

1995-128350

N95-34771

Quality Assurance and Risk Management: Perspectives on Human Factors Certification of Advanced Aviation Systems

Robert M. Taylor¹ & Iain S. MacLeod²

¹RAF Institute of Aviation Medicine

²Aerosystems International

Summary

This paper is based on the experience of engineering psychologists advising the U.K. Ministry of Defense (MoD) on the procurement of advanced aviation systems that conform to good human engineering (HE) practice. Traditional approaches to HE in systems procurement focus on the physical nature of the human-machine interface. Advanced aviation systems present increasingly complex design requirements for human functional integration, information processing, and cognitive task performance effectiveness. These developing requirements present new challenges for HE quality assurance (QA) and risk management, requiring focus on design processes as well as on design content or product.

A new approach to the application of HE, recently adopted by NATO, provides more systematic ordering and control of HE processes and activities to meet the challenges of advanced aircrew systems design. This systematic approach to HE has been applied by MoD to the procurement of mission systems for the Royal Navy Merlin helicopter. In MoD procurement, certification is a judicial function, essentially independent of the service customer and industry contractor. Certification decisions are based on advice from MoD's appointed Acceptance Agency. Test and evaluation (T&E) conducted by the contractor and by the Acceptance Agency provide evidence for certification. Certification identifies limitations of systems upon release to the service. Evidence of compliance with HE standards traditionally forms the main basis of HE certification and significant non-compliance could restrict release.

The systems HE approach shows concern for the quality of processes as well as for the content of the product. Human factors certification should be concerned with the quality of HE processes as well as products. Certification should require proof of process as well as proof of content and performance. QA criteria such as completeness, consistency, timeliness, and compatibility provide generic guidelines for progressive acceptance and certification of HE processes. Threats to the validity of certification arise from problems and assumptions in T&E methods. T&E should seek to reduce the risk of specification non-compliance and certification failure.

This can be achieved by creative and informative T&E as an integrated component of the design process. T&E criteria for HE certification should be directly linked to agreed on systems measures of effectiveness (MOE). HE risk should be managed principally through iterative T&E and progressive acceptance. Integrated and iterative HE T&E procedures linked to MOE criteria should feed progressive acceptance and provide confidence of compliance with specification and QA criteria. Certification should also include human behavior as an integral part of total systems functioning.

Traditionally, the risk for human performance in systems has been a customer responsibility. Recent initiatives in procurement policy however seek to provide a more integrated approach in which human resource issues, including operator/maintainer capability and training, are considered at all stages of the procurement process. The success of this initiative will depend on the ability to measure and predict human competencies in systems operations. It may be possible to successfully specify requirements for skill and rule-based behavior, but uncertainties inherent in the performance of knowledge based behavior present difficulties for system specification and certification.

Background

Experience with human factors (HF) aspects of various MoD air systems acquisition programs from the late 1970s through the 1980s revealed a number of general problems with the process of procuring systems to conform with good HE practice (Taylor, 1987). These problems may be summarized as follows:

- HF requirements were poorly defined in system specifications.
- HE design standards focused on the physical characteristics of the human-machine interface and not on the design process nor the performance and effectiveness of functions, tasks, and operating procedures.
- Increasing systems complexity amplified the impact of HF on operator performance and mission effectiveness.
- Poor systems integration increased human information processing and operator workload and reduced situational awareness.
- Responsibility for HF was shared between the customer and the supplier.
- The demand for human factors advice was increasing beyond that which could be supplied by customer HF advisors.
- Contracting policy (fixed price) encouraged rigid adherence to specifications and reduced the flexibility of changing HF requirements during system design and development.
- Acceptance procedures for HE quality assurance based on ergonomic checklists and late demonstration evaluation were ineffective and not directly related to mission effectiveness criteria.
- Problems with operating complex systems were difficult and costly to resolve through in-service modification and rectification.
- Unacceptable HF risk was carried by the customer.

The Human Engineering Approach to Systems Design

In 1985, discussions with North American HF colleagues in the ASCC and NATO military aircrew systems and cockpit standardization fora revealed similar problems in HE procurement. U.S. human factors personnel made substantial inroads into HE procurement problems during the Navy F/A-18 aircraft acquisition program. The procurement was based on extensive application of the principles of U.S. Department of Defense (DOD) Military Specification MIL-H-46855, "Human Engineering Requirements for Military Systems, Equipment and Facilities." MIL-H-46855 concentrates on the importance of timeliness of key HE activities, traceability, and on performance of critical tasks. It highlights the importance of early "front-end" analysis techniques (mission and scenario analysis, functional analysis, functional allocation, task analysis, and performance prediction) in reducing subsequent system development costs and risks. The progressive nature of these stages in human engineering analysis is illustrated in Figure 1. The design/development process is iterative. Analyses are repeated several times during the course of design/development. MIL-H-46855 promotes the value of an agreed on, tailored, and systematic Human Engineering Program Plan (HEPP) with traceability of the required HE effort from initial analysis, design and development, to final system test and evaluation including activities, responsibilities, time-scales, products, and deliverables. The HEPP specifies detailed contractor HE responsibilities and requires full consideration of resourcing, cost, and risk implications during contract tendering. Application of the HEPP is coupled with U.S. Military Standard MIL-STD-1472, "Human Engineering Design Criteria for Military Systems, Equipment and Facilities," which provides detailed equipment design requirements for good HE practice. Canadian HF colleagues who used the same principles verified that, used properly, MIL-H-46855 provided an excellent approach.

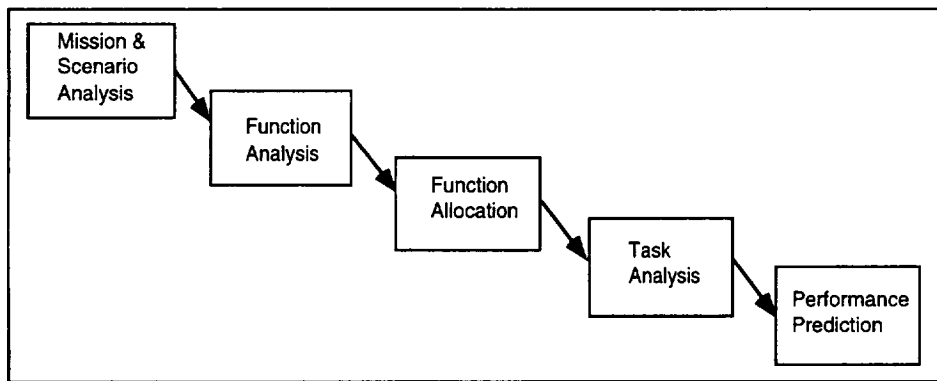


Figure 1. Stages of Human Engineering Analysis (From Beevis, 1992)

In 1985, NATO and ASCC cockpit design standards were concerned with relatively specific technologies, equipment, and individual controls, displays, layout, and lighting requirements. There was no statement of integrating policy, however. Based on the North American experience, it was decided there was a need to generate international standards similar to MIL-

H-46855 and MIL-STD-1472 in order to specify human engineering activities during aircrew systems acquisition. The derivative NATO and ASCC standards have been available since 1990. The sequence of NATO STANAG 3994 activities is illustrated in Figure 2. Similar activities are identified in the tri-service MoD Defense Standard DEF-STAN-00-25, "Human Factors for Designers of Equipment: PART 12: Systems," published in 1989. This MoD standard provides "*permissive guidelines*" in accordance with the "systems" approach without explicitly defining the requirement for a structured plan (i.e., no HEPP). Other initiatives aimed at wider integration of human resource considerations in systems acquisition, including manpower, personnel, training, and safety requirements, such as the U.S. Army Manpower and Personnel Integration (MANPRINT) program recently adopted by the U.K. MoD Army, incorporate similar systems HEPP procedures based on MIL-H-46855. Detailed MANPRINT HE procedures are described in Army Material Command Pamphlet AMC-P 602-1, "MANPRINT Handbook for RFP Development" (Barber, Jones, Ching, & Miles, 1987).

Test and Evaluation in Systems Human Engineering

According to STANAG 3994/MIL-H-46855 philosophy, the aim of HE T&E is to verify that the human-machine interface and operating procedures are properly designed so that the system can be operated, maintained, supported, and controlled by user personnel in its intended environment. The following guidance is derived from the STANAG with extracts from DOD-HDBK-763, "Human Engineering Procedures Guide" (U.S. Department of Defense, 1987).

Identification of Test Parameters

System performance requirements need to be identified for verification during HE T&E. Identification of HE T&E parameters should be based on Mission Analyses in conjunction with Critical Task Analyses and Loading Analyses. The criteria for selecting system performance requirements should be the same as those for identifying critical tasks. These requirements should be used to develop an HE test plan for approval by the procuring agency.

Test Plan

The HE Test Plan (HETP) should specify the type of test and evaluation techniques, rationale for their selection, the procedures to use, data to gather, number of trials, number and training of trial subjects, trial conditions, and criteria for satisfactory performance. The relationship with other T&E activities should also be indicated. The HETP should be specified to ensure that human performance requirements of the system are met and reported to the customer. Areas of non-compliance and their consequences should be identified with justification provided. The information should enable the customer to determine operators' and maintainers' performance and their influence on total system effectiveness and reliability. It should also indicate how the test program results will influence the design and apply to follow-on equipment or similar systems.

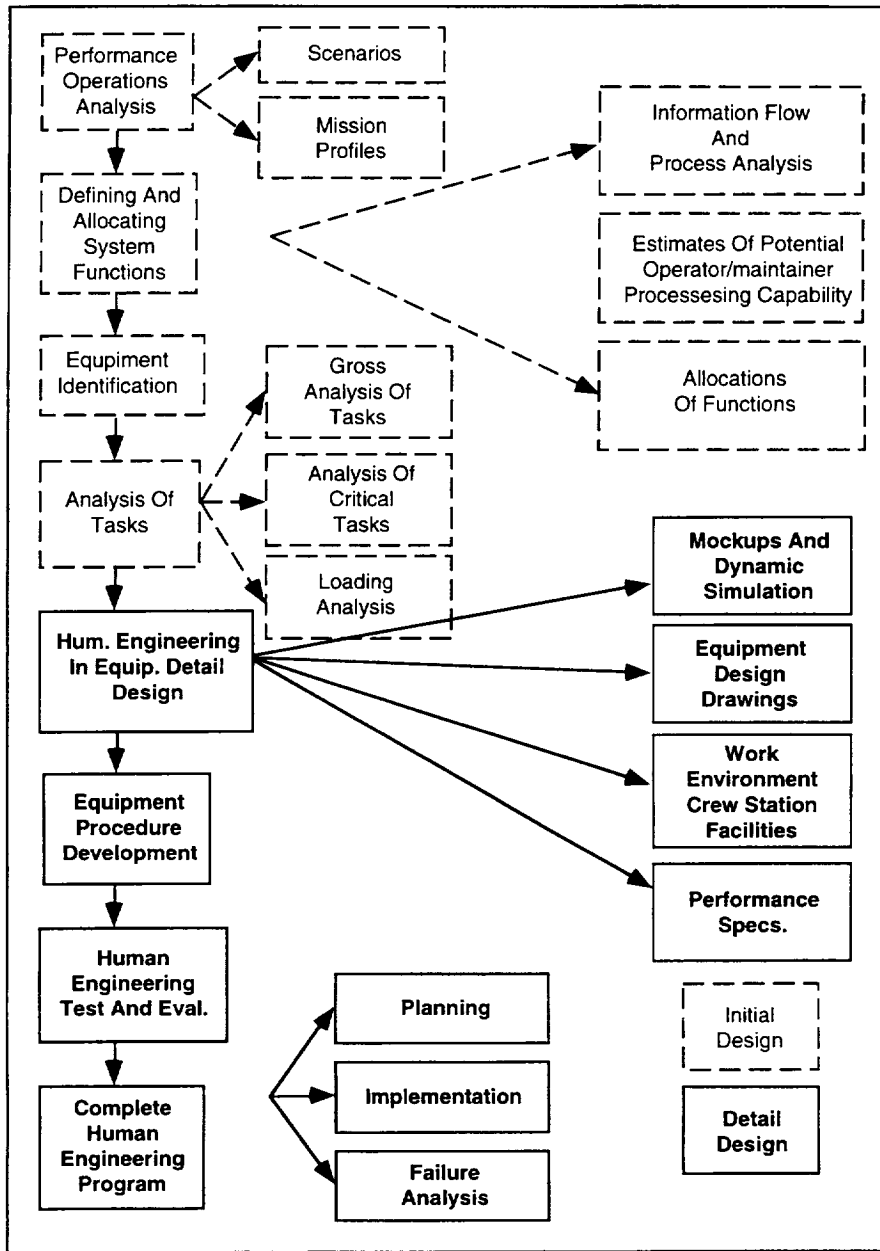


Figure 2. STANAG 3994 Activities

Quality Assurance Compliance

In indicating how HETP data will be used the plan should describe if the collected data will be used as formal proof of quality assurance compliance. Proof of compliance should be indicated as by either analysis, inspection, demonstration, or measurement. MIL-H-46855 reporting requirements call for Data Item Descriptions (DIDs) which include a Human Engineering Test Report or HETR. Formal compliance may be provided by the HETR.

NATO DRG Endorsement

The systems approach to HE was reviewed and endorsed recently by NATO Defense Research Group (DRG), Panel 8, RSG 14, "Analysis Techniques For Man-Machine Systems Design." The report by RSG 14 (Beevis, 1992) offers the following observations:

- The concept of a system may have been established prior to consideration of HF issues. As a result, designers and engineers have difficulty understanding the need for analyzing systems from a functional point of view. Therefore HE analyses of function allocation are of little value.
- The importance of the approach is that it permits engineers and designers to examine the system concept in new ways by identifying functions which must be performed rather than identifying subsystems which may be required.
- The function-oriented point of view facilitates development of novel system designs and encourages revolutionary as well as evolutionary changes.
- Increasing levels of automation and complexity in advanced mission systems magnify the importance of detailed analysis of the roles and functions of human operators.
- The effectiveness of HE analysis techniques is based on separating the system design problem into functions, subsystems, or states which are defined and validated.
- The subsystems are then recombined to predict system performance and operator/maintainer workload.
- It is generally assumed that the prediction of system performance is valid if it is based on the validated performance of sub-systems.
- Quality assurance aspects of the various techniques needs to be better understood.
- The link from HE analyses to system performance requirements must be made explicit.
- In most analyses, particularly for function allocation, the link is indirect and can only be provided by further analyses of system performance.

Merlin Human Engineering

In the U.K. we have experience with applying MIL-H-46855 principles by citing STANAG 3994 as a mandatory reference on several air systems acquisition programs. We have been particularly keen on raising the profile and effectiveness of HE and emphasize shifting more HE risk in procurement to contractors while maintaining HE quality assurance. STANAG 3994 is

perceived as a potentially valuable aid both for maintaining HE quality assurance and for managing HE risk in the procurement of complex mission systems. Also the risk for HE is perceived as particularly important during complex mission systems procurement. For complex systems, situation assessment and mission performance effectiveness are functions of the integration and interaction between the operator and the equipment's information processing and cognitive decision-making capabilities. The U.K. program which provides the most advanced example of STANAG 3994 application is the procurement of the Royal Navy Merlin (formerly EH101) Anti Submarine Warfare (ASW) helicopter. This project is known as the Merlin Prime Contract (MPC). The RAF Institute of Aviation Medicine (IAM), DRA Farnborough, and Aerosystems International have acted as HE technical advisors on the program. This paper is largely based on the HEPP acceptance/compliance assurance issues that have arisen on the MPC program.

Merlin Specification Rationale

The development of the U.K. Royal Navy (RN) Merlin helicopter evolved from the RN EH101 development program by transferring responsibility for the RN EH101 helicopter to a prime contractor (IBM/ASIC). In the process the helicopter was renamed Merlin. To aid the submission and assessment of bids by potential prime contractor candidates, the Merlin aircraft was specified according to design, functionality, and its Operational Performance and Acceptance Specification (OPAS). The Technical Requirement Specification (TRS) lists standards and rules governing design. The OPAS dictates the trials, their types and formats, and methods required for acceptance of Merlin by the RN. Figure 3 shows the basic contents of the Merlin specification.

Operational Performance and Acceptance Specification (OPAS)

The OPAS trials occur in two forms. Single Task Trials assess the operational performance of individual equipment. Stressing Mission Trials on the other hand assess the operational performance of multiple systems within a realistic flight trial and operational scenario. The requirements for trial aircrews are specified and where a need for trained service aircrews is identified, appropriate qualifications, experience, and conversion training are established. The means of assessing trial performance is also specified. One of the primary criteria for assessment are measures of effectiveness (MOE). The MOE are based on specific high level functions that are progressively isolated to MOE levels depicting specific performance characteristics that must be demonstrated over a series of trials. Pass/fail acceptance criteria are agreed on for the deterministic Single Task Trials. The operator-in-the-loop stressing missions will be performed on a test and declare basis (i.e., with no pass/fail criteria). Current judgment assumes that service crew competence is not a contractor responsibility. Thus, crew performance is considered to be an uncontrolled and unpredictable variable. The contractor's intention is to reduce risk in the stressing missions by additional operator-in-the-loop simulations prior to OPAS.

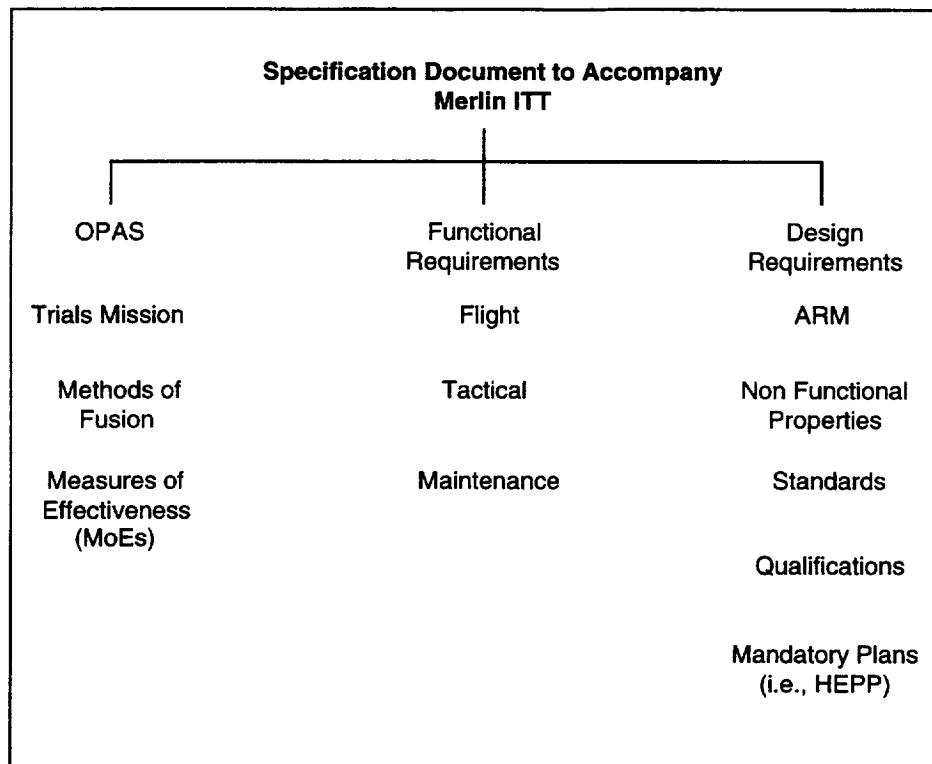


Figure 3. The Contents of the Merlin Specification

Merlin Human Engineering Program Plan

The application of human engineering to the Merlin is governed by a mandated HEPP, in accordance with STANAG 3994. The HEPP is managed by Westlands Helicopters Ltd. (WHL) on behalf of IBM/ASIC. The coordinated HEPP is a tailored implementation of STANAG 3994 and is applicable to all new or modified equipment and systems delineated by the Merlin specification (essentially an updated EH101 specification), namely: Active Dipping Sonar (ADS), Data Link (DL), Identification Friend or Foe (IFF), Global Positioning System (GPS), and Digital Map. Figure 4 illustrates the concept of the HEPP and T&E binding together Merlin high level functionality.

The weakness of the HEPP is its limited influence on equipment or systems which were developed for RN EH101 without a mandated HEPP and will remain largely unmodified. The plan focuses on extended mission systems human machine interfaces (HMI) in the rear cabin where the Merlin specification is of primary influence. Aircraft HE integration issues pertaining to the flight deck exert little influence on the Merlin HEPP, as they have been addressed

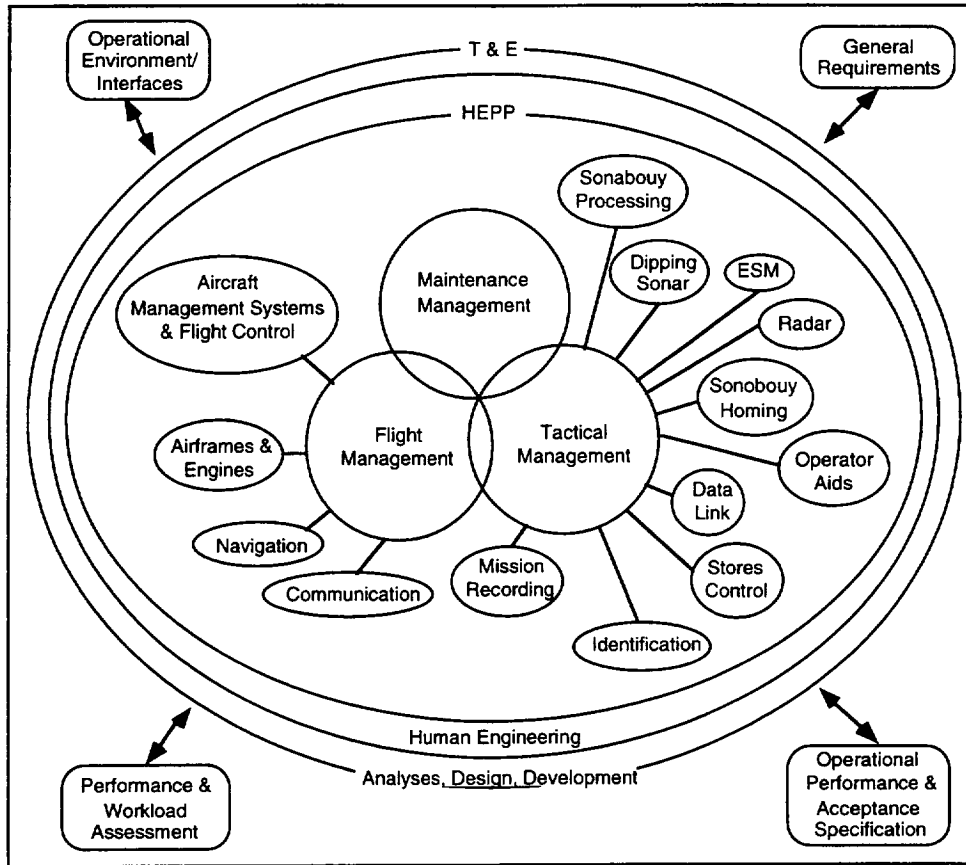


Figure 4. Merlin High Level Functionality

through RN EH101 development. OPAS fulfills the mission analysis requirement. Also, system functions are based largely on the existing EH101 definition and allocation and are amplified by the Merlin Functional Requirements Definition (FRD). Further functional analysis is rendered either unnecessary or potentially ineffective as a result. Notwithstanding the requirements of the new Merlin equipment, the HEPP largely concerns post activities equipment identification, from task analysis to equipment detail design, with the traditional emphasis on HMI. The primary focus is to ensure that as new features are added operator HMI workload remains manageable. Also early identification of workload and design challenges reduces the risk of future cost and scheduling problems. Consequently, the HEPP embodies a strong workload emphasis. It specifies the analyses, simulation assessments, workload measurement trials, and tools for HMI development. In summary, through extended HMI the HEPP and associated T&E linked with OPAS MOEs can be conceived as the means of delivering HE for required TRS and FRD high level functionality. Figure 5 shows the HE testing sequence in relation to the system life cycle.

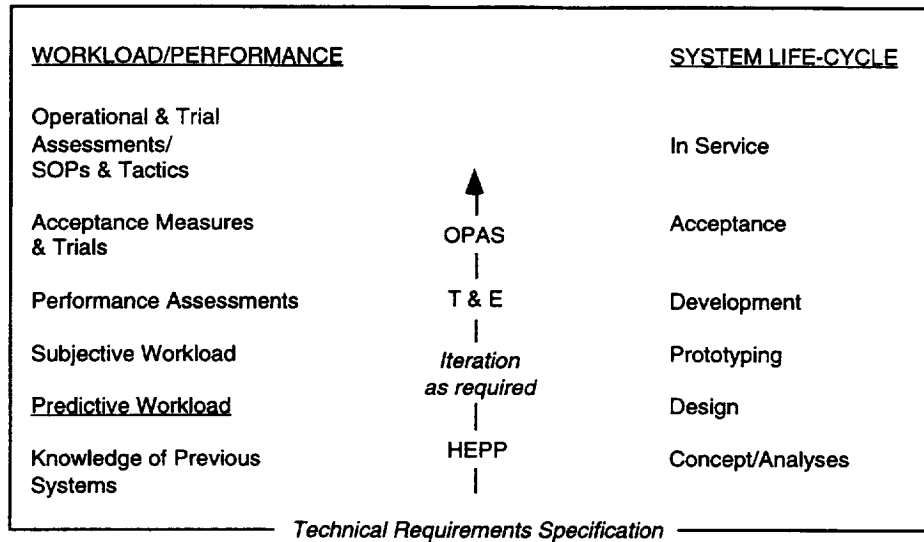


Figure 5. HE Testing Sequence in Merlin Life Cycle

Merlin Predictive Analysis

A key feature of the Merlin HEPP is its inclusion of predictive analyses of workload and decision-making to aid design assessment, to support progressive HE acceptance, and to anticipate future simulation and flight trials (MacLeod, Biggen, Romans, & Kirby, 1993). Critical mission segments were selected from OPAS. Mission “story-lines” were created for the segments based on interviews with Subject Matter Experts (SMEs). These story-lines were transformed into Operational Sequence Diagrams (OSDs) at the aircrew sub-task activity level and the OSDs were the basis for workload and decision analyses. The sequencing and relationship of the analyses are depicted in Figure 6.

Workload Analysis

In workload analysis, detailed task timelines were generated from empirical observation and published task-time data. Attentional demand loadings were created from SME loading estimates using VACP (visual, auditory, cognitive, psychomotor) workload model criteria recommended by MoD (Taylor, 1990), and were subsequently validated by the contractor (Biggen, 1992). Results were used to indicate workload peaks and troughs, to determine their causes, and to suggest solutions for ameliorating unwanted workload. The data generated to date indicate predicted task-time overruns on critical mission segments as compared with baseline intended times. The overruns were addressed largely with reference to the efficiency of proposed operating procedures. The predicted workload data obtained so far indicate some short transient areas of multi-task conflict during continuous monitoring tasks, leading to reduced situational awareness due mostly to the demands of simultaneous intercom tasks. There

were also indications of imbalance in workload distribution between the two rear-operator positions (observer and air crewman).

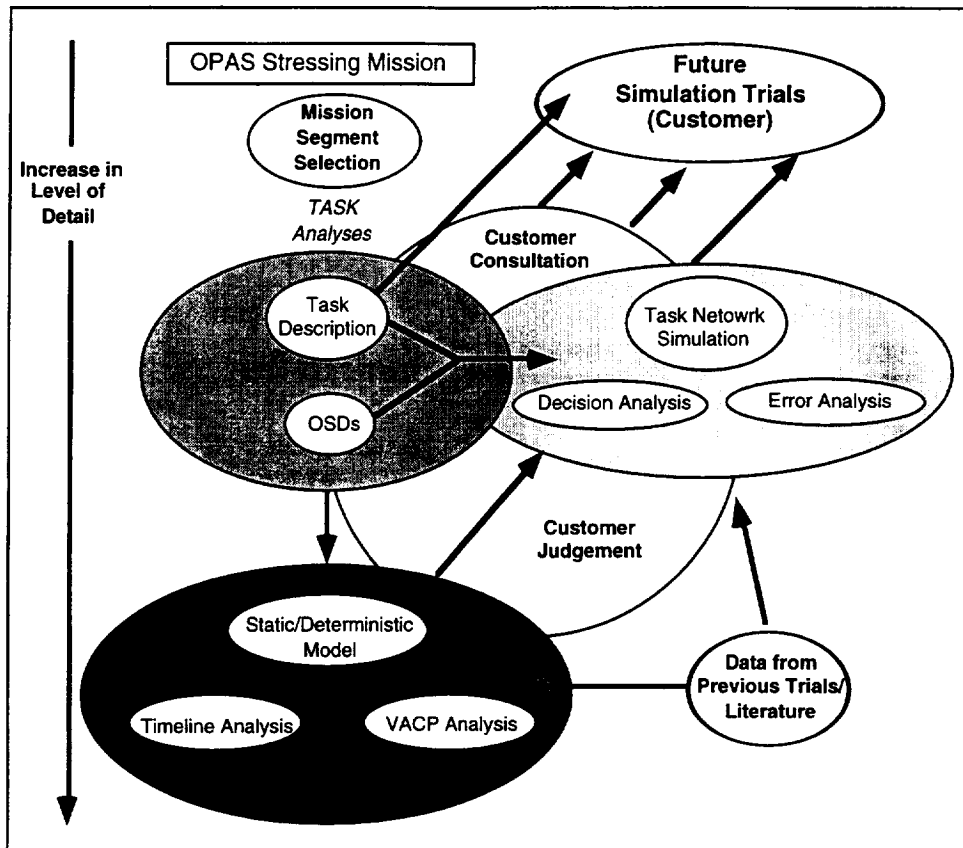


Figure 6. Relationship of Merlin HE Predictive Analyses

On the whole, predictions were judged by the contractor as indicating manageable workload problems, with amelioration evidenced through procedure development and crew training. Further modeling prediction and examination would occur during simulator workload validation. The initial analysis was static and deterministic. However future analyses using dynamic and stochastic network simulation are planned. Maintaining and refining the workload prediction model and keeping it up-to-date with new equipment and task requirements is an important responsibility for progressive HE acceptance.

Decision Error Analysis

The decision analysis used a novel technique to examine task related decision processes and their associated errors. The TRS called for particular attention to the cognitive aspects of Merlin

HE. The quality of situation assessment and decision-making were considered key factors in determining operational effectiveness of the Merlin mission system. This consideration influenced the choice of Stressing Missions for OPAS. Stiles and Hamilton (1987) point out that interdependency of mission goals means there are often decision points which permit the operator to modify intentions according to assessment of the situation. Options associated with goals are controlled at these points. The designer must therefore ensure that option paths are clearly presented at these junctures within the situation context. Decision analysis could become the controlling activity for the design process, complementing information analysis. It was necessary to develop a novel technique because decision analysis is a relatively new activity. Several attempts at developing a task analysis technique for decision making have been reported in the literature. But, as noted in the RSG 14 reported (Beevis, 1992), no single most promising technique has emerged. The form of decision analysis used on Merlin is described in detail by MacLeod, Biggen, Romans, and Kirby (1993).

In summary, based on the OPAS mission story-line OSDs, human error probabilities associated with performance of task segments were generated based on the literature or SMEs. The effects of errors on subsequent decision processes were estimated by SMEs in terms of error probability and error severity. The error influences on critical tactical decisions were then mapped against estimated task times through dynamic stochastic network simulation in MicroSAINT for Windows™ (MSW). MSW provided dynamic simulation of critical decisions and errors through various decision paths to operator task completion using Monte Carlo rules. The results provided traceable evidence of the efficacy of tactical decisions on the probability of mission success and identified critical decision points affecting mission performance. The critical decision points were correlated with the workload analysis. They could also be used to guide design activity through improved information availability, option clarification and highlighting, and procedure modification and training.

Certification

By definition, to certify is to endorse or guarantee that certain required standards have been met. Certification is “the act of certifying” or “the state of being certified.” The word “certify” has its roots in the Latin *certus* (certain) and *facere* (to make). “To be certain” means to be positive and confident about the truth of something. In law, certification is a document attesting to the truth of a fact or a statement.

The requirements for the act of certification are that the system should fit its intended purpose and meet specific requirements of reliability, safety, and performance. Certification is more than endorsing compliance with the system specification, a contracting authority concern, because the specification may not include all the necessary requirements.

Government Functions

In government management of systems design the role of certification can be considered as a judicial function rather than a legislative or executive function. Certification is a judgment on the

design standard of the system and carries with it major implications for program risk and cost. The following are further notions of how these functional distinctions can be applied:

- **Legislative Functions:** Staff requirement generation, system technical requirements specification, design standards definition, acceptance standard definition, technical transfer agreement, and contracting.
- **Executive Functions:** Contract management, program planning, concept analysis, prototyping, design, development, documentation, and production.
- **Judicial Functions:** Test and evaluation, compliance demonstration, acceptance, concession negotiation and agreement, audit, quality assurance, and certification.

Legislative functions are responsibilities of the customer, task sponsor, or contracting authority (MoD) and its project/program office. Executive functions are largely responsibilities of the contractor/manufacturer, in consultation with the customer authority. Separation of the judicial function from the legislative and executive functions is essential to preserve judicial effectiveness. Failure to achieve certification has major implications for both the customer and the contractor. It follows, then, that in the interests of independence and impartiality, HE certification needs to be independent from both legislative and executive functions. Certification of the overall testing and acceptance plan should ultimately be the responsibility of an independent agency appointed by the customer authority and recognized by the contractor/manufacturer.

Certification Authority

Certification is the end product of successful test and evaluation. Logic dictates that test and evaluation follows analysis and design. In the U.K., the ultimate endorsement for military aircraft systems is the Release to Service granted by the MoD Controller Aircraft (CA), namely, the CA Release. Certification for civil aircraft is issued by the Civil Aviation Authority (CAA). CAA certification must be particularly stringent because of the responsibility for carrying passengers. The object of CA Release is to provide a statement to the Service Department that the aircraft will perform its intended in-Service role with acceptable levels of safety and effectiveness. The statement includes any limitations or restrictions to observe in operating the aircraft at the defined build standard. All systems should be safe to operate and fully effective under all specified environmental conditions. CA Release covers the performance of mission systems and vehicle engineering systems, as well as basic handling qualities of the aircraft. CA Release is a progressive activity, beginning with an Initial Temperate Functional CA Release covering the temperature environment for initial aircraft delivery for flight testing. Subsequent stages of release extend the scope of clearances for flight testing of early production aircraft through the activities leading to formation of the first operational squadron.

MoD's current policy is to appoint an Acceptance Agency to ensure that the system produced is adequately tested to prove that it satisfies specification requirements. The Acceptance Agency interfaces directly with the contractor on behalf of the MoD Authority in order to endorse trial plans, monitor trials, and assess results against contractual performance criteria and recommends acceptance or rejection by MoD. Responsibility for trial planning and control rests with the contractor. A MoD Trials Agency may be appointed to assist the contractor with trial planning and control details involving MoD facilities and to provide advice on operational and support requirements. The MoD Aeroplane and Armament Experimental Establishment

(A&AEE) at RAF Boscombe Down is the MoD agency for aircraft operational trials and acceptance testing. A&AEE provides the aircrew for the Merlin contractor T&E progressive acceptance demonstrations and flight trials. CA Release is based on recommendations by A&AEE. A&AEE assessments are governed by requirements of the aircraft technical specification and relevant MoD Defense Standards, MIL Specifications and MIL Standards, particularly DEF-STAN-00-970, "Design and Airworthiness Requirements for Service Aircraft." DEF-STAN-00-970 includes chapters on general HE requirements for cockpit vision, controls, displays, layout, and lighting. These chapters are referenced in the system specification and are used by the manufacturer to guide design activities. The manufacturer is required to provide evidence of qualification for compliance to assist the certification process. Avionics systems rigs with representative human-machine interfaces are used by A&AEE to support the process of CA Release. Data generated by the contractor during developmental trial testing also contribute to CA Release. A&AEE does not employ HE specialists, therefore weakening A&AEE's ability to act as an Acceptance Agency for HE. There is merit in having a single Acceptance Agency responsible for all aspects of aircraft acceptance. DRA and IAM provide A&AEE with technical advice and scientific support for HE Acceptance. As the demand for HE Acceptance increases and becomes more sophisticated, the need may arise for A&AEE to employ HE specialists as an integral part of its acceptance function.

Certification Validity

The credibility or trustworthiness of certification depends on the validity of the evaluation on which it is based. Careful attention must be paid to threats to validity for particular evaluations and design decisions. Sherwood-Jones (1987) provides a summary of the threats to quality in evaluations using quasi-experimental designs; behavioral scientists and HE specialists will find them familiar. There are nine threats to internal validity:

- History – events, other than those studied between pre-test and post-test, that could provide an alternative explanation of effects.
- Maturation – processes within the system producing changes as a function of time passage.
- Instability – unreliability of measures, fluctuations in sampling.
- Testing – the effect of taking a test on the scores of a second test.
- Instrumentation – changes in calibration, observers, or scores that produce changes in obtained measurements.
- Regression artifacts – pseudo-shifts from subject or treatment selection based on extreme scores.
- Selection – bias from differential recruitment of comparison groups leading to different mean levels on measure of effects.
- Experimental mortality – differential loss from comparison groups.
- Selection maturation interaction – bias from different rates of "maturation" or "autonomous change".

Six threats to external validity can be identified pertaining to problems with interpreting experimental results and generalizing to other settings, treatments, and measures of the effect:

- Interaction effects of testing – for example, pretesting effects-sensitivity to variables.

- Interaction of selection and experimental treatment – non-representative responsiveness of the treated population.
- Reactive effects of experimental arrangements – artificiality in the experimental setting that is atypical of the normal application environment.
- Multiple treatment interference – effects of multiple treatments as distinct from separate treatments.
- Irrelevant responsiveness of measures – all complex measures have irrelevant components that may produce apparently relevant effects.
- Irrelevant replicability of treatments – complex replications failing to reproduce the components responsible for the effects.

Quality Assurance

In accordance with the emphasis in MIL-H-46855/STANAG 3994 on functional effectiveness, certification of criteria for HE acceptance should provide a broad endorsement of *quality assurance* (QA) or fitness for purpose. The word “quality” is defined as “the totality of features and characteristics of a product or service that bear on its ability to satisfy a given need.” The definition of quality assurance is “all activities and functions concerned with the attainment of quality.” MoD Defense Standard DEF-STAN-05-67, “Guide to Quality Assurance in Design,” emphasizes that those concerned with a given project can contribute to and are involved with maximizing and assuring its quality. QA organizations undertake specific activities in measuring quality and ensuring that appropriate contributions are made by all personnel to quality assurance. But responsibility for the final product’s quality rests with line managers who are responsible for design and production, including performance over the system life cycle. This is a basic tenet of *Total Quality Management* (TQM).

HE can support the TQM approach by helping to identify characteristics of system users and their requirements, as well as features of operator/maintainer performance which contribute to variance in the system product or output. The RSG 14 Report (Beevis, 1992) notes that distinction is made between *quality of design*, meaning “the process of task recognition and problem solving with the objective of creating a product or a service to fulfill given needs,” and *quality of conformance*, meaning “the fulfillment by a product or service of specified requirements.” HE QA is a function of how well it contributes to the design of an effective system (quality of design) and how well it provides accurate, timely, and usable information for the design/development team (quality of conformance). The following indices or criteria were proposed by RSG 14 (Beevis, 1992) as providing evidence for HE QA:

- Schedules which show that analyses will be timely
- Organization charts which indicate that the HE effort will be integrated with other systems engineering and Integrated Logistical Support (ILS) activities
- Use of metrics and measures of effectiveness that are compatible with each other and with other engineering activities
- Compliance with a relevant specification

Scheduling and charting HE activities are key MIL-H-46855/STANAG 3994 tenets. On the basis of a critique of HE analysis techniques, RSG 14 (Beevis, 1992) recommends considering the following QA criteria during development of a HEPP:

- Completeness
- Consistency with preceding analyses
- Timeliness
- Compatibility with other engineering analyses

Consideration of QA draws attention to the need for concern for both the design process and content of the product. Advanced systems employ new interface technologies and concepts. Existing HE standards for detailed equipment design are losing relevance and influence as new technologies and concepts are introduced. Currently the nature of the design process is assuming greater importance in products' overall quality. HE certification for advanced aviation systems needs to be concerned more with proof of process than proof of content, according to the philosophy of MIL-H-46855/STANAG 3994.

Creative Evaluation

The certifying authority might wish to conduct some form of human factors or ergonomic audit for QA certification purposes. Indeed, the U.S. General Accounting Office (1981) provides guidelines for this purpose by identifying questions to help assess whether or not human factors were considered during the weapon system acquisition process. But such an audit would not serve to inform the design process. Evaluation should be useful, informative, and preferably, creative. The need for useful evaluation was addressed by Patton (1978). Evaluation can be either "*formative*," aimed at improving the design, or "*summative*," aimed at deciding whether or not to proceed with a design. There are two fundamental requirements for making evaluation useful:

- Relevant decision makers and information users, rather than an abstract target audience, must be identified.
- Evaluators must react, adapt, and actively work with identified decision-makers so as to make informed judgments about the evaluation; i.e., focus, design methods, analysis, interpretation, and dissemination.

Progressive Acceptance

Both in common engineering practice and in the formalized approach advocated by MIL-H-46855 and STANAG 3994, HE acceptance testing is embedded as an integral part of the design process. HE involves a logical sequence of mostly iterative activities, each involving the application and testing of design and performance criteria and associated standards. Like software QA, T&E for HE acceptance needs to be phased or progressive. Progressive acceptance T&E should be embodied in the different stages and levels of the system design and development process. The T&E could be referred to as technical rather than operational. Higher levels of HE QA concerned with functionality and effectiveness are the most significant and yet the most difficult to check. Consequently, there is a danger that verifying integrated functional effectiveness of the total system, with the operator/maintainer in the loop, will be fully addressed only in final operational acceptance testing. Relying only on final operational T&E for full HE acceptance is risky, particularly with complex mission systems that require major engineering integration activity and are designed to prevent potentially high operator workload.

In theory, the system should be designed to pass operational T&E without any uncertainty. Progressive HE acceptance testing is needed during integration on rigs, simulation facilities, and development aircraft to ensure that the lower level requirements are being dealt with correctly. Otherwise it is unlikely the higher levels will be acceptable. It is emphasized that the process must address in particular depth the operational performance of complex mission systems to guarantee functional integrity and effectiveness. Progressive acceptance is a key contributor to proof of process.

Certification of Human Behavior

The GFE Approach

Formal acknowledgment of human functioning as an integral component of systems, together with equipment operation, is a relatively recent development. Certification of systems where the human is considered as a system component presents new challenges for systems engineering. The traditional approach to systems engineering focuses on equipment operation. It treats the human operator/maintainer as a given quantity, over which the contractor has little or no control or responsibility, often “jokingly” referred to as Government Furnished Equipment (GFE). The traditional design objective is to provide a system fit for a purpose that can be reliably, safely, and effectively operated by the “average” operator/maintainer. Unfortunately, “average” is ill-defined and becomes a quantity left to the judgment of the MoD A&AEE test aircrew. The danger in the GFE approach to human capability is that it implicitly assumes that treating the performance of the average operator/maintainer in a deterministic, predictable, and mechanistic manner is adequate, when in fact the uniquely human characteristics in systems are flexibility, adaptability, and unpredictability. Consequently, traditional HE analyses have tended to be “physicalistic” (anthropometry, ingress/egress, workspace layout, visibility and reach, lighting, and task timeline analysis) rather than cognitive (situation assessment, decision-making, errors of judgment, expertise, intentions, application of knowledge, tactics, strategy, and goals). The consequences of the physicalistic/cognitive distinction are discussed in detail in the second ASI position paper by the authors (MacLeod & Taylor, 1994). The GFE approach prevents the Merlin OPAS Stressing Missions from being more than a test and declare process. The customer still bears the risk of total integration failure since this can be attributed to GFE variables. MANPRINT procedures, introduced since the EH101 procurement, seek to address the problem on future programs by procuring manpower, personnel and training, and human engineering.

Cognitive Functions

The traditional HE assumptions about human design requirements are at best limited in scope, and at worst invalid, if they are based on inappropriate models of human interaction in systems. They may lead to inaccurate, unrealistic, and optimistic assessments of overall system capability and effectiveness. Recent U.K. procurement experience indicates a tendency to be over optimistic with predictions of future operational performance of complex advanced systems under development. With the GFE approach, the risk for human functionality in total system performance is carried by the customer rather than by the contractor. Failure to achieve systems

performance targets in T&E can be ascribed to human capability or performance variability. The problem then becomes one of the human not matching the machine rather than the converse, and it needs to be solved by improved customer-provided training or by enhanced customer selection standards, not by in-service system upgrades. This is increasingly untenable in a procurement climate seeking to minimize the risk to the customer. It is particularly inadmissible for procurement of complex advanced mission systems where system performance effectiveness is increasingly a function of operator-equipment integration and cognitive level interactions dealing with information processing, situation assessment, and decision-making. The RSG 14 report (Beevis, 1992) concludes that while it is generally assumed that new advanced systems place increasingly high demands on the cognitive aspects of operator/maintainer behavior, most HE techniques on the other hand lend themselves to the description of skilled behavior, not cognitive behavior. It seems that certification of HF in advanced future systems will require better resolution, analysis, and engineering of cognitive functions than presently available with HE techniques. Stiles and Hamilton (1987) describe how a cognitive engineering approach to functional analysis will be needed for identifying a pilot's intentions during his or her interface with the system, as well as for providing a design (information and/or control) to help achieve the intentions. The requirement for improved resolution of cognitive functionality is discussed further in the second position paper by the authors (MacLeod & Taylor, 1994).

Aircrew Certification

Certification procedures for aircrew selection/training might provide some of the missing human cognitive functional concepts and behavioral parameters needed for advanced aircrew systems HF certification. However, aircrew selection and training criteria are not yet firmly based on an understanding of cognition and behavior theory. Criteria for certifying aircrew ability as "adequate" for civil flying or "above average and not requiring further training" for military flying are largely based on performance of instrument flying tasks and knowledge of rules and procedures for air safety. The required standards of *airmanship* are still highly subjective and largely the responsibility of experienced assessors/flying instructors. However, it is possible that the mystery surrounding airmanship will dissipate. MIL-H-46855 and STANAG 3994 call for a Potential Operator Capability Analysis to provide data for defining and allocating functions. Also, MANPRINT requirements for Target Audience Description (TAD) demand a more explicit, objective, and theoretically consistent approach for defining aviator performance.

The problems of measuring and developing competence in the cockpit are major concerns of training technologists. Brown (1992) notes the increasing concern with cognitive decision-making competencies for combat aircrews in addition to traditional requirements for flying skills and knowledge. In the systems approach to training, competency is viewed as an outcome of a system and an integral part of its overall operation. Recent procurement policy for "turn-key" training systems has created the need for more functional and performance-based specifications rather than formerly equipment-based specifications (Brown & Rolfe, 1993). The customer must therefore define the operating constraints and the training outcomes required, including the activities to be learned on a device, the rate of learning, and the performance standard. Thus there is increased emphasis on the quality of the task and training analysis performed by the supplier in determining that equipment will satisfy task demands. Attention is

also focused on the role of evaluation in acceptance testing; evaluation may need to be extended into the system life-cycle to demonstrate that a device actually instructs.

A recent review of the requirements for operator and automation capability analysis, in the context of advanced aircrew system design and "human-electronic crew" teamwork, points to the key role of human performance modeling for predicting human system performance (Taylor & Selcon, 1993). The embedded human performance model for cockpit performance prediction and pilot intention inferencing in the U.S. Air Force Pilot's Associate indicates some of the necessary HE elements (Lizza, Rouse, Small, & Zenyuth, 1992). There is a need for a common performance-resource model and associated taxonomy for systematically linking human resource capabilities to mission performance task demands that incorporate features required for HE analysis and relevant human competence parameters (Taylor, 1991).

SRK Taxonomy

The taxonomy of skill, rule, and knowledge-based (SRK) behavior provides a potentially useful way of thinking about HF certification issues. In *skill-based behavior*, exemplified by the performance of controlling tasks, performance is relatively easily measured, demand is relatively easily predicted, and the capability requirement can be specified and verified. Hence, skill-based behavior is a strong candidate for HF certification. More or less the same can be said for *rule-based behavior*, exemplified by supervisory and monitoring tasks. Difficulty arises with the certification of *knowledge-based behavior*, exemplified by planning and decision-making tasks. By definition, knowledge-based behavior is novel, measurement of performance is qualitative and at best nominal (e.g., correct or incorrect decision), and demand is stochastic and probabilistic rather than predictable and deterministic. The capability requirement for knowledge-based behavior is the most difficult to anticipate, specify and verify.

It is difficult to conceive of a contractor being prepared to guarantee, say, that incorrect decisions concerning uncertainty would be made less than five percent of the time. Traditionally, analysis of decision points where the operator changes goals, alters information, and controls requirements, is omitted from the design process. Some progress can be made, though, through decision analysis (MacLeod, Biggen, Romans, & Kirby, 1993; Stiles & Hamilton, 1987). Metzler and Lewis (1989) report that the procurement of the Airborne Target Handover System/Avionics Integration (ATHS/AI) for the Apache (AH-64A) aircraft specified a 30 percent reduction in crew task time for each task (60 percent overall), 90 percent mission reliability, and no more than five percent of the mission aborts *attributed to human error*. The Merlin decision analysis explored the impact of decisions on the probability of mission success; the findings however are considered indicative rather than definitive.

Ideally, the design goal is to provide systems that are totally predictable and reliable. This must mean avoiding, if possible, the need for knowledge-based behavior, but probably the provision of totally automated systems. However, it is in the nature of the military environment that human situation assessment, hostile intention inferencing, and unbounded knowledge-based behavior applied through flexible adaptation of goals, tactics, and strategy often provide the "combat winning edge." Systems that are intended to operate in uncertain environments need to provide the unrestricted scope for appropriate knowledge-based behavior. The recent debate about providing situational awareness in highly automated systems is an example of this problem. Arguably for certain military systems where effectiveness depends on flexibility, adaptability, and unpredictability it is the limitless capacity for knowledge-based behavior that needs to be certified.

Conclusions

Notwithstanding system life cycle considerations (i.e., maintenance, in-service modification, up-dating), certification marks a formal end to the system design, development, and production process. It is the last operational endorsement of the proof of concept, proof of process, and proof of product. It is the final sanction of the solution to the design problem. The threat of non-certification and a severely restricted release to service is a potentially powerful device. It could help ensure that HF considerations maintain their rightful place at the center of the design process. Consideration of the ability to certify HF aspects of system design is a sign of the maturation and acceptance of HF methodologies and standards. But, realistically most HF issues are a long way from being assigned sufficient importance to become potential "show stoppers" for certification. With power comes a risk of abuse. The preceding could be a problem if certification is seen as an end in itself. What happens if, in assessing novel technology and a revolutionary new system concept, existing certification criteria are wrongly focused, invalid, and fail to measure true impacts on operators' health and safety? The certification authority should find an incumbent obligation of concern that necessitates continual self-evaluation. Care must be taken not to assign blind trust to existing certification procedures. Certification alone is not generative or creative. Front-end analysis, iterative design and testing, and progressive acceptance provide the methods and tools for generating confidence and HE quality assurance necessary for certification. There is a danger of certification encouraging "rear-end analysis." As such, it carries many of the characteristics and weaknesses of traditional, 1970s style late ergonomic assessments, as identified at the beginning of this paper. Neither is certification a panacea, capable of remedying the ills of poor design methodology. It can only be as good as the front-end analysis and T&E that feeds it. It is probably essential to ensure that HF considerations, HE processes, and HE standards are contractually mandated as an integral part of the design process using MIL-H-46855/STANAG 3994 procedures. HF certification then can be added to endorse compliance with these contractually binding requirements.

The uncertainty of human reliability is a fundamental problem for HF certification. Certification also concerns matters which are certain and true. Obviously, one cannot be certain about matters which are variable. Certification cannot be obtained for design concepts or prototypes tested only in the abstract or by simulation. Certification can only be valid for the real product tested in the real operational environment. Progressive acceptance rather than certainty is all that can be obtained for concepts and prototypes. Certification can guarantee that specific absolute HF design standards are met and that necessary design and test processes and activities have taken place. However when a human is an integral system component, it is difficult to conceive of contractually meaningful expressions of certainty about total system fitness for purpose, system performance, and functional effectiveness. Human performance, whether skill, rule, or knowledge-based, is inherently uncertain. All that can be expected with certainty is an endorsement or guarantee that sometimes the required standards of human-systems performance will not be met. Levels of confidence in human systems performance could be provided in probabilistic rather than absolute terms. Probabilistic certification of human-systems operation might provide the basis for a form of limited release to service, perhaps associated with additional supervisory, performance monitoring, and training safeguards. In advanced systems, the role of humans is increasingly one of dealing with the uncertainty that cannot be handled automatically, or the variability that cannot be predicted and

controlled. The human component is responsible for generating the required system performance and for achieving the intended system effectiveness goals under circumstances that cannot be entirely predicted and anticipated. Probabilistic descriptions of the intended and expected system operation, performance, and effectiveness are likely to become more common as specification goals and certification norms. Certainty is perhaps too absolute a term for many HF certification requirements. Confidence, acceptance, and perhaps *certitude* may be more appropriate terms for describing the relative uncertainties of human-machine systems performance.

References

- Barber, J. L., Jones, R. E., Ching, H. L., & Miles, J. L. (1987, September). *MANPRINT Handbook for RFP Development* (AMC-P 602-1). HQ U.S. Army Material Command.
- Beevis, D. (1992, July). *Analysis Techniques for Man-Machine System Design*. NATO, AC/243 (Panel 8) TR/7.
- Biggen, K. (1992, March). *EH101 Mission Workload Simulation Validation Trials Report* (Westlands Helicopters Report No. ER02Q002W).
- Brown, H. M. (1992). Competency in the cockpit. In D. Saunders & P. Price (Eds.), *Developing and Measuring Competence. Aspects of Educational and Training Technology XXV*. London.
- Brown, H. M., & Rolfe, J. M. (1993). *Training requirements or technical requirements*. Paper submitted for publication.
- Lizza, C. S., Rouse, D. M., Small, R. L. & Zenyuth, J. P. (1992). Pilot's associate: An evolving philosophy. In T. E. Emerson, M. Rienecke, J. Riesing, & R. M. Taylor (Eds.), *The human electronic crew: Is the team maturing?* (U.S. Air Force Wright Laboratory Report No. WL -TR-92-3078). Wright-Patterson Air Force Base, OH: U.S. Air Force Wright Laboratory.
- MacLeod, I. S., Biggen, K., Romans, J. & Kirby, K. (1993). Predictive workload analysis--RN EH101 helicopter. *Contemporary Ergonomics 1993*. London: Taylor & Francis.
- MacLeod, I. S., & Taylor, R. M. (1994). Does human cognition allow human factors (HF) certification of advanced aircrew systems? In J. A. Wise, V. D. Hopkin, & D. J. Garland, (Eds.), *Human Factors Certification of Advanced Aviation Technologies*. Daytona Beach: Embry-Riddle Aeronautical University Press.
- Metzler, T. R., & Lewis, H. V. (1989, June). *Making MANPRINT count in the acquisition process* (Army Research Institute Note 89-37). U.S. Army Research Institute.
- NATO. *The application of human engineering to advanced aircrew systems* (STANAG 3994 AI).
- Patton, M. Q. (1978). *Utilization-focused evaluation*. Beverley Hills: Sage.
- Sherwood-Jones, B. (1987). Human-factors audits and fitness for purpose. *Proceedings of the CAP Scientific Conference*.
- Stiles, L., & Hamilton, B. E. (1987). Cognitive engineering applied to new cockpit designs. *Proceedings of the American Helicopter Society National Specialists Meeting: Rotorcraft Flight Controls and Avionics*. Cherry Hill, PA.

- Taylor, R. M. (1987). *Some thoughts on the future of engineering psychology in Defense*. Position Paper for the British Psychological Society Conference on the Future of the Psychological Sciences, Harrogate.
- Taylor, R. M. (1990). *Merlin MPC Workload Acceptance Criteria* (IAM Letter Report 016/90). RAF Institute of Aviation Medicine.
- Taylor, R. M. (1991). Human operator capability analysis for aircrew systems design. *Proceedings of a panel session at the British Psychological Society 1991 Occupational Psychology Conference* RAF Institute of Aviation Medicine Letter Report No. 004/91. RAF Institute of Aviation Medicine.
- Taylor, R. M., & Selcon, S. J. (1993). Operator and automation capability analysis: Picking the right team. *Combat Automation for Aircraft Weapon Systems: Man/Machine Interface Trends and Technologies*. Neuilly Sur Seine: NATO AGARD CP 520.
- U.K. Ministry of Defense. (1989). *Human factors for designers of equipment* (DEF-STAN-00-25).
- U.K. Ministry of Defense. *Design and airworthiness requirements for service aircraft* (DEF-STAN-00-970).
- U.K. Ministry of Defense. *Guide to quality assurance in design* (DEF-STAN-05-67).
- U.S. Department of Defense. (1987). *Human engineering procedures guide* (DOD-HDBK-763).
- U.S. Department of Defense. *Human engineering design criteria for military systems, equipment and facilities* (MIL-STD-1472).
- U.S. Department of Defense. *Human engineering requirements for military systems, equipment and facilities* (MIL-H-46855).
- U.S. General Accounting Office. (1981). *Guidelines for assessing whether human factors were considered in the weapon system acquisition process* (GAO FPCD-82-5).