
COGNITIVE SYSTEMS ENGINEERING (CSE) FRAMEWORK FOR EVALUATING NEW COCKPIT INTERFACES

Kamilla Run Johannsdottir
Jo-Anne LeFevre, Ph D.
Chris M. Herdman, Ph D.

Centre for Applied Cognitive Research,
Carleton University, Ottawa, Canada

With an increased role of modeling and simulation new cockpit technologies can be thoroughly evaluated before being implemented in an aircraft. This process of evaluation, however, needs to be guided by a framework based on our knowledge of the strengths and limitations of the human operator. Situation awareness (SA) has become the dominant construct used in evaluating new technology. However, it is unlikely that a single construct accurately captures the complexity of the cognitive processes under study. Rather than using a single construct we propose a cognitive systems engineering (CSE) framework for guiding the modeling and simulation process in evaluating new cockpit technologies. The CSE framework uses converging measures of three central constructs (i.e., SA, workload, and task-relevant performance) to operationalize the relevant cognitive processes underlying the pilot-machine interaction. It is argued that converging measures of these central constructs are essential for providing a comprehensive perspective on the impact of new interfaces on the pilot and the crew.

Introduction

In the past, the implications of pilot-machine interactions were rarely explored thoroughly when new technology was added to existing systems. With the development of affordable simulation environments, however, the feasibility of testing new technologies before they are installed in aircraft has increased substantially. Savings in terms of human costs and technology retrofits are potentially enormous. Furthermore, there have been recent advances in both our knowledge of human psychology and the capabilities of simulation environments, thus supporting a greater role for modeling and simulation. However, it is important that the evaluation process of modeling and simulation be guided by a proper operationalization of the cognitive processes under study.

It has proven difficult to properly define the relevant cognitive processes for empirical evaluation of the pilot-machine interface. One approach to this problem is for researchers to focus on a single construct that captures a substantial and/or relevant portion of the pilots’ performance. The construct of situation awareness (SA) has been used in this way, both in the aviation field and more broadly when researchers have explored the relations between technology and human performance. SA refers to the pilots’ conscious comprehension of the environment and their ability to project future scenarios (Endsley, 1995a, 2000). Hence, SA is most commonly measured with various questionnaires whereby pilots either subjectively evaluate their SA or the reported knowledge of the pilot is compared to the actual state of the system and the environment (Endsley, 1995b; Pew, 2000). Less commonly, SA is also evaluated by measuring task-relevant performance, which is assumed to give an indirect indication of pilots’ comprehension of the situation (Vidulich, 2000). Subjective and objective measures are rarely used together to evaluate the impact of new technology on the pilot or the crew.

Measuring SA has provided some important insight into the impact that a new technology may have on pilots’ comprehension of a system and their ability to selectively process relevant information. However, the pilot-machine interaction depends on a complex, dynamic model of the system and the situation that includes much more than selective attention and comprehension. For example, a pilot’s ability to interact with a system is also highly dependent on perceptuomotor coordination and action planning. It is therefore unlikely that a single construct can sufficiently capture the behavioural variance produced by the complex pilot-machine interaction. Furthermore, a single construct is unable to effectively represent cognitive processes that underlie pilot-machine interactions. It is important that the constructs that are chosen to evaluate the pilot-machine interaction be
comprehensively operationalized. Using a single operation seriously limits the nature and the meaning of the phenomena under study (Gardner, Hake and Eriksen, 1956). Due to the complexity of the pilot-machine interaction and the underlying cognitive processes, it is unlikely that a single measure will provide sufficient information to allow for a general conclusion as to how a new technology affects the pilot or the crew. Therefore, in order to properly operationalize the chosen constructs, multiple operations or measures should be used.

In the present paper we propose a cognitive systems engineering (CSE) framework for guiding the modeling and simulation process in evaluating new cockpit technologies. The CSE framework uses converging measures of three central constructs (i.e., SA, workload, and task-relevant performance) to operationalize the relevant cognitive processes under study. We argue that converging measures of these central constructs are essential for providing a comprehensive perspective on how new interfaces influence pilots’ responses in the cockpit.

**The CSE framework**

As shown in Figure 1, the CSE framework includes a theoretical construct, the dynamic mental model, and three empirical (i.e., measurable) constructs: situation awareness, workload, and task-relevant performance. Our review of the literature indicated that these three empirical constructs capture a significant amount of the variance in the pilot-machine interface.

![Figure 1](image-url)

The proximal goal of the CSE framework is to provide a context within which to interpret the dependent variables that are assessed in a modeling and simulation evaluation. The CSE framework is not intended to represent a complete model either of the human operator or of the situation, although further developments of the framework will be designed to expand the theoretical and predictive power of the model.

The central theoretical construct in the CSE framework is the dynamic mental model. It captures the notion that the human operator constantly creates and maintains an internal representation of the ongoing situation. When experimental methods are used to measure performance in an evaluation of new technology, all of the measurements indirectly index the pilot’s dynamic mental model. The three empirical constructs provide a comprehensive (although not exhaustive) assessment of the dynamic mental model. Workload indexes the cognitive effort required by the pilot, SA captures the pilot’s perception of events and his integration of those events into a coherent understanding of the situation, and task-relevant performance captures the behaviors associated with the interactions between the pilot (or crew) and the machine.

SA is an empirical construct in the CSE framework that can be defined simply as “knowing what is going on around you” (Endsley, 2000, p. 5). The term was originally used to capture why some fighter pilots were more successful (and therefore lived longer) than others (Spick, 1988). SA is closely tied to knowing how to distinguish important information in the environment from less important information (selective attention), as well as the ability to quickly comprehend and predict the importance of changes to elements in the environment (Adams, Tenney and Pew, 1995; Durso and Gronlund, 1999; Endsley, 1995a; 2000; Sarter and Wood, 1995).

Measuring SA provides important insight into how well the pilot comprehends a system’s functionality and how well the pilot is able to integrate the information presented on the instrumentation into a coherent picture of the system and the environment.

The second empirical construct that forms the core of the CSE framework is workload. Workload is an important construct in aviation and other complex cognitive tasks because humans are limited in their ability to process information and to respond appropriately (Hancock and Desmond, 2001). A helicopter pilot on a search-and-rescue mission who is flying with night vision goggles near the ground during a rain storm is likely to be in a situation of heavy workload. A large amount of rapidly changing
information has to be monitored and the pilot must constantly update his or her dynamic mental model of the environment. In contrast, a pilot flying a routine transit leg in good weather is probably in a low workload situation.

Workload has proven to be a very useful construct for understanding changes in pilots’ behaviour under different situational demands and constraints (Flach and Kuperman, 2001; Wickens, 2001). Technology “improvements” should hypothetically decrease workload, but in practice, a technology that adds information to the pilot’s environment and/or requires the pilot to perform additional tasks may actually increase workload, at least in the short term (Vidulich, 2000). Thus, measuring changes in workload as a function of technology changes is crucial to understanding how technology influences pilots’ performance.

The third empirical construct in the CSE framework, task-relevant performance, refers to the actions of the pilot or the crew (in relation to mission demands) that are potentially affected by the new technology. Adding a new technology to the cockpit can affect pilots’ performance in various ways. For example, the addition of a new display screen to the F-18 cockpit that has enhanced information about approaching threats should improve pilots’ ability to manoeuvre in a threatening environment. However, other aspects of pilots’ behaviour might be impaired or affected in ways that decreases their SA and/or interferes with how they interact with other systems. For example, adding the new display screen to the F-18 cockpit should result in the pilot spending time looking at that screen and interacting with it in certain ways. Concomitantly, the pilot may spend less time using other sources of information, or may use that information differently. Thus, defining and measuring task-relevant performance is an important aspect of understanding the impact of a new technology on performance in the cockpit.

A detailed discussion of the relations among SA, workload, and task-relevant behaviour is beyond the scope of the present paper. In brief, however, it is clear that both SA and workload represent outcomes that may not be directly realized in overt behaviour. Instead, we see these as a product of the pilot’s creation and use of the dynamic mental model. Therefore, under some conditions, we would expect to find a high correlation between SA and workload. If, for example, a decrease in workload allows pilots to spend more time scanning the environment and detecting dangerous situations more quickly, then SA will increase as workload decreases. In our view, a complete disconnection between workload and SA would be evidence against the proposed framework since both constructs are assumed to be based on the pilots creating and updating their mental model.

Good SA does not always lead to good performance and high workload does not always predict poor performance. For example, highly trained pilots may function very well under high workload situations because of their extensive training and experience such that they continue to function effectively despite increased task demands. Nevertheless, we would predict that some other aspect of their performance (such as SA) might decrease under heavy cognitive demands. The CSE framework is based on the assumption that the three empirical constructs will typically be related such that, in many situations good SA (or low workload) will predict good performance. Hence, if task-relevant performance is operationalized appropriately as a specific and direct measurement of the behaviour that is likely to be affected by the technology change, then changes in task-relevant behaviour should be correlated with SA. On this view, if task-relevant behaviour becomes worse with new technology, then SA must necessarily decrease.

In summary, according to the CSE framework, the three experimental constructs (SA, workload, task-relevant performance) are second-order reflections of the pilot’s dynamic mental model of the situation. Because it is impossible in practice to directly measure the contents of the dynamic mental model, defining performance relative to multiple constructs that access the mental model is likely to provide more useful information than focusing on a single construct.

In accord with the CSE framework, we propose that the evaluation of a new cockpit technology should include at least one, and preferably multiple measures of situation awareness, workload, and task-relevant performance respectively. Use of multiple measures will allow for a richer and more accurate answer to the question “how does the new technology affect the human-machine interaction”? 

The value of multiple converging measures within the CSE framework

The pilot’s internal model is a complex theoretical construct that is not directly measurable (as it is not possible to directly measure “memory” or “thinking”), and therefore the only measures we can use are behavioural in the sense that the pilot must perform some action that is then assessed. It is worth
technology and as such can be extremely informative. Degree of comfort and acceptance of the new subjective measures allow us to evaluate pilots’ performance decrement. This is particularly true for discomfort may increase significantly and later cause performance. However, their level of acceptance and system does not significantly affect pilots’ subjective measures, for example, might be better measuring the pilots’ head position as an index of maneuver and what information they noticed. A report on where they were looking during a provides an evaluation. Pilots might be asked to assess the three central constructs, we propose that researchers collect multiple measures for each construct.

An important distinction that is often overlooked in aviation research is between subjective and objective measures of constructs such as SA or workload. A subjective assessment, for example, of situation awareness requires that the pilot makes a judgment or provides an evaluation. Pilots might be asked to report on where they were looking during a maneuver and what information they noticed. A researcher could objectively measure SA by measuring the pilots’ head position as an index of where the pilot is attending. It is critical to distinguish between objective and subjective measures because an individual’s perception of their behavior or memory for a situation can be wrong. Subjective measures, for example, might be better characterized as perceived workload or perceived SA.

Subjective measures allow us to evaluate pilots’ degree of comfort and acceptance of the new technology and as such can be extremely informative. For example, it is possible that adding a new cockpit system does not significantly affect pilots’ performance. However, their level of acceptance and discomfort may increase significantly and later cause performance decrement. This is particularly true for high stress situations where workload is suddenly increased (Andre, 2001). Similarly, pilots may subjectively prefer a new system to existing system whereas their performance is being impaired by the new technology. Therefore, using both subjective and objective measures will provide a more comprehensive evaluation of the impact of new technology on the pilot-machine interaction.

In a recent study, we demonstrated the benefits of using multiple measures to evaluate the implementation of a direct voice input (DVI) system for controlling the on-board computer in the CH146 Griffon helicopter (Herdman et al., 2001; Lessard et al., 2001). In this study, heads-up time was identified as a key to enhancing SA. Increased heads-ups time should improve pilots’ ability to detect and respond to events in the external scene. Heads-up time was measured by tracking the pilots’ head position throughout simulated reconnaissance missions. Heads-up time in the DVI condition was compared to a standard manual input condition that was configured based on how the Griffon crew currently enters commands into the CDU. As predicted, heads-up time increased with DVI relative to the manual condition (by an impressive 42%), indicating that the technology change had at least one of the expected and desirable outcomes on pilots’ behaviour.

However, it was recognized that introducing the DVI system to the Griffon cockpit has a variety of other potential effects. First, DVI had the potential to change the crew interactions in that the flying pilot now had the opportunity to control the on-board computer (i.e., the CDU), whereas in the manual input situation only the non-flying pilot can enter commands on the CDU. So the workload or performance of the flying pilot might also be affected by the new technology. Second, if looking at the CDU to manually enter commands was a workload-intensive activity for the non-flying pilot, then DVI might decrease his or her overall workload. Other aspects of the pilots’ task performance, however, were unlikely to be affected in the types of missions that were flown.

Herdman et al. (2001) included objective measures to assess the workload demands of the DVI versus manual input systems. Objective workload was measured using detection of auditory and visual stimuli (i.e., targets) by both pilots. The targets (auditory tones or visual flashes in the external scene) were presented randomly in the course of the simulated missions. Pilots were instructed to respond as quickly as possible when they detected a target by pressing a key. Target detection as an index of workload has been used extensively and thus has
both empirical and theoretical support. Essentially, the speed and accuracy with which pilots respond to the auditory and/or visual targets is used as a measure of their available attentional capacity.

Herdman et al. (2001) found that the workload of the non-flying pilots was less in the DVI condition than in the manual input condition. It was concluded, therefore that for the non-flying pilots, the DVI system should improve SA and lower workload. Interestingly, it was found that the workload of the flying pilots increased significantly in the DVI condition. This increase in workload occurred even though the flying pilots used the DVI capability infrequently (i.e., less than 1 minute DVI time per each 20 minute mission). Subjective measures of the flying pilot’s workload and SA did not differ across the DVI versus manual input condition, however. These results support our contention that a broad assessment of multiple constructs is necessary to achieve a comprehensive understanding of the impact of a new technology.

The example from the DVI study emphasizes the importance of using converging measures to properly evaluate the impact of new technology on the pilot and the crew. Research using the CSE framework and the principles of broad assessment will test the usefulness of this approach. These techniques are not complicated to apply. For example, to assess the use of a new altimeter, altitude maintenance could be used as an index of the pilot’s adherence to the flight plan (task-relevant behaviour), the simulation can be frozen and pilots could be asked to report altitude information (situation awareness). Their workload could be measured (using target detection) with the new and old instruments.

Systematic assessment of all three constructs with both subjective and objective indices would allow for a comprehensive picture of how the new technology influences the pilot-machine interaction. By including multiple measures of the three behavioural constructs a more complete picture can be inferred about the underlying cognitive processes.

In summary, the CSE framework encourages researchers to develop measures that assess pilots’ behaviour from multiple perspectives. The framework brings different measures and different definitions of the pilot-machine interaction together in a single framework that will allow us to more thoroughly evaluate the implications of new cockpit technology for the pilot. The central assumption is that using converging measures of these three central constructs (i.e., SA, workload, and task-relevant performance), will provide a comprehensive perspective on how new interfaces influence pilots’ responses in the cockpit.

Conclusions

The increasing complexity of the modern cockpit calls for the development of tools and methods that allows us to evaluate the impact of new cockpit technology on the pilot and the crew. However, such evaluation tools must be guided by a proper operationalization of the relevant cognitive processes. The present paper proposes a cognitive systems engineering (CSE) framework that uses converging measures of central constructs (SA, workload, and task relevant performance) to evaluate how the pilot-machine interaction is affected by new technology.

A central tenet of the CSE framework is that the pilot-machine interaction in the cockpit is too complex for a single construct to provide sufficient information to evaluate the impact of an interface on pilots’ overall behaviour. By using converging measures of the three central constructs it is more likely that we are capturing the relevant cognitive processes we want to evaluate. Furthermore, by distinguishing among SA, workload, and task-relevant performance, and the underlying mental representation (the dynamic mental model), researchers can more clearly operationalize the concepts for empirical purposes.

In particular, the CSE framework proposes the use of both subjective and objective measures for each of the three empirical constructs of SA, workload and performance. This is because subjective and objective measures of the same construct can produce different outcomes. For example, individual’s perception of their behaviour or memory for a situation can diverge considerably from their actual performance. It is argued within the CSE framework that using multiple operations of both subjective and objective measures will provide a more comprehensive operationalization of SA, workload and performance and as such will provide a more complete evaluation of how a new technology affects the pilot and the crew.

The CSE framework is expected to provide important support to the modelling and simulation process in evaluating new cockpit technology. Research that is conducted within the CSE framework should allow for comprehensive and valid assessment of human factors aspects of new aviation technology. Systematic application of the framework in the evaluation of new technology for the cockpit will
allow researchers to evaluate the results of assessments produced by different labs under different conditions to be compared more easily.

ACKNOWLEDGEMENTS. This research was funded by the Department of National Defence, Canada. We thank the members of the ACE lab at the CACR for their valuable input. Correspondence should be directed to Dr. Chris M. Herdman, Centre for Applied Cognitive Research, Department of Psychology, Carleton University, Ottawa, Ontario, Canada, K1S 5B6.
(email: cherdman@ccs.carleton.ca)

References


