, while group comparisons using t-tests or ANOVAs will examine differences by role or experience level. Exploratory factor analysis will assess the underlying structure of the decision variables and evaluate construct validity, and cluster analysis will be used to identify natural groupings of information types based on rating patterns across decision variables and flight phases. Cross-tabulations and chi-square tests will further explore associations between information types and selected phases of flight, highlighting when particular information is most often judged as safety-critical.

Where sample size allows, multivariate analyses such as regression or MANOVA may also be conducted to evaluate how combinations of decision variables predict phase-of-flight associations or criticality scores. Together, these results will refine the decision-support framework by pinpointing the variables that most reliably predict perceived safety-criticality, clarifying role- and context-specific differences, and guiding calibration of the tool for distinct operational phases and use cases.

The results will inform refinement of the decision-support framework by identifying which variables most consistently predict perceived safety-criticality, how perceptions differ by role and context, and how the tool can be calibrated for specific operational phases or use cases. Should future research expand the stakeholder base to include other operational roles (e.g., pilots, dispatchers, flight service personnel), the methodology will require re-validation to ensure accurate mapping of role-specific barriers, responsibilities, and risk interactions, in order to support a more comprehensive systems-level view of SCI handling across the NAS.

7.3. Phase II: Simulation Testing and Assessment of Human Factors Threats

Phase II transitions the research from qualitative elicitation to empirical evaluation by implementing human-in-the-loop simulations to examine how delivery method, timing, and operational context influence ATC performance, workload, and error rates during high-consequence, real-time operations. This phase is designed to reflect the operational realities of controller roles and other SCI-interfacing positions that must manage complex and dynamic communication environments.

Scenarios will incorporate realistic conditions derived from Phase I findings and will be structured to model variability in airspace class, traffic volume, operational phase, and environmental constraints. While ATCs will serve as the primary participants in high-fidelity simulation trials, pre-operational support roles – such as Pre-Flight Briefers, Inflight Specialists, Flight Data Coordinators, Weather Analysts, and System/Safety Analysts – will contribute to scenario development and parallel testing of coordination tools and delivery mechanisms that feed into controller workflows.

The goal of Phase II is to identify which combinations of delivery method and timing optimize task success and minimize cognitive burden, thereby generating empirical evidence for the prioritization logic and decision-support modeling to be developed in Phase III.

7.3.1 Research Questions

1. How do timing, delivery method, and operational context affect controller workload across different phases of flight operations?
2. What delivery methods (e.g., voice, CPDLC, automated tools) are perceived as most usable and effective under varying levels of cognitive demand?
3. Which delivery conditions result in the most timely, accurate responses and lowest error rates when processing safety-critical information?

7.3.2 Hypotheses

- H1: SCI delivered closer to the relevant operational decision point will result in lower workload and higher task performance.
- H2: Delivery methods requiring fewer manual interactions (e.g., voice delivery) will reduce task load compared to digital/app-based delivery, particularly under high workload conditions.
- H3: Perceived usability, trust, and performance will vary significantly by delivery modality and timing, with human-mediated channels expected to outperform lower-integration digital systems in high-complexity scenarios.

7.3.3 Participants

Participants for Phase II will be drawn from a targeted sample of operational roles identified as central to the communication, coordination, and interpretation of SCI within flight operations. These include FAA and FAA Contract Tower ATCs—across Tower, TRACON, ARTCC, and CERAP facilities—as well as Pre-Flight Briefers, Inflight Specialists, Flight Data Coordinators, Weather Analysts, and Safety/System Operations Analysts. While the core simulation trials will focus primarily on real-time control environments involving ATCs, the additional roles will contribute through parallel scenario reviews, structured tool-based simulations, and communication-chain exercises designed to capture the broader context of SCI delivery.

This inclusive, role-specific sampling approach ensures that Phase II findings are grounded in the operational realities of those who directly or indirectly influence SCI prioritization and uptake under high-tempo conditions. Each role provides distinct insight into the flow and filtration of SCI at different stages of the operational timeline – from preflight briefing to real-time control and post-event safety monitoring. Sampling targets for each group are provided in Appendix E.

7.3.4 Independent Variables

The independent variables (IVs) are organized into two broad categories: SCI Information Characteristics and Contextual Complexity Variables. Each IV was chosen based on established human factors theory, aviation guidance, and empirical findings related to cognition, communication, and risk perception in time-sensitive operational environments.

The three primary SCI-related variables, volume, relevance, and timeliness, will be manipulated within controlled scenarios to simulate realistic variations in how information is delivered. These dimensions represent key cognitive load drivers, and their thresholds will be informed by SME input and literature on memory capacity, workload interference, and signal-to-noise ratio modeling. Contextual complexity variables encompass real-world operational factors such as aircraft type, weather, airspace class, and phase of flight, and should be incorporated either as manipulated elements in scenario design or as covariates in data analysis to reflect naturalistic complexity gradients.

All independent variable levels are described below, with full detail and thresholds listed in Appendix D. The two major categories of IVs include:

1. Information Characteristics

Volume of Information

Volume is defined as the number of discrete safety-critical information elements presented to participants within a decision window, relative to cognitive processing capacity and available time. This variable is grounded in working-memory research, with thresholds informed by Cowan's (2008) estimate of 3 – 5 cognitive chunks and Miller's (1956) well-known 7 ± 2 range. High-volume conditions are designed to exceed these limits to simulate overload. These thresholds span below, near, and above established human capacity constraints and are consistent with visual display guidance recommending partitioning or filtering when data exceed the capacity of a single frame (Mejdal et al., 2001).

Relevance of Information

Relevance reflects the degree to which information elements are actionable, accurate, and appropriate for the operator's task, role, and timing. High relevance supports quick and accurate decision-making, whereas irrelevant or low-priority information can increase workload and delay appropriate responses (Endsley & Jones, 1997; Orasanu et al., 2011). Relevance bands should be selected to create strong contrasts for experimental inference and are aligned with human factors recommendations that discourage unnecessary or distracting information in high-risk environments (U.S. DoD, 2012).

Timeliness of Information

Timeliness refers to whether the SCI is delivered at a moment when it can be cognitively and operationally processed before the decision point. Even accurate information becomes ineffective or counterproductive if delivered too late under time pressure (Endsley, 1995; Orasanu et al., 2002). Manipulations reflect principles from task-analysis-based scenario design and cognitive workload theory (Wickens & Hollands, 2013). Timing windows for each scenario should be pre-defined by SMEs based on the task phase.

2. Contextual Complexity Variables

Contextual complexity refers to the dynamic operational factors that affect how safety-critical information is interpreted and acted upon. These include real-world variability in operations, aircraft systems, airspace, and environmental conditions. Such factors are known to compound cognitive workload and affect communication efficacy (Orasanu et al., 2011; Rasmussen, 1997). The categories listed below are based on regulatory definitions, airspace operations data, and NTSB-reported factors associated with communication breakdowns and workload spikes. Each factor has been shown to influence either the likelihood or clarity of SCI transmission and uptake (Orasanu et al., 2011; FAA, 2023a–d). Future work may also consider pilot experience and communication proficiency as additional dimensions of contextual complexity, particularly as these factors may influence the clarity, timing, and coordination of SCI exchanges in dynamic operational environments. Variables include:

- Operation Type
- Aircraft Type
- Airspace Class
- Geographic Location
- Phase of Flight
- Weather Conditions
- Traffic Density
- Traffic Volume

These contextual features should either be manipulated in the scenario design or recorded as covariates during data analysis. Their inclusion allows for ecological validity and supports the generalizability of the findings to real NAS conditions.

7.3.5 Measures

Phase II builds upon the findings of Phase I by empirically testing how timing, delivery method, and operational context influence task performance, cognitive workload, and other safety-related outcomes in a controlled simulation environment. Measures are aligned with the constructs identified in Phase I and assessed using standardized instruments, observational coding, and participant feedback.

This phase uses controlled, scenario-based simulations to create realistic operational environments while allowing precise measurement of participant behaviors, subjective perceptions, and errors. The measures and instruments in Phase II are designed to capture both objective and subjective outcomes aligned with the constructs identified in Phase I, ensuring continuity across research phases.

To assess the constructs defined in the preceding section, the proposed work employs a combination of validated, standardized, and custom-developed instruments. The measures listed below and detailed in Appendix G were selected or designed to align closely with the study's measures, ensuring both conceptual relevance and operational feasibility. Each measure was chosen to capture either objective behavioral data, subjective perceptions, or qualitative insights in a way that reflects the demands of realistic operational contexts. The following subsections describe each instrument, its purpose, what it measures, and how it is administered within the research. Together, these tools enable a comprehensive evaluation of how SCI delivery affects task performance, workload, usability, and decision-making.

1. Cognitive Workload

To evaluate the way SCI impacts task execution under varying workload conditions, this study incorporates the Aviation Task Workload Impact Tool (ATWIT), a structured workload estimation framework adapted from validated models in air traffic and flight operations research (Bainbridge & Dorneich, 2016; Loft et al., 2007; Roth et al., 2021). The ATWIT supports scenario-based quantification of how SCI affects timing alignment, task prioritization, and decision-making capacity in real time. Scoring criteria include the degree of task disruption, immediacy of required action or coordination, and cognitive interference based on operational context.

ATWIT prompts participants to rate their perceived workload at fixed intervals or in response to specific scenario events using a single-dimension scale. This approach is well-suited for dynamic and time-sensitive environments such as ATC, where workload fluctuates in response to unpredictable traffic patterns, coordination demands, and sector complexity. By capturing fine-grained temporal workload patterns, the tool enables mapping of cognitive load relative to SCI presentation events and supports validation of the timing-alignment criteria central to this study's methodology.

i. Procedure

ATCs will receive auditory or on-screen prompts every 2–3 minutes during the scenario or immediately following a key SCI event. Participants will respond via a keypad or touchscreen interface without pausing the task. Responses will be timestamped and synchronized with scenario timelines to allow for precise alignment with SCI delivery characteristics.

ii. Scoring

ATWIT scores will be collected in real time during simulation trials, with participants providing single-dimension workload ratings on a scale ranging from 1 (no workload) to 7 (extremely high workload) at fixed intervals or immediately following salient SCI delivery events. Each response will be timestamped and synchronized with the scenario timeline to allow precise alignment between perceived workload and key operational events (Stein, 1985).

iii. Data Analysis

To analyze these data, workload ratings will first be plotted as individual time-series curves, enabling visualization of temporal patterns and identification of workload peaks and valleys relative to information delivery characteristics. For statistical analysis, repeated-measures ANOVA will be conducted to examine within-subject effects of SCI variables (e.g., volume, timing, modality) on workload scores across timepoints. Mixed-effects models may also be employed to account for random effects associated with individual differences and scenario variation. This approach supports detailed, context-specific interpretation of workload dynamics and provides empirical insight into the cognitive impact of different SCI configurations in high-tempo operational environments. These patterns will be used to identify high-risk conditions for overload and degradation of performance.

2. Situation Awareness

Situation awareness refers to an operator's ability to perceive relevant elements in the operational environment, comprehend their meaning, and project their status into the near future (Endsley, 1995). SA is critical to safe and effective performance in dynamic domains such as aviation, particularly when operators must rapidly interpret and act on SCI under time pressure and varying workload. This research will assess SA using a dual-method approach: a subjective post-scenario rating and objective performance-based indicators.

i. Procedure

The Situation Awareness Rating Technique (SART) will be administered immediately after each simulation scenario concludes. Participants will be prompted to complete the SART before engaging in any debriefing or completing other measures, to preserve the immediacy and integrity of their perceptions. The instrument consists of nine items presented on a 7-point Likert scale, each corresponding to one of three dimensions: attentional demand, attentional supply, and situational understanding. Participants will complete the ratings individually and confidentially using paper forms or digital survey software (e.g., Qualtrics), depending on the simulation setting. Researchers will ensure consistent administration by providing standardized written and verbal instructions, and no discussion among participants will be allowed during completion to avoid bias or influence. The process takes approximately 3–5 minutes per scenario. Once completed, SART data will be collected and linked to scenario metadata, such as condition type (e.g., SCI modality, volume), to support analysis of how situational factors influence perceived awareness.

To complement self-reported SA, trained observers will code behavioral indicators of SA using structured logs during real-time observation and scenario playback. Observers will identify specific behaviors aligned with Endsley's (1995) three levels of SA: (1) perception of relevant information (e.g., recognition of hazards), (2) comprehension of meaning (e.g., identifying system status or conflicts), and (3) projection of future status (e.g., proactive actions based on expected developments). Indicators will include timeliness and correctness of responses, accuracy of verbal or system-based cues, and anticipatory procedural steps. This method reflects validated techniques for performance-based SA measurement in aviation and other safety-critical domains (Salmon et al., 2006; Stanton et al., 2005).

Observers will receive extensive scenario-specific training using coded examples and calibration exercises. Independent ratings will be obtained from at least two observers per scenario, and inter-rater reliability will be assessed using Cohen's Kappa (κ), with $\kappa > 0.75$ considered the threshold for acceptable agreement (Bakeman & Gottman, 1997). Observed SA performance will be categorized and scored as "correct," "partially correct," or "incorrect" and aggregated across scenarios. These ratings will be analyzed alongside subjective SART scores and performance data to provide a multidimensional picture of SA under varying SCI conditions.

ii. Scoring

Responses on the SART will be scored using the established formula developed by Taylor (2017), where the composite score is calculated as:

SA Composite = Understanding (U) + Supply of Attention (S) – Demand on Attention (D).

Each of the nine items corresponds to one of the three core dimensions: Attentional Demand (e.g., instability, complexity), Attentional Supply (e.g., concentration, spare mental capacity), and Situational Understanding (e.g., information quality, clarity of overall situation). Items are rated on a 7-point Likert scale, and scores for each dimension will be summed and used to compute the overall composite score. This structure has been validated in a range of operational domains, including aviation (Selcon et al., 1991), and its alignment with Endsley's (1995) three-level SA model makes it well-suited for assessing perceived awareness in dynamic environments. Internal consistency will be assessed via Cronbach's alpha for each SART dimension to evaluate reliability.

iii. Data Analysis

To assess perceived SA across conditions, descriptive statistics such as means, medians, standard deviations, and interquartile ranges will be calculated for the composite and dimension-level SART scores.

To examine the influence of SCI characteristics (e.g., volume, modality, and timeliness) on SA, linear mixed-effects models will be employed, with fixed effects for condition and random intercepts for participants to account for repeated measures within individuals (West et al., 2015). In addition, correlational analyses will be performed to explore relationships between SART scores and other dependent variables, including cognitive workload (e.g., ATWIT), trust in information source, and task performance metrics. The triangulation of subjective and objective data will strengthen construct validity, as recommended in SA literature (Endsley, 2015; Salmon et al., 2009). Exploratory regression and interaction analyses may also be conducted to identify potential moderation effects of workload on SA.

3. Shared Situation Awareness

Shared Situation Awareness is the degree to which team members possess the same SA on shared SA requirements, that is, the overlapping portion of situation awareness required across team roles (Endsley & Jones, 1997).

In high-tempo, safety-critical environments such as aviation, SSA supports coordinated action by aligning team members' mental models, threat assessments, and task priorities, thereby reducing communication breakdowns, errors, and delays (Endsley, 1995a; Endsley, 1998).

This work adopts a triangulated measurement approach to assess SSA from subjective, objective, and behavioral standpoints. Specifically, SSA will be evaluated using:

- Team Situation Awareness Global Assessment Technique (TSAGAT)
- SSA Accuracy Scores (from freeze-probe agreement across teammates)
- Communication Alignment Coding

This mixed-method design enables a comprehensive analysis of both perceived and observed SSA under varying safety-critical information delivery conditions.

i. Procedure

SSA will be assessed using a multi-method approach following each experimental scenario, combining subjective self-report measures, objective freeze-probe convergence scores, and structured behavioral coding of team communication. First, participants will complete a post-task questionnaire adapted from the Team Situation Awareness Global Assessment Technique (TSAGAT), a validated instrument developed for high-stakes domains such as aviation and trauma care (Endsley, 1995a; Endsley & Jones, 1997). Items are presented on a 7-point Likert scale and assess perceived alignment with teammates in understanding current operational states, anticipating future events, and prioritizing actions. Sample items include statements such as "My teammates and I had the same understanding of the current situation," and "We anticipated the same future events and outcomes." Responses will be collected electronically and analyzed both at the item level and in aggregate to assess perceived SSA.

In addition to self-reports, objective measures of SSA accuracy will be captured through freeze-probe events embedded within each simulation scenario. These brief pauses prompt each participant to independently respond to targeted questions about the current operational state, such as hazard locations, system status, or likely next steps. Pairwise agreement scores will be computed for each team using weighted metrics of correctness (alignment with scenario ground truth) and consistency (alignment between teammates' responses), providing an in-scenario measure of mental model convergence (Gorman et al., 2005; 2006).

Finally, audio recordings from each scenario will be transcribed and analyzed using a structured communication coding scheme to identify behavioral indicators of SSA. Coders will review verbal exchanges for markers such as confirmation statements, shared terminology, anticipatory coordination, and frequency of corrective interactions. Each exchange will be rated for alignment status, and the percentage of aligned interactions will be calculated for each team. Coders will be trained on annotated exemplars, and inter-rater reliability will be assessed using Cohen's Kappa, with values ≥ 0.75 considered acceptable (Bakeman & Gottman, 1997). This triangulated approach to measuring SSA, incorporating subjective perceptions, objective accuracy, and behavioral evidence, ensures a robust evaluation of team-level situational understanding in dynamic, high-stakes environments.

ii. Scoring

TSAGAT Scores: Each participant's responses will be averaged across items to yield a composite SSA perception score. Higher scores indicate stronger perceived alignment. Subscale scores (e.g., shared threat perception, procedural coordination) may be extracted through factor analysis if warranted.

SSA Accuracy Scores: For each freeze-probe event, pairwise comparisons between team members will be scored using a 3-point system: exact match (2), partial match (1), and mismatch (0). These scores will be averaged per scenario to yield a quantitative measure of convergence in SA.

Communication Alignment: The total number of aligned exchanges will be divided by the total number of exchanges to compute a percentage alignment score. High scores reflect fluent, efficient communication indicative of shared mental models and mutual awareness.

iii. Data Analysis

To analyze shared situation awareness, a multi-tiered data analysis strategy will be employed. Descriptive statistics, including means, standard deviations, and distribution patterns, will be calculated for each SSA metric across experimental conditions to provide a foundational understanding of performance trends and variability among participants and teams.

For inferential analysis, mixed-effects regression models will be used to evaluate the impact of safety-critical information characteristics, specifically modality, volume, and timeliness, on SSA outcomes. These models are well-suited for repeated-measures data and will include participant and team as random effects to control for intra-group variation. Fixed effects will assess the direct influence of SCI variables and their interactions with cognitive workload.

To explore interrelationships between constructs, correlational analyses will be conducted to examine associations between SSA measures and other dependent variables, including ATC cognitive workload (ATWIT), and task performance metrics. Depending on the distribution of each variable, Pearson's or Spearman's correlation coefficients will be applied.

An exploratory multivariate pattern analysis (MPA) will be performed to detect combinations of SCI characteristics and team behaviors that are predictive of SSA breakdowns. This approach will focus on identifying complex interaction effects, such as the compounding impact of modality-switching, delayed SCI delivery, and ambiguous cues on team-level situational alignment.

Finally, qualitative thematic analysis will be used to interpret communication alignment data. Transcribed scenario recordings will be clustered thematically to identify recurring linguistic patterns, coordination strategies, and failure modes related to SSA. This qualitative layer of analysis will enrich the quantitative findings by providing context around how shared understanding is built, or breaks down, during dynamic task execution.

4. Trust in Information Source

Automated Source

Trust in an automated system is a critical determinant of whether operators rely on its outputs, particularly in dynamic, high-risk environments such as aviation.

In the context of this work, Trust in Information Source refers to the degree to which pilots and ATCs perceive the SCI delivery system as competent, reliable, transparent, and supportive of their goals. This construct draws from Lee and See's (2004) theoretical foundation for trust in automation and is operationalized using a psychometrically validated measure developed by Koerber et al. (2005).

Unlike interpersonal trust, which depends on relational cues and social dynamics, trust in automated systems is shaped by perceptions of performance consistency, system integrity, transparency, and goal alignment. Over- or under-trust can lead to inappropriate reliance behaviors, such as rejecting valid alerts or failing to notice faulty guidance, especially under workload or time pressure (Lee & See, 2004). Thus, measuring trust in the information source is essential for evaluating whether users are likely to use or disregard the SCI system during operations.

The Koerber scale is specifically designed for dynamic environments where operators must decide, often under pressure, whether and to what extent to rely on automation. This aligns closely with the needs of this research, where high cognitive workload, shifting priorities, and distributed coordination make calibrated trust essential for safety (Koerber et al., 2005; Lee & See, 2004).

The Koerber Trust in Automation Scale is a 15-item questionnaire grouped into three subscales:

1. Performance-based Trust: Perceptions of system competence, reliability, and consistency.
2. Process-based Trust: Beliefs about the system's transparency, explainability, and observability.

3. Purpose-based Trust: Degree to which the system is perceived as aligned with the operator’s goals.

Participants will rate their agreement with each statement using a 5-point Likert scale ranging from 1 (Strongly disagree) to 5 (Strongly agree). All items will be presented in randomized order to reduce order effects.

i. Procedure

Participants will complete the measure immediately after completing their assigned scenarios and associated workload measures. This placement is intentional to capture perceptions of system trust following interaction with the information delivery method under task conditions simulating real operational stressors.

Instruction should be given to focus on their experience with the system, providing the safety-related information, not on human colleagues or traditional channels. Instructions will clarify that "system" refers to the automated SCI display or interface that filtered, sorted, and prioritized information during the scenario.

ii. Scoring

Scoring of the Koerber Trust in Automation Scale involves calculating individual subscale scores for Performance-based Trust, Process-based Trust, and Purpose-based Trust by averaging responses within each subscale. A total trust score is derived by averaging all 15 items, providing an overall measure of the participant’s trust in the automated SCI delivery system. Higher scores reflect greater trust in the system’s competence, transparency, and alignment with operator goals. Internal consistency for each subscale will be evaluated using Cronbach’s alpha ($\alpha \geq 0.70$ deemed acceptable), and any subscales failing to meet this threshold will be flagged for potential revision in subsequent phases.

iii. Data Analysis

Internal consistency reliability will be assessed for each subscale using Cronbach’s alpha ($\alpha >$ .70 acceptable). Descriptive statistics (means, standard deviations) will be reported for each subscale and total trust score. Comparative analyses (e.g., ANOVAs or t-tests) will be used to assess differences in trust scores across conditions that vary by SCI delivery modality, timing, or relevance. Correlational analyses will examine relationships among trust and other key constructs such as perceived workload (ATWIT), shared situation awareness (TSAGAT, convergence scores), and information usability (PSSUQ), providing insight into the role of trust as a mediator or moderator of SCI effectiveness in operational contexts. Between-group comparisons will be tested using independent samples t-tests or ANOVAs, depending on the final study design. If SCI delivery methods are varied across scenarios, trust levels will be compared across conditions to evaluate how features such as information timeliness, volume, modality, or relevance influence operator trust in the system.

Human Source

To assess participants' trust in human sources of safety-critical information (e.g., pilots, ATCs, dispatchers, and other operational collaborators), this research adapts McAllister's (1995) dual-dimensional trust scale. The original scale differentiates between cognition-based trust (belief in competence, reliability, and role-based integrity) and affect-based trust (emotional connection and willingness to be vulnerable in communication).

This distinction is particularly relevant in distributed aviation operations where personnel must rapidly coordinate across organizational boundaries, often without prior relationships.

The adapted measure has been contextualized for safety-critical environments to reflect aviation-specific terminology and operational realities (e.g., verbal coordination, NOTAM delivery, PIREPs, etc.). Items have been modified to focus on participants' perceptions of the professionalism, reliability, and interpersonal responsiveness of the human source who delivered or communicated the SCI during the task scenario.

The full list of adapted items appears in Appendix G and includes statements such as:

- I believe this individual communicates operational information with professionalism and accuracy.
- If I raised a concern or question, I believe they would respond in a thoughtful and supportive way.

iv. Procedure

Participants will complete the questionnaire immediately following each task scenario. Each scenario will be structured to include at least one human-communicated instance of safety-critical information (e.g., reroute request, hazard advisory, clearance correction), allowing participants to evaluate a specific interpersonal source based on their real-time experience.

The questionnaire will be administered digitally via a secure Qualtrics platform, using a 5-point Likert scale for each item (1 = Strongly Disagree, 5 = Strongly Agree). Items will be randomized to reduce order effects. Separate versions will be used depending on the role of the human information source (e.g., ATC, CIC), but core content and scale structure will remain consistent across participant types.

v. Scoring

Responses will be grouped into two subscales reflecting McAllister's (1995) original constructs:

- Cognition-Based Trust Subscale (6 items)
- Affect-Based Trust Subscale (5 items)

Subscale scores will be computed by averaging item responses. A composite trust score may also be used in exploratory analyses. Reverse-coded items (if retained) will be adjusted during scoring to ensure directional consistency.

Higher scores indicate greater trust in the human source of information across cognitive and emotional domains.

vi. Data Analysis

Descriptive statistics (mean, SD, range) will be reported for each subscale and the total trust score. Reliability analysis (Cronbach's alpha) will be conducted separately for cognition-based and affect-based subscales to assess internal consistency.

To examine the impact of trust in human sources on SCI use and decision-making, trust scores will be used as independent variables or mediators in regression models predicting outcomes such as:

- Time to action
- Accuracy of response
- Perceived information relevance
- Situation awareness scores

In Mixed-Methods Analyses, trust data may be triangulated with qualitative comments from structured interviews about perceptions of credibility, communication quality, or breakdowns in coordination.

5. Usability of Delivery Methods

To evaluate participant perceptions of system usability following scenario completion, this research will employ the Post-Study System Usability Questionnaire (PSSUQ; Lewis, 1992, 1995). Widely validated across domains, the PSSUQ is a standardized post-use instrument designed to assess users' subjective evaluation of a system's ease of use, information presentation, and interface responsiveness. It is especially well-suited for evaluating complex systems used in high-stakes operational environments, such as air traffic control and aviation navigation systems (Fruhling & Lee, 2005; Lewis, 1995).

The PSSUQ comprises three subscales:

1. System Usefulness (SYSUSE): assesses the general effectiveness and efficiency of the system
2. Information Quality (INFOQUAL): evaluates the clarity, format, and supportiveness of the information provided
3. Interface Quality (INTERQUAL): captures the intuitiveness and consistency of the interface design

Each item is rated on a 7-point Likert scale (1 = Strongly Agree to 7 = Strongly Disagree), with an option for "Not Applicable." Lower scores reflect higher perceived usability. This structure allows fine-grained analysis of specific usability factors and their potential impact on operator performance in the context of SCI delivery.

i. Procedure

Participants will complete the PSSUQ immediately following each scenario. This timing ensures that responses reflect immediate impressions of the system's usability, contextualized by the specific scenario just completed. Instructions should be given to reflect only on the interface and information systems used during the most recent scenario. The questionnaire will be administered via a digital form that logs response time and preserves item order to avoid ordering effects, and responses will be stored anonymously for analysis. No items will be omitted or modified, ensuring comparability with previous validation efforts.

ii. Scoring

PSSUQ responses will be scored by calculating the mean score for each of the three subscales—SYSUSE, INFOQUAL, and INTERQUAL—as well as an overall total usability score. Scores range from 1 (Strongly Agree) to 7 (Strongly Disagree), with lower scores indicating more favorable usability perceptions. "Not Applicable" responses will be excluded on a per-item basis, and participants with more than 20% missing data will be removed from subscale analysis to maintain psychometric validity. Internal consistency of each subscale will be assessed using Cronbach's alpha, with $\alpha \geq 0.70$ indicating acceptable reliability.

iii. Data Analysis

Descriptive statistics (means, standard deviations) will be reported for all usability measures across participant groups. Group differences (e.g., pilot vs. controller) and between-scenario comparisons will be evaluated using independent-samples t-tests or MANOVA, depending on normality assumptions.

Where appropriate, scores will also be correlated with ATWIT workload scores to examine whether usability perceptions are systematically related to workload experience, as prior research suggests usability may mitigate perceived workload under certain conditions (Fruhling & Lee, 2005). This analysis will help determine whether higher usability mitigates perceived workload under complex operational conditions.

6. Risk Perception

Participants' perception of risk in operational scenarios involving safety-critical information will be measured using an adapted version of the Flight Risk Perception Scale (FRPS) developed by Winter et al. (2019). The FRPS was specifically designed to measure risk perception among pilots in aviation environments and has demonstrated strong reliability and construct validity. For the purposes of this work, the FRPS will be modified to reflect risks related to timeliness, relevance, volume, and modality of SCI for controllers, rather than focusing solely on physical flight conditions.

In dynamic operational settings, these judgments are shaped by available information, past experiences, and cognitive workload. Measuring risk perception is critical for understanding how individuals prioritize and respond to SCI under uncertainty, and how mismatches between perceived and actual risk may influence decision-making.

i. Procedure

To evaluate perception of operational risk in response to SCI, an adapted version of the FRPS will be administered immediately following each experimental scenario. This measure conceptualizes and captures real-time judgments about risk likelihood, severity, and controllability in response to SCI-related events (e.g., delayed PIREPs, conflicting NOTAMs, late weather updates). The adapted version will also tailor item content to reflect hazards emerging from delayed, irrelevant, or overwhelming information, as well as unclear or mismatched modalities (e.g., auditory versus visual). The adaptations ensure alignment with the constructs under investigation and preserve the theoretical structure of the original scale.

Each scenario will be followed by a brief vignette or probe item summarizing the critical SCI delivery event(s) presented during the task. Participants will be asked to rate:

- Perceived Likelihood of the scenario leading to negative operational impact (e.g., miscoordination, increased controller workload, pilot confusion)
- Perceived Severity of potential consequences should the event escalate
- Perceived Controllability – the extent to which the controller felt able to mitigate or manage the risk

The adapted FRPS will be delivered via a digital tablet interface or secure desktop form, immediately after participants complete workload and trust measures (e.g., ATWIT, PSSUQ). Instructions will emphasize that participants should base their ratings solely on the scenario just experienced, reflecting their real-time perceptions in context.

ii. Scoring

Each of the three subcomponents—likelihood, severity, and controllability—will be rated on a 5-point Likert scale:

1 = Very Low

2 = Low

3 = Moderate

4 = High

5 = Very High

For likelihood and severity, higher scores reflect higher perceived risk. For controllability, higher scores reflect greater perceived ability to manage the situation. For consistency, controllability scores will be reverse-coded for total risk index calculations, so that all higher composite values indicate greater perceived risk (i.e., low controllability = high risk).

A composite risk perception score will be calculated for each scenario by averaging the three item scores (with controllability reversed). Scores may also be retained as separate subscales to explore differential effects on perceived risk dimensions.

iii. Data Analysis

Descriptive statistics will be computed for each of the three FRPS subcomponents (likelihood, severity, controllability) and for the total composite score. To evaluate the impact of SCI variables (e.g., volume, modality, timing), ANOVA or mixed-effects modeling will be used to compare risk perception across different experimental conditions.

Additional analyses will explore how risk perception correlates with:

- Cognitive workload
- Trust in the SCI system
- Usability perceptions

This analysis will provide insight into how SCI presentation features influence perceived operational risk and whether these perceptions align with workload stressors or diminished system trust.

Where relevant, scenario characteristics (e.g., airspace class, flight phase, SCI modality) will be included as covariates in regression models to assess whether operational context mediates or moderates the effect of SCI characteristics on perceived risk.

7. Task Performance

Task performance is defined as the accuracy, timeliness, and procedural compliance of participants' actions in response to safety-critical information. In high-reliability domains such as aviation, effective task performance is essential for mitigating hazards and ensuring safe outcomes (Wickens & Hollands, 2013; FAA Human Factors Design Standard, 2016).

The ability to complete tasks accurately and in accordance with procedures, particularly under varying information conditions and cognitive demands, reflects both individual and team-level operational effectiveness (Salas et al., 2006).

This work adopts an error classification framework that disaggregates performance into four key categories:

1. Omission: Failure to perform a required task or respond to a relevant SCI cue.
2. Commission: Execution of an inappropriate or unnecessary action (e.g., acting on outdated or irrelevant SCI).
3. Timing: Delays or premature responses that reduce operational relevance or increase risk.
4. Procedural Deviation: Failure to follow established standard operating procedures (SOPs), even if the outcome was otherwise correct.

These error types allow for a nuanced comparison across delivery conditions (e.g., low vs. high volume, visual vs. auditory modality) and cognitive workload levels to assess how SCI characteristics influence operational reliability and task success

i. Procedure

Task performance will be assessed using a dual-method approach to ensure both precision in measurement and richness in behavioral interpretation. Performance is defined as the extent to which participants, individually and as part of the system, achieve safe and effective management of safety-critical information while meeting operational demands. This includes the ability to maintain situation awareness, shared situation awareness, and acceptable cognitive workload while operating within established safety risk boundaries (FAA Human Factors Design Standard, 2016; Endsley, 1995).

First, automated system logs will capture the latency between the delivery of SCI and the participant's corresponding acknowledgment or action. These logs will provide time-stamped records of task execution, offering high-resolution data on decision latency and allowing comparison across different information delivery conditions, workload levels, and roles.

Second, structured observer logs will be used to document and classify participant responses in real time and through post-scenario video review. Observers will record task execution patterns, identify errors, and assess adherence to established standard operating procedures. To support this effort, all observers will undergo scenario-specific training on the predefined error taxonomy used in the study.

Observer ratings will provide behavioral context to the system-generated performance data, allowing cross-validation of response timing, accuracy, and compliance. Inter-rater reliability will be assessed using Cohen's kappa (κ), with $\kappa > 0.75$ indicating acceptable agreement among raters (Bakeman & Gottman, 1997). Error frequency and patterns will be coded during and after scenario playback to identify relationships between specific delivery conditions (e.g., SCI modality or timing) and degraded performance.

This comprehensive, multi-method approach enables robust assessment of both human-level and system-level task performance across key dimensions: accuracy of operational actions, timeliness of response, procedural adherence, and error classification.

Together, these data streams allow for an integrated evaluation of SCI delivery effectiveness in dynamic, safety-critical environments.

ii. Scoring

Latency will be measured in seconds and computed as the interval between the system timestamp of SCI delivery and the moment of participant acknowledgment or initiation of action. This continuous variable provides a high-resolution indicator of decision latency, enabling evaluation of how information characteristics (e.g., volume, modality, and timing) and cognitive workload influence response speed across different operational roles.

Error frequency will be coded dichotomously (1 = present, 0 = absent) for each of the four predefined error types, omission, commission, timing, and procedural deviation, within each scenario. These binary classifications allow straightforward quantification of performance deviations and support direct comparison across delivery conditions. Coded errors will reflect both real-time and retrospective observer assessments to ensure accuracy.

The clarity of this scoring approach allows for integration with mixed-effects modeling and multivariate pattern analysis, ensuring that latency and error metrics are treated consistently as either continuous (latency) or categorical (error type presence) variables in downstream analyses. This structure also supports cross-referencing with system log data and inter-rater reliability metrics from observer training procedures.

iii. Data Analysis

Data analysis will include a combination of descriptive and inferential statistical techniques to examine task performance outcomes across SCI delivery conditions. Descriptive statistics, including means, medians, standard deviations, and error frequencies, will be calculated for core performance metrics: latency, error rates, and procedural adherence. These summaries will provide a foundational understanding of performance distributions and trends across participants and experimental scenarios.

To assess the influence of SCI volume, modality, and timeliness on performance, mixed-effects regression models will be employed. This approach is well-suited to the study's repeated-measures design, as it accounts for within-subject variation by treating participants as a random effect, while modeling the fixed effects of delivery condition variables. Separate models may be constructed for each error type (omission, commission, timing, procedural deviation) and for latency outcomes, depending on observed distributional properties.

In addition, multivariate pattern analysis will be used to identify constellations of delivery and task conditions that are predictive of degraded performance. This includes interaction effects (e.g., high SCI volume and delayed delivery) and cumulative workload burdens that may increase error susceptibility. Specific attention will be paid to patterns associated with increased frequency of procedural deviations and omissions, as these error types are especially safety-relevant in operational environments.

This combined analytical strategy enables the detection of both linear relationships and complex performance trends, ensuring a nuanced understanding of how information-delivery characteristics impact decision latency and operational accuracy in high-consequence environments.

8. Mental Models and Decision Strategies

Understanding how participants conceptualize, interpret, and respond to safety-critical information is essential for identifying the cognitive mechanisms that influence task performance and operational decision-making. In dynamic environments such as air traffic control, mental models shape how individuals detect, prioritize, and act on information under time-pressure, uncertainty, and variable workload conditions.

This method offers a direct, context-sensitive approach to capturing how mental models function in real-time decision-making and complements the study's broader set of behavioral and subjective measures. By systematically coding participants' verbal responses, the interviews will provide insight into the cognitive processes underlying performance variation and support the integration of qualitative reasoning patterns with quantitative workload, performance, and risk data.

i. Procedure

Following completion of the simulation scenarios and surveys, participants will engage in structured debrief interviews designed to elicit information about mental models, decision-making heuristics, and contextual reasoning used during the tasks. Interviewers will follow a standardized script to reduce bias, while allowing participants to describe their thought processes, priorities, and situational interpretations in their own words. Prompts will target detection, interpretation, prioritization, and timing of responses, as well as strategy adjustments due to workload, uncertainty, or information modality (Cooke & Gorman, 2005; 2006).

Interviews will be audio-recorded, transcribed, and analyzed thematically to identify recurring patterns and unique insights. This qualitative data complements the quantitative findings by revealing the “why” behind observed behaviors and perceptions.

This measure captures:

- The way participants conceptualize and interpret information during tasks
- Heuristics and strategies used to prioritize and act on information
- Contextual factors influencing decisions (e.g., workload, phase of flight)
- Alignment (or misalignment) between participant expectations and system behavior
- Narratives explaining why certain actions were taken (or not)

ii. Data Analysis

Audio-recorded interviews will be transcribed verbatim and analyzed using thematic analysis following Braun and Clarke’s (2006) six-phase framework: familiarization, initial coding, theme development, theme review, definition/naming of themes, and final reporting.

A deductive-inductive hybrid coding approach will be used. Initial codes will be guided by the predefined categories in the Mental Models & Decision Strategies Codebook (e.g., DS-Prioritize, MM-Update), reflecting theoretical constructs from cognitive task analysis and naturalistic decision-making research (Cooke & Gorman, 2005; 2006; Klein, 1997). Coders will also remain open to emergent codes that arise from the data, allowing for context-specific insight into participant reasoning under uncertainty, high workload, or ambiguous SCI.

To ensure reliability and rigor, at least two independent coders will apply the codebook to an initial subset of transcripts. Inter-coder reliability will be assessed using Cohen’s kappa, with $\kappa \geq 0.75$ considered acceptable (Bakeman & Gottman, 1997). Any discrepancies will be discussed and resolved through consensus, and the codebook will be refined iteratively.

Thematic saturation will guide the completion of coding across all transcripts.

Following initial coding, cross-case matrices will be generated to compare cognitive strategies by role, modality of SCI, information volume, and timing of delivery. Particular attention will be given to:

- Patterns of decision strategy shifts across scenarios.
- Mismatch between participant mental models and actual scenario events.
- Reported influence of workload and modality on interpretation and prioritization.

These qualitative insights will be integrated with quantitative performance, workload, and SA metrics to form a triangulated interpretation of how cognitive strategies shape response to SCI in safety-critical aviation operations.

7.4. Phase III: Development and Validation of the Decision-Support Framework

Phase III builds directly on the empirical results and operational insights generated during Phase II. In that phase, simulation-based activities involving air traffic and system operations personnel tested how SCI is interpreted and acted upon under varying delivery conditions. Findings from Phase II identified key variables influencing performance, workload, and usability – including timing of delivery, communication modality, and contextual complexity – which now inform the foundational logic of the SCI prioritization model.

This phase focuses on the validation, refinement, and feasibility assessment of a decision-support model designed to sort and prioritize SCI based on urgency, workload, and operational context. The model may take the form of a decision tree, rule-based logic, or scoring algorithm that ranks information according to time sensitivity, consequence severity, and modality effectiveness. Evaluation will include scenario-based testing, retrospective application to real-world cases, and structured review by SMEs from ATC, flight data coordination, weather, and system operations.

In addition to empirical testing, Phase III will generate technical guidance and operational procedures to support future integration of the model into ATC workflows, automation tools used by system operations analysts, and data presentation formats relevant to weather and coordination roles. This includes recommendations for display logic, timing thresholds, escalation protocols, and communication workflows.

Scalability and adaptability of the model will be evaluated across controller specialties (Tower, TRACON, ARTCC, CERAP) and supporting analytical roles such as weather and safety analysts, to ensure relevance across high-tempo operations, distributed coordination environments, and asynchronous decision support. The phase will conclude with an assessment of the model's readiness for field testing and implementation planning in Phase IV.

7.4.1 Research Questions

1. Can a structured decision-support model (e.g., decision tree or rule-based logic) accurately prioritize safety-critical aeronautical information across varying operational conditions?

2. Does model-driven delivery improve performance and reduce workload compared to baseline or alternative conditions?
3. Is the model adaptable across aircraft types, airspace classes, and mission profiles without degradation in usability or effectiveness?

7.4.2 Hypotheses

- H1: Model-driven delivery will result in lower error rates and improved task performance in high-complexity scenarios.
- H2: Subjective workload (NASA-TLX) and usability scores will be significantly more favorable under model-informed delivery conditions.
- H3: The model will perform consistently across variable operational contexts, including flight phases, delivery modalities, and regional differences.

7.4.3 Participants

Phase III includes participation by ATCs, Flight Data Coordinators, Weather Analysts, and Safety/System Operations Analysts, as these roles are directly responsible for managing or influencing real-time operational decisions involving SCI. The primary aim of this phase is to validate and refine the decision-support framework and prioritization logic under operationally realistic conditions – conditions that require participants who routinely operate in high-tempo, safety-sensitive environments and are equipped to critically evaluate the timing, modality, and usability of SCI delivery strategies under such constraints.

While other roles, such as dispatchers, flight service specialists within the lower 48 contiguous United States, and UAS operators, contribute to the broader information ecosystem, they typically do not engage in the synchronous, real-time decision-making processes that characterize the operational contexts targeted in this phase. Accordingly, participation is limited to roles whose responsibilities entail immediate interpretation, filtering, or coordination of SCI during dynamic air traffic operations. This targeted inclusion ensures that model validation is grounded in the lived experience of those most affected by SCI delivery and is aligned with the operational demands of the NAS.

7.4.4 Independent Variables

Phase III is not designed as an experimental manipulation phase but rather as a validation and feasibility assessment of the decision-support framework derived from earlier empirical findings. Unlike Phases I and II, which systematically varied independent variables such as information volume, modality, and timing, Phase III aims to assess how the refined prioritization model performs under realistic operational conditions. The focus is on real-world applicability and implementation feasibility, rather than isolating causal effects. As such, Phase III utilizes naturalistic, scenario-based evaluations and expert review rather than experimental control. This decision reflects the phase’s objective to determine whether the model’s logic, structure, and guidance align with actual workflows, reduce error potential, and support user trust and usability without introducing new manipulated conditions. Therefore, no independent variables are formally tested in this phase.

7.4.5 Measures

Phase III serves as a validation and refinement phase, using the same suite of psychometrically sound instruments and observational methods applied in Phase II. These include measures of workload (ATWIT), situation awareness (SA, SSA), trust in information systems (Koerber Scale), usability (PSSUQ), and risk perception (adapted FRPS), task performance, and mental models & decision strategies as shown in Appendix G. No new variables are introduced; instead, the goal is to determine whether previously observed patterns are reproducible, whether the prioritization model behaves as expected under realistic task conditions, and whether participant perceptions and performance outcomes support implementation readiness.

7.4.6 Data Analysis

The Phase III validation analysis will focus on confirming the reliability, sensitivity, and predictive validity of the dependent measures established in Phases I & II. Analyses will assess whether previously observed relationships among cognitive workload, SA, SSA, usability, trust, risk perception, performance, and decision-making replicate under new participant samples, refined manipulations, and adjusted scenario parameters. The goal is to validate the measurement framework, confirm operational effect sizes, and identify any residual measurement gaps before finalizing the methodology for operational deployment.

i. Procedure

Following each scenario, participants will complete the same post-scenario instruments as in Phase II. All tools will be administered electronically in the same order, using a digital interface that timestamps responses. Observational data and freeze-probe responses for SSA will again be collected during scenario execution using the same event markers and pause intervals defined in earlier phases. Audio recordings will be captured for all scenarios and transcribed for behavioral coding using the validated Phase II scheme.

ii. Scoring

All measurement instruments used in Phase III will retain the original scoring protocols established in Phase II to ensure consistency for cross-phase comparison. ATWIT workload ratings, scored on a 1–7 scale, will be timestamped and synchronized with SCI delivery events to capture temporal workload fluctuations and scenario-specific response patterns. PSSUQ responses, rated on a 1–7 Likert scale, will be reverse-coded where necessary so that lower scores indicate higher perceived usability; subscale and total scores will be computed for System Usefulness, Information Quality, and Interface Quality.

For trust measurement, Koerber Trust Scale items will be rated on a 1–5 scale and aggregated into three subscales: Performance-based, Process-based, and Purpose-based trust. These subscales capture different dimensions of operator perceptions toward the SCI system. For risk perception, the adapted FRPS will be used to evaluate perceived likelihood, severity, and controllability of SCI-related risk events, with each item scored on a 1–7 scale. Situation awareness accuracy will be quantified by calculating the percentage of correct responses to freeze-probe questions, along with inter-participant agreement on shared mental model items. Additionally, communication alignment in SSA will be coded and scored based on the frequency of observed alignment behaviors using the structured protocol developed in Phase II. Inter-rater reliability for coded data will be rechecked using Cohen's Kappa to ensure scoring consistency and stability.

iii. Data Analysis

Quantitative Data Analysis

Quantitative analysis will begin with validation of the reliability and structural integrity of all multi-item scales. Internal consistency will be re-evaluated using both Cronbach's alpha and McDonald's omega for each construct to ensure robust psychometric properties across participant groups. Confirmatory Factor Analysis (CFA) will be applied to multidimensional instruments, including the PSSUQ and Koerber Trust Scale, to test whether the factor structures observed in Phase II remain stable in the Phase III sample. Convergent and discriminant validity will be assessed through Pearson or Spearman correlations between conceptually related (e.g., workload and SA) and unrelated constructs, helping confirm measurement precision and construct independence.

Replication of Phase II findings will be tested using mixed-effects models and repeated-measures ANOVA to evaluate whether previously observed patterns, such as increased workload during SCI volume surges or in severe weather contexts, are reproduced in Phase III. Effect sizes and confidence intervals will be compared across phases to assess the stability and generalizability of relationships. To evaluate SSA convergence, intraclass correlation coefficients (ICCs) and percentage agreement will again be calculated. Acceptability thresholds established during Phase II, such as a minimum of 75% alignment for shared situation awareness, will be retained as operational benchmarks to judge acceptable model performance.

Qualitative Data Analysis

Qualitative analysis in Phase III will use the Phase II codebook and thematic framework as the basis for evaluating new transcripts, communication logs, and debriefing notes. Coders will apply predefined labels to communication strategies, mental models, and prioritization behaviors using the same structured protocol. Inter-coder reliability will be recalculated using Cohen's Kappa, with a minimum threshold of 0.75 required to confirm scoring consistency. Observed decision strategies and workload adaptations will be compared directly against Phase II thematic patterns to determine whether key behaviors and communication approaches are replicable under different operational constraints. Instances where new or contradictory patterns emerge will be flagged for discussion with subject matter experts and integrated into final model revisions if warranted.

Integrated Validation and Triangulation

An integrated joint display analysis will be used to cross-reference quantitative outcomes with qualitative themes, allowing for identification of areas where data streams converge, indicating strong empirical and operational support for the model, and areas of divergence that may suggest required refinement. Strong convergence will signal that the model logic is well-founded and generalizable across tasks and conditions. Partial alignment may suggest areas requiring adjustment to rules or timing thresholds, while significant divergence will highlight critical structural or conceptual gaps that must be resolved before operational integration. This triangulation process enhances the model's ecological validity and strengthens the foundation for eventual field testing.

Outcome and Model Finalization

The outcome of Phase III will be a validated, operationally grounded decision-support framework for sorting and delivering safety-critical aeronautical information. Only those measures and relationships demonstrating adequate reliability, stability across scenario variations, and alignment with subject matter expert judgment will be retained for Phase IV. The finalized model will include rule-based logic for delivery modality, timing, and prioritization thresholds based on scenario urgency, operator workload, and risk characteristics. Survey scores, workload data, and performance metrics will be cross-referenced with qualitative patterns to ensure the model supports consistent decision-making across high-complexity operational contexts. Risk thresholds and error tolerance levels, such as delays exceeding defined operational boundaries, will be reviewed in collaboration with FAA stakeholders to confirm their acceptability for field deployment. This fully integrated model will serve as the foundation for broader implementation in Phase IV.

7.5. Phase IV: Implementation Planning and System Integration

Phase IV focuses on applying the validated decision-support model developed in earlier phases to real-world operational aviation environments. This phase serves as a quasi-validation step using structured walkthroughs with subject matter experts (SMEs) who are not currently active controllers but serve in adjacent operational roles and regularly engage with safety-critical information.

Their input is not a substitute for controller feedback from Phases I-III. Rather, it provides a supplemental lens to identify potential edge cases, surface overlooked assumptions, and assess transferability before broader FAA or cross-domain applications are considered. This step is especially valuable for testing the robustness of the methodology outside the constraints of live operations.

In contrast to the empirical testing in Phases I-III, Phase IV emphasizes model feasibility, usability, and integration readiness. It involves close coordination with FAA representatives, industry partners, system developers, and operational users to ensure the model aligns with real-world workflows and cognitive demands across aviation environments.

7.5.1 Recommendations for Phase IV Integration

1. Use Research Questions

Although Phase IV is exploratory and implementation-focused, it is recommended to address the following research questions to fully address possible constraints and develop guidance for implementation and operational use:

- What technical, procedural, or organizational barriers exist to implementing the prioritization model into operational aviation systems?
- How do pilots, controllers, system developers, and regulators perceive the model's feasibility, usability, and potential safety impact?
- Can the model be effectively integrated into existing communication workflows, cockpit tools, and controller interfaces without introducing additional workload or usability burdens?

2. Hypotheses

No formal hypotheses are tested in this phase, given the focus on translational and implementation outcomes rather than empirical hypothesis testing.

3. Participants

Participants in Phase IV should include SMEs, system developers, human factors specialists, and representative operational users engaged in stakeholder workshops, usability walkthroughs, and limited pilot testing. These participants are selected for their expertise in operational workflows, human-systems integration, and regulatory compliance. Unlike earlier phases, Phase IV does not include a broad operational sample but instead focuses on stakeholders with direct experience designing, deploying, and regulating aviation technologies. Other operational roles, such as weather analysts, dispatchers, and flight service specialists, may be consulted in an advisory capacity during co-design workshops but are not central to the formal evaluation activities.

4. Measures

Measures in Phase IV should focus on assessing the feasibility, usability, and integration potential of the decision-support model. Recommendations for areas of focus include measures previously used and described:

- Perceived Feasibility. Participants' judgments on how readily the model could be implemented within existing FAA systems, controller interfaces, and cockpit tools.
- Usability Ratings: Assessments of ease of use, clarity, fit within workflows, and trust in the model's logic and delivery decisions.
- Integration Latency. System-level metrics quantifying any delays between information generation and delivery during pilot tests or walk-throughs.

- Barriers and Enablers. Qualitative identification of technical, procedural, and organizational factors facilitating or hindering implementation.
- Trust in Decision Support Framework. User confidence in the model's appropriateness and timing under realistic conditions.
- Instruments. Several qualitative and quantitative instruments should be deployed to evaluate the implementation process:
- Stakeholder Interviews. Structured interviews conducted with pilots, controllers, developers, and regulators to capture perceptions of the model's feasibility, usability, and potential safety impact.
- Usability Walkthroughs. Hands-on sessions in which participants interact with prototype interfaces or decision logic, providing direct feedback on usability and integration.
- Co-Design Workshops. Collaborative sessions involving SMEs, system developers, and end users to refine integration plans and address procedural or technical barriers.
- Field Testing or Pilot Studies. Where feasible, small-scale deployment of the model in operational or high-fidelity simulated environments to gather initial integration data.
- User Surveys. Post-engagement Likert-type questionnaires measuring satisfaction, trust, perceived usefulness, and potential cognitive impact.

Limitations

While this study offers a rigorous and empirically grounded approach to developing a framework for prioritizing and delivering SCI, several limitations should be acknowledged.

Scope of Operational Roles Included

Due to funding constraints and project scoping requirements, this research is limited to a subset of SCI users: FAA and FAA Contract Tower ATCs, Flight Data Coordinators, Weather Analysts, and Safety/System Operations Analysts. Other critical user groups – particularly pilots, dispatchers, pre-flight briefers, and UAS operators – are excluded from direct participation in this phase. As a result, the findings may not fully capture the interactional or downstream effects of SCI delivery across the broader operational ecosystem. While the selected roles represent high-value nodes in the SCI chain, further validation with excluded groups will be essential to ensure full-system integration and to minimize unintended consequences.

Geographic Scope Limitations

This study is geographically limited to the contiguous United States. Operational environments in Alaska, Hawaii, and U.S. territories differ significantly in terms of airspace structure, infrastructure availability, weather phenomena, and communication modalities. These regions require tailored SCI delivery strategies that fall outside the scope of this research. Future studies should be designed specifically to evaluate SCI prioritization and delivery practices in these areas, particularly Alaska, given its high operational risk and unique aviation landscape.

Partial Validation of the Conceptual Framework

The bow-tie risk model was used extensively in the early design phases to map hazards and identify pathways of SCI degradation; however, it is not directly validated within the simulations. Although its influence is visible in the structure of variables, prioritization logic, and scenario design, the model's full potential as a predictive or diagnostic tool has not yet been realized. Future phases may consider closing this gap through back-mapping of real-time data to bow-tie constructs or integrating quantitative risk scoring within the delivery logic.

Simulation Limitations and Ecological Validity

Although the high-fidelity simulations are carefully constructed to reflect realistic operational scenarios, they cannot fully replicate the pressures, distractions, and organizational constraints present in live environments. This may affect the generalizability of observed behavior, especially in terms of timing precision, trust calibration, or communication dynamics. Moreover, some SCI delivery characteristics – such as cross-system latency, interface inconsistency, or procedural ambiguity – may not be fully captured in the experimental environment.

Measurement Strategy Trade-offs

To enable cross-phase comparisons and maintain continuity across user roles, this research relies on well-established but generalized measurement tools (e.g., NASA-TLX, ATWIT, PSSUQ, Trust in Automation). While these tools offer psychometric reliability, they may lack the granularity needed to detect role-specific variations in SCI interaction. For example, the workload experienced by Safety Analysts may be influenced by different decision timelines or alert thresholds than those used by front-line controllers. In addition, some adapted measures (e.g., FRPS for SCI risk perception) may require further refinement and validation for non-pilot roles.

Need for Broader Collaboration and Funding Pathways

The long-term impact of this work depends on collaboration across organizational boundaries and funding mechanisms. Because SCI spans multiple domains – including flight operations, air traffic management, meteorology, and safety oversight – a comprehensive solution must eventually integrate perspectives from all user groups. This will require strategic alignment of priorities, shared data access, and funding partnerships that support broader system-wide validation and eventual field implementation. Without this coordinated effort, the framework risks becoming siloed or misaligned with real-world deployment environments.

Conclusion

This proposal provides a human-centered empirical methodology for developing a validated framework to sort and prioritize SCI within NAS operations. The four-phased design addresses both systemic and human vulnerabilities by integrating the bow-tie risk assessment model with core human factors constructs, including cognitive workload, SA, SSA, and trust in information sources. The bow-tie model played a central role in early design phases by framing SCI-related hazards in a structured, scenario-relevant manner. While not directly tested in simulations, its influence is embedded in the selection of variables, risk logic, and prioritization algorithms that now form the foundation of the decision-support framework.

It is important to recognize that the current effort focuses on a subset of operational roles – specifically, ATCs, Flight Data Coordinators, Weather Analysts, and Safety/System Operations Analysts. Although these roles represent critical nodes in the SCI delivery chain, additional users such as pilots, dispatchers, and pre-flight briefers were not included due to funding constraints. As a result, the current framework may not fully capture the downstream or interdependent effects of SCI delivery across the entire NAS. Expanding participation in future phases will be essential for system-wide integration.

Phase III offers a rigorous validation of the proposed framework in operationally realistic environments, using measures carried forward from Phase II to confirm stability, usability, and predictive relevance. However, the ecological limits of simulation, the absence of certain user roles, and the adaptation of some instruments for non-traditional contexts must be considered when interpreting the findings. These constraints underscore the need for broader collaboration and sustainable funding mechanisms to continue this work at scale.

Ultimately, this research lays the groundwork for a deployable, operationally informed SCI delivery model that improves usability, trust, and cognitive efficiency. Future efforts should focus on expanding the stakeholder base, refining the risk logic through integration with real-time system data, and ensuring that the final framework is adaptable across mission profiles and regulatory domains. The methodological integrity and multi-perspective design of this project position it well for the Phase IV transition, provided that system-wide validation and implementation support are pursued jointly by stakeholders across the NAS.

8. References

Aven, T. (2016). Risk assessment and risk management: Review of recent advances on their foundation. *European Journal of Operational Research, 253*, 1-13. https://doi.org/10.1016/j.ejor.2015.12.023

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual review of psychology*, *63*(1), 1-29.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*. Cambridge University Press.

Bailey, L. L., Willems, B. F., & Peterson, L. M. (2001). *The effects of workload and decision support automation on enroute r-side and d-side communication exchanges* (No. DOTFAAAM0120).

Bainbridge, L. (1983). Ironies of automation. In *Analysis, design and evaluation of man–machine systems* (pp. 129-135). Pergamon.

Bainbridge, L., & Dorneich, M. (2016). Processing underlying human performance. J A. Wise, V. D. Hopkin, & D. J. Garland, (Eds.), *Handbook of aviation human factors*, (2nd ed., pp. 7-1 – 7-63). CRC Press. https://doi.org/10.1201/b10401

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*. Cambridge university press.

Barshi, I., & Farris, C. (2016). Misunderstandings in ATC communication: Language, cognition, and experimental methodology. Routledge.

Beringer, D. B. and Schvaneveldt, R. (2002). Priorities of weather information in various phases of flight. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 46*(1), 86-90. https://doi.org/10.1177/154193120204600

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brimner, S. S., Cochran, S. P., Colvin, J. D., Durfee, A. F., Page, E. M., Street Jr, T. D., Shattuck, L. G., Nicholson, M. C., & Matsangas, P. (2023, September). Assessing flight information priority by flight phase for US Army future vertical lift platforms. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 67, No. 1, pp. 27-31). Sage CA: Los Angeles, CA: SAGE Publications. https://doi.org/10.1177/21695067231192421

Cardosi, K. M. (1993). *An analysis of en route controller-pilot voice communications* (No. DOTVNTSCFAA932).

Cardosi, K. M. (2017). Runway safety. In *Human factors impacts in air traffic management* (pp. 63-90). Routledge.

Cardosi, K. M., Brett, B., & Han, S. (1996). *An Analysis of TRACON (Terminal Radar Approach Control) Controller-Pilot Voice Communications* (No. DOTVNTSCFAA967).

Cardosi, K., Falzarano, P., & Han, S. (1998). *Pilot-Controller Communication Errors: An Analysis of Aviation Safety Reporting System (ASRS) Reports* (No. DOTVNTSCFAA984).

Center for Chemical Process Safety. (2001). *Layer of protection analysis: Simplified process of risk assessment*. Wiley-AIChE.

Center for Chemical Process Safety. (2008). *Guidelines for hazard evaluation procedures* (3rd ed.). Wiley-AIChE.

Center for Chemical Process Safety. (2016). *Human factors in barrier management: A White Paper by the Chartered Institute of Ergonomics & Human Factors*. Wiley-AIChE.

Center for Chemical Process Safety (CCPS). (2020a). *Bow-Ties in Risk Management: A Concept Book for Process Safety.* Hoboken, NJ: Wiley/American Institute of Chemical Engineers.

Center for Chemical Process Safety (CCPS). (2020b). *Guidelines for barrier management in process safety*. Center for Chemical Process Safety, American Institute of Chemical Engineers.

Cohen, T. N., Wiegmann, D. A., & Shappell, S. A. (2015). Evaluating the reliability of the human factors analysis and classification system. *Aerospace medicine and human performance*, *86*(8), 728-735. https://www.doi.org/10.3357/AMHP.4218.2015

Cooke, N. J., & Gorman, J. C. (2006). Assessment of team cognition. *International encyclopedia of ergonomics and human factors*, 271-276.

Cushing, S. (1994). *Fatal words: Communication clashes and aircraft crashes*. University of Chicago Press.

Damos, D. L. (1997). The influence of experience on pilot workload. International Journal of Aviation Psychology, 7(4), 375–388. https://doi.org/10.1207/s15327108ijap0704_5

De Dianous, V., & Fiévez, C. (2006). ARAMIS project: A more explicit demonstration of risk control through the use of bow-tie diagrams and the evaluation of safety barrier performance. *Journal of Hazardous Materials, 130*(3), 220–233. https://ineris.hal.science/ineris-00961900/file/2006-035_post-print.pdf

Degani, A., & Wiener, E. L. (1994). *Philosophy, policies, procedures and practices: The four 'P's' of flight deck operations.* In Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Degani, A., & Wiener, E. L. (1997). Procedures in complex systems: The airline cockpit. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *27*(3), 302-312. https://doi.org/10.1109/3468.568739

Dekker, S., Hollnagel, E., Woods, D., & Cook, R. (2008). Resilience Engineering: New directions for measuring and maintaining safety in complex systems. *Lund University School of Aviation*, *1*, 1-6.

Dismukes, R. K., Berman, B. A., & Loukopoulos, L. D. (2007). The Limits of Expertise: Rethinking Pilot Error and the Causes of Airline Accidents. https://doi-org.ezproxy.lib.ou.edu/10.4324/9781315238654

Dornheim, M. A. (2000). GA aircraft lag in safety systems. *Aviation Week & Space Technology, 152*(23), 70–72.

Drew, T., Võ, M. L. H., & Wolfe, J. M. (2013). The invisible gorilla strikes again: Sustained inattentional blindness in expert observers. *Psychological science*, *24*(9), 1848-1853.

Drijvers, L., & Holler, J. (2023). The multimodal facilitation effect in human communication. *Psychonomic Bulletin & Review*, *30*(2), 792-801.

Durso, F. T., & Manning, C. A. (2008). Air traffic control. *Reviews of human factors and ergonomics*, *4*(1), 195-244. https://doi.org/10.1518/155723408X342853

Durso, F. T., Rawson, K. A., & Girotto, S. (2017). Comprehension and situation awareness. In *Situational awareness* (pp. 77-108). Routledge.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Endsley, M. R. (2015). *Situation awareness for operations*. CRC Press.

Endsley, M. R. (2017). Direct measurement of situation awareness: Validity and use of SAGAT. In *Situational awareness* (pp. 129-156). Routledge.

Endsley, M. R. (2021). Situation awareness. *Handbook of human factors and ergonomics*, 434-455.

Endsley, M. R., Bolté, B., & Jones, D. G. (2003). *Designing for situation awareness: An approach to user-centered design.* CRC Press.

Endsley, M. R., & Connors, E. S. (2008, July). Situation awareness: State of the art. In *2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century* (pp. 1-4). IEEE.

Endsley, M. R., & Garland, D. J. (Eds.). (2000). *Situation awareness analysis and measurement*. CRC press.

Endsley, M., & Jones, W. M. (1997). *Situation Awareness Information Dominance & Information Warfare* (No. ALCFTR19970156). https://apps.dtic.mil/sti/pdfs/ADA347166.pdf

Federal Aviation Administration. *Human Factors Design Standard* (2016) U.S. Department of Transportation.

Federal Aviation Administration. (2023a). Pilot's Handbook of Aeronautical Knowledge (FAA-H-8083-25C). Flight Standards Service. Retrieved from https://www.faa.gov/regulations_policies/handbooks_manuals/aviation/faa-h-8083-25c.pdf

Federal Aviation Administration. (2023b). Aeronautical Information Manual (AIM). https://www.faa.gov/air_traffic/publications/media/aim_basic_dtd_4-20-23.pdf

Federal Aviation Administration. (2023c). *Aerospace Forecast FY 2022–2042*. U.S. Department of Transportation.

Federal Aviation Administration. (2023d). *Instrument Procedures Handbook (FAA-H-8083-16B)*. U.S. Department of Transportation.

Federal Aviation Administration. (2023e). *FAA Order 8040.4B: Safety Risk Management Policy*. U.S. Department of Transportation. Retrieved from https://www.faa.gov/documentLibrary/media/Order/FAA_Order_8040.4C.pdf

Federal Aviation Administration. (2025). *FAA Order JO 7110.65BB: Air traffic control*. https://www.faa.gov/regulations_policies/orders_notices/index.cfm/go/document.current/documentnumber/7110.65

Federal Aviation Administration. (2025, June 9). *FAA Order 7930.2U: Notice to Airmen (NOTAM).* Washington, DC: U.S. Department of Transportation.

Federal Aviation Administration. (2025, February 11). *NOTAM Governance Overview.* Retrieved from FAA NOTAM webpage.

Federal Aviation Administration. (2025, September 30). *What is a NOTAM*: Retrieved from the FAA NOTAM webpage. https://www.faa.gov/about/initiatives/notam/what_is_a_notam

Fischhoff, B., Hope, C., & Watson, S. R. (2013). Defining risk. In *Readings in risk* (pp. 30-42). RFF Press.

Fruhling, A., & Lee, S. (2005). Assessing the reliability, validity and adaptability of PSSUQ. *AMCIS 2005 proceedings*, 378.

Gao, F. (2016). *Modeling human attention and performance in automated environments with low task loading* (Doctoral dissertation, Massachusetts Institute of Technology).

Geitner, C., Biondi, F., Skrypchuk, L., Jennings, P., & Birrell, S. (2019). The comparison of auditory, tactile, and multimodal warnings for the effective communication of unexpected events during an automated driving scenario. *Transportation research part F: traffic psychology and behaviour*, *65*, 23-33.

Gonzalez, C. (2004). Learning to make decisions in dynamic environments: Effects of time constraints and cognitive abilities. *Human Factors*, *46*(3), 449-460.

Gorman, J. C., Cooke, N. J., Pederson, H. K., Olena, O. C., & DeJoode, J. A. (2005, September). Coordinated awareness of situation by teams (CAST): Measuring team situation awareness of a communication glitch. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 49, No. 3, pp. 274-277). Sage CA: Los Angeles, CA: SAGE Publications.

Gorman, J. C., Cooke, N. J., & Winner, J. L. (2006). Measuring team situation awareness in decentralized command and control environments. *Ergonomics*, *49*(12-13), 1312-1325.

Gutzwiller, R. S., & Clegg, B. A. (2013). The role of working memory in levels of situation awareness. *Journal of Cognitive Engineering and Decision Making*, *7*(2), 141-154.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology*, *8*(1), 23. https://doi.org/10.20982/tqmp.08.1.p023

Hansman, R. J., & Davison, H. J. (2000). The effect of shared information on pilot/controller and controller/controller interactions. *Proceedings of the AIAA/IEEE Digital Avionics Systems Conference*, 1–9. https://doi.org/10.1109/DASC.2000.884903

Harris, D., & Li, W. C. (2017). *Decision making in aviation*. Routledge.

Hernandez, C. S., Ayo, S., & Panagiotakopoulos, D. (2021, October). An explainable artificial intelligence (xAI) framework for improving trust in automated ATM tools. In *2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC)* (pp. 1-10). IEEE.

Hilburn, B. (2004). Cognitive complexity in air traffic control: A literature review. EUROCONTROL.

Histon, J. M., & Hansman, R. J. (2008). Mitigating complexity in air traffic control: The role of structure-based abstractions. *Journal of Air Traffic Control, 50*(3), 10–14.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors*, *57*(3), 407-434.

Hollnagel, E. (2017). *Safety-II in practice: Developing the resilience potentials*. Routledge. https://doi.org/10.4324/9781315201023

Hollnagel, E., & Bye, A. (2000). Principles for modelling function allocation. *International Journal of Human-Computer Studies*, *52*(2), 253-265.

International Civil Aviation Organization (ICAO). (2013). *Safety Management Manual (SMM)*. Doc 9859, 3rd Ed. Montréal, Canada: International Civil Aviation Organization.

International Civil Aviation Organization (ICAO) (2018). *Annex 11: Air Traffic Services*. Montréal, Canada: International Civil Aviation Organization.

Israel, G. D. (1992). *Determining Sample Size*.

Jones, L. A., & Sarter, N. B. (2008). Tactile displays: Guidance for their design and application. *Human factors*, *50*(1), 90-111.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making.

Lai, H. Y., Chen, C. H., Khoo, L. P., & Zheng, P. (2019). Unstable approach in aviation: Mental model disconnects between pilots and air traffic controllers and interaction conflicts. *Reliability Engineering & System Safety*, *185*, 383-391.

Lee, J. R., Fanjoy, R. O., & Dillman, B. G. (2005). The effects of safety information on aeronautical decision making. *Journal of Air Transportation, 10*(3).

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, *35*(10), 1243-1270. https://doi.org/10.1080/00140139208967392

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors*, *46*(1), 50-80. https://doi.org/10.1518/hfes.46.1.50_30392

Leveson, N. G. (2011). Engineering a safer world: Systems thinking applied to safety. MIT Press.

Lewis, J. R. (1992, October). Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ. In *Proceedings of the human factors society annual meeting* (Vol. 36, No. 16, pp. 1259-1260). Sage CA: Los Angeles, CA: Sage Publications.

Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, *7*(1), 57-78.

Loft, S., Sanderson, P., Neal, A., & Mooij, M. (2007). Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications. *Human Factors, 49*(3), 376–399. https://doi.org/10.1518/001872007X197017

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: an integrative review. *Theoretical Issues in Ergonomics Science*, *8*(4), 277-301. https://doi.org/10.1080/14639220500337708

McAllister, D. J. (1995). Affect-and cognition-based trust scale. *PsycTESTS*, *38*, 24-59. https://doi.org/10.1037/t55229-000

Mejdal, S., McCauley, M. E., & Beringer, D. B. (2001). *Human factors design guidelines for multifunction displays* (No. DOTFAAAM0117).

Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human factors*, *50*(2), 194-210.

Mosier, K. L., & Fischer, U. M. (2017). Judgment and decision making by individuals and teams: Issues, models, and applications. In *Decision making in aviation* (pp. 139-198). Routledge.

Mosier, K. L., Fischer, U., Burian, B. K., & Kochan, J. A. (2017a). Autonomous, context-sensitive, task management systems and decision support tools I: Human-autonomy teaming fundamentals and state of the art (No. NASA/TM-2017-219565).

Mosier, K. L., Fischer, U., & Orasanu, J. (2017b). Collaboration and coordination in the cockpit: The crew as a distributed system. In F. Durso, R. Nickerson, S. Dumais, S. Lewandowsky, & T. Perfect (Eds.), *Handbook of applied cognition* (2nd ed., pp. 451–480). Wiley. https://doi.org/10.1002/9781119137506.ch17

National Transportation and Safety Board. (2001a). *Aviation accident final report: Greenville, Maine; Piper PA-60-700P; N407SA; October 3, 1999.* (Accident No. NYC00LA003). National Transportation Safety Board. https://data.ntsb.gov/carol-repgen/api/Aviation/ReportMain/GenerateNewestReport/57926/pdf

National Transportation and Safety Board. (2001b). *Aviation accident final report: Hilliard, Florida; Gulfstream American AA-1B; N9888L; August 1, 1999.* (Accident No. MIA99FA208). National Transportation Safety Board. https://data.ntsb.gov/carol-repgen/api/Aviation/ReportMain/GenerateFactualReport/46947/pdf

National Transportation and Safety Board. (2013). *Aviation accident final report: San Marcos, Texas; Piper PA-34-200T; N1162X; April 3, 2012* (Accident No. CEN12LA229). National Transportation Safety Board. https://data.ntsb.gov/carol-repgen/api/Aviation/ReportMain/GenerateNewestReport/83306/pdf

National Transportation and Safety Board. (2016). *Educating Controllers on Two Midair Collisions Report ASR-16-06.* (Accident Nos. ERA15MA259A/B; WPR15MA243A/B). National Transportation Safety Board. https://www.ntsb.gov/investigations/AccidentReports/Reports/ASR1606.pdf

National Transportation and Safety Board. (2018a). *Safety Recommendation Report A-18-24.* https://www.ntsb.gov/safety/safety-recs/recletters/A-18-023-029.pdf.

National Transportation and Safety Board. (2018b). *Safety Recommendation Report AIR-18-01.* https://www.ntsb.gov/investigations/AccidentReports/Reports/AIR1801.pdf.

National Transportation and Safety Board. (2018c). *Taxiway overflight, Air Canada Flight 759, Airbus A320-211, C-FKCK, San Francisco, California, July 7, 2017.* (NTSB/AIR-18/01). National Transportation Safety Board. https://www.ntsb.gov/investigations/Pages/DCA17IA148.aspx

Nielsen, J. (1994). Report from a 1994 Web usability study. *Jacob Nielsen's Alertbox*.

OpenAI. (2024). *ChatGPT* (May 24, GPT-4-turbo version) [Large language model]. https://chat.openai.com/

Orasanu, J. (1991). Shared mental models and crew decision making (NASA Technical Memorandum 103884). NASA Ames Research Center. https://ntrs.nasa.gov/citations/19920062313

Orasanu, J., Fischer, U., & Davison, J. (2002). Risk perception: A critical element of aviation safety. *IFAC Proceedings Volumes*, *35*(1), 49-58. https://doi.org/10.3182/20020721-6-ES-1901.01639

Orasanu, J., Fischer, U., & Davison, J. (2011). Distributed crew performance in emergencies. *The International Journal of Aviation Psychology, 21*(3), 214–234. https://doi.org/10.1080/10508414.2011.582452

Orasanu, J., Martin, L., & Davison, J. (2001). Cognitive and contextual factors in aviation accidents: Decision errors. In *Linking expertise and naturalistic decision making* (pp. 209-225). Psychology Press.

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors*, *39*(2), 230-253. https://doi.org/10.1518/001872097778543886

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, *30*(3), 286-297.

Patterson, R. D. (1982). Guidelines for auditory warning systems on civil aircraft. Instituut voor Perceptie Onderzoek, RP/ne 82/01. *Manuscript no*, *413*.

Rausand, M. (2013). *Risk Assessment: Theory, Methods, and Applications.* (Vol. 115). John Wiley & Sons.

Rasmussen, J. (1997). Risk management in a dynamic society: a modelling problem. *Safety science*, *27*(2-3), 183-213.

Rasmussen, J., Pejtersen, A. M., & Goodstein, L. P. (1994). *Cognitive systems engineering*. John Wiley & Sons, Inc..

Reason, J. (1990). Human error. Cambridge University Press.

Roth, E., Klein, D., & Ernst, K. (2021). *Aviation decision making and situation awareness study: Decision making literature review*. (Report No. USAARL-TECH-CR--2022-17). U.S. Army Aeromedical Research Laboratory. https://apps.dtic.mil/sti/pdfs/AD1160556.pdf

Salas, E., Prince, C., Baker, D. P., & Shrestha, L. (1995). Situation awareness in team performance: Implications for measurement and training. *Human Factors*, *37*(1), 123–136.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors, 50*(3), 540–547. https://doi.org/10.1518/001872008X288457

Salas, E., Rosen, M. A., Burke, C. S., & Goodwin, G. F. (2006). The wisdom of collectives in organizations: An update of the teamwork competencies. *Team effectiveness in complex organizations*, 39-79.

Salmon, P. M., Stanton, N. A., & Jenkins, D. P. (2009). *Distributed situation awareness: Theory, measurement and application to teamwork*. CRC Press.

Salmon, P. M., Stanton, N. A., Walker, G. H., & Green, D. (2006). Situation awareness measurement: A review of applicability for C4i environments. *Applied Ergonomics*, 37(2), 225-238.

Schvaneveldt, R., Beringer, D. B., Lamonica, J., Tucker, R., & Nance, C. (2000). *Priorities, organization, and sources of information accessed by pilots in various phases of flight* (No. DOT/FAA/AM-00/26). United States. Department of Transportation. Federal Aviation Administration. Office of Aviation. Civil Aerospace Medical Institute.

Selcon, S. J., Taylor, R. M., & Koritsas, E. (1991). Workload or situational awareness?: TLX vs. SART for aviation SAR task analysis. *Proceedings of the Human Factors Society Annual Meeting*, 35(2), 62-66.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1986). The psychometric study of risk perception. In *Risk evaluation and management* (pp. 3-24). Boston, MA: Springer US.

Stanton, N. A., Salmon, P. M., Walker, G. H., & Jenkins, D. P. (2005). *Human factors methods: A practical guide for engineering and design*. Ashgate.

Stein, E. S. (1985). *Air Traffic Controller Workload: An Examination of Workload Probe*. FAA Technical Center. Report No. DOT/FAA/CT-TN84/24.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.

Taylor, R. M. (2017). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In *Situational awareness* (pp. 111-128). Routledge.

Thalmann, M., Souza, A. S., & Oberauer, K. (2019). How does chunking help working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(1), 37.

Todd, P. M., & Gigerenzer, G. (1999). What we have learned (so far). In *Simple heuristics that make us smart* (pp. 357-365). Oxford University Press.

US Department of Defense. (2012). Department of Defense design criteria standard: Human engineering. *(MIL-STD-1472G)*.

Van Benthem, K., & Herdman, C. M. (2020). The importance of domain-dependent cognitive factors in GA safety: Predicting critical incidents with prospective memory, situation awareness, and pilot attributes. *Safety Science*, *130*, 104892. https://doi.org/10.1016/j.ssci.2020.104892

Van Erp, J. B., & Van Veen, H. A. (2004). Vibrotactile in-vehicle navigation system. *Transportation Research Part F: Traffic Psychology and Behaviour*, *7*(4-5), 247-256.

von Thaden, T. L. (2008). Distributed information behavior: A study of dynamic practice in a safety critical environment. *Journal of the American Society for Information Science and Technology*, *59*(10), 1555-1569. https://doi.org/10.1002/asi.20842

Vu, K. P. L., Strybel, T. Z., Kraut, J., Bacon, P., Minakata, K., Nguyen, J., Rottermann, A., Battiste, V., & Johnson, W. (2010, October). Pilot and controller workload and situation awareness with three traffic management concepts. In 29th digital avionics systems conference (pp. 4-A). IEEE.

West, B. T., Welch, K. B., & Galecki, A. T. (2015). *Linear mixed models: A practical guide using statistical software*. CRC Press.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, *3*(2), 159-177.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors, 50*(3), 449–455.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). *Engineering psychology and human performance* (4th ed.). Pearson Higher Ed.

Wickens, C. D., & Carswell, C. M. (2021). Information processing. *Handbook of human factors and ergonomics*, 114-158.

Wiegmann, D. A., & Shappell, S. A. (2017). *A human error approach to aviation accident analysis: The human factors analysis and classification system*. Routledge.

Wilson, J., Scielzo, S., Nair, S., & Larson, E. C. (2020). Automatic gaze classification for aviators: Using multi-task convolutional networks as a proxy for flight instructor observation. *International Journal of Aviation, Aeronautics, and Aerospace*, *7*(3), 7. https://doi.org/10.15394/ijaaa.2020.1499

Winter, S. R. Truong, D., & Keebler, J. R. (2019). The flight risk perception scale (FRPS): A modified risk perceptions scale for measuring risk of pilots in aviation. *International Journal of Aviation Research, 11*(01), 57-72.

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: mental workload in ergonomics. *Ergonomics*, *58*(1), 1-17. https://doi.org/10.1080/00140139.2014.956151

9. Appendix A. NTSB Recommendation A-18-024

The following Safety Recommendations were issued as a result of NTSB Aviation Incident Report “Taxiway Overflight Air Canada Flight 759 Airbus A320-211, C-FKCK, San Francisco, California, July 7, 2017.” Report Number AIR-18-01. Adopted on September 25, 2018. Published on October 11, 2018.

Safety Recommendation Number	Safety Recommendation Recipient	Safety Recommendation Text
A-18-024	Federal Aviation Administration	Establish a group of human factors experts to review existing methods for presenting flight operations information to pilots, including flight releases and general aviation flight planning services (preflight) and aircraft communication addressing and reporting system messages and other in-flight information; create and publish guidance on best practices to organize, prioritize, and present this information in a manner that optimizes pilot review and retention of relevant information; and work with air carriers and service providers to implement solutions that are aligned with the guidance.

9.1. FAA Response to NTSB Safety Recommendation A-18-24

U.S. Department of Transportation
Federal Aviation Administration

Office of the Administrator

800 Independence Ave., S.W.
Washington, DC 20591

March 31, 2021

The Honorable Robert L. Sumwalt
Chairman, National Transportation
Safety Board
490 L'Enfant Plaza East, S.W.
Washington, DC 20594

Dear Chairman Sumwalt:

This is in further response to Safety Recommendations A-18-24 issued by the Board on October 11, 2018. The Board issued this safety recommendation as a result of the incident that occurred on July 7, 2017, in which Air Canada flight 759, an Airbus A320-211, Canadian registration C-FKCK, was cleared to land on runway 28R at San Francisco International Airport, San Francisco, California, but instead lined up with parallel taxiway C. Four air carrier airplanes (a Boeing 787, an Airbus A340, another Boeing 787, and a Boeing 737) were on taxiway C awaiting clearance to take off from runway 28R. The incident airplane descended to an altitude of 100 feet above ground level and overflew the first airplane on the taxiway. The incident flight crew initiated a go-around, and the airplane reached a minimum altitude of about 60 feet and overflew the second airplane on the taxiway before starting to climb. None of the 5 flight crewmembers and 135 passengers aboard the incident airplane were injured, and the incident airplane was not damaged. The incident flight was operated by Air Canada under Title 14, Code of Federal Regulations Part 129 as an international scheduled passenger flight from Toronto/Lester B. Pearson International Airport, Toronto, Canada. An instrument flight rules flight plan had been filed. Night visual meteorological conditions prevailed at the time of the incident.

A-18-24. Establish a group of human factors experts to review existing methods for presenting flight operations information to pilots, including flight releases and general aviation flight planning services (preflight) and aircraft communication addressing and reporting system messages and other in-flight information; create and publish guidance on best practices to organize, prioritize, and present this information in a manner that optimizes pilot review and retention of relevant information; and work with air carriers and service providers to implement solutions that are aligned with the guidance.

FAA Comment. The Federal Aviation Administration's (FAA) Air Traffic Organization and Civil Aviation Medical Institute formulated a plan to address this recommendation, which includes a multi-level review to:

- Determine what information is most critical during each flight operation phase. This would include:
  - A search of the Aviation Safety Reporting System (ASRS) database to determine where incidents have occurred because of a lack of information;
  - Interviews/mission walkthroughs with flight crews to discuss what critical information is needed at what points in the flight, combined with discussions of how that information might most effectively be presented; and
  - A survey provided to pilots in the part 91, 135, and 121 communities to provide additional clarification on the findings of the ASRS database search and the interviews.

- Determine the timing of delivery:
  - Delivering information during planning or earlier in a flight to make planning/preparation more effective; and
  - Early presentation combined with just-in-time information delivery to reduce the prospective memory burdens on the crews.

- Develop recommendations for how air traffic control (ATC) distributes the necessary information and how services will be employed. Flight crews and ATC personnel in simulator studies could explore the most promising interventions. These could include:
  - A search of the Air Traffic Safety Action Program database to determine where incidents have occurred because of a lack of information;
  - Interviews/cognitive walkthroughs with ATC to discuss potential changes in procedures and/or use of technology for communicating the critical information identified through the process described above;
  - A survey provided to ATC to collect additional information;
  - Laboratory simulations of flight deck and air traffic missions/operations may be conducted to empirically evaluate these recommendations; and
  - Delivery mechanisms might incorporate graphical presentation on iPads/electronic flight bags or multi-function displays, controller–pilot Data Link Communications messages, or heads-up display overlays, where available.

The final research products will include a list of requirements and proposed methodology for identification, sorting, and delivery of information by phase of flight, type of operation, current conditions, relevant notice to airmen, etc., to increase the probability of delivering critical information in an easily understandable, actionable format. Any proposed requirements for data display, sorting, etc., on one or more systems will be discussed with the appropriate program offices for implementation, addressing what can readily be implemented with few changes and what might be possible as systems evolve in the future.

I will keep the Board informed of the FAA's progress on this safety recommendation and anticipate providing an update by October 31, 2021.

Sincerely,

Steve Dickson
Administrator

9.2. FAA Further Response to NTSB Safety Recommendation A-18-24

U.S. Department
of Transportation
**Federal Aviation
Administration**

Office of the Administrator

800 Independence Ave., S.W.
Washington, DC 20591

July 5, 2022

The Honorable Jennifer Homendy
Chair, National Transportation
Safety Board
490 L'Enfant Plaza East, S.W.
Washington, DC 20594

Dear Chair Homendy:

This is in further response to Safety Recommendation A-18-24 issued by the Board on October 11, 2018, and supplements our previous letters. The Board issued this safety recommendation as a result of the incident that occurred on July 7, 2017, in which Air Canada flight 759, an Airbus A320-211, Canadian registration C-FKCK, was cleared to land on runway 28R at San Francisco International Airport (SFO), San Francisco, California, but instead lined up with parallel taxiway C. Four air carrier airplanes (a Boeing 787, an Airbus A340, another Boeing 787, and a Boeing 737) were on taxiway C awaiting clearance to take off from runway 28R. The incident airplane descended to an altitude of 100 feet above ground level and overflew the first airplane on the taxiway. The incident flight crew initiated a go-around, and the airplane reached a minimum altitude of about 60 feet and overflew the second airplane on the taxiway before starting to climb. None of the 5 flight crewmembers and 135 passengers aboard the incident airplane were injured, and the incident airplane was not damaged. The incident flight was operated by Air Canada under Title 14, *Code of Federal Regulations* Part 129 as an international scheduled passenger flight from Toronto/Lester B. Pearson International Airport (YYZ), Toronto, Canada. An instrument flight rules flight plan had been filed. Night visual meteorological conditions prevailed at the time of the incident.

A-18-24. Establish a group of human factors experts to review existing methods for presenting flight operations information to pilots, including flight releases and general aviation flight planning services (preflight) and aircraft communication addressing and reporting system messages and other in-flight information; create and publish guidance on best practices to organize, prioritize, and present this information in a manner that optimizes pilot review and retention of relevant information; and work with air carriers and service providers to implement solutions that are aligned with the guidance.

FAA Comment. In our previous letter, we noted that the Federal Aviation Administration's (FAA) Air Traffic Organization and Civil Aviation Medical Institute (CAMI) formulated a multi-level plan to address this recommendation. We have completed a review of the National Aeronautics and Space Administration's Aviation Safety Reporting System Database and developed draft surveys intended for pilot and air traffic control respondents who use or handle aeronautical information. Due to the impact of the COVID-19 pandemic, we have been unable to conduct in-person interviews/mission walk-throughs with flight crews to discuss what critical information they need at specific points in the flight and how to present that information effectively. However, the FAA has communicated with flying organizations in Alaska and the contiguous United States to discuss challenges encountered with Notices to Airmen and has incorporated these findings in our pilot surveys. Our next step is to begin coordinating the surveys with labor relations groups, which we anticipate completing by the end of the summer 2022.

Additionally, the Paperwork Reduction Act requires the FAA to seek approval from the Office of Personnel Management to survey general aviation pilots. While the FAA seeks this approval, researchers at CAMI are planning discussions with additional subject matter experts on the next steps concerning the range of observations and operations.

I will keep the Board informed of the FAA's progress on this safety recommendation and anticipate providing an update by April 30, 2023.

Sincerely,

Billy Nolen
Acting Administrator

10. Appendix B: Example Types of Aeronautical Information

Type of Aeronautical Information	Acronym	Definition
Notice to Airmen	NOTAM	A Notice to Air Mission (NOTAM) is a notice containing information essential to personnel concerned with flight operations but not known far enough in advance to be publicized by other means. It states the abnormal status of a component of the NAS, not the normal status (FAA, 2025, September 30).
Pilot Report	PIREP	PIREPs are reports of observed in-flight weather conditions, provided by pilots and given to flight service or ATC, or submitted electronically through an Electronic Flight Bag (EFB) application or the Aviation Weather Center website, and then disseminated to other pilots.
Meteorological Aerodrome Report	METAR	Meteorological Aerodrome Report (METAR), also known as Meteorological Terminal Aviation Routine Weather Report, Meteorological Terminal Air Report, or Meteorological Airfield Report, is a format for reporting weather information.
Airman's Meteorological Information	AIRMET	An AIRMET (AIRman's METeorological Information) advises of weather that may be hazardous, other than convective activity, to single-engine, other light aircraft, and Visual Flight Rule (VFR) pilots
Significant Meteorological Information	SIGMET	Significant Meteorological Information affects all aircraft. These denote more severe weather conditions than AIRMETs. SIGMETs are generally issued for shorter periods than AIRMETs. They usually expire after four hours.
Project-relevant information within each of the types of aeronautical information will be identified and documented by SMEs in Phase I of the research.		

11. Appendix C: Background Resources Reviewed

Topic Area	Title	Relevance to Study	Constructs or Variables Supported
Safety-Critical Information Management	Safety Recommendation Report AIR-18-01 https://www.ntsb.gov/investigations/AccidentReports/Reports/AIR1801.pdf	Identifies failure in safety-critical info delivery (e.g., AC759 NOTAM handling).	Safety-critical info, NOTAMs, etc.
	Recommendation A-18-24 https://www.ntsb.gov/safety/safety-recs/recletters/A-18-023-029.pdf	Specific recommendation to FAA on improving aeronautical info delivery.	
	FAA ARC on Human Factors in NOTAM Modernization– Executive summaries and public docket materials. https://www.faa.gov/air_traffic/flight_info/aeronav/acf/media/Briefings/NOTAM_Modernization.pdf	Addresses challenges in NOTAM interpretation and system modernization.	
	NOTAM Decoder Reference Guide. https://www.scribd.com/document/422473689/Jeppesen-notam-decoder	Practical tool for parsing and understanding safety-critical NOTAM content.	
	FAA Order JO 7110.65X. https://www.faa.gov/regulations_policies/orders_notices/index.cfm/go/document.current/documentnumber/7110.65	Governs the real-time relay of safety-critical info by controllers.	
Human Factors in Info Delivery	• Barshi, I., & Farris, C. (2016). *Misunderstandings in ATC communication: Language, cognition, and experimental methodology*. Routledge.	All of these highlight how communication and delivery affect safety and decisions.	Usability, overload, modality

Topic Area	Title	Relevance to Study	Constructs or Variables Supported
	• Lukić, J. (2014). The impact of information and communication technology on decision making process in the big data era. *Megatrend review*, *11*(2), 221-233. http://scindeks-clanci.ceon.rs/data/pdf/1820-3159/2014/1820-31591402221L.pdf • Molesworth, B. R. C., & Estival, D. (2015). Miscommunication in general aviation: The influence of external factors on communication errors. http://dx.doi.org/10.1016/j.ssci.2014.11.004 • Prinzo, O. V. (2001). *Data-linked pilot reply time on controller workload and communication in a simulated terminal option* (No. DOTFAAAM018)..		
Cognitive Workload and Processing	• Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7 • Wickens, C. D., Helton, W. S., Hollands, J. G., & Banbury, S. (2021). *Engineering psychology and human performance*. Routledge. • Rausand, M. (2013). Risk assessment: Theory, methods, and applications. (Vol. 115). John Wiley & Sons.	These address how humans process, become overloaded by, or respond to information.	ATWIT Scale, MRT, CLT

Topic Area	Title	Relevance to Study	Constructs or Variables Supported
Situation Awareness and SSA	• Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32–64. https://doi.org/10.1518/001872095779049543 • Rasmussen, J. (1997). Risk management in a dynamic society: a modelling problem. *Safety science*, *27*(2-3), 183-213. https://doi.org/10.1016/S0925-7535(97)00052-0	These are foundational texts on SA, SSA, and risk perception in dynamic systems.	Endsley SA Levels, SSA
Risk Perception and Risk Assessment	• Rasmussen, J. (1997). Risk management in a dynamic society: a modelling problem. *Safety science*, *27*(2-3), 183-213. https://doi.org/10.1016/S0925-7535(97)00052-0 • Rausand, M. (2013). *Risk assessment: Theory, methods, and applications*. (Vol. 115) John Wiley & Sons.	These directly inform the bow-tie model and study's approach to measuring risk.	Bow-Tie Framework, FAA SMS
Information Characteristics & Modality	FAA (2021). Flight Information Services–Broadcast (FIS-B) Program Status Memo. https://www.faa.gov/sites/faa.gov/files/2022-01/PL_115-254_Sec_502_Air_Traffic_Control_Modernization_NextGen_COMPLETE.pdf	Describes systems used for real-time information delivery, relevant to modalities and timing.	Volume, relevance, timing
Trust in Information & Automation	• Prinzo, O. V. (2001). Data-linked pilot reply time on controller workload and communication in a simulated terminal option (No. DOTFAAAM018).		Trust in systems & automation

Topic Area	Title	Relevance to Study	Constructs or Variables Supported
	• Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors*, *46*(1), 50-80. https://doi.org/10.1518/hfes.46.1.50_30392		
Contextual Complexity in Aviation	• FAA (2022). FAA Aerospace Forecast FY 2022–2042. https://www.faa.gov/sites/faa.gov/files/2022-06/FAA_Aerospace_Forecasts_FY_2022-2042.pdf • Federal Aviation Administration. (2017). Instrument Procedures Handbook (IPH) (FAA-H-8083-16B). Flight Standards Service. Retrieved from https://www.faa.gov/sites/faa.gov/files/regulations_policies/handbooks_manuals/aviation/instrument_procedures_handbook/FAA-H-8083-16B.pdf • Federal Aviation Administration. (2023). Pilot's Handbook of Aeronautical Knowledge (FAA-H-8083-25C). Flight Standards Service. Retrieved from https://www.faa.gov/regulations_policies/handbooks_manuals/aviation/faa-h-8083-25c.pdf • U.S. Department of Transportation, Federal Aviation Administration. (2023).	These provide operational demand and complexity projections influencing future workload. They also describe real-world operating conditions across airspace, flight phases, and aircraft types.	Phase, airspace, aircraft, region

Topic Area	Title	Relevance to Study	Constructs or Variables Supported
	Aeronautical Information Manual (AIM). https://www.faa.gov/air_traffic/publications/media/aim_basic_dtd_4-20-23.pdf		
Measurement and Instrument Development	• Stanton, N. A., Salmon, P. M., Rafferty, L. A., Walker, G. H., Baber, C., & Jenkins, D. P. (2017). Human Factors Methods: A Practical Guide for Engineering and Design. https://www.academia.edu/download/30211217/9781409457541_us.pdf • Brooke, J. (1996). SUS: A quick and dirty usability scale. http://www.tbistafftraining.info/smartphones/documents/b5_during_the_trial_usability_scale_v1_09aug11.pdf • Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. Human Factors, 37(1), 32–64. https://doi.org/10.1518/001872095779049543	These are foundational tools or guides for designing and validating study instruments.	Surveys, interviews, scenario probes
Aviation Safety and Regulatory Frameworks	• FAA Order JO 7110.65X (cross-listed) • AIM, IPH, PHAK (cross-listed) • NTSB Recommendations (cross-listed) • ICAO (2022). Human Performance Manual (Doc 10151).	Define formal frameworks and regulatory reference points that guide safety expectations.	FAA SMS, ICAO, NTSB

Topic Area	Title	Relevance to Study	Constructs or Variables Supported
	https://www.icao.int/safety/OPS/OPS-Section/Documents/Advance-unedited.Doc.10151.alltext.en.pdf		
Info Prioritization & Decision Support	• FAA ARC on Human Factors in NOTAM Modernization (cross-listed) • Jeppesen NOTAM Decoder (cross-listed) • Lukic (2014); Barshi et al. (2016); Molesworth (2015) (cross-listed)	Contributes to understanding evaluation of and actions regarding information under pressure.	Filtering models, prioritization
Team Dynamics and Coordination	Dismukes, R. K., Berman, B. A., & Loukopoulos, L. (2017). *The limits of expertise: Rethinking pilot error and the causes of airline accidents*. Routledge. https://doi.org/10.4324/9781315238654	Explores pilot error in team contexts; highlights misalignments and human variability in complex operations.	Role coordination, cross-team SSA

12. Appendix D: Contextual Complexity Dimensions

12.1. Information Characteristics

1. Volume

Definition

Volume is the count and density of discrete SCI elements presented within a decision window, relative to human short-term/working-memory capacity and available processing time. Human capacity reliably hovers around 3–5 chunks with the classic 7±2 range reflecting task- and chunking-dependent upper bounds (Cowan & Chen, 2008; Thalmann et al., 2019). Aviation standards further advise partitioning when a display contains too much data for presentation in a single frame, to avoid clutter and overload (Mejdal et al., 2001).

Manipulation: Volume will be varied by adjusting the number of discrete information elements presented in a scenario.

Volume	Description	Note
Low	~3–5 discrete items	At or just under the Cowan capacity
Medium	~8–10 items	Near the Miller 7±2 band, requiring some chunking/visual structuring
High	~15–20 items	Well beyond working-memory capacity, deliberately requiring paging/filters per MIL-STD guidance to create measurable overload pressure
Volume thresholds (3–5; 8–10; 15–20) are a priori manipulations designed to span below / near /well above established working-memory limits, grounded in Cowan's 3–5-chunk estimate and Miller's 7±2, and consistent with MIL-STD guidance to partition when data exceed a single frame. These should be validated with SMEs in pilot runs before main data collection.		

These cut points are design choices intended to span below/at/above capacity and will be confirmed in SME pilot runs; they are consistent with cognitive limits and display-clutter guidance.

2. Relevance

Relevance refers to the degree to which delivered information is meaningful, actionable, and appropriate to the operator's current context, task, and role. In high-stakes operational settings, irrelevant or low-priority information can obscure critical cues, increase workload, or delay appropriate decision-making (Endsley & Jones, 1997; Orasanu et al., 2011).

Manipulation: Signal-to-noise (US DoD, 2012).

Signal	Noise	Description
High Relevance	≥80% actionable items	high signal, low distractors
Low Relevance	≤40% actionable items	low signal, high distractors
Relevance bands (≥80% vs ≤40% actionable) are experimental signal/noise conditions aligned with HF standards to minimize non-essential information; they create strong contrast needed for inference and should be SME-checked for realism.		

These percentage bands are experimental manipulations set to produce clear high/low conditions while aligning with HF guidance to minimize non-essential information; final cut points will be checked with SMEs for face validity before deployment. These percentage bands are experimental manipulations set to produce clear high/low conditions while aligning with HF guidance to minimize non-essential information; final cut points will be checked with SMEs for face validity before deployment.

3. Timeliness

Timeliness refers to the degree to which information is delivered at a point in time when it can effectively support the operator's decision-making and action. Untimely information, even if accurate, may be unusable or counterproductive in time-critical operations (Orasanu et al., 2001; Orasanu et al., 2002; Endsley, 1995). Manipulation: (Wickens & Hollands, 2013; Endsley, 1995)

- Optimal: Delivered during low/moderate workload before the decision point.
- Poorly Timed: Delivered during peak workload or after decision point.

Operational timing windows should be derived from task analysis (phase-specific actions) and will be pre-specified by SMEs for each scenario.

Measurement:

- Subjective: Likert scale from "too late to use" to "perfectly timed."
- Behavioral**:** Missed actions or delayed responses.

Timeliness windows are tied to phase-anchored task analyses (AIM/NTSB) and SA measurement traditions that control probe timing; windows should be specified per scenario with SME review.

12.2. Contextual Complexity

Traffic complexity refers to the interaction of multiple operational variables – such as aircraft density, heterogeneity, trajectory convergence, equipage differences, pilot proficiency, and sector geometry – that shape controller workload and influence SCI processing. It is not solely a function of traffic volume but includes dynamic elements such as coordination demands, rate of change, and pilot response variability. In this framework, traffic complexity is embedded across multiple dimensions of contextual complexity, including traffic volume, operation type, airspace class, weather conditions, and SCI urgency. Its influence is particularly salient during high-consequence phases of flight or in sectors with diverse aircraft performance and communication capabilities (Eurocontrol, 2006; FAA, 2023d).

Definition

Contextual complexity refers to the dynamic and interacting operational variables that shape how tasks are performed and information is processed.

In aviation, this includes phase of flight, geographic location, airspace class, aircraft type, weather conditions, and traffic volume. High contextual complexity increases the likelihood of information overload and miscommunication (Orasanu et al., 2011; Rasmussen, 1997).

1. **Operation Type**

Definition

Operation type categorizes the nature and purpose of the flight based on regulatory and operational frameworks. Common classifications include commercial airline (Part 121), charter or on-demand operations (Part 135), corporate/business aviation (Part 91K), general aviation (Part 91), agricultural operations, UAS (Part 107), and experimental or test flights. Each type presents different communication environments, SCI expectations, and safety oversight levels (FAA, 2023c).

- Commercial Airline
- General Aviation
- Rotorcraft Operations
- UAS Operations.

2. **Aircraft Type**

Definition

Aircraft type refers to the specific model or series of an aircraft characterized by its design features, performance capabilities, and operational limitations. This includes differences in avionics, propulsion (e.g., piston, turboprop, jet), weight class, and certification category, all of which affect pilot workload, required procedures, and communication needs (FAA, 2023b).

- Fixed-Wing Jet

- Fixed-Wing Propeller
- Rotorcraft
- UAS.

3. Airspace Class

Definition

Airspace class defines the regulatory and operational environment in which aircraft operate, structured by altitude, traffic density, and required separation services. Classes A through G are distinguished by levels of ATC involvement, equipment requirements, and pilot certification standards, each influencing the flow of SCI and pilot-controller interactions (FAA, 2023a).

Class	Altitude Range	Who Flies Here?	Helicopters	Key Features
A	18,000 feet MSL to FL600 (60,000 feet MSL)	Commercial airline pilots, corporate jet pilots, military pilots, IFR general aviation pilots	Rare for helicopters due to altitude limitations, but some IFR-capable helicopters (e.g., military, offshore transport, or corporate helicopters) may operate here under Instrument Flight Rules (IFR).	Only IFR flights allowed; ATC clearance required; Mode C transponder and ADS-B Out required
B	Surface to 10,000 feet MSL	Airline pilots, corporate jets, general aviation (with clearance), military aircraft	Helicopters can operate in Class B airspace (e.g., near major cities like New York, Los Angeles), but require ATC clearance, a transponder, and ADS-B Out. Special helicopter corridors (e.g., Hudson River VFR Corridor) allow VFR operations without full clearance.	ATC clearance required; Two-way communication, Mode C transponder, and ADS-B Out required

Class	Altitude Range	Who Flies Here?	Helicopters	Key Features
C	Surface to 4,000 feet AGL	Regional airlines, general aviation, corporate aircraft, flight schools, military aircraft	Helicopters operate in Class C for medical transport (EMS), corporate flights, and news/media operations. Two-way communication and a transponder are required. ATC provides separation for IFR helicopters but allows flexibility for VFR helicopter routes.	Requires two-way communication with ATC; Mode C transponder and ADS-B Out required
D	Surface to 2,500 feet AGL	General aviation, flight training, corporate aircraft, helicopters, cargo pilots	Common for police, EMS, corporate, and training flights. Helicopters require two-way radio communication with ATC but can often receive special routing due to their ability to hover and land in confined areas.	Two-way communication required; No transponder requirement unless specified
E	Starts at 1,200 feet AGL (sometimes 700 feet or surface) up to Class A	IFR commercial pilots, general aviation, gliders, military training flights	Helicopters commonly transition through Class E airspace en route between cities, hospitals, or remote locations. IFR helicopter operations also occur here in poor weather conditions.	Controlled airspace for IFR flights; VFR flights allowed, but do not need ATC clearance
G	(Uncontrolled airspace, surface to 700/1,200 feet AGL in most	General aviation, student pilots, crop dusters, helicopters,	Helicopters frequently operate in Class G for EMS, law enforcement, agricultural, utility, and private operations. No ATC clearance or transponder is	Uncontrolled airspace; No ATC clearance required; Pilots responsible for their own separation

Class	Altitude Range	Who Flies Here?	Helicopters	Key Features
	areas)(varies by location)	gliders, ultralights	required, and pilots operate under see-and-avoid rules.	

4. Geographic Location

Definition

Geographic location refers to the spatial environment in which aviation operations occur and encompasses physical, meteorological, and infrastructural characteristics. These include terrain, regional weather variability, facility capabilities, and traffic density, each of which can impact communication demands, SCI delivery, and operational risk (FAA, 2023a; Orasaanu et al., 2011).

FAA Region	States Included			
Alaskan*	Alaska			
Central	Iowa	Kansas	Missouri	Nebraska
Eastern	District of Columbia	Detroit	Maryland	New Jersey
	New York	Pennsylvania	Virginia	West Virginia
Great Lakes	Illinois	Indiana	Michigan	Minnesota
	North Dakota	Ohio	South Dakota	Wisconsin
New England	Connecticut	Massachusetts	Maine	New Hampshire
	Rhoad Island	Vermont		
Northwest Mountain	Colorado	Idaho	Montana	Oregon
	Utah	Washington	Wyoming	
Southern	Alabama	Florida	Georgia	Kentucky
	Missouri	North Carolina	Puerto Rico*	South Carolina
	Tennessee	Virgin Islands*		
Southwest	Arkansas	Louisiana	New Mexico	Oklahoma
	Texas			
Western-Pacific	Arizona	California	Hawaii	Nevada
	American Samoa*	Guam*	Marshall Islands*	
Table shows FAA designated regions; does not take into account topography differences (e.g., high-altitude, coastal and oceanic airspace, mountains, high-density & complex airspace, commercial space operations). *Not included in this research study.*				

5. Phase of Flight

Definition

Phase of flight refers to the discrete segments of a flight's lifecycle, each with distinct communication demands, operational task loads, and information salience for controllers (ICAO, 2013; FAA, 2023d). In the ATC context, the type, urgency, and format of SCI that must be communicated can vary significantly between phases. For example, reroutes and runway configuration changes are often most time-sensitive during approach and landing, while convective weather deviations may require earlier coordination during cruise or descent.

This study does not attempt to catalogue all operational details of each phase but instead uses phase of flight as a contextual factor to examine the way SCI salience and timing requirements shift with operational context. These contextual shifts are directly tied to workload and decision-making demands, which may impact SCI prioritization and trust in information sources.

Phase	Definition
Standing	Prior to pushback or taxi, or after arrival, at the gate, ramp, or parking area, while the aircraft is stationary.
Pushback/Towing	Aircraft is moving in the gate, ramp, or parking area, assisted by a tow vehicle (tug).
Taxi	The aircraft is moving on the aerodrome surface under its own power prior to takeoff or after landing.
Takeoff	From the application of takeoff power, through rotation, and to an altitude of 35 feet above runway elevation.
Initial Climb	From the end of the Takeoff subphase to the first prescribed power reduction, or until reaching 1,000 feet above runway elevation or the VFR pattern, whichever comes first.
Enroute	Instrument Flight Rules (IFR): From completion of Initial Climb through cruise altitude and completion of controlled descent to the Initial Approach Fix (IAF). Visual Flight Rules (VFR): From completion of Initial Climb through cruise and controlled descent to the VFR pattern altitude or 1,000 feet above runway elevation, whichever comes first.
Maneuvering	Low altitude/aerobatic flight operations.
Approach	Instrument Flight Rules (IFR): From the Initial Approach Fix (IAF) to the beginning of the landing flare. Visual Flight Rules (VFR): From the point of VFR pattern entry, or 1,000 feet above the runway elevation, to the beginning of the landing flare.

Phase	Definition
Landing	From the beginning of the landing flare until aircraft exits the landing runway, comes to a stop on the runway, or when power is applied for takeoff in the case of a touch-and-go landing.
Emergency Descent	A controlled descent during any airborne phase in response to a perceived emergency situation.
Uncontrolled Descent	A descent during any airborne phase in which the aircraft does not sustain controlled flight.
Post-Impact	Any of that portion of the flight which occurs after impact with a person, object, obstacle or terrain.
Unknown	Phase of flight is not discernible from the information available.

6. Weather Conditions

Condition	Manipulation	Measurement
VFR	Clear to good visibility, ceilings above VFR minima, routine winds, no significant convective activity. Scenario cues: high visibility on displays, few weather calls, standard ATC spacing.	Weather tag in scenario script, pilot/controller reports of visibility and ceiling, absence of IFR-only procedures.
IFR	Reduced visibility and/or ceiling triggers instrument procedures, approach minima in play, increased ATC coordination. Scenario cues: instrument approach clearances, missed approach briefings, tighter altitude and routing constraints.	Use of IFR procedures logged, clearances issued, instrument approach flown.
Marginal VFR	Variable ceiling and visibility, intermittent light precipitation or haze, frequent updates to pilot reports. Scenario cues: borderline weather calls, optional approach type decisions, reroute advisories.	Frequency of weather updates and rebriefs, changes of intended procedure, workload comments tied to weather ambiguity.
Severe Weather	Convective SIGMETs, wind shear alerts, embedded thunderstorms, significant turbulence or icing. Scenario cues: urgent advisories, reroutes, holds, ground stops.	Count of hazard advisories, number of weather-driven reroutes or holds, latency to comply with weather avoidance.

7. Traffic Volume

Level	Definition	Manipulation	Measurement
Low	Sparse traffic with minimal conflicts and low communication frequency.	Few aircraft on frequency or scope, long gaps between calls, wide separation.	Radio/transmission counts per minute, conflict probe count, controller task load notes.
Medium	Moderate traffic with occasional conflicts that require routine sequencing.	Steady call cadence, periodic vectoring, standard miles-in-trail, occasional speed control.	Conflicts resolved per 10 minutes, average handoff rate, number of clearances requiring readback.
High Traffic Density	Dense traffic with frequent conflicts, compressed timelines, and sustained coordination.	Rapid-fire transmissions, simultaneous requests, tight separation, metering or flow restrictions.	Calls per minute, conflicts queued, reroute or delay directives, interruption counts.

8. SCI Risk Level

Definition

Hazard potential of a piece of information along the bow-tie pathway, defined by the likelihood that a miss or misinterpretation leads toward the top event and the severity if it does. (Aven, 2016; FAA, 2016; CCPS, 2020a; CCPS, 2020b).

Level	Definition	Manipulation	Measurement
Low	If missed or misunderstood, minimal impact on safety margin, usually recoverable within routine procedures.	Non-time-critical advisories, minor route notes, convenience updates.	SME risk rating matrix, absence of safety-critical consequences in answer key.
Medium	If missed or misunderstood, measurable reduction in safety margin that requires corrective action but remains controllable.	Runway change with ample time, moderate weather build-ups, temporary altitude amendments.	Required corrective steps in script, added workload and minor delay recorded.
High	If missed or misunderstood, credible path to incident or accident unless promptly mitigated.	Windshear alert, runway closure in use, critical altitude crossing, immediate conflict alert,	Time thresholds for safe response embedded in answer key, binary success

		severe convective penetration risk.	criteria, SME high-risk tag.
Hazard potential per bow-tie model risk pathways.			

9. Modality

Modality refers to the channel or sensory pathway through which information is delivered, such as visual (e.g., display alerts), auditory (e.g., radio communications), or tactile (e.g., vibration cues). In aviation, multimodal systems are used to reduce cognitive load, improve redundancy, and increase the likelihood of correct information interpretation under time pressure (Wickens & Hollands, 2013; Geitner et al., 2019).

Modality	Definition	Manipulation	Measurement
Audio	Information delivered through the auditory channel, typically ATC or intercom voice, aural alerts, tones.	Deliver clearances and advisories by radio or synthetic voice, optionally under high frequency congestion or overlapping audio. Add or remove aural alert tones to test masking and priority.	Copy accuracy and readback errors, requests for repeats, missed or stepped-on calls, response latency from audio onset.
Visual	Information delivered through visual channels, such as EFB pages, cockpit displays, maps, symbology, text messages.	Present NOTAMs, weather layers, route amendments, and alerts on EFB or avionics; vary visual density and placement, require or remove paging/scrolling to test discoverability.	Dwell time on pages, page changes, fixation or selection logs if available, accuracy of actions tied to visual prompts, time from on-screen cue to action.
Tactile	Information delivered through touch-based cues (e.g., vibration, or haptic alerts) designed to capture attention when visual and auditory channels are saturated, supporting rapid detection of urgent conditions without adding to communication or display load.	Present urgent SCI using vibration or haptic alert cues, either alone or paired with visual/auditory alerts; vary timing, intensity, and redundancy to test whether tactile delivery improves noticeability during high workload or high channel saturation.	Response latency from tactile cue onset, acknowledgment rate, correct identification of alert meaning, missed or ignored tactile alerts, and participant ratings of usefulness, intrusiveness, and compatibility with workload.

Example Contextual Complexity

Complexity Level	Weather Conditions	Traffic Density	Coordination Demands	Information Asymmetries
Low	Clear weather, minimal disruption	Light traffic, ample airspace	Minimal coordination required (single role, stable conditions)	All participants have identical and complete information
Medium	Moderate weather (e.g., scattered storms, turbulence)	Moderate traffic, routine reroutes	Some cross-role coordination needed (pilot–controller, or controller–dispatcher)	Minor discrepancies in available information requiring clarification
High	Severe weather (e.g., convective storms, icing, diversions)	High traffic, constrained airspace	Intensive multi-role, multi-facility coordination under time pressure	Significant gaps or delays in information, requiring inference or confirmation

13. Appendix E: Sample Size Approach

13.1. Sample Size Formulas

Basic Sample Size Formula for Populations Over 10,000:

$$n = \frac{Z^2 p(1-p)}{e^2}$$

Where:

Z = 1.96 (Z-score for 95% confidence)

p = 0.5 (most conservative estimate of proportion)

e = Desired margin of error (expressed as a decimal: 0.05 for (±5%), 0.07 for (±7%)

Formula with Finite Population Correction for Small Populations:

$$n_{adj} = \frac{n_0}{1 + \left(\frac{n_{0-1}}{N}\right)}$$

Where:

η_0 = initial sample size for an infinite population (384)

N = population size

13.2. Sample Size Calculations

Position	Estimated Population Size	Sample Size (±5%)	Sample Size (±7%)
Air Traffic Controllers			
TRACON	4,000	350	183
Tower	4,756	355	185
En Route (ARTCC)	4,328	353	184
Center Radar Approach (CERAP)	100	80	65
FAA Contract Tower	1,571	309	174
Sub Total	**14,755**	**1,447**	**791**

Note. Sample sizes assume a 95% confidence level, $p = 0.05$, and desired margins of error of ±5% or ±7%, with finite population correction applied for small samples (Cochran, 1977; Israel, 1992).

* Population size estimate is approximate. It is important to include contract tower participants alongside FAA tower participants in sample size calculations because they operate under different structures, resources, and constraints. Their inclusion enhances the generalizability of the findings across the full range of tower environments, captures variability in procedures and communication practices, and ensures that safety recommendations are relevant across both FAA-operated and non-FAA-operated towers. Excluding them risks overlooking context-specific challenges that could influence how safety-critical aeronautical information is processed, prioritized, or acted upon.

14. Appendix F: Phase I Data Collection Instruments

14.1. Decision Support Variables

14.1.1 Criticality of Information

Definition

The importance of specific information in ensuring the safety and effective decision-making during flight. High-criticality information often addresses immediate risks to safety.

Relationship to Aviation Safety

Information deemed critical is central to the user's ability to address threats, manage system failures, and maintain safe operation. Missing critical data can directly lead to accidents or fatalities.

Benchmark Rating Scale

Value	Rating	Description
1	Not critical	The information has no impact on flight safety or operations.
2	Low criticality	The information has a minor impact on safety but does not require immediate attention.
3	Moderate criticality	The information impacts operations or safety under certain conditions.
4	High criticality	The information significantly impacts safety and requires prompt attention.
5	Extremely critical	The information is essential for preventing immediate harm or catastrophic outcomes.

Theoretical Foundations

Endsley, 1995; Orasanu et al., 2002; Orasanu et al., 2001; von Thaden, 2008; Lee et al., 2005

14.1.2 Clarity/Interpretability of Information

Definition

The extent to which the information is clearly presented and unambiguous.

Relationship to Aviation Safety

Misinterpreted data leads to increased workload and potential safety hazards.

Benchmark Rating Scale

Value	Rating	Description
1	Very unclear	Difficult or impossible to interpret.

2	Somewhat unclear	Key parts were ambiguous.
3	Neutral	Neither easy nor difficult to interpret.
4	Clear	Easy to understand in real-time.
5	Very clear	Instantly and unambiguously understood.

Theoretical Foundations

Endsley, 1995; Degani & Wiener, 1997

14.1.3 Preventive Value of Information

Definition

The extent to which a piece of information could help prevent a hazardous, unsafe, or undesirable event if received, understood, and acted upon. *Mitigation Potential* refers to its capacity to reduce the severity or downstream effects of an emerging threat or system failure. This construct captures the safety-buffering utility of information.

Relationship to Aviation Safety

In aviation, timely and contextually relevant information often serves as a first line of defense against operational risk. Data that enables early detection of hazards (e.g., pilot reports, SIGMETs, or NOTAMs) can either prevent an incident or mitigate its effects. The higher the preventive value, the more critical the information is to proactive safety management.

Benchmark Rating Scale

Value	Rating	Description
1	No value	This information has no potential to prevent or reduce harm.
2	Low value	May provide minor context, but unlikely to change outcome.
3	Moderate value	Helps shape awareness or adjust behavior to reduce risk.
4	High value	Directly enables actions that reduce threat or error severity.
5	Critical value	Without this information, a harmful event likely will occur.

Theoretical Foundations

Endsley, 1995,; Reason, 1990; Rasmussen, 1997; Hollnagel, 2017; Van Benthem & Herdman, 2020

14.1.4 Scope of Operational Impact

Definition

Refers to how many aspects of a specific operation (e.g., flight plan, routing, separation, crew procedure) are influenced by the information. Unlike systemic scope, this construct is internal to a particular control task and measures how broadly a single operation is affected.

Relationship to Aviation Safety

Information with high operational scope can alter navigation, fuel planning, communication sequences, or emergency planning within a single mission or shift. This increases task complexity and introduces higher cognitive workload, increasing the potential for error if not managed appropriately.

Benchmark Rating Scale

Value	Rating	Description
1	Minimal impact	Affects one task or decision point (e.g., frequency change).
2	Low impact	Affects one or two aspects of the current operation.
3	Moderate impact	Affects several tasks or required brief plan updates.
4	High impact	Requires coordination of multiple tasks or procedural changes.
5	Pervasive impact	Requires reconfiguration of entire operational flow (e.g., reroute, delay, new clearance).

Theoretical Foundations

Endsley, 1995; Wickens & Carswell, 2021; Sweller, et al., 2011; FAA, 2023b

14.1.5 Severity of Consequences

Definition

The potential damage or harm caused if critical information is not provided, is delayed, or is misunderstood.

Relationship to Aviation Safety

Severity assessments help prioritize the development of systems and protocols for managing the most dangerous situations, such as fire or loss of cabin pressure.

Benchmark Rating Scale

Value	Rating	Description
1	No consequences	The safety event has no measurable consequences.
2	Minor consequences	The event results in negligible effects on safety or operations.

3	Moderate consequences	The event results in operational challenges or minor injuries.
4	Major consequences	The event causes major injury or system failure.
5	Catastrophic consequences	The event results in loss of life, major system failure, or total operational breakdown.

Theoretical Foundations

Dekker et al., 2008; ICAO, 2013; Rasmussen, 1997

14.1.6 Magnitude of Change Needed

Definition

The perceived need for changes to the information content, delivery method, or interface.

Relationship to Aviation Safety

Feedback on necessary system improvements guides system iteration and prevents recurring design failures.

Rating Scale

Value	Rating	Description
1	No changes	System is optimal as-is.
2	Minor tweaks	Aesthetic or small usability adjustments.
3	Moderate changes	Rewording or improved timing needed.
4	Major revisions	Significant restructuring or modality shift required.
5	Complete overhaul	Information failed to meet operational needs.

Theoretical Foundations

Nielsen, 1994; FAA Human Factors Design Standard, 2016

14.1.7 Systemic Scope of Impact

Definition

The breadth of effect that a given piece of information has across the larger air traffic or aviation system, including multiple aircraft, airspace sectors, control centers, or operational functions. It represents how far-reaching the consequences of that information are beyond a single actor.

Relationship to Aviation Safety

Information with high systemic impact affects coordination, synchronization, or safety across multiple flight crews, ATC units, or systems. Examples include region-wide weather alerts, TFRs, or NOTAMs that change traffic flow. Miscommunication or delay in such information can propagate system-wide disruptions or hazards.

Rating Scale

Value	Rating	Description
1	Isolated impact	Affects only my individual aircraft or area of control.
2	Limited local impact	Affects my team or position, but not the broader system.
3	Moderate spread	Influences nearby sectors, aircraft, or functions.
4	Broad system impact	Requires coordination across multiple teams or sectors.
5	Widespread systemic effect	Affects entire operational areas or networks (e.g., region-wide reroutes).

Theoretical Foundations

Leveson, 2011; Reason, 1990; ICAO, 2013; Dekker et al, 2008; Hollnagel, 2017

14.1.8 Time Sensitivity

Definition

The length of time an operator can safely delay taking action on a piece of information after it has been received. It reflects the temporal flexibility available before inaction becomes unsafe, ineffective, or leads to operational disruption.

Relationship to Aviation Safety

This construct captures the decision delay tolerance or "response window", the amount of time available to evaluate and act on information after it is received. Even if the information doesn't feel immediately urgent, a delayed response can compromise the user's ability to manage unfolding threats.

Rating Scale

Value	Rating	Description
1	No tolerance	Requires immediate action; any delay would be unsafe or unacceptable.
2	Low tolerance	Action needed within seconds or very short time to remain effective.
3	Moderate tolerance	Could have a slight delay, but any delay beyond a short window may cause problems.
4	High tolerance	There is flexibility, action could be deferred without major consequence.

5	Very high tolerance	Substantial time available to act; delay would not meaningfully affect outcome.

Theoretical Foundations

Mosier et al., 2017a; Endsley, 1995; Gonzalez et al., 2004; Parasuraman et al, 2000

14.2. Structured Interview Protocol Example

Purpose: To capture operator perspectives on the clarity, timing, volume, relevance, and contextual factors of SCI delivery, as well as perceived workload and decision-making impacts.

Section	Example Prompts
Scenario Recall & Key Events	“Can you walk me through the sequence of events from your perspective?” “Which pieces of information stood out as most important?”
Information Characteristics	Volume: “Did you feel the amount of information you received was too little, about right, or too much? Why?” Relevance: “How useful or actionable was the information provided?” Timeliness: “Was the information provided at a time you could act on it effectively?”
Cognitive & Operational Impact	“How did the information delivery affect your workload?” “Did it change your mental model of the situation?”
Contextual Modifiers	“Did any environmental, operational, or traffic conditions affect your ability to use the information?”
Suggestions & Improvements	“What changes would make the information delivery more effective in similar scenarios?”

15. Appendix G: Measures

15.1. Cognitive Workload

Definition

Cognitive workload refers to the mental effort required to perceive, interpret, and act on information while performing operational tasks. In aviation contexts, these tasks include monitoring system states, communicating with other operators, managing automation, and making time-sensitive decisions under uncertainty (Wickens, 2002; Wickens, 2008; Stein, 1985). Cognitive workload influences information processing capacity, situation awareness, and decision latency, and is therefore a critical factor in evaluating the timing and prioritization of safety-critical aeronautical information.

Air Traffic Workload Input Technique (ATWIT)

Scoring Rubric and Data Collection Approach

Element	Description
Rating Scale	1 = Very Low to 7 = Very High perceived workload
Prompt Type	Time-based prompts (e.g., every 2–3 mins) and event-triggered prompts (e.g., immediately after SCI delivery)
Data Collection Tool	Digital entry via tablet, keyboard, or on-screen prompt
Timestamping	All ATWIT entries synchronized with scenario clock and SCI delivery events
Aggregation	Mean scores, standard deviations, and time series workload profiles, by scenario condition, participant, and role.

Application Example:

During a simulated approach phase, a SIGMET injection triggers an event-based ATWIT prompt. A controller rates workload as a 6/7 due to concurrent traffic coordination. This rating is time-aligned with system logs and qualitative behavior coding.

Justification:

ATWIT is widely used in aviation and complex sociotechnical domains to capture rapid workload fluctuations without interrupting task performance (Stein, 1985; Loft et al., 2007; Roth et al., 2021).

Qualitative Coding Framework for Workload Indicators

This coding scheme is applied within ±30 second windows surrounding ATWIT prompts to contextualize workload ratings.

Code	Description
INT-COMP	Evidence of cognitive integration difficulty (e.g., combining multiple information sources)
MOD-MIS	Modal mismatch or switching burden (e.g., conflicting visual and auditory inputs)
ATT-SPL	Split attention across tasks or displays
ACT-BURD	Difficulty coordinating or acting on SCI
OVERLAP	Multiple codes present, indicating sustained overload

Reliability:

Inter-rater reliability will be assessed using Cohen's Kappa, with $\kappa \geq 0.75$ indicating acceptable agreement (Hallgren, 2012; Cohen et al., 215; Wilson et al., 2020).

Mixed-Methods Integration

Step	Approach
Temporal Overlay	Align ATWIT timestamps with SCI delivery and behavioral codes
Triangulation	Identify convergence between ATWIT spikes, behavioral indicators, and SCI events
Pattern Tracking	Compare workload trajectories across scenarios, roles, and phases of flight
Validation Strategy	Assess stability of workload patterns across Phase II and Phase III testing

Supplemental Post-Scenario Dimensional Probes

Although ATWIT is a single-item real-time workload measure, post-scenario probes will be used to capture workload dimensions aligned with the adapted Multiple Resource Theory framework (Hilburn, 2004).

Dimension	Sample Probe Item
Cognitive Integration Complexity	I had to mentally combine information from multiple sources.
Modal Load Compatibility	The delivery method interfered with other tasks.
Attentional Demand Type	I had to divide attention between multiple tasks or people.
Actionability and Coordination Burden	I had to coordinate or confirm actions before responding.

Use a Likert scale (1–7) with anchors like:

1 = Strongly Disagree

2 = Disagree

3 = Somewhat Disagree

4 = Neutral

5 = Somewhat Agree

6 = Agree

7 = Strongly Agree

Mixed-Methods Coding Guide for Cognitive Workload Assessment

Adapted Workload Dimension	ATWIT-Relevant Subconstruct	Quantitative Scoring Approach	Qualitative Indicators
Cognitive Integration Complexity	Task Complexity	High ATWIT during multi-source SCI integration	Clarification requests, pauses
Modal Load Compatibility	Channel Switching	ATWIT spikes after multimodal SCI	Display toggling, message repetition
Attentional Demand Type	Divided Attention	ATWIT fluctuations during coordination	Frequency juggling, strip switching
Actionability and Coordination Burden	Action Planning	Elevated ATWIT when multi-role coordination required	Calls to other controllers or escalation

15.2. Situation Awareness

Definition

The operator's perception of relevant elements in the environment, comprehension of their meaning, and projection of their status in the near future (Endsley, 1995).

Measured using:

1. Situation Awareness Rating Technique (SART)
   - Purpose: Subjective assessment of SA immediately after scenario completion.
   - Components:
     - Demand on Attentional Resources (perceived complexity, instability, variability)
     - Supply of Attentional Resources (familiarity, concentration, spare mental capacity)
     - Understanding of the Situation (information quantity, quality, and integration)
   - Scoring: Composite SA score calculated as (Understanding + Supply) − Demand.
   - Administration: Administered post-trial via Likert-type items (1 = very low, 7 = very high).

Citations: Taylor, 2017

2. Performance-Based SA Indicators
   - Purpose: Objective behavioral markers tied to scenario events.
   - Examples:
     - Accuracy of identifying relevant hazards or traffic.
     - Correct sequencing of procedural steps.
     - Timeliness of hazard recognition and response.
   - Collection: Derived from task logs, audio transcripts, and simulator playback.

Theoretical Foundations

Durso et al., 2017; Endsley, 1995; Gorman et al., 2005; Gorman et al., 2006

15.3. Shared Situation Awareness

Definition

The degree to which team members possess the same SA on shared SA requirements, the overlapping portion of SA required across team roles (Endsley & Jones, 1997)

Measure using:

- Team Situation Awareness Global Assessment Technique (TSAGAT): Developed to assess shared SA by capturing team members' responses to SA probes. It was validated in trauma team training, showing that shared SA metrics correlate with team performance
  - Purpose: Self-report measure of perceived alignment between participants' understanding and their teammates'.
  - Sample Items:
    - "My teammates and I had the same understanding of the current situation."
    - "We anticipated the same future events and outcomes."
  - Format: 7-point Likert scale; administered post-scenario.

Theoretical Foundations

Endsley, 1995a; Endsley, 1998; Gorman et al., 2005; Gorman et al., 2006

- SSA Accuracy Score (Objective Measure)
  - Purpose: Compare individual SA probe responses across team members for convergence.
  - Calculation:
    - Compute pairwise agreement scores on freeze-probe responses (correctness and consistency).
    - Higher agreement indicates stronger SSA.
- Communication Alignment Coding
  - Purpose: Analyze recorded communications for alignment cues.
  - Indicators:
    - Repeated confirmation statements ("copy," "affirmative")
    - Shared terminology for hazards or traffic
    - Minimal corrective exchanges or clarifications
  - Scoring: Percentage of exchanges rated as "aligned" by coders using a predefined scheme.

Triangulation: Combine subjective ratings (SART, Shared SA Questionnaire) with objective indicators (performance accuracy, probe convergence, communication coding).

Link to Hypotheses: SA and SSA scores will be compared across modalities, traffic volume, weather conditions, and SCI risk levels to evaluate their influence on decision-making and prioritization effectiveness.

15.4. Trust in Information Source

Definition

The extent to which the user considers the information source to be reliable and credible.

Relationship to Aviation Safety

Trust influences system use and compliance. Low trust can lead to rejection of correct safety information.

Rating Scale

Value	Rating	Description
1	No trust	Assumed the information was unreliable or incorrect.
2	Low trust	Questioned the information’s accuracy or intent.
3	Moderate trust	Trusted only under specific conditions.
4	High trust	Considered the source reliable in most contexts.
5	Complete trust	Fully relied on the source without hesitation.

Theoretical Foundations

Lee & See, 2004; Merritt & Ilgen, 2008

15.5. Perceived Usefulness of Information

Definition

The extent to which the information supports decision-making, situation awareness, or task performance.

Relationship to Aviation Safety

Usefulness drives information retention and appropriate response. If perceived as useful, safety-critical information is more likely to be trusted and acted upon.

Benchmark Rating Scale

Value	Rating	Description
1	Not useful	Did not assist with decision-making or flight safety.
2	Slightly useful	Offered limited or tangential support.
3	Moderately useful	Provided relevant information in some situations.
4	Very useful	Contributed significantly to operational decisions.
5	Extremely useful	Essential to safe and effective performance.

Theoretical Foundations

Parasuraman et al., 2000; Wickens et al., 2015

15.6. Performance

Definition

The extent to which the system and its human operators achieve safe and effective management of SCI within the operational demands of the NAS.

Specifically, performance encompasses both human-level and system-level outcomes, reflecting ability to maintain situation awareness, shared situation awareness, and acceptable cognitive workload while operating within established safety risk boundaries (FAA, 2016; Endsley, 1995).

Measured using:

- System logs
- Observer logs
- Accuracy of actions
- Latency (time from delivery to acknowledgment/action)
- Procedural adherence

This measure captures:

- Accuracy of operational actions (correct vs. incorrect)
- Timeliness of responses (latency between information delivery and acknowledgment or action)
- Adherence to standard operating procedures
- Types and frequency of errors:

- Omission: Definition (failure to act when required)
- Commission: Definition (incorrect action taken)
- Timing: Definition (delayed or premature action)
- Procedural deviation: Definition (action inconsistent with procedures)

Timing and Errors

This measure captures:

- Objective response time (latency) from delivery to acknowledgment/action
- Frequency of:
    - Omission errors
    - Commission errors
    - Timing errors
    - Procedural deviations
- Patterns of error occurrence across different conditions (timing, modality, workload)

15.7. Mental Models & Decision Strategies

Definition

Mental Models & Decision Strategies refer to participants' cognitive representations of the system and the task environment, and to how they prioritize and act on safety-critical information. This measure captures not only what decisions were made but also why they were made, including the influence of workload, context, and expectations.

Measures

Structured post-scenario interviews will elicit qualitative insights about how participants interpreted the scenarios, prioritized information, and adapted their strategies to workload and context. Interviews will be audio-recorded, transcribed, and analyzed thematically. Structured post-scenario interviews will be used to elicit qualitative data on:

- How participants conceptualize and interpret information during tasks
- Heuristics and strategies used to prioritize and act on information
- Contextual factors influencing decisions (e.g., workload, phase of flight)
- Alignment (or misalignment) between participant expectations and system behavior
- Narratives explaining why certain actions were taken (or not)

Theoretical Foundations

Orasanu, 1991; Lai et al., 2019

15.8. Risk Perception

Definition

Risk perception refers to the subjective evaluation by operators of the likelihood and severity of adverse outcomes. It is shaped by personal experience, cognitive workload, trust in systems, information quality, and cultural or organizational norms (Rasmussen, 1997; Reason, 1990).

Flight Risk Perception Scale (FRPS)

Example Items:

Perceived Likelihood of Risk

Definition: How likely is it that the situation described would lead to an operational issue or hazard?

1. The delayed delivery of weather information would likely cause pilot confusion or misjudgment.
2. The absence of timely PIREPs would likely increase my sector's operational risk.
3. Late SCI delivery would likely result in degraded situation awareness.
4. If NOTAMs were inaccurate or missing, a coordination breakdown would be likely.

Perceived Severity of Consequences

Definition: If an issue occurred, how serious would the consequences be?

5. If the SCI was misinterpreted, the operational consequences would be severe.
6. A breakdown in communication about this SCI would have serious safety implications.
7. The delayed information in this scenario could result in cascading risks across sectors.
8. Failure to act on this SCI in time would result in high workload and coordination burden.

Perceived Controllability / Risk Management Confidence

Definition: To what extent did you feel able to control or manage the risk in the scenario?

9. I felt confident in my ability to compensate for the SCI delay.
10. The situation was manageable with existing procedures and coordination.
11. I had sufficient tools and support to mitigate the risk presented.
12. I could have easily regained control if the SCI had been further delayed.

Note: Items 9–12 are reverse-coded during composite score calculation so that higher total scores consistently reflect higher *perceived risk.*

Scoring Table for Adapted FRPS

Domain	Item Numbers	Score Range Per Item	Interpretation	Notes
Likelihood	1–4	1 (Very Unlikely) to 5 (Very Likely)	Risk expectancy	Higher = more likely risk
Severity	5–8	1 (Very Low) to 5 (Very High)	Risk consequence	Higher = greater perceived consequence
Controllability	9–12	1 (Very Low) to 5 (Very High)	Risk mitigation capacity	Reverse-coded in composite score
Composite Score	All 12 items	Avg of all (with 9–12 reversed)	Overall risk perception index	Higher = greater perceived risk

Use a Likert scale (1–7) with anchors like:

1 = Very Low / Very Unlikely

2 = Low / Unlikely

3 = Neither Likely nor Unlikely

4 = High / Likely

5 = Very High / Very Likely

Digital Delivery & Conditional Logic (for Qualtrics or Tablet Interface)

Integrate the FRPS in your simulation debrief platform using conditional logic, particularly in Qualtrics or a tablet-based app.

Logic Flow Example:

1. Display Instructions:

   "You have just completed a scenario in which safety-critical information (e.g., a NOTAM, weather update, or PIREP) was presented under varying conditions. Please respond to the following items based on your immediate impressions."

2. Embed Vignette Prompt:

   "During the scenario, an urgent weather update was received late in the flight phase. The pilot's actions following the update were delayed."

3. Conditional Branching (Optional, by Modality):

   - If SCI delivery method = voice radio: "Rate your perceptions based on the radio-delivered information."
   - If SCI delivery method = digital interface: "Rate your perceptions based on the automated SCI system."

4. Randomize Item Order (to reduce order bias)

5. Reverse-code Controllability Items Automatically
6. Timing:
    - Trigger immediately after ATWIT completion.
    - Store timestamp to calculate duration of response.

Theoretical Foundations

Winter et al. (2019)