NASA/TM—20205010988

Identification of Scenarios for System Interface Design Evaluation
CAST SE-210 Output 2
Report 5 of 6

Randall J. Mumaw
San Jose State University Foundation

Dorrit Billman
San Jose State University Foundation

Michael S. Feary
NASA Ames Research Center

March 2019

NASA STI Program…in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at http://www.sti.nasa.gov

- E-mail your question to help@sti.nasa.gov

- Phone the NASA STI Information Desk at 757-864-9658

- Write to:
  NASA STI Information Desk
  Mail Stop 148
  NASA Langley Research Center
  Hampton, VA 23681-2199

NASA/TM—20205010988

Identification of Scenarios for System Interface Design Evaluation
CAST SE-210 Output 2
Report 5 of 6

Randall J. Mumaw
San Jose State University Foundation

Dorrit Billman
San Jose State University Foundation

Michael S. Feary
NASA Ames Research Center

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California

March 2019

Available from:

NASA STI Program
STI Support Services
Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199

This report is also available in electronic form at http://www.sti.nasa.gov
or http://ntrs.nasa.gov/

Table of Contents

Acronyms and Definitions

AC.............................Advisory Circular
ASA ....Airplane State Awareness
ATC ....air traffic control
CAST ....Commercial Aviation Safety Team
CFR............................Code of Federal Regulations
CG..............................center of gravity
CIAIAC......................Comisión de Investigación de Accidentes e Incidentes de Aviación Civil
ConOps ....Concept of Operations
CTA ....cognitive task analysis
FAA ....Federal Aviation Administration
FMEA ....Failure Modes and Effects Analysis
FMS ....flight managment system
FPA ....flight path angle
FTA............................Fault Tree Analysis
LOC ....loss of control
LOC-I.........................Loss of Control In Flight
MCP...........................mode control panel
NASA ....National Aeronautics and Space Administration
NPP ....nuclear power plant
PFD ....primary flight display
SAE............................Society of Automotive Engineers
SE...............................Safety Enhancement
STAMP ....Systems-theoretic accident model and processes
TLX............................task load index
VS ....vertical speed (automation mode)

Identification of Scenarios for System Interface Design
CAST SE-210 Output 2
Report 5 of 6

Randall J. Mumaw, Dorrit Billman, and Michael S. Feary

Executive Summary

This report is one in a series of reports describing research to develop enhanced approaches to design and evaluation of flight deck interaction. The research was conducted to support the Commercial Aviation Safety Team (CAST) response to incidents and accidents caused by failure of the flight crew to maintain aircraft attitude or energy state awareness.

This report partially responds to Safety Enhancement (SE) 210.2 (CAST, 2014) by describing development of operational scenarios for evaluation of flight crew interaction. Evaluations of human performance with an interface rely on asking humans to perform operational tasks in an appropriate operational context. The report includes the major sources of material for developing operational scenarios—which include system state, operational tasks, operational context, and safety events—and a method for developing an appropriate set of operational scenarios for interface evaluation.

SE-210 Project Overview

The Commercial Aviation Safety Team (CAST) created a team to analyze a set of incidents and accidents associated with the flight crew's loss of awareness of aircraft attitude or energy state. These events are referred to more broadly as a loss of Airplane State Awareness (ASA), and they are a substantial subset of loss of control (LOC) accidents. A subsequent CAST ASA team developed a set of mitigation strategies—referred to as Safety Enhancements (SEs)—to reduce the likelihood of ASA events occurring in the future. Six of the SEs (SE 200, 207 through 211) requested further research on mitigation strategies. Our work was specifically intended to address research identified in SE 210 Output 2 (see https://www.skybrary.aero/bookshelf/books/ 2540.pdf).

SE-210 Output 2 addresses the contributions from the flight deck interface in shaping pilot awareness. More specifically, the focus is on assessing or *evaluating the flight deck interface to determine how well it supports ASA*. We have produced a series of reports on this topic:

1) In a report titled "Overview of research approach and findings," we introduce our research approach and compile our key observations and findings. This provides a summary of how our research method developed and what we found.

2) Part of our work was a more-detailed analysis of the role of awareness in the ASA events. In a report titled "Factors that influenced Airplane State Awareness accidents and incidents," we describe a number of factors that contributed to the apparent loss of awareness or to the resulting loss of control. This analysis demonstrates that pilot attention and understanding of the system are important elements of awareness. This report also offers proposals for modifications of the interface to mitigate those

factors, and then, describes how you might evaluate the effectiveness of those proposed modifications.

3) In a related report, titled "The role of alerting system failures in loss of control accidents," we analyze how alerting for LOC-related hazards, such as low airspeed, unreliable airspeed, and approach to stall, can fail to lead to an upset recovery. Alerting is the last line of defense against flight path management hazards; it is there to ensure awareness when pilot-driven attention and awareness fail. This report looks at why alerting does not always save the day.

Through our work, we had the opportunity to become more familiar with current evaluation and certification rules, guidance, and practices that define the process for the applicants (equipment manufacturers) and the Federal Aviation Administration (FAA). Evaluation and certification of flight deck interface elements consider a broad range of flight crew performance topics. We narrowed the focus of our work to flight crew awareness, attention, and understanding, and specifically examined these aspects of human performance in relation to relevant rules (e.g., 14 CFR 25.1302) and advisory material (e.g., AC 25.1302-1). This new material offers a more complete description of flight crew performance issues in the context of the flight deck interface; however, no consistent approach for application has been established.

4) In a report titled "Evaluation issues for a flight deck interface," we attempt to describe the broader scope of flight crew performance issues to show how awareness and attention issues fit within the larger set. We also do an inventory of FAA certification rules to demonstrate that there are not rules that apply to every issue. AC 25.1302 has improved guidance for addressing evaluation of awareness, attention, and understanding, and we hope that our work can contribute to future updates of the guidance material.

5) The current report, titled "Identification of scenarios for system interface design evaluation," focuses on the operational scenarios that can be used in the context of interface evaluation. It offers several perspectives on how to ensure that pilot or flight crew performance is evaluated in an important operational context. Because it is unlikely that evaluation can be performed for the full range of operational settings, this report offers a method for selecting appropriate scenarios.

Finally, the bulk of our work in this project was focused on methods for evaluating a flight deck interface for how well it supports awareness and its critical elements: attention and understanding.

6) A report titled "Best practices for evaluating flight deck interfaces for transport category aircraft with particular relevance to issues of attention, awareness, and understanding" focuses on evaluation techniques and metrics. It considers opportunities to evaluate the interface from early to late stages of development; it considers the various ways in which the interface can fail to support awareness, attention, and understanding; and, it summarizes appropriate evaluation methods for different issues. This 6th report draws on the characterization of issues and of scenario selection presented in other reports that are relevant to awareness.

1. Introduction

This report describes guidance for evaluation of aircraft flight deck interfaces; that is, the technology—displays, controls, automation, etc.—used by the flight crew to operate the airplane. Specifically, the report focuses on evaluating changes to flight deck interfaces, including:

- new or modified system functions
- new system information
- a change to information in the form of its organization, its integration, or the way it is represented
- new interface technology or a new "widget"
- new operational procedures for flight crews

Any evaluation activity involving human performance requires the use of an operational task or set of tasks to create a setting for evaluation. We refer to a combination of these operational tasks as a scenario.

2. What is a Scenario?

Figure 1 illustrates the context for defining scenarios. On the left, there is a "system," or in this case, an airplane [Note that "system" is used to describe only the hardware and software part]. The figure shows a Concept of Operations (ConOps) that defines which system functions will be performed by the technology, which functions will be performed by the pilots, and which will be shared between the two. An example of a function currently performed by technology (i.e., an airplane system) is determining the appropriate flight control inputs to climb to an altitude target when the autopilot is flying the airplane. An example of a function performed by pilots is setting the flaps appropriately for the airplane's speed. System functions may be represented on the flight deck interface (hence, the arrows from each to the system interface). The pilot functions are performed through the interface; this is a primary purpose of the interface. But, some of the system functions should also be represented on the interface to the extent that the pilots need to maintain an awareness of the performance of those functions, or when performance is shared. The center of this figure shows that the flight deck interface, plus the operational procedures, plus the training the pilots receive need to support the set of tasks that pilots need to perform.

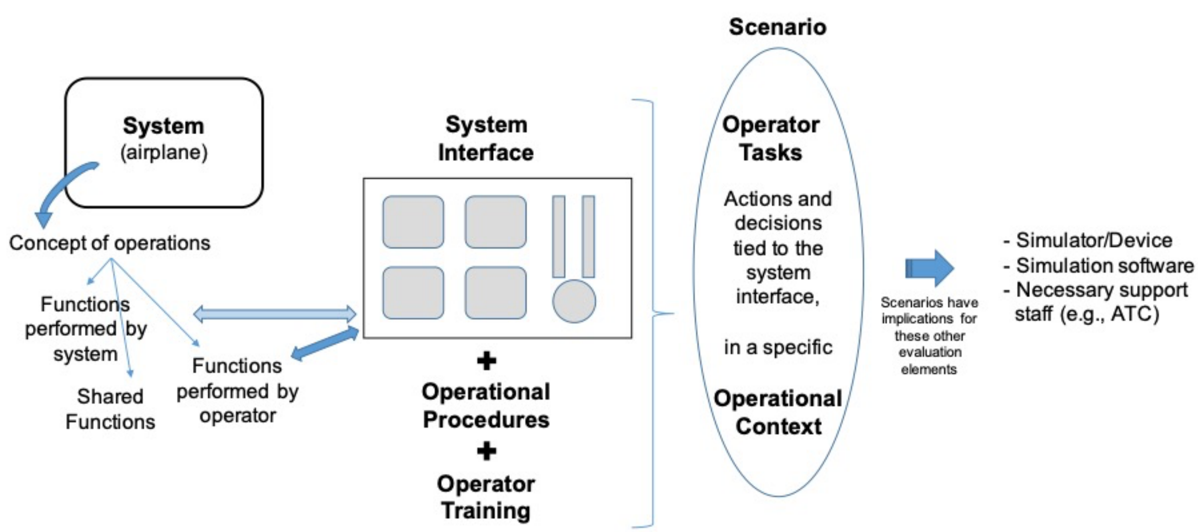
Also, illustrated on Figure 1 is the importance of operational context; context refers to at least the following:

- phase of flight (for airplanes) or operating regime
- environmental conditions and objects outside the airplane, such as weather, day/night, turbulence, obstacles, etc.
- operating conditions, such as time pressure, degraded information, dense traffic, quality of the written operational documents, changes to the airspace infrastructure, etc.
- demands from outside the airplane, such as the air traffic control (ATC) clearance
- airplane system condition/health
- crew status (e.g., health, experience, background)

Operator tasks need to be performed at an appropriate time in the appropriate conditions, and the scenario needs to specify these, as well. Therefore, a scenario, for our purposes, is a set of flight crew tasks performed in a specific operational context.

Also, note that the scenario, when defined this way, has implications for other elements of the interface evaluation. Specifically, as shown in Figure 1, it has implications for the fidelity and completeness of the simulator or device on which the evaluation will be conducted, as well as the fidelity of the software running the simulation. Also, the scenario has implications for the number and type of support staff; e.g., there may be a need for a dispatcher, air traffic control, or other pilots.

Figure 1. Context for Defining Scenarios



3. Sources of Scenarios

The following sections describe, in more detail, the various sources for identifying tasks and contexts for developing operational scenarios.

3.1. System State

System state captures whether the system is operating in a nominal state, in some uncommon regime, or in some degraded or compromised state (e.g., due to a system component failure). Primarily, there is value in evaluating how the interface communicates the full range of system state and supports operators in managing the system in each state since there is risk if the interface does not reveal degradations in functionality appropriately. Ideally, interface evaluation scenarios provide a thorough and systematic exploration of normal conditions as well as system failures (both hardware and software failures) and unexpected conditions. Considerations should include:

- failure of an airplane system (e.g., a hydraulic system), leading to a shutdown or loss (e.g., of some hydraulics)
- degradation of an airplane system component, leading to substandard performance of that system
- a software failure, leading to airplane system performance different than was expected

In failure cases, actions at the level of the total flight deck interface should be evaluated, not just actions at the level of the new interface element. However, the selection of scenarios should consider the potential role that the new interface element may have in managing non-normals, especially for non-normals tied to the system functions represented in the interface element. Also, the evaluation should consider failures in the functions that drive the interface element.

There are two general approaches for identifying the starting point for these types of evaluations: bottom-up and top-down. One approach, characterized by Failure Modes and Effects Analysis (FMEA) (SAE International, 2012), considers failures of system components and then determines how each failure would affect system performance and, especially, the likelihood of an undesired outcome (e.g., loss of an airplane function). This is a bottom-up process, from component failures to undesired system outcomes. For thinking about interface evaluation, it might be helpful to ensure the following system elements are considered to determine how the interface's effectiveness may be changed.

- Failures of system components, specifically when airplane system components are being represented in the interface (e.g., engines for an engine display), or failures of the physical components of the interface itself (e.g., the display).
- Failures of system sensors or other ways that system data are lost or are no longer valid. How does the interface indicate that the data are erroneous (not valid) or their validity is suspect (e.g., due to a disagreement among individual sensors)? How does the interface show that data were lost? Loss of certain types of data (e.g., airspeed) can further result in the overall system dropping into a less-automated or less-capable state.
- Failures in data processing or integration (e.g., the placement and movement of the flight director), which would affect the presentation of data/ information on the interface. How does the interface show that an interface function was lost?
- Failures that lead to a degraded level of system performance so that the systems or controls may operate differently, such as in a degraded mode of automation in a commercial airplane.

A very different version of this bottom-up approach is the STAMP approach (Systems-theoretic accident model and processes)(Leveson, 2012) that starts from describing the system as a set of controlled processes and then identifies the ways in which those controls can fail to manage the system as intended. STAMP goes beyond the traditional approach to failures, which tends to focus on hardware failures. STAMP considers failures in hardware, software, organizations, and human performance to manage a controlled process, and, in this way, identifies potential system failure states.

A second general approach, which is called Fault Tree Analysis (FTA) (Lee et al., 1985), is a "top down" process that identifies undesired system states and attempts to identify all of the ways in which system and operator failures could produce those states. This complementary process can be a good check on the completeness of the FMEA. Typically, the faults are defined at a system level, but it is possible to think of interface-oriented faults, such as:

- A loss of valid interface data. What are the ways in which system data can be lost or made invalid?
- A failed or lost display device. What are the ways in which the physical components of the interface can be compromised so as to limit interface access?

- A loss of an interface function. There are often a number of functions in the interface (e.g., computations, alerting) and failures can lead to a loss or degradation of those functions.
- A loss of interaction capability. One specific function is the ability to use the interface to interact (e.g., with data, with system controls), and a loss of this interaction can remove methods for changing the airplane state.

It is also worth considering operation in extreme conditions. Typically, scenarios are developed to explore the behavior of the interface within the normal operational envelope; for airplanes, items such as expected maximum altitude and expected maximum crosswinds are considered as limits. Ideally, the flight deck interface functions effectively throughout all possible conditions that the system can encounter. (We understand that this space of all possible conditions is potentially incredibly large and has to be limited in some way.) If the system is exposed to some condition that was not expected, it should still provide a valid representation of the world and/or the state of the system, or, at worst, indicate that the representation is not valid.

For airplanes, extreme conditions might transition into a part of the envelope that is difficult to anticipate fully. For example, transitions into an aerodynamic stall are poorly simulated. The following types of interface issues can arise:

- Ambiguous indications, such as a non-moving indications at an extreme value, make it difficult to know if a sensor has failed (and this is an indication of failed state) or if there is a valid indication somewhere beyond the limits of the indicator's scale.
- An alarm is triggered on some specific condition and that alarm will not trigger (again) under other conditions that also should trigger it.
- Loss of sensor data that requires an alternate means to perform tasks or make decisions
- Control actions may not have the same effect in the extreme conditions
- Limitations in how well the interface allows one to see the larger picture of what is happening. Displays are likely to be more focused on smaller system elements and it may be difficult to represent the larger set of problems in an integrated way.

The following types of questions should be used to guide the interface evaluation:

- What is the full range on any indication scale that is possible (although some values may be unlikely) and how will the interface adapt to show values outside the normal range?
- Is there software that will fail when a parameter goes outside of an expected range, and how is that failure represented on the interface?
- Generally, how are sensor failures revealed by the interface, and how are erroneous (or suspicious) parameter values represented by the interface?

3.2. Operational Tasks

The primary focus of a scenario is on an operational task or set of tasks, which should be broadly defined to include activities such as taking actions on the system, monitoring the interface to understand system state, diagnosing a failure, or coordinating actions with a crew member. In this section, we describe multiple ways to identify the operator tasks that need to be supported by the interface.

A good starting place in thinking about the set of operational tasks that are meant to be supported by an interface element is a statement of its "intended function" (see the AC 25.1302-1 and 14 CFR 25.1301 (FAA, 2013)). The objective of the intended function statement is for the applicant to specify what operational tasks the function is intended to support so that the regulator can apply the appropriate evaluation. This specification of an intended function can also drive the interface evaluation. According to AC 25.1302-1, the intended function should pull together the following types of information:

- The intent of each interface feature or function.
- The operator tasks that are associated with each interface feature or function; more specifically, any assessments, decisions, or actions that operators are expected to make based on information provided by the interface.
- Additional system information that should be used in combination with the information provided by the interface.
- The phases of flight (operating states or regimes) in which the interface is to be used.
- Other operating conditions (e.g., weather) that help determine how or when the interface should be used.

An example of an intended function specification for a new interface element that presents a selector for an airplane's autobrakes is the following:

- The tasks that should be supported are:
  - allow the pilot to see all of the levels of autobraking that can be selected
  - allow the pilot to see the current selected (commanded) level of autobraking
  - allow the pilot to select a level of autobraking
  - allow the pilot to see the current actual level of autobraking
  - allow the pilot to see when autobraking cannot be selected (e.g., due to a failure)
  - allow the pilot to see specific autobraking levels that cannot be selected (e.g., due to operating conditions)
  - aid the pilot in detecting when the autobraking level has been changed by the automated controller
- The phases of flight or operating conditions that are relevant are:
  - anytime the airplane is on the ground
  - when the airplane is on descent or approach and an autobraking level needs to be set
- No other system information is required to support these tasks.

According to FAA AC 25.1302-1 (section 5-3(c)), a well-specified intended function should provide a set of tasks that form the foundation for the evaluation scenarios. The intended function should also provide information on relevant operational context, such as task performance triggers, and on other elements of the interface that may be needed for the evaluation. Note that the task statements in the intended function also provide a specification of the relevant aspects of pilot performance: see, select, detect (for the example above). Although these will require further specification, they start to hint at the ways in which the interface needs to support operators, and therefore, the types of performance measures that will be most appropriate.

Identifying operator tasks will, in some cases, be as straightforward as developing a good intended function description. For some systems, more comprehensive task documents will be available. Ideally, complex, safety-critical systems are developed by a thorough task analysis (sometimes in the form of a Hierarchical Task Analysis). The task analysis can be derived from an existing system (if there is one) or is developed from a formal analysis technique (e.g., Annett, Duncan, Stammers, & Gray, 1971; Annett & Stanton, 2000; Miller, 1962). The intent of task analysis is to identify the full set of operator tasks as a driver to the design of the system and its interface (and later, as the primary input to operator training). It is meant to be a fairly complete specification of how the interface elements are used in operating the system. Ideally, the task analysis creates a link to interface elements so that one can determine which interface elements are used for which tasks. In this way, it is possible to identify the full set of tasks that touch the interface element being evaluated.

Because task analysis is typically focused on interface actions—that is, the touches on interface controls define the set of system tasks—tasks that do not require interface actions (e.g., monitoring tasks, decision-making tasks, awareness or assessment tasks) may not be represented in the task analysis. Cognitive task analysis (CTA) was developed to capture the non-action elements of performance (see Crandall, Klein, & Hoffman, 2006 for a review) and can identify tasks such as:

- knowing when to initiate actions
- developing an understanding of the state of the system (or the interface)
- identifying unexpected indications and diagnosing problems

These are tasks that the interface needs to support, as well.

Another form of task analysis is from the “jobs to be done” literature (Ulwick, 2016). This technique, developed out of a user experience/usability framework, is intended to specify what tasks a technology user wants to perform.

A useful distinction when thinking about the full set of tasks for an interface element is the distinction made by Rasmussen (1983) between skill-based, rule-based, and knowledge-based operator performance.

- Skill-based performance refers to actions that are highly practiced or “automatic,” such as catching a ball, or steering a car through a series of turns. For a skilled driver, steering through a right-hand turn can be done without consciously thinking about the actions being taken.
- Rule-based performance is typically a set of actions determined during system design (e.g., a checklist has been developed) as a response to a specific event or situation. For example, specific non-normal procedures have anticipated specific system failures and scripted out a set of operator actions to be taken in response. Rule-based performance also describes well-trained or rehearsed sets of actions applied to complete a task. For these cases, the operator is working through an easily articulated, well-established set of steps for completing a task.
- Knowledge-based performance is generated by the operator at the time it is needed. The operator has to determine that no procedure exists and that he/she needs to understand what the operational goals are and how they might be achieved.

Tasks, as documented for system operation, can be skill-based or rule-based sets of actions. However, it is important to also consider how the system interface supports knowledge-based performance in situations for which there may be no procedure, or in situations for which multiple procedures seem to apply. It may be harder to identify the tasks that require knowledge-based performance because they are often situations in which the expected approach breaks down or there is an unanticipated system failure.

The generation of tasks through these different techniques can produce a large task set, perhaps larger than can be accommodated in the evaluation scenarios. When task sets are too large, it may be possible to reduce the set by setting selection criteria. A common feature of a task analysis (Annett & Stanton, 2000) is a characterization of task difficulty, frequency, and importance, where:

- difficulty reflects how easy it is to perform the task
- frequency reflects how frequently the task is performed
- importance reflects how important a task is for successfully completing a job or for operating the system

Urgency can also be added to this set of ratings, where urgency reflects how quickly the task must be performed.

Another method for selecting tasks is to define a representative sample of system functions. For example, a flight planning tool might have several functions, such as planning for alternate destinations or developing multiple flight plans for successive flights to be flown in the same day. A representative sample of tasks should ensure that it touches on each of these functions.

3.3. Operational Context

Operational context captures a range of operational factors that can occur around the task performance at the heart of the scenario. Task performance can potentially be affected by many aspects of the operational environment and identifying and including relevant factors can create a more realistic scenario. The operational context should also cover more extreme cases (with acceptable risk; e.g., very high workload) that allow the evaluation of operator/pilot performance near the edges of operations or human performance.

Some of the typical issues that are used to shape an evaluation scenario are the following:

- *Time pressure*. Evaluation of task performance should account for appropriate time-pressure. Time pressure can occur as a result of normal operational pressures; for example, hurrying to make a take-off slot. In other cases, time pressure is driven by the time available before a bad outcome results from a system failure or emergency (e.g., Kourdali & Sherry, 2017).
- *Time of day/fatigue*. Airline operations occur around the clock and across time zones, especially for international operations. Evaluation scenarios that include pilots at various levels of fatigue and at various points in the sleep-wake cycle can better represent operational conditions.
- *Workload*. Similarly, additional workload can remove resources from task performance, and increasing levels of workload can lead to performance decrements. Note that workload has different components; for example, the NASA task load index (TLX) scale (Hart & Staveland, 1988) assesses perceived mental demand, physical demand, temporal demand (or time pressure), effort, and frustration. These different components can be manipulated independently.

- *Distraction*. Operating a complex system, such as a commercial jet transport, requires managing many activities at the same time, such as communications from others in the system, unexpected changes to the plan, and system failures (e.g., see Loukopoulos et al., 2009). A major part of the difficulty in operations is managing the various demands and changes that occur as a natural part of operating in a complex environment. Introducing a distraction—either totally unrelated to operations or tightly linked to operations—allows the evaluation to focus on how well the interface supports task management. Specifically, it introduces issues such as recovering from an interruption, determining priorities when there are multiple tasks to perform, and managing crew resources to ensure that someone continues to focus on system operation.
- *Environmental factors*, such as vibration, turbulence, temperature, lighting. Performance with a display or controls or other interface elements should be considered in the context of the types of environmental conditions that can typically occur. In a flight deck, vibration and turbulence are not uncommon. In maintenance environments, light level, temperature, and humidity may affect performance negatively.
- *Weather/visibility*. For aviation, these factors affect the ability to get information from outside the airplane and can especially affect tasks that are performed near the ground. In addition to visibility, weather can increase the number of tasks that need to be managed by the flight crew. For example, weather or visibility conditions at the planned destination can increase decision making regarding overall mission planning.

The set of operational factors can also expose areas of complexity or difficulty in the operating environment. A study by Roth et al. (1994) looked at the types of scenarios being developed for nuclear power plant (NPP) operators and for air traffic controllers during their simulator-based training. One conclusion drawn was that the formal training curriculum tended to emphasize a progression toward "many and fast"—that is, scenarios progressed in difficulty by adding more events or by making the events occur at a faster pace. These are largely workload-related factors, and while these are useful methods for developing a learning progression, they are not the only forms of complexity/difficulty occurring in a realistic operational environment. Roth et al. supplemented this set by including issues they identified in a range of system safety events.

While the emphasis of the work by Roth et al. was on introducing complexity/difficulty into the training progression, this work can also inform the types of scenarios being developed for interface evaluation. Ideally, the interface design makes it easier to understand and manage complexity and difficulty. Certainly, not every form of complexity/difficulty is relevant to the evaluation of a specific interface element. However, there is value in considering a fairly broad list of issues to determine which of them may be relevant to a feature or function that the airplane interface element provides. The following is a list of forms of complexity/difficulty to be considered in developing evaluation scenarios:

- *Expectation/Anticipation*. When an evaluation study participant is expecting or anticipating some event, his/her performance will be better than when the event is unexpected (Casner et al., 2013). It can be difficult to create an unexpected event in an evaluation setting since, generally, in that setting, the participant expects certain types of events that may not occur frequently in actual operations (e.g., a system failure), and it is, therefore, difficult to simulate the onset of truly unexpected events.

One approach for inserting unexpected events is to create strong expectations for other events; that is, tell the study participant that the scenario is about operational context and task A when it is actually about B. Another approach is to create surprises relevant to the interface element being evaluated. For example, if a new navigation and weather display is being evaluated, useful scenarios might specify environment changes in possible, but unlikely ways. However, generally, it is probably difficult to create the same level of "startle" or surprise in an evaluation setting. Unfortunately, this can mean that only a single trial is possible for each study participant: after one "black swan"[1] event has been presented, the participant will be alerted and in a very different state (Wickens et al., 2009).

Landman et al. (2017) discuss the idea of a surprise as an event that does not fit with the pilot's current understanding of the situation. They use the term "frame" to capture the pilot's understanding of the situation. When there is a mismatch—e.g., the airplane is banked right when the pilot thought it was banked left—the pilot has to work through a "reframing" process to develop a new understanding. This surprise increases workload and leads to a degradation in performance. Inconsistent indications, gradual shifts, or indirect evidence of a system malfunction are also circumstances likely to require reframing. The need to identify and shift to a different frame may be particularly difficult for "black swan" situations. In these situations, operators are working both to understand and to mitigate problems, and technology may increase rather than relieve operator burdens (Woods & Patterson, 2000).

A misplaced expectation can also be tied to specific system indications or parameters. There have been safety events (e.g., the Spanair 5022 accident; CIAIAC, 2011) in which the pilot thought he/she knew the value of an indication and perhaps even looked at the indication but did not perceive or process it. A very strong expectation for a certain value can prevent thorough processing of the actual indication (e.g., Simons & Chabris, 1999). Especially when the indication changes in a subtle way (no salient alerting for the change), changes to unexpected values can go unnoticed.

- "*Garden path" situations*, which is a scenario in which the early indications strongly suggest a familiar system state or condition, but the scenario ends up in a different state or condition. This can entice the participant into seeing the indications through that frame. Thus, the operator is led "down the garden path" to an incorrect interpretation of system state.
- *Discounting*. Scenarios that present situations in which some system indications can be (inappropriately) discounted or "rationalized away" because they are considered to be the result of a known fault or of an action from an automated system response. For example, the airplane is rolling away from wings level due to an autopilot failure, but it is interpreted as the airplane turning to follow the flight management system (FMS) flight path.
- *Uncertainty*. Scenarios that lead to loss of data or uncertainty regarding the usual set of system indications (e.g., loss of a display, loss of power that drives other systems).
- *Masking*. Scenarios in which a problem in one system is masked by another system. For example, the autopilot can compensate for a failure of a control surface or for a misalignment of the thrust levers.

[1] Black swan refers to a unique, rarely seen event.

- *Effects at a distance*. Scenarios in which one system fault propagates through interconnected systems to produce effects at a distance, or loss of a resource (such as power) common to many systems; that is, indications appear in multiple systems that are normally not associated (e.g., failure of an electrical bus).
- *Violate expectations*. Scenarios that deviate from standard operator assumptions or expectations. For example, faulty indications, such as unreliable airspeed, or systems that fail to work as demanded.
- *Sleight of hand*. Scenarios that create a need for attending to one part of the interface while equally or more important events are occurring on another part of the interface. For example, pulling attention away from the primary flight display (PFD) while airspeed is dropping below the target airspeed.
- *Special knowledge*. Scenarios that require operators to access knowledge and/or resources beyond those that are a part of the standard interface presentation. For example, having to refer to non-routine paper or electronic documents, such as system schematics.
- *Judgment*. Scenarios that require operators to step outside of the formal guidance/procedure and use judgment or take a discretionary action. Or, scenarios that do not fit perfectly with the formal guidance/procedure and require it to be adapted. (See also Knowledge-based performance, described above.)
- *Teamwork*. Scenarios in which a complete understanding of system state requires integrating knowledge distributed across multiple crew members, perhaps in multiple locations. For example, the event requires that the flight crew communicates with the cabin crew.
- *Unfamiliar*. Scenarios that require actions not often practiced (typically, actions for non-normal or emergency tasks).
- *Goal conflict*. Scenarios that force system operators to choose between conflicting goals. For example, shutting off a system as part of completing a non-normal checklist but that system has the potential to play a role in continued safe flight and landing.

3.4. Safety Events

A special case for identifying tasks and operational contexts for evaluation scenarios focuses on looking at safety events that have occurred in actual operations. Indeed, a new/modified interface element is sometimes developed to address a specific operational risk that was revealed from one or more safety events. Especially when the analysis of several aviation accidents points to a potential interface issue, such as missing information or confusing information, there is a desire to make changes to the interface as a means to reduce the risk of future mishaps.

An example of interface changes that were driven by safety events (in commercial aviation) is the flight path angle (FPA) vs vertical speed (VS) mode indication. An Airbus 320 accident in Strasbourg, France by Air Inter suggested that the flight crew was confused by how the flight mode was indicated. There was a small difference between the flight path angle mode and the vertical speed mode on the interface. The flight crew seemed to have selected the wrong mode and descended much more steeply than planned, hitting the ground well before the airport. The presentation of the modes was changed to reduce confusability. (for an overview, see http://lessonslearned.faa.gov/ ll_main.cfm?TabID=2&LLID=57&LLTypeID=2)

When these types of changes are made to the interface, the primary objective is to avoid the undesired outcome (e.g., confusion, inappropriate control actions) when similar circumstances occur in the future. The approach to identifying relevant scenarios in this case is to analyze the accident/incident situations to determine what interface issues may have contributed to the events. Specifically, there is value in understanding the following:

- What was the flight crew trying to do?
- Did the interface design contribute to inappropriate or inadequate pilot understanding or actions?

A companion report ("Factors that influenced airplane state awareness accidents and incidents," Mumaw et al., 2019) provides examples of this form of analysis. In that report, we started with the results of the Commercial Aviation Safety Team (CAST) Airplane State Awareness (ASA) analysis group, who analyzed 18 events tied to low-energy state awareness and attitude awareness. That analysis identified a number of "themes," which were common characteristics across events. Specifically, some of the themes were:

- ineffective alerting (in 18 events)
- distraction (18)
- lack of external visual references (17)
- crew resource management (16)
- automation confusion (14)
- safety culture (12)
- inappropriate control action (12)
- training (9)

From this analysis, we were able to identify both the types of interface enhancements that might improve flight crew performance, and the scenarios that could be used to assess those interface enhancements.

Belcastro et al. (2014), also interested in identifying how loss of control (LOC) accidents develop, reviewed a set of 275 LOC accidents that occurred between the years 1996 and 2010. For each event, they catalogued the types of causes and contributing factors that preceded the accident, and then identified those that were occurring more frequently. The following are the types of events that were occurring more frequently (on more than 100 of the 275 accidents):

- vehicle upset conditions (188 occurrences)—conditions such as an abnormal attitude, low airspeed or low energy, abnormal flight trajectory, uncontrolled descent, and stall
- inappropriate crew action or inaction (160 occurrences)—includes items such as pilot loss of awareness, mode confusion, improper recoveries, improper procedures, inadequate monitoring, and abnormal control inputs
- system or component failures/malfunctions (117 occurrences)—this covers various airplane system failures

Further, Belcastro et al. looked at the sequence of these contributing factors and noticed that the most common initiating events were:

- system or component failures/malfunctions

- inclement weather and atmospheric disturbances
- inappropriate crew action or inaction
- vehicle impairment (such as contaminated airfoil, weight or center of gravity [CG] issues)

The CAST ASA study developed accident descriptions in greater detail. Note that inappropriate control actions were common to both of the CAST and Belcastro et al. analyses. Also note that system/component failures were much less frequent in the CAST work.

4. Summary Guidance for Developing a Specific Scenario

The following outlines the necessary elements for pulling together a set of evaluation scenarios. Appendix A provides a more proceduralized version.

4.1. Exploit Diversity

A single evaluation scenario should draw from several of these sets of considerations. For example, a scenario can be about performing specific recovery tasks when the system is in a degraded state and there are performance pressures on the operators due to the potential for a catastrophic outcome. The framework presented here is intended to make the case that any set of evaluation scenarios should at least consider the very large set of potential scenarios.

4.2. Cover a Range of Tasks

It is likely not possible to evaluate all operational tasks. Decisions will need to be made about which operational tasks to study in which contexts. For any interface or interface element, you want to know that operators can perform the full set of operational tasks that are required. However, in many cases, it will not be possible to include all tasks in a formal evaluation.

In these situations, there are several techniques for selecting a subset of tasks for evaluation.

- *System functionality.* Especially when the interface element being evaluated introduces new system functionality—for example, generating options in a decision-making task—it is important to select a set of tasks that provide a representative sample of the functionality (tasks that touch all the elements of the new functionality). In many cases, the interface element's new functionality actually adds tasks that did not exist in the previous interface.
- *Interface functionality*. When the interface element being evaluated is more about a change or upgrade in interface technology—for example, a new interaction widget—then the tasks selected for evaluation should provide a representative sample of the interface functions. Examples might be navigating between displays, taking actions on display-based controls, or entering data.
- *Critical tasks.* In some cases, the purpose of the new interface element is to better support task performance that is linked to critical system operations or, in the past, was linked to a safety event. In these cases, an analysis of the situations that led to those events can identify tasks for evaluation (see the companion report "Factors that influenced airplane state awareness accidents and incidents," Mumaw, Billman, & Feary [2019] for examples of this approach).
- *Frequency, importance, and urgency*. Another method for down-selecting tasks is to identify the tasks that are performed more frequently or the tasks that are rated

as the most important tasks for operator performance. While this approach ensures a more complete evaluation of tasks that are probably trained and more frequently performed, it may not sample as widely across all the display elements and functions. A different approach to task selection is to think about the tasks that have to be performed under the most demanding conditions, such as tasks that require urgency.

4.3. Identify the Most-important Performance Issues

A companion document ("Evaluation issues for a flight deck interface," Mumaw, Haworth, Billman, & Feary, 2019) describes a range of interface evaluation issues. This report describes 40 broad issues that may have relevance for evaluating an interface element. For a newly designed, complete interface for a complex system, it may make sense to address every one of these issues. However, not all of these evaluation issues will be required for an interface element that represents additional functionality or some refinement or upgrade in interface technology.

Many elements of the interface will have been evaluated when the larger system interface was developed and certified and will not need to be revisited when a new display or control is added. For example, issues concerning viewability and reach to the interface displays are unlikely to change with the addition of a new interface element. Similarly, some basic usability issues, such as the design of interaction widgets or display navigation, are likely to have been evaluated previously, assuming they do not change significantly in the new interface element. However, when there are relevant issues that were not addressed in previous evaluations, such as, perhaps, support for crew coordination, those issues should be addressed for this interface element.

One way to focus on specific evaluation issues is to try to anticipate the ways in which interface design trade-offs and operator performance might lead to system performance issues or system safety issues. For example, the development of more task-oriented displays can lead to a greater need to manage displays to ensure that the appropriate displays can be seen at a critical time. Or, new task-specific alarms can lead to poor integration across the full set of alarms. Try to determine how new uses of technology can create unintended consequences for human performance through high-level reviews of these evaluation issues.

4.4. Add Operational Context

Identify aspects of the operational context that can improve the evaluation setting. The section above described a wide array of modifications that could be used to further develop a scenario. Context modification can add realism to:

- *Settings.* A setting should include realistic ATC, weather, or distractions from the cabin crew.
- *Focus on interface elements under evaluation*. For example, if a role of the interface element is to provide flight guidance, then context modification could create situations in which flight guidance is critical for flight path management.
- *Degraded conditions.* For example, reducing mental resources available for additional tasks by adding secondary tasks. This is done to assess the interface element's ability to support operators in these conditions.
- *Complexity.* For example, some interface elements are used in combination with many simultaneous tasks.

For each scenario that is evaluated, the evaluator needs to determine the extent of the operational context provided. This can vary in two ways that are linked to the evaluation's objectives (research questions being asked):

- *Scope.* This specifies the extent to which the full system interface is required. At one extreme, the evaluation includes the entire interface, the operational procedures, full operator staffing, and support personnel (e.g., ATC, dispatch). At the other extreme, the evaluation focuses on an isolated display that supports a small set of narrowly defined tasks.
- *Integration*. Integration is the degree to which the evaluation covers realistic operational tasks. In some evaluations, it may make sense to evaluate performance on isolated tasks, such as the ability to make the correct judgment from a representation on a display, or the ability to discriminate colors, or the strength required to move a controller. In other cases, the tasks being performed rely on having a higher level of integration with the operational system.

Scope and integration are generally highly correlated. However, there are cases in which they can be separated. For example, an evaluation of the ability to view or reach the elements of the system interface requires a complete interface mock-up to assess viewing angles or reach distances from a seated operator position. But, the tasks being performed do not require integration with the operational system.

If the full set of evaluation issues has been identified, try to identify opportunities to combine issues into a single scenario or reduced set of scenarios. For example, it may be possible to evaluate interface usability issues in the same scenario as the one developed for evaluating workload or task management. For more guidance on enhancing efficiencies in the overall evaluation process, see the companion report, "Best practices for evaluating flight deck interfaces for transport category aircraft with particular relevance to issues of attention, awareness, and understanding." (Billman et al., 2019).

5. Summary and Conclusions

In this report,we described the important role of a scenario for evaluating a system interface by looking at human performance with that interface. A scenario was defined as an operational task or set of tasks performed in an operational setting. We argued that the scenarios used for flight deck evaluations can push beyond the normal tasks at the heart of the interface. Although it is important for an evaluation to look at normal tasks, it is also important to consider how the interface performs when the system state is not normal and for the types of safety events that have occurred. Further, we described a number of factors that can alter the operational context for task performance, and we discussed how those context factors can be used for interface evaluation. The last section offered a set of guidelines for developing operational scenarios for interface evaluation.

References

Annett, J., Duncan, K. D., Stammers, R. B., & Gray, M. J. (1971). Task analysis. Department of Employment Training Information Paper No. 6. London, UK: Her Majesty's Stationary Office (HMSO).

Annett, J., & Stanton, N.A. (2000). Task analysis. London: Taylor & Francis.

Belcastro, C.M., Foster, J., Newman, R.L., Groff, L., Crider, D.A., & Klyde, D.H. (2014). Preliminary analysis of aircraft loss of control accidents: Worst case precursor combinations and temporal sequencing, AIAA Guidance, Navigation, and Control Conference, AIAA SciTech Forum, (AIAA 2014-0612).

Billman, D., Mumaw, R.J., & Feary, M. (2019). Best practices for evaluating flight deck interfaces for transport category aircraft with particular relevance to issues of attention, awareness, and understanding. CAST SE-210 Output 2. NASA/TM-2019-220390.

Casner, S.M., Geven, R.W., & Williams, K.T. (2013). The effectiveness of airline pilot training for abnormal events. *Human Factors*, 55, pp. 477–485.

CIAIAC (2011). Accident involving a McDonnell Douglas DC-9-82 (MD-82) aircraft, registration EC-HFP, operated by Spanair, at Madrid-Barajas Airport, on 20 August 2008. Report A-032/2008.

Commercial Aviation Safety Team (2014). Airplane State Awareness Joint Safety Implementation Team: Final report, Analysis and recommendations. https://www.skybrary.aero/bookshelf/books/3000.pdf

Crandall, B., Klein, G., & Hoffman, R.R. (2006). Working minds: A practitioner's guide to cognitive task analysis. Cambridge, MA: MIT Press.

Federal Aviation Administration (2013). AC 25.1302-1 – Installed systems and equipment for use by the flight crew.

Hart, S., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P.A. Hancock & N. Meshkati (eds.), *Human Mental Workload: Advances in Psychology*, 52. Amsterdam: North Holland.

Kourdali, H., & Sherry, L. (2017). Available Operational Time Window (AOTW): A method for evaluating and monitoring airline procedures*, Journal of Cognitive Engineering and Decision Making, 11*(4), pp 371–381.

Landman, A., Groen, E.L., van Paassen, M.M., Bronkhorst, A. W., & Mulder, M. (2017). Dealing with unexpected events on the flight deck: A conceptual model of startle and surprise. *Human Factors, 59*(8), pp 219–228.

Lee, W.S., Grosh, D.L., Tillman, F.A., Lie, C.H. (1985). Fault tree analysis, methods, and applications: A review. *IEEE Transactions on Reliability, R-34*(3), pp. 194–203.

Leveson, N. (2012). Engineering a safer world: Applying systems thinking to safety. Cambridge, MA: MIT Press.

Loukopoulos, L.D., Dismukes, R.K., & Barshi, I. (2009). The multitasking myth: Handling complexity in real-world operations. Surrey, England: Ashgate.

Miller, R.B. (1962). Task description and analysis. In R.M. Gagne (ed.), *Psychological principles in system development*. New York: Holt, Rinehart, and Winston.

Mumaw, R.J., Billman, D., & Feary, M. (2019). Factors that influenced airplane state awareness accidents and incidents. NASA TM-20205010985.

Mumaw, R.J., Haworth, L., Billman, D., & Feary, M. (2019). Evaluation issues for a flight deck interface. NASA TM-20205010986.

Rasmussen, J. (1983). Skills, rules, knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man and Cybernetics, 13*, pp. 257–266.

Roth, E.M., Mumaw, R.J., & Lewis, P.M. (1994). *An empirical investigation of operator performance in cognitively demanding simulated emergencies.* (NUREG/CR-6208). Washington, DC: Nuclear Regulatory Commission.

SAE International (2012). Recommended failure modes and effects analysis (FMEA) practices for non-automobile applications. ARP5580.

Simons, D.J., & Chabris, C.F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception, 28*, pp. 1059–074.

Ulwick, A.W. (2016). Jobs to be done: Theory to practice. Idea Bite Press

Wickens, C.D., Hooey, B.L., Gore, B.F., Sebok, A., & Koenicke, C.S. (2009). Identifying black swans in NextGen: Predicting human performance in off-nominal conditions. *Human Factors, 51*, pp. 638–651.

Woods, D.D., & Patterson, E.S. (2000). How unexpected events produce an escalation of cognitive and coordinative demands. In P. Hancock & P. Desmond (Eds.), *Stress Workload and Fatigue*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Appendix A. Scenario Selection Procedure

The following is a series of steps for identifying appropriate operational scenarios to be used to evaluate a new interface element.

Step 1. Determine the objectives for the evaluation.

An interface evaluation can occur at different points in development and be driven by a number of different objectives. Many of these objectives are tied to how much the evaluation requires an integrated interface.

Early in development, the focus of evaluation may be on a new display element, indication, or control, and the objective is to determine how usable/intuitive it is for pilots. At this point, the new display element can be evaluated in a stand-alone setting. That is, there may be no need to integrate it into the full flight deck interface.

Later in development, perhaps as part of the certification process, these new display elements will need to be evaluated in the context of the full flight deck interface. That is, the evaluation needs to consider how that display element influences overall workload, salience, priorities, consistency with other flight deck interface components, etc. The set of evaluation measures expands and becomes more complicated as further integration occurs.

Step 2. Select tasks for evaluation.

When selecting a subset of tasks for evaluation (outside of tasks you are required to select), strive to sample interface tasks broadly to include multiple perspectives; specifically, select tasks to include:

- System assessment, both in normal and non-normal conditions. Can pilots quickly understand the state of the system? Can pilots understand what they can do and cannot do with the interface? Can pilots easily see how to initiate a task?
- Interface interactions. Look at interface inputs and data entry, as well as interface navigation. For physical controls, look at reach and strength requirements.
- System functions. Especially when the interface element being evaluated introduces new system functionality—for example, generating options in a decision-making task—it is important to select a set of tasks that provide a representative sample of the functionality (tasks that touch all the elements of the new functionality).
- Salience. Which elements of the interface attract the most attention or are most easily noticed when they change? Which elements are neglected?
- Critical tasks. In some cases, the purpose of the new interface element is to better support task performance that is linked to critical system operations or, in the past, was linked to a safety event. These may include tasks that need to be performed very rapidly.

Other inputs to task selection consider:

- Giving high priority to tasks rated well for frequency and importance. This emphasis pushes the evaluation toward tasks that are probably trained and more frequently performed.

- Looking for a constellation of tasks that are performed together so that there are efficiencies in running the evaluation.
- Potential design trade-offs and novel system issues. Try to anticipate the ways in which interface design trade-offs and operator performance might lead to system performance issues or system safety issues. For example, the development of more task-oriented displays can lead to a greater need to manage displays to ensure that the appropriate displays can be seen at a critical time. Or, new task-specific alarms can lead to poor integration across the full set of alarms. Try to determine how new uses of technology can create unintended consequences for human performance.

You should avoid these selection methods:

- A sole focus on normal tasks (tasks expected for routine operations) performed in the expected order. This approach can do a poor job of taking into account the unexpected but realistic operational contexts.
- A sole focus on tasks rated low for frequency and importance. While it may be useful to sample from this set, these tasks can represent a very small part of operations.
- There may be a tendency to sample tasks that are the easiest to simulate but it is important to sample in other ways and then find a simpler simulation technique to ensure all tasks are represented. That is, it is more important to sample broadly than to ensure a high level of simulation fidelity.

Keep in mind also that tasks may reappear at different phases of evaluation (see Billman et al., 2019). That is, an interface element could be evaluated early in development with a focus on a narrowly defined task, such as making a simple judgment, and then it could appear much later in development in a more integrated operational setting where there is more interest in overall pilot attention and decisions about task management. Tasks, especially the degree to which they are integrated into an operational setting, are to a large extent, determined by the stage of interface development.

Step 3. Identify the tasks that can be performed with the interface element.

A. Identify "touch" tasks.

For the new interface element being evaluated, identify (make a list of) all related tasks that include some interface interaction: touching a display-based control, navigating a menu, turning a knob, flipping a switch, etc. These tasks are relatively easy to identify.

B. Identify cognitive tasks.

For the new interface element being evaluated, identify (make a list of) all related tasks that involve one of the following:

- Awareness of the cu:rrent value. The interface needs to support awareness of (as appropriate):
  - the current actual value/setting
  - the potential range of values/settings
  - a commanded or expected value/setting
  - deviations from expected or "normal"
  - context variables, which aid in determining what the expected value should be

– a predicted value (if relevant to task performance)
– any relevant thresholds (e.g., the upper bound of normal range, operational decision points) that may be near the current value or that represent an abnormal state

- Feedback on operator actions. The interface needs to support awareness of (as appropriate):
  - a clear indication in the interface that an action was taken
  - a clear indication of changes to system state
- Awareness of the system state. The interface needs to support awareness of (as appropriate):
  - an effective representation of system state, especially relative to system objectives or current operating objectives
  - a representation of which agents are currently controlling the system and how they are controlling it
  - information about which system components may be inoperable or unavailable
  - settings or states or interactions that are not currently available
- Decision making triggers. The interface needs to support awareness of:
  - important changes to system state that are decision points for action, including
    - the existence of system or system component failures that need to be addressed by operator action
    - transitions to undesired system states
  - information that aids in understanding which operator actions have the highest priority
  - information that aids in understanding what course of action is the best

C. Identify tasks that are tied to non-normal system states.

For the new interface element being evaluated, identify (make a list of) all related tasks that are needed only when non-normal system states occur; consider:

- The airplane entering a non-normal situation, such as moving outside of the normal flight envelope (e.g., an upset) or loss of sensor information
- System component or system function failures and the potential need to perform non-normal procedures or checklists
- The need to assess system state when system components fail (i.e., the ease of understanding the changes to system state); specifically, the ability to determine that a system is in a degraded mode that may not have the same functionality
- Loss of displays and the potential need to:
  - restore displays
  - perform tasks with fewer displays
  - use a "compacted" form of the same information
- Loss of data and the potential need to:
  - understand there has been a change to data validity
  - determine whether any data are valid

– perform a task with other data

- Loss of indications and the potential need to rely on back-up indications or proxy indications for task performance
- Loss of interaction capability and the potential need to interact with the interface using an alternate input

The result of these identification activities should be a fairly complete set of tasks that may need to be performed with the new interface element.

Step 4. Determine available requirements and resources for evaluation.

With unlimited time and budget, it would be possible to evaluate pilot performance on all of the tasks identified in Step 1. However, it is most likely that you will need to select a sample of tasks for evaluation.

First, determine if there are any requirements from the regulator (or from others) regarding the need to evaluate certain tasks. The certification process may require certain tasks to be included, or perhaps, an overall system safety analysis requires the inclusion of certain tasks.

Second, determine what resources will be available for evaluating the new interface element; consider simulators/mock-ups, measurement technology, evaluation personnel, access to appropriate pilots, and time (or schedule gates that will drive timing).

Ideally, from these considerations, you can determine the overall scope of the evaluation.

Step 5. Identify appropriate Scope and Integration for each task.

For each task that is evaluated, you need to determine the appropriate level of operational context. As described above, this can vary in two ways:

- Scope. This specifies the extent to which the full system interface is required. At one extreme, the evaluation includes the entire interface, the operational procedures, full operator staffing, and support personnel (e.g., ATC, dispatch). At the other extreme, the evaluation focuses on an isolated display that supports a small set of narrowly defined tasks.
- Integration. Integration is the degree to which the evaluation covers realistic operational tasks. In some evaluations, it may make sense to evaluate performance on isolated tasks, such as the ability to make the correct judgment from a representation on a display, or the ability to discriminate colors, or the strength required to move a controller. In other cases, the tasks being performed rely on having a higher level of integration with the operational system.

Scope and integration are generally highly correlated. However, there are cases in which they can be separated. For example, an evaluation of the ability to view or reach the elements of the system interface requires a complete interface mock-up to assess viewing angles or reach distances from a seated pilot position. But, the tasks being performed do not require integration with the operational system.

Step 6. Identify appropriate operational context factors for each task.

Each task should be considered with regard to specific operational context factors that may be relevant. The section on operational context above described a wide array of modifications that could be used to further enhance a scenario for evaluating task performance. Generally, these modifications address one of two evaluation characteristics: increased operational realism or increased pressure on pilot performance.

Operational context factors that increase realism include:

- The evaluation setting, which could include realistic ATC, weather or other environmental factors, or distractions from the cabin crew.
- Realistic forms of complexity that can interact with the task; a range of these was described in the operational context section
- Techniques to reduce artificial expectations tied to the evaluation setting

Operational context factors that place greater pressure on pilot performance:

- Artificial or operationally realistic time pressure
- Artificial or operationally realistic workload
- Artificial or operationally realistic distraction from primary flight tasks
- Degraded conditions: Environmental factors (such as vibration) that realistically degrade the operational environment