N95-34773

Certify For Success: A Methodology For Human-Centered Certification of Advanced Aviation Systems

Ronald L. Small & William B. Rouse

Search Technology

Introduction

This position paper uses the methodology in *Design for Success* (Rouse, 1991) as a basis for a human factors certification program. The *Design for Success* (DFS) methodology espouses a multi-step process to designing and developing systems in a human-centered fashion. These steps are as follows:

- Naturalizing – Understand stakeholders and their concerns.
- Marketing – Understand market-oriented alternatives to meeting stakeholder concerns.
- Engineering – Detailed design and development of the system considering tradeoffs between technology, cost, schedule, certification requirements, etc.
- System Evaluation – Determining if the system meets its goal(s).
- Sales and Service – Delivering and maintaining the system.

Because the main topic of this paper is certification, we will focus our attention on step 4, System Evaluation, since it is the natural precursor to certification. Evaluation involves testing the system and its parts for their correct behaviors. Certification focuses not only on ensuring that the system exhibits the correct behaviors, but *only* the correct behaviors. Before we delve into evaluation and certification issues, however, some brief explanations of the other key DFS steps are necessary to put the system evaluation step and the subsequent certification step (outlined herein) in context with the overall methodology.

Naturalizing

The main purpose for naturalizing is to understand the purpose of the system to be certified and to understand the concerns of the various system stakeholders. From a human-centered perspective, the system's purpose should be described in a way that explains why and how the

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland

system supports the human operator in accomplishing his or her goals. For example, if we define the airline pilot's job as safely and efficiently moving passengers from origin to destination, then the purpose of the airliner and all its parts are to *support* the pilot. (By "parts" we mean electric, hydraulic and engine subsystems; flight management and other software modules; and, individual components such as radios, circuit breakers, throttle levers and switches.) Note that we are *not* stating that the pilot's job is to fly the airplane. Nor are we stating that the airplane transports people.

Rather, the emphasis is on the human, the pilot, whose job it is to transport passengers by *using* the airplane. The subtle distinction of such a statement of system purpose is a key to thoroughly understanding and properly executing human-centered design, development and certification of aviation systems. This distinction becomes clearer with practice and is at the heart of naturalizing.

Defining the system's purpose requires understanding the history of the domain and the environment in which the to-be-certified system is to operate. Questions for identifying these issues include:

- Is this a new system or upgrade?
- If new, what was done previously? Why?
- What is the purpose of the system? (Answers should be stated in a human-centered format, as in the above airplane's purpose.)
- What problems are there with the existing system?

(Note: If the system is completely new (no predecessor), the risk is too great and the system is not suitable for certification.)

The reasons for asking these questions are to understand the system's purpose and operational goals, and to begin defining the set of measurements for evaluation and certification. Other measurement issues surface during discussions with stakeholders which must be recorded for use during the evaluation step. Typical stakeholders and their concerns are described next.

Stakeholder Concerns

Before the system stakeholder concerns can be addressed, the various stakeholders must first be identified. Typically system stakeholders are designers, developers, users, maintainers, purchasers, and certifiers of the system. Groups of stakeholders as well as individuals should be identified so that questionnaires can be devised and interviews can be scheduled. It is important to pay special attention to groups or individuals knowledgeable in the current certification processes of similar systems, since the emphasis of this paper is on certification. Questions asked of stakeholders include:

- What is the purpose of the system from your perspective?
- What behaviors are expected during normal operations?

- What behaviors are expected during abnormal (degraded system) situations?
- What are the expected roles of the human operator in both of the above conditions?

The purpose for asking these questions is to understand the various stakeholder concerns so that the certification can proceed along a well-defined path; after all, typical certification budget and schedule resources are limited. This well-defined path is derived from the measurement issues identified during this naturalist phase; therefore, stakeholder concerns must be expressed in quantifiable and measurable terms. These stakeholder-defined metrics then combine with the system metrics (defined earlier) to form the set of measurement issues which are the basis of system evaluation and certification.

The stakeholders should have representatives on that certification team which actually conducts the system evaluation. This team concept ensures that all relevant stakeholder concerns are properly addressed during the evaluation and certification process. System evaluation is the subject of the next section.

System Evaluation

The first step in system evaluation is to define human-centered metrics based upon the system's goals and purpose, and based upon the stakeholders' concerns gathered during naturalizing. Human-centered measurements are those that evaluate system performance and behavior from the human operator's perspective. For example, a software function may be able to execute in five milliseconds; but the system operator may only be able to comprehend that function's outputs at a 1Hz rate. There is no reason to test that software function at an execution speed faster than 1Hz (*from a human-centered certification standpoint;* however, other system engineering reasons may exist for testing that function at the 200Hz. rate).

Quantifiable metrics must be defined not only for the whole system, but for subsystems, modules, and components in order to evaluate their performance and behavior as the system is constructed. While the certification authority is concerned with the system-level performance and behavior of the completed system, it is important that the certification team have confidence in the underlying parts of the system. Therefore, this team should have access to developmental testing metrics, methods and results; additionally, they should independently verify a subset of those earlier tests.

Also, for human-centered certification purposes, the parts of the system should be evaluated as they interact to form operator-observable behaviors. These *threads of interaction* allow an operator representative on the certification team to focus on specific behaviors under specific circumstances – something that is difficult to do when evaluating the entire system because repeatable conditions are harder to generate as the system grows in complexity.

Another consideration for the certification team is to evaluate subjective as well as objective metrics. Subjective metrics include those that measure operator performance, workload, situational awareness, tendencies to commit errors (due to memory overload, operational stresses, mode confusion, a faulty mental model of the system, etc.), and the appropriate task mixes between automation and the operator.

Methods for objective and subjective evaluation are presented in the next sub-section.

Evaluation Methodology

The guiding principle for system evaluation is to test the system and its parts in such a manner as to yield results that can be compared against the metrics determined earlier. Analyses and evaluation methods include:

- Paper and pencil (mathematical) analyses
- Modeling of the system and/or its parts from a human-centered view
- Operator-in-the-loop experiments for even greater fidelity.

Each method is further discussed below, amplified with examples from our experiences

Mathematical Analyses. An envisioned airport safety system is being designed to detect and help prevent runway incursions and have minimal false alarms (a typical engineering trade-off between increasing system sensitivity and minimizing false alarms). Airport tower controllers are also responsible for detecting and preventing runway incursions (among their many other duties), so we performed a signal detection comparison between the automation's specified detection performance and the historical controller detection performance. Since runway incursions happen so infrequently, and since controllers detect and act to prevent most impending runway incursion accidents, we wanted to know if an automated runway incursion prevention system would boost the *overall* detection and prevention of incursions.

Using a statistical distribution analysis, we found that the automated safety system is not likely to improve the overall detection and prevention of runway incursions. This result is mainly due to the fact that controllers are already very good detectors of impending incursions, and so their signal detection performance distribution vastly dominates the specified signal detection performance of the automated system. Obviously, we made some very broad assumptions, but even with this fairly inexpensive evaluation method, we were able to recommend that the automated system's detection rate should be somewhat modified. Another recommendation was to further analyze the result using higher fidelity analysis methods, such as modeling, which is described next.

Modeling. Modeling is useful for testing hypotheses about the real system under conditions that the real system cannot be exposed to – for cost, safety or other reasons. Digital models also allow for testing system behaviors in faster-than-real-time, thus enabling many replications under specified conditions which can yield statistically significant results.

For example, we developed a digital simulation of Atlanta's Hartsfield International Airport to test hypotheses about the effects of various features of the airport automation system described above. While there were many simplifying assumptions needed to develop a model of this complex environment in a reasonable amount of time, we were able to make some recommendations about controller communication workload under varying conditions. We could never have done such an analysis on the real system because it would have interfered with airport operations. Plus, we ran the model for replications of 40 simulated days in just a few minutes which enabled us to quickly obtain statistically significant results.

Another benefit to system modeling is that analytical results help fine-tune higher fidelity analyses such as simulation studies (described next), thus making these more expensive evaluation methods more cost effective.

Simulation Experiments. System simulations are the next step increase in fidelity over digital modeling. Simulation experiments with real system operators participating are useful when human operator interactions are required to evaluate the system (or some part of it) and yet the real system cannot be used because it does not exist yet, or because safety, cost or operational reasons preclude using the actual system.

For example, we were involved in the design, development and evaluation of the Pilot's Associate (PA), an electronic copilot for a next-generation single-seat tactical fighter. A simulation of the fighter's cockpit was needed to conduct utility testing of PA. This testing compared PA and non-PA conditions and used metrics ranging from fuel consumption to kill ratios to situational awareness. A method chosen for evaluating this range of metrics was pilot-in-the-loop simulation experiments because pilot opinion and performance comparisons were of vital importance to many of PA's stakeholders (Cody, 1992). (Incidentally, the PA program also used digital models to focus the piloted simulation experiments on the metrics and conditions where the greatest performance differences were expected.) While operator-in-the-loop simulation experiments have greater costs than the previous evaluation methods, their credibility is also greater. It is usually the case that higher fidelity (more expensive) evaluation methods are also more credible; but, that does not usually detract from the conclusions reached by the less expensive methods.

Methodology Summary

The goals for system evaluation are to analyze the system's performance (and all earlier intermediate results) relative to the set of metrics defined during naturalizing, and then to formulate conclusions and recommendations for system modification. In accomplishing these goals, the evaluation team must define follow-up analyses and tests where performance results do not meet expectations. The team also determines if new metrics are needed. If so, they refine metrics, as appropriate, then conduct additional analyses and tests, and iterate as needed until all metrics are satisfied.

As the system is being designed, developed and produced, the evaluation team should be the system's designers and developers. Test results are then made available to the final evaluation team. It is important to emphasize that each analysis method helps define the higher fidelity evaluations. That is, the results from each method must be analyzed relative to previously-defined metrics, and they must be used to refine any subsequent evaluation methods, or the next iterations of previous methods. A human-centered evaluation and certification process is necessarily iterative.

Now that we have described system evaluation, we shall next highlight the distinctions between it and certification.

Certification Issues

While certification can be described as a more formalized evaluation process, it is distinct from the evaluation process described above in that it must *independently* analyze the system. This independent analysis can be very structured in the sense that different systems or components have to pass differing levels of scrutiny during certification.

For example, the RTCA (Requirements and Technical Concepts for Aviation, an industry group that devises standards for aviation systems) advocates different categories for certifying a system and its parts. The categories are based upon the criticality of failure conditions, namely:

- "Catastrophic – Failure conditions which would prevent continued safe flight and landing"
- Hazardous – Failure conditions which reduce safety margins, cause physical distress and such high air crew workload that tasks may not be completed accurately
- Major – Failure conditions which increase crew workload thereby impairing crew efficiency
- Minor – Failure conditions which slightly increase crew workload
- "No effect — Failure conditions which do not affect the operational capability of the aircraft or increase pilot workload" (Struck, 1992, page 5)

These categories can serve to guide the human-centered certification process, described next.

Certification Process

How should a human-centered certification be conducted? The RTCA seems to emphasize crew workload levels in its definitions, and so should a human-centered certification methodology. Of course, workload levels are not the *only* human-centered measure. A certification team must also address the following concerns:

- What are the error conditions and the likelihood of the human operators committing those errors?
- What are the normal and abnormal operator procedures, and their likelihood of being performed correctly under varying conditions?
- What training is required for the system operators and maintainers?
- What screening for skills and physical or physiological attributes is required?
- What is the tendency for the system's human-machine interface to promote the development of accurate mental models by operators in typical operational environments?

Answering these questions is a non-trivial exercise, but the methodology for answering them is similar to the evaluation methodology described earlier. The gist of the distinctions between evaluation and certification is that *certification ought to analyze failure conditions and their consequences*, whereas *evaluation examines correct or expected system behaviors.*

Other differences between evaluation and certification relate to rules of development that are designed to minimize the system's dynamic response to conditions. Certifiable systems should

not have unpredictable failure conditions. For example, when we built a certifiable knowledge base development tool, we had to pay special attention to some specific software engineering issues, including:

- Pointers – Introduce the potential for directing software execution to places in computer memory that may not be available for normal computations.

- Dynamic memory allocation – Introduces the potential for allocating memory that is already being used for other purposes.

- Compilers – The compiler used for development *must* be the same as that used for creating the actual executable code and for certification. The effect of this rule is that it inhibits the use of software development environments that typically have debuggers or other enhancements that enable more efficient software development, but that also greatly increase the amount of executable code loaded into a mission computer, for example. Consequently, a sparse environment must be used for development, which is bad for software development efficiency, or two compilers must be used – one for development and one for pre-certification compilation – an expensive proposition (Hammer, Skidmore & Rouse, 1993).

Another major difference between evaluation and certification is the composition of the certification team. As mentioned earlier, the evaluation team should initially be the system's designers and developers. The human-centered certification team *must* be independent, although it should examine the metrics, tests and analyses used by the evaluation team to ensure that the metrics are suitable and provide complete coverage for the entire system and its parts.

Certification Team Composition

One last set of questions in this paper concerns the composition of the human factors certification team:

- Do the members of this team need to be certified in the human-centered certification of systems?

- If so, what should be done to determine the certification team member's qualifications?

In order to answer all the previous questions during the certification process, the certification team must be competent in a wide range of human-centered issues. In fact, we think that the certification team members should be certified by the certification authority in accordance with some professional standards and formal training (the training curriculum also requires certification then). Determining a person's or group's competency in human-centered system design was one project's task that we recently accomplished. We devised a set of questions whose answers could be weighed and scored according to the needs of the system's stakeholders (we also recommended scoring guidelines). While the questions are too numerous to present here, they are based on the decomposition of human-centered system design competencies into four major topics and twenty specific issues (Figure 1). A human-centered certification team should have individuals competent in, and certified for, evaluating a system in

terms of the specific issues enumerated in Figure 1. A team approach seems necessary because there are too many issues for one individual to be responsible for during the certification process.

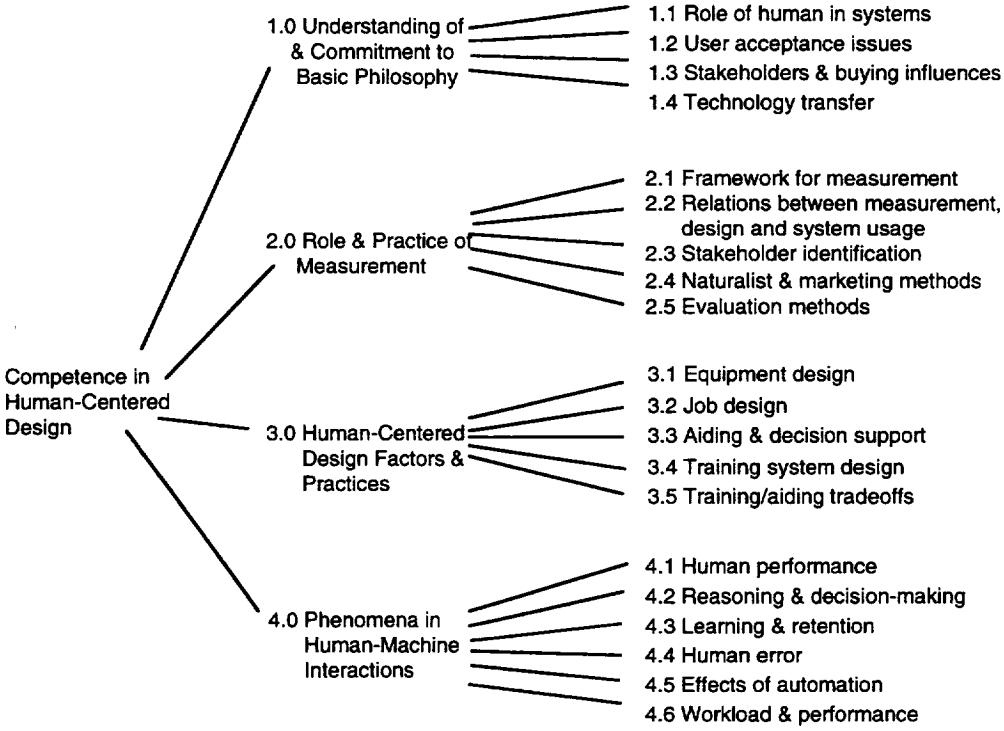

Figure 1. Competencies in human-centered system design (after Cody, 1993, page C-3)

Conclusion

As implied by the RTCA categories listed earlier, not every system or component should be certified to the same level. The extent of certification should relate to the component's or system's safety criticality. The extent of human factors certification should relate to the component's or system's level of interaction with the human operator. Criticality and level of interaction also affect which system stakeholders and issues require the most focus for the certification process, which brings us back full circle to our initial naturalizing step and the

analysis of stakeholder concerns. Typical human-centered stakeholder concerns are reflected in the topics and issues contained in our list of human-centered system design competencies.

It is important to remember that human-centered issues comprise only one set of issues among the many that must be considered when certifying aviation systems. From our perspective though, the human-centered issues are the most important because if the human cannot safely and effectively operate the system, all the other issues may be rendered irrelevant; and, by considering the human-centered issues, all the other critical issues are likely to be considered due to the up-front stakeholder analysis.

Acknowledgments

The authors gratefully acknowledge the inputs of John Hammer, Paul Frey, and Monica Skidmore from Search Technology, and the many workshop attendees, especially Rene Amalberti, Vince Galotti, Dick Gilson, Lew Hanes, Hartmut Koelman, Paul Stager and Ron Westrum who helped focus and coalesce the ideas presented in this paper.

References

Cody, W. J. (1992, October 4). *Test report for the pilot's associate program manned system evaluation.* (DARPA/USAF contract number F33615-85-C-3804, CDRL sequence number 4). Atlanta, GA: Search Technology, Inc. and Lockheed Aeronautical Systems Company

Cody, W.J. (1993, April). *Competencies in human-centered system design.* Appendix C. In R. L. Small, W. B. Rouse, P. R. Frey, & J. M. Hammer (Eds.), *Phase I Report: Understanding the Airspace Manager's Role in Advanced Air Traffic Control System Concepts* (contract number DTFA01-92-C-00028). Washington, DC: Search Technology, Inc. for the Federal Aviation Administration, ARD-210.

Hammer, J. M., Skidmore, M. D., & Rouse, W. B. (1993, April). *Limits identification and testing environment* (contract number NAS1-19308). Hampton, VA: Search Technology, Inc. for the National Aeronautics and Space Administration, Langley Research Center.

Rouse, W. B. (1991). *Design for Success.* New York: John Wiley & Sons, Inc.

Struck, W. F. (1992). *Software considerations in airborne systems and equipment certification.* Requirements and Technical Concepts for Aviation, Washington, D.C., Draft 7 of DO-178A/ED-12A. RTCA Paper number 548-92/SC167-177, July 27.

Selection and Training