CHAPTER 8

Human System Measurements and Trade-offs in System Design

MICHAEL BARNES and DAVID BEEVIS

8.1 INTRODUCTION

The aim of human systems integration (HSI) for military systems is to ensure that "human considerations . . . shall be effectively integrated into the design effort for defense systems to improve total system performance and reduce costs of ownership" [U.S. Department of Defense (DoD), 1991a]. The purpose of this chapter is to discuss human performance measurement issues related to system design, particularly those issues that require trade-off decisions. The basic approach is to measure performance in terms of system goals and to choose measurement processes that reflect the context of the environment the system is being designed to operate within. This is similar to *use-centered* and *ecological interface design* philosophies that are currently being investigated in the cognitive engineering domain (Flach et al., 1998). That is to say, measurement is not concerned with human limitations per se or even environmental or technological problems but rather how these problems in concert affect the accomplishment of overall system goals (cf. interface design issues; Rasmussen and Vicente, 1989). The measurement problem is difficult because of the complexity of the design space, which includes not only volatile operational environments but also changing technological and human requirements as well. The solution is to develop a flexible measurement strategy that addresses system goals and top-level requirements while it adjusts to the various contingencies of the evolving design process. In effect, a successful measurement paradigm allows the design team to answer the "show me the payoff" question implicit in all design decisions.

8.2 HUMAN SYSTEM MEASUREMENT

Human system measurement processes consist of assigning numbers to events that have important design implications. The ability of the process to predict the effects of design

Handbook of Human Systems Integration, Edited by Harold R. Booher.
ISBN 0-471-02053-2

outcomes crucial to operational objectives determines the value of a particular measurement procedure. By their very nature, HSI issues that pertain to complex systems have multiple operational requirements as well as future operating environments that are not well understood much less well defined. Because of this complexity, there is no unique measurement solution (Barnes et al., 1996). Various approaches, from traditional human engineering analyses to laboratory experiments to large-scale computer simulations, each having its own advantages and limitations, are required to measure those aspects of system performance impacting HSI during various phases of the design process. The criteria for choosing a specific approach depend on the operational context in which the system will be used. This includes its ecological validity, ease of use, cost, and insight it brings to particular design or crew issues in relation to functional requirements (Flach et al., 1998; Meister, 1986). In evaluating military systems, an important distinction that underlies all assessment is that between measures of performance (MOPs) and measures of effectiveness (MOEs). Starting with this distinction and adapting measurement paradigms developed by Meister (1985) and Erickson (1984), a general measurement strategy will be described below that allows the analyst to consider the ecological context and inherent complexity of future military systems while maintaining an appropriately eclectic approach for each design phase. Three topics of human system measurement help provide context for the rest of the chapter:

- human measures of performance,
- HSI measures of effectiveness, and
- human systems measurement problems.

8.2.1 Human Measures of Performance

The basic human-related MOPs are time to perform a task and the accuracy with which the task is completed, or its complement, the level of human error. The latter variable may be expressed in other terms, such as the probability of detecting a target. Most often, human-related MOPs are collected in laboratory experiments. The drawback to using such measures is that test results may have little impact because their effect on overall system performance is unknown. For example, Erickson (1984) argues that if the measurement process focuses on performance issues rather than overall systems effectiveness, the HSI design process is in danger of becoming marginalized. This is because the focus on local performance issues often fails to show a link between a proposed design change and overall system requirements. Indicating an operator's ability to resolve resolution lines on a display or measuring his or her comfort level may have little impact on the design process. The crucial question is not whether the operator is comfortable or the display is optimal but rather what effect these conditions have on system goals. In particular, what is the system cost of a poor interface or an uncomfortable seat in terms of overall effectiveness?

Human performance measures must be interpreted in the context of the operational situation, the state of the individual performing the task, and the implications of the measured performance on systems effectiveness (Charlton, 1996). Thus, human performance measures must be translated into what used to be called *systems-relevant criteria* but now more commonly are called *measures of effectiveness*. The MOEs permit the comparison of alternative systems in terms of functional objectives and mission needs.

Examples of MOEs include loss exchange results, systems saved, and tons delivered per day (DoD, 1991b). Chapanis (1960) provides early examples of the differences between human factors MOPs and MOEs. For example, he refers to a case where data on the human detection of a radar target (MOP) were translated into the detection range of the radar (MOE). Another example is translating measurements of human performance in an air traffic control task (MOP) into criteria such as the amount of fuel consumed by aircraft, the length of time the aircraft would be under control, and the separations that would be maintained between aircraft (MOE).

8.2.2 HSI Measures of Effectiveness

The corollary of the kinds of translations described by Chapanis (1960) is that the measures of human performance must be compatible with the system of measurement used to express MOEs. Erickson (1984) argues that system component and operator performance requirements are not explicit in the upper levels of any system analysis that is conducted early in concept development. There may be no direct relationship between operator task performance and system performance criteria unless the connection is made explicit by analysis. This becomes increasingly difficult as systems become more automated. Erickson describes an approach to developing a "capability hierarchy" starting with a functional analysis and decomposing the performance requirements from that level. He notes that it is necessary to go down at least two levels in the hierarchy before operator performance criteria become apparent (see Fig. 8.1).

Figure 8.1 depicts a useful paradigm for measuring the impact of design factors on system effectiveness. Erickson's paradigm captures the spirit of much of the above discussion. The process starts with mission goals (block 1) and uses a step-by-step process that encompasses a variety of measurement methods to build a utility model of the developing system within its intended military environment. Implicit in Erickson's discussion is the importance of the derivation of various combat costs and benefits that will be used to create MOEs within this environment. After determining mission goals, the analyst defines the system and develops the MOEs in relation to how a system interacts within the total operational environment. Descriptive modeling and task-analytic approaches are used to define the system and begin to understand the role of the crew within the operational milieu. Naikar and Sanderson (2000) have extended Erickson's approach by using work domain analysis (WDA), which is derived from an analysis of system goals and provides a framework for (1) evaluating the implications of detailed technical proposals for system overall functionality and (2) aggregating the implications of the lower level MOPs.

It is important that blocks 2, 3, and 4 be done as a team effort. The MOEs must be specified, and the critical system factors must be identified. The process of generating operational and systems data requires a team approach because no one engineering discipline or operational expert can understand all the ramifications of a complex system. When these steps are completed, we can then move to blocks 5, 6, and 7 to complete the evaluation of system effectiveness. Of particular note is the use of human performance data as inputs into constructing the framework for the modeling environment and human performance experimentation as an integral part of exercising the model. Erickson does not argue that human performance data are unimportant, only that data must be understood within its full operational context (i.e., or more generally its ecological context; Flach et al., 1998). Realistic simulation methods are the most likely confirmation of an effectiveness

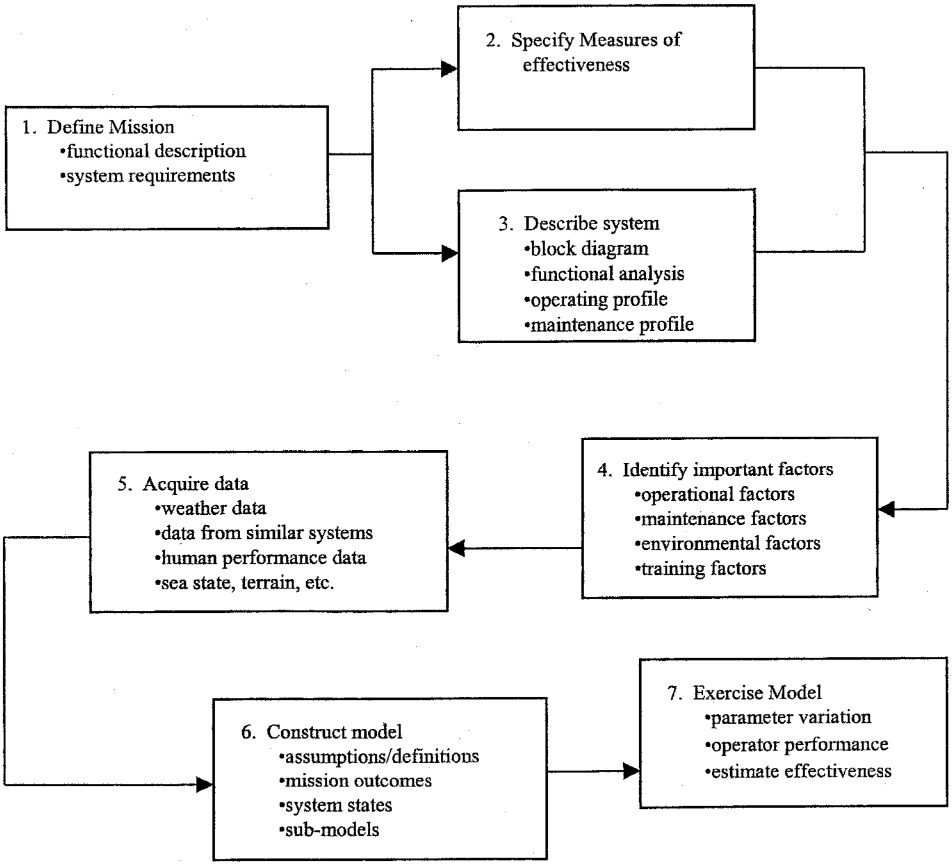

Figure 8.1 Principal activities required to evaluate system effectiveness (after Erickson, 1984).

model given the current state of the art. The multistep, carefully integrated measurement process was considered necessary by Erickson because military environments are too complex to be represented by simple performance or analytical methods isolated from their effect on the total system.

8.2.3 Human Systems Measurement Problems

The principal drawback to systematic approaches such as Erickson's (1984) paradigm is the implication that we must know a great deal about the system before we can start the HSI process. The opposite is true; understanding the human role should proceed in parallel with the engineering design. In most military systems, the human role is paramount either in execution or decision making, and this role must be defined at least loosely before a working concept can be developed. Moreover, the more accurate the conceptualization of the human role is within the design process, the more likely the system is to meet milestones and cost goals. In attempting to understand the human role during the concept stage, the same basic approach can be used—certain overarching goals must be developed and human performance issues must be identified in terms of these goals. Methods must be used to try to measure or perhaps more realistically approximate the impact of the proposed human component on the nascent system even before detailed engineering specifications are available. Understanding the human role before knowing the exact parameters of the proposed system has always been a difficult problem; however, the advent of more cost-effective modeling techniques and simulation paradigms should make this problem more tractable.

In most cases, predecessor systems exist, but rapid changes in technology make direct comparisons of human performance issues for a new generation of systems difficult. For example, researchers at Fort Huachuca investigated generic crew roles for a family of unmanned aerial vehicles (UAVs) that were in various stages of conceptual development (Barnes et al., 2000). The training and doctrine system manager (TSM) was concerned with the skill level necessary to operate these new UAVs because this variable would influence all future UAV design decisions. Two flight operators were evaluated: the external pilot (EP) who flies the UAV for take-off and landing using a radio-controlled hand device and the air vehicle operator (AVO) in the ground shelter who flies the UAV at all other times. The crucial question was whether the flight crew needed to be flight certified. When a variety of test instruments and interviews with over 70 subject matter experts were used to answer the question, it was found that most of the flight safety problems took place during take-off and landing. As an example, Figure 8.2 shows that most of the skill-loading problems related to safety issues for the EP positions were the result of differences between experienced and inexperienced operators. In effect, experienced EPs used conceptualization skills about the qualities and characteristics of the air vehicle to anticipate problems for take-off and landing whereas the inexperienced operator used perceptual and psychomotor skills associated with direct control of the system to react to problems. This suggested that the problem was most likely a training issue. Accident reports and training performance data supported this hypothesis. The result of the analyses suggested a different training strategy rather than flight certification as the answer to improving EP performance. An indirect result was a greater emphasis on developing safe automatic take-off and landing technologies to circumvent this problem altogether.

The serial nature of the data collection effort in Figure 8.2 reflects the linear, compartmentalized design philosophy of the 1980s, which was both time consuming and cumbersome. Improved simulation methods and a determined effort by the military to infuse experimentation and concept exploration earlier in the design process have resulted in more of a spiraling concurrent engineering paradigm. Early approximations of the system are modeled and tested and even field tested before any mature design concept exists. This approach is highly iterative; the paradigm is closed loop with iterations often starting at the “define mission” stage because the design process is also being used to define mission elements. This process is still new, but the establishment of battlefield laboratories by all U.S. services, extensive use of warfighter experiments to validate new concepts, and an emphasis on using modeling and simulation during the design process all point to an evolving design philosophy. This approach, with emphasis on modeling and simulation tools, requires a more adaptable measurement paradigm. The same Erickson processes still occur but in parallel. Since a single model or simulation that is in any sense complete is usually not available early in the design process, the HSI practitioner must rely on a combination of modeling, human experimentation, and simulation for each stage of development. Erickson’s approach still provides a valuable measurement framework. The same general philosophy is used in current programs, but the measurement process takes place in a more dynamic and iterative environment.

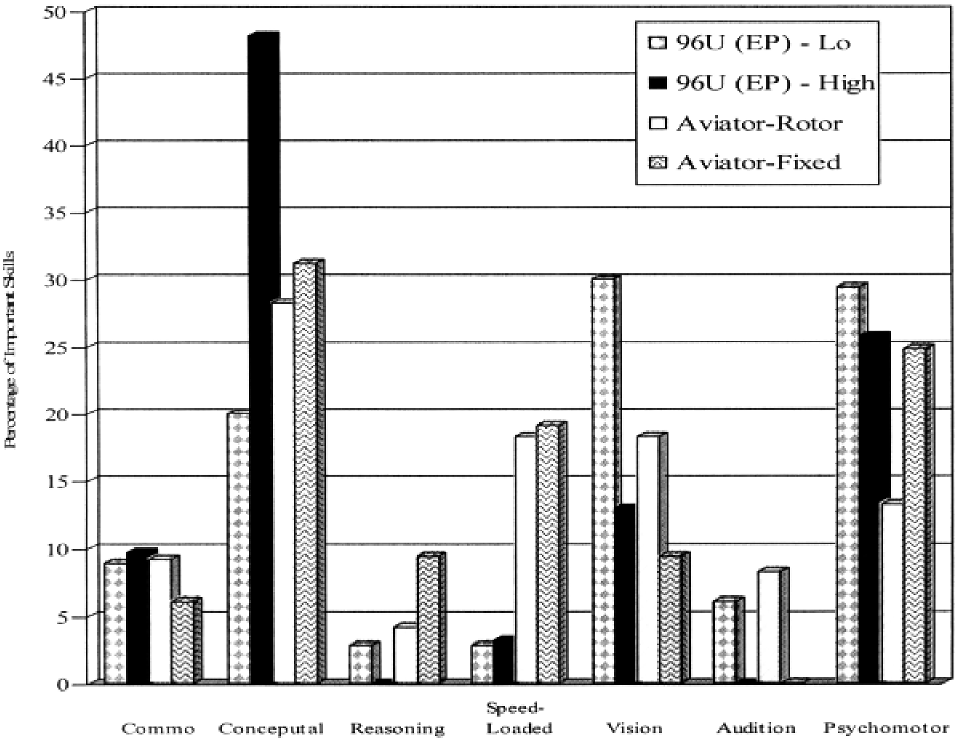

Figure 8.2 Percentage of important skills used during emergencies, shown by external pilots (EPs) with high and low experience levels. Aviator information is shown for comparison.

8.3 GENERAL MEASUREMENT MODEL FOR HSI

The measurement process should be motivated by a cost–benefit paradigm. Measurement procedures and scales need to be chosen carefully to mirror important systems requirements and costs (Meister, 1985). All successful strategies have the same general components: some way of defining the system and its general impact on the stated requirements and methods for evaluating its impact to ensure the end products meet system goals (Gould, 1988; Meister, 1985; Whiteside et al., 1988). Data collection strategies start with system objectives (see Fig. 8.3), which lead to the definition of top-level system requirements and top-level definition of MOEs. From the bottom up there are three basic measurement protocols:

- analytical methods and models,
- human performance experiment, and
- realistic validation methods.

Each of these measurement methods is useful in helping to understand the impact of design decisions at various stages of the process. The MOPs developed from this combination of protocols feed into the MOEs that help determine critical design decisions.

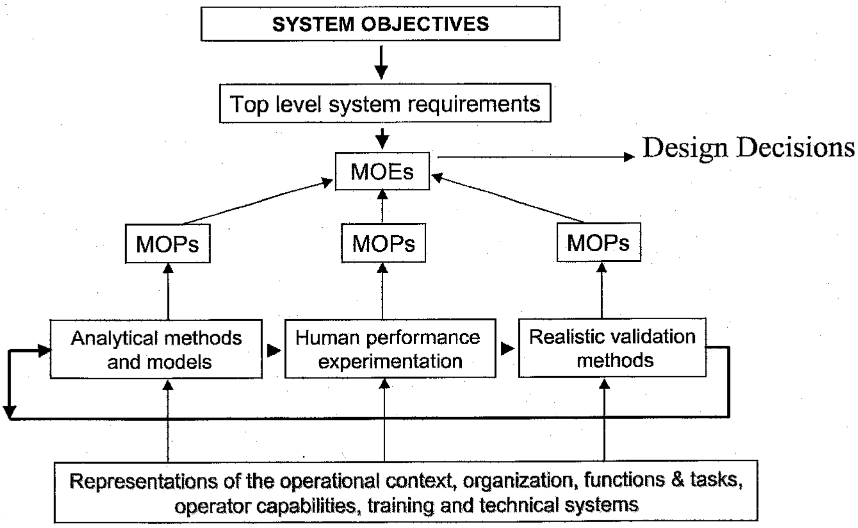

Figure 8.3 General measurement model for HSI.

The importance of using these three classes of measurement is that they perform separate functions related to assessing system performance during the design process. The descriptive models and analytical techniques give the designer an understanding of system dynamics and MOEs and permit analysis of concepts before any actual design decisions are made. Human performance experimentation is the mechanism for evaluating specific design decisions or trade-offs important enough to be evaluated early to midway through the design process. Also, well-controlled experiments can be used to verify or update early modeling efforts and to understand the cognitive and performance issues evinced by the operational environment. The realistic simulations and exercises are usually the first time the design prototypes are exposed to complex operational environments. These exercises can be part of the formal testing cycle, but the exercises are usually done early enough in design to be sensitive to important operational and engineering issues. The purpose is not only to validate the design but also to allow the engineers to evaluate design options and trade-off decisions in their intended environment. Validation methods differ from human performance experimentation because of the greater emphasis on realism and the extensive use of closed-loop exercises to measure performance. The price of attaining realism is usually a decrease in experimental control and an increase in cost.

The various techniques provide different types of measurement scales, from nominal scales to ratio scales. (Nominal scales identify distinct entities, e.g., A not B; ordinal scales express differences as greater or less without reference to units; interval scales use equal intervals of measurement, such as degrees Fahrenheit, but without defining the end points; and ratio scales can be treated by all mathematical manipulations because they have defined zero points, such as a measurement of length or speed.) In the initial requirements definition and concept development stages, some MOEs may be clear and expressed on a ratio scale such as system range or payload, whereas others and human MOPs may be expressed on an ordinal scale of measurement. Since the various HSI techniques are applicable at different points in the acquisition design and development cycle, the scale of measurement that can be used changes throughout the system design and development process. For example, mission analyses, which are used to define the requirements for new systems, describe events and sequences of events that are nominal or ordinal data, yet mission analyses also include distances, times, and probabilities, which are ratio-scale data. In contrast, functional analyses, which are derived from mission analyses, provide descriptions of the functions to be performed but may not include performance requirements (Beevis, 1999).

To guide the HSI specialist in developing MOPs, the more generic HSI techniques, along with their measurement parameters and their relationships to system performance, are given in Table 8.1. The techniques are classified by domains (manpower, personnel, etc.) as described in Chapter 1. Table 8.1 shows some domain-specific considerations when measuring the human impact on complex systems.

The key to measurement is to match a specific technique to a particular problem during various phases of system development. The use of several of the above methods in concert is critical in the design process because there is no single research method that is flexible enough, cost effective, well controlled, and sufficiently realistic to address the multitude of problems associated with complex systems (Wickens and Hollands, 2000). In selecting a method, users should remember that many techniques are complementary and should be selected for their contribution to subsequent stages of measurement. For example, field observations, function and task analysis, and experimentation can be used successfully to define the requirements for complex human-in-the-loop simulation of advanced defense systems (Greenley et al., 1998). The literature suggests general guidelines for choosing methods based on past research and engineering experience (Barnes et al., 2000; Gould, 1988; Knapp, 1996; Meister, 1985, 1987):

- Evaluation early in the process is necessary because fixed designs are nearly impossible to change. Inexpensive analytic techniques (especially in the early phases) can uncover many HSI problems before they impact the design process.
- Evaluation must be conducted in a system context.
- Evaluation is iterative; results should verify previous analyses as well as suggest future exercises.
- Well-controlled simple procedures should always precede complex evaluation methodologies.

Early evaluation is important. For example, identifying the personnel and training requirements is crucial to early design decisions because the personnel skill mix necessary for a system is not only an important life-cycle cost but also an issue that affects the military organization as a whole. The supply of the necessary operators and maintainers has to be initiated years ahead of the delivery date for a new system. Therefore, it is important to identify and evaluate design features that affect such issues early in the project. Inexpensive analytic techniques often have a very high payoff if used early. To ensure that these methods do impact the design, they need to be tailored to specific design issues early enough in the process to be useful to the designer. Also, the HSI effort must be tightly coupled with the other engineering efforts. For example, operator workload analysis can be very useful as a design tool provided the results are applied to design problems such as crew task allocation or automation decisions (Barnes et al., 2000; Beevis and Essens, 1997) and the HSI team works closely with other design engineers to understand the

TABLE 8.1 HSI Measurement Techniques and Their Relationship to System Performance

Technique	Measurement Parameters	Relationship to System Performance
	Manpower Domain	
Analysis and modeling—parametric estimates based on displacement, weight, power, etc.	Numbers of personnel	Assumes acceptable system performance and thus availability; may not reflect performance of future systems
Analysis by analogy (e.g., early comparability analysis)	Staffing estimates based on similar systems	Assumes acceptable system performance; may not reflect performance of future systems
Computer simulation of operator tasks	Numbers of personnel required at different points through a mission	Estimates can be related to system availability in normal and degraded modes of operation or maintenance.
	Personnel Domain	
Analysis by analogy (e.g., early comparability analysis)	Inventories of aptitudes, skills, and experience levels required and special physical requirements based on similar subsystems	Assumes acceptable performance based on skills required with existing systems
Analysis of skills	Inventories of aptitudes, skills, and experience levels and special physical requirements	Assumes acceptable performance based on analysis of categories of skills required
Computer simulation (e.g., FOOTPRINT)	Data on manpower, personnel, and training characteristics of each military occupation speciality.	Assumes acceptable performance based on analysis of skills required and available
Analysis and modeling—personnel cost estimates	Costs of operational, maintenance, and training personnel	Can be used as input to cost trade-offs for given level of system performance
	Training Domain	
Analysis by analogy with existing systems	Estimates of training system requirements based on existing systems	Defines training performance goals to be met for system to be effective
Analysis—training needs	Estimates of training system requirements	Defines performance goals for operator tasks required for system to be effective

(*continued*)

TABLE 8.1 (*Continued*)

Technique	Measurement Parameters	Relationship to System Performance
Computer simulation (IMPRINT)	Estimates of system performance as a function of skill levels	System performance is simulated as a function of workload and operator skills and task performance.
	Health Hazards Domain	
Analysis—health hazards assessment	Exposure times for personnel for health hazards	System effectiveness may be constrained by exposure times, e.g., limits to load carriage and number of rounds fired by a weapon.
	System Safety Domain	
Analysis—risk assessment	Probabilities of injury or system failure	System effectiveness can be predicted as a function of personnel and subsystem availability as predicted by failure analysis.
	Human Factors Engineering (HFE) Domain	
Analysis—missions and scenarios	Sequences of events, distances, ranges, times and environmental conditions	Scenarios for HFE should be developed from the scenarios in the mission needs statement and the operational requirements document.
Functional analysis of system[a]	Sequences and flows of functions required to perform system missions	Functional analyses define system and subsystem goals. System effectiveness is implicit in meeting those goals. Can be used to identify MOEs as suggested by Erickson (1984).
Task analysis[a]	Sequences and times of operator tasks, task tolerances and performance requirements and operator interface design requirements	Analyses reflect system performance requirements and establish criteria for experiments, rapid prototype evaluations, and field trial performance.

TABLE 8.1 (*Continued*)

Technique	Measurement Parameters	Relationship to System Performance
Cognitive task analysis	Diagrams showing the relationship of system goals to cognitive processes (e.g., information processing) and to lower level cognitive and physical tasks	Analyses reflect goals and knowledge-oriented behavior rather than data-oriented and interface manipulation behavior (Rasmussen, 1986). Emphasis is on top-down analysis.
Task network models and simulations	Times and probabilities of completing sequences of tasks	Simulation outputs must be translated to relate them to system effectiveness. Can produce estimates of workload and some MOEs and confirm assumptions about manning and personnel issues.
Laboratory experiments	Times, errors, measures of comprehension, retention of information, etc.	Results must be translated to be relevant to system effectiveness.
Rapid prototyping of operator: machine interface	Times, errors, measures of comprehension, and ease of use	Results must be translated to be relevant to system effectiveness.
Complex human-in-the-loop simulation	Performance of specific tasks, times, errors, measures of comprehension, retention of information, and ease of use	System MOEs can be measured directly if the simulation fidelity permits. Human MOPs must be translated into system MOEs.
Complex field trials and warfighter exercises	Times, errors, measures of comprehension, retention of information, and ease of use	MOEs can be collected directly, as well as MOPs.

[a] A more detailed review of the links between HFE analyses and system performance is provided in Beevis, 1999.

various design and crew trade-offs necessary to reduce crew workload before a mature design is in place.

System context is vital—the measurement scale must reflect important system dynamics. For example, Knapp (1996) points out that for command and control (C2) systems the measure of interest is the information flow. Specifically, she felt it was important to measure the impact of operator workload and the resulting message flow in terms of its effects on command decision making and execution. Knapp was able to model the information flow for a number of C2 systems showing the costs and benefits of various design options for proposed future systems. Her measures were closely related to important system parameters. If she had simply reported her results in terms of system latency or overall workload, it is doubtful her results would have been as well received.

Knapp also delineated these effects in terms of the cognitive skills necessary to perform a particular task giving the design team additional insight into the quality of personnel needed for the various positions as different automation and design options were considered (Muckler et al., 1991).

Evaluation must be iterative because the overall design process is one of successive approximations. Once the design options are set in concrete, HSI analysis tends to be ignored. The good news is that the variety and usefulness of HSI tool sets have increased remarkably in the last 10 years, especially in their ability to evaluate iterative design options. Human systems integration modeling packages such as improved performance research integration (IMPRINT) tools allow multiple analyses using the same software environment. The chief advantage of these new modeling environments is that as more data are collected and new design options are proposed, it is relatively easy to revise the original model and rerun the analyses.

Well-controlled simple procedures can reduce the amount of effort required to investigate HSI issues. A simple analysis or experiment can uncover obvious flaws and permit the analyst to direct future efforts toward investigating the more subtle and complex trade-off issues. Simply requiring extensive HSI analysis or modeling does not guarantee that these tools or their results will be used in the design process. But well-controlled procedures can negate requirements for further complex techniques that add little to the design process.

8.4 ANALYTICAL AND MODELING TECHNIQUES EARLY IN DESIGN PROCESS

Most activities in the early stages of system design and development involve the analysis and synthesis of a design solution. Defining system components, operations, and crew issues is the initial step of any analytical effort for a new system. There are a number of traditional methods and data sources that will make the analyst's job easier. Various documents that are part of the design process such as mission needs statements, operational requirements, etc., are good starting points for the analyst to begin to understand the purpose and intended uses of the system early in the development cycle.

In many of the military programs, there are early exercises and "rock drills" (walk-through simulations of various doctrinal concepts) that define the doctrine being developed to counter future-threat profiles. Being part of these exercises is extremely useful because the exercises emphasize and elaborate the operational issues that the system is being designed to address. The importance of understanding the military purpose and intended environment before any analysis is attempted cannot be overemphasized.

Analytical techniques that are part of the traditional system engineering approach include functional-flow diagrams (Meister, 1985) and requirements modeling (Hatley and Pirbhai, 1987). These approaches are used during the design process to model the system and its various components and their interrelationships. These methods are quite helpful in developing a "blueprint" of the overall system and as such provide a good reference point for HSI analyses. However, if the models are developed solely for engineering guidance there are a number of drawbacks:

- These models often do not exist in the early conceptual stages and yet important HSI issues need to be addressed before a well-defined design is developed.

- System component and operator performance requirements are not explicit in the upper levels of any system analyses that are conducted early in concept development (Erickson, 1984).
- Often the human role in the system is not well defined in these representations (Beevis, 1987).
- Functional-flow diagrams that focus on engineering issues can unintentionally foster a myopic view of the system, especially in terms of how the system interacts as part of the total combat system.

After understanding whatever engineering documentation exists at that point (including HSI documents), the next step is to choose a method to represent the human component as part of the system. The purpose of these analytical and modeling methods is to understand and predict the impact of various important crew functions on early design concepts when there is little or no performance data for the new system. Choosing the correct method and measurement scale depends on the design issue.

8.4.1 HSI Analysis Techniques

Human factors engineering technology includes a suite of analysis techniques that are compatible with a number of systems engineering analysis methods. The generic forms of analysis correspond to those recommended in US-MIL-HDBK 46855 and include mission and scenario analysis, function analysis, function allocation, task analysis, performance prediction, and interface and workspace design (Beevis, 1999). These analyses can make a major contribution to the implementation of human factors in a project, particularly if coupled with other techniques such as experimentation, rapid prototyping, and human-in-the-loop simulation. For example, in the development of the F-18 aircraft, mission analyses were used to identify the likely use of aircraft systems, their operational modes, pilot tasks, and control and display requirements to establish an overall concept for the operator–machine interface. The results of the analytical effort were then refined and validated in a very extensive manned simulation of the aircraft (Merriman and Moore, 1984).

As described earlier, the various analytical techniques provide a range of predictive performance measures, some of which can be related to MOEs and some of which require additional analysis or quantification through experimentation, simulation, or trials. Function-flow diagrams provide the basis for developing performance specifications related to each system function as recommended by Erickson (1984) and for developing requirements specifications for the operator–machine interface. Because they describe the functions that must be performed by a system without reference to hardware, software, or humans, the early stages of human factors engineering (HFE) analysis can be reused and updated as technology improves.

Early analysis can reduce the potential range of design solutions to a point where options can be evaluated through user trials or experimentation. Early analysis can also identify areas where performance may be a significant factor that requires confirmation by experimentation, modeling, or simulation. For example, in the development of a targeting device for a shoulder-launched, ground-to-air missile, 18 possible combinations of display devices and formats for displaying gross and fine azimuth to the operator (a design option decision tree) were analyzed to identify the advantages, disadvantages, and performance implications of each one (see Fig. 8.4). From the analysis the two most

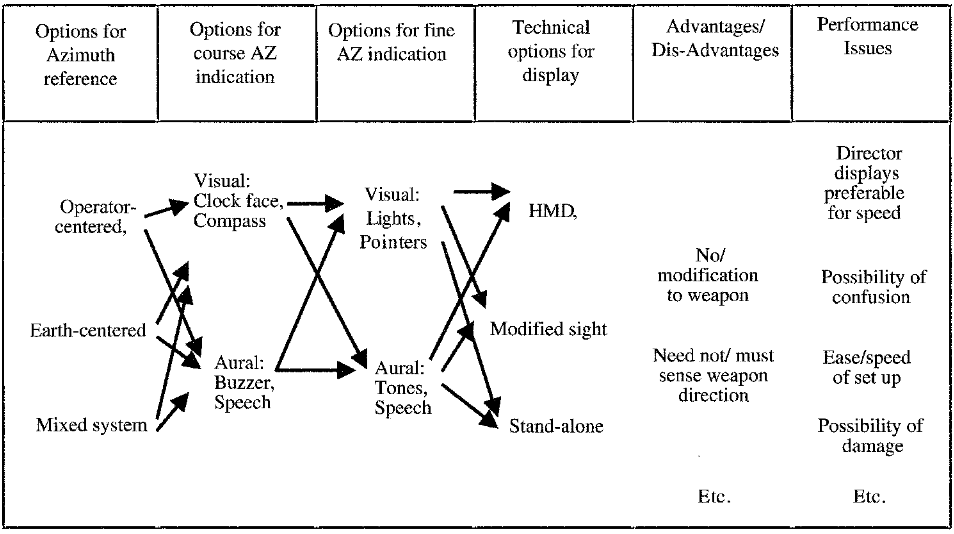

Figure 8.4[1] Design options decision tree used to identify most promising candidates for a display.

promising concepts were selected for a simple simulation experiment that confirmed the performance characteristics predicted by the analysis.

8.5 HUMAN PERFORMANCE EXPERIMENTATION

Measures of human performance are components of all the measurement techniques under discussion. The use of traditional experimental methods (Kirk, 1982) is an important source of obtaining these data, and these methods allow for a more precise understanding of the processes that affect system design. The problem with many of the more subjective methods and realistic simulations and field exercises is the lack of experimental control (McBurney, 1998). The causal relationships among variables may be impossible to untangle during a large field exercise, and many of the analytic techniques lack rigorous statistical verification. Experimental methods can be used both for parameter estimation and hypothesis testing. The problem with these methods is that the control conditions inevitably introduce a degree of artificiality into the measurement process. This is especially true because cost and pragmatic considerations limit the number of factors and ambient conditions that can be considered. For example, a 10-factor between-subject design to ensure no sequence effects would require a minimum of 2048 subjects if each factor had two levels to use conventional analysis-of-variance techniques.

More flexible experimental approaches are available (Williges, 1995) and are discussed below; however, in general, the tighter the experimental control, the more likely constraints are to be placed on realism. This is not an argument against experimental control; rather, it is a reminder that the true power of a good experimental approach is in hypothesis testing and not in measuring real-world processes.

The measurement model presented here is predicated on the use of initial modeling and realistic validation procedures being used in concert with true experiments. The role of the controlled experiment should be dictated by the initial analysis, which should point to the critical design decisions that need to be considered: the important crew-related factors that need to be investigated and the MOEs that need to be addressed. A good experiment is similar to a laser; it covers a small area of the response surface (the surface defined by the multiple regression equation that describes the relationship among all experimental variables), but it is very effective if it is directed toward the most critical factors. As any good experimentalist knows, it is the relevance of the questions asked and the experimental procedures used that determine the value of an experiment.

8.5.1 Person-in-the-Loop Simulation Experiments

Small-scale elegant experiments are rare during the design process because few of the problems being investigated for complex military systems lend themselves to such a paradigm. An example might be investigating a number of control-display options in the laboratory before narrowing the field to a few that can be evaluated during a more realistic simulation exercise. It is cost effective to eliminate deficiencies early in the design cycle using simple experimental methods wherever possible.

The initial experimentation involving either part-task simulations or more limited scenarios should be designed to optimize experimental control. These exercises are a bridge between the laboratory and more realistic simulations and field testing that follow. For example, a series of simulation experiments were performed at Redstone Arsenal to investigate fatigue and equipment factors for a variety of UAV options requiring day and night crews to rotate duty cycles every 12 of 24 hours over a 72-hour operational tempo (Barnes and Matz, 1998). The investigators were able to control most experimental factors and ensure that all participants received the same target sets, rest conditions, etc., to the point of controlling their rest periods when they were not on 12-hour shifts. The ground control station, flight parameters of the simulator, mission taskings, and target sets were all realistic. Important design requirements related to manning and console design were addressed by using measures of target acquisition and safety. The results were useful in that they indicated serious problems with single crew configurations and suggested crew fatigue problems related to circadian rhythms. Although the exercise was well controlled and realistic, it could not capture the actual stress, unpredictability, and interrupted sleep patterns of combat. Again, important compromises between experimental control and realism had to be made by the data collectors.

8.5.2 Experimental Designs for Complex Spaces

A number of experimental approaches such as response surface methodologies, confounded designs, and quasi-experimental methods have been developed specifically to measure complex environments (Box et al., 1978; Cook and Campbell, 1979). Response surface and confounded designs allow the investigator to measure a restricted portion of the response surface to estimate behaviors over the entire response surface.

A simple example is an experiment designed to model the effect of display size on 10 operational and sensor factors for a navy attack aircraft. The initial investigation used a screening technique (supersaturated fractional factorial design; Barnes, 1978) that identified four experimental factors and their second-order interactions as accounting for 50 percent of the variance in a target acquisition task. These four factors were used to create an orthogonal regression equation using conventional experimental designs. The fractional design permitted the investigation of a large response surface with few subjects and experimental conditions, and the results were used to investigate the most crucial factors in an unconfounded data space.

These designs are not panaceas; their utility depends on the particular measurement problem addressed. The measurement of complexity often results in the loss of experimental control, but this loss is worth the price if the response surface is better understood and the results lead to more predictive models or better experimental precision using so-called pure experimental designs.

The following sections describe a variety of simulation and virtual methods that are capable of reproducing much of the realism of field exercises and actual operational conditions. Complete control is impossible if for no other reason than the closed-loop nature of the real world and the bewildering array of variables involved. Under these conditions, the use of quasi-experimental methods and statistical controls is as important as the use of control techniques in classical experimental paradigms. The more sources of variation that exist, the more crucial it is to control as many of them as possible.

8.6 MODELING AND SIMULATION

A model is a representation of critical aspects of objects or situations. In the context of HSI, modeling refers to mathematical models of human performance and crew workload. Simulation is a method for implementing a model over time. Simulations can be “constructive” based on mathematical or parametric models: they can be “virtual” simulations in which operators use representations or rapid prototypes of systems and interfaces or they can be live simulations with users conducting trials with actual equipment. One advantage of the use of modeling and simulation is that work done for project planning or concept development can be carried over and reused and exploited in project development, project definition, and implementation. This permits MOPs and MOEs to be refined throughout the system’s development process.

Operations analysis (OA) makes extensive use of models and simulation. However , it has proven very difficult to establish a link between the OA modeling activities and HSI modeling, despite the Military Operations Research Society having held several conferences on the subject. One reason for this is that combat models tend to focus on the outcome of engagements, whereas human factors models focus on performance of specific tasks. McMillan and Martin (1994) suggested that the human factors models can be used off-line to generate statistical distributions and performance shaping functions for use in OA models. It is not clear how the output of OA models can be used to focus HSI efforts without going through the kind of decomposition of performance requirements that is recommended by Erickson (1984).

Compared with other engineering disciplines, a review of the human factors and human engineering literature reveals a limited amount of human performance data to support modeling and constructive simulation. While it is true that much more effort is required in this area, a variety of models of human performance are available (McMillan et al., 1991). Some of these models are parametric, but several have been developed from first principles. Many of these models can be expressed as task network models for use in systems design and development. Task network modeling (TNM) represents the complex pattern of operator tasks as an interlinked network of simple human performance task models. One such task model reported by Card et al. (1986) is an information-processing representation of the human operator. The model, comprising perceptual, cognitive, and motor systems, includes times for a variety of types of information processing. Despite their seeming simplicity, such models can represent quite complex applications. For example, using an information-processing model and general principles of human performance, including assumptions of single-channel processing and task completion times estimated from the literature, a network model was developed for the tasks associated with air traffic control. Run as a constructive simulation, the model produced results for most performance parameters that were close to experimental observations of human performance obtained in a manned simulation of the same tasks (Burbank, 1994).

Probably the most adaptable approach to modeling human factors aspects of systems, TNM produces descriptive models of the human tasks and interrelationships in systems, and such models can be used as the basis for simulations. Such simulations can address a wide range of problems and are probably the most common application of TNM. They can be applied to very simple multitask models or to complex multioperator systems. The validity of such models is sometimes questioned. Because they contain less than complete detail, all models have limited validity, but many models can be constructed that are useful in the design and development process. The key to successful TNM simulations is the level of detail of the analysis on which the model is based. Function-flow diagrams must be decomposed to at least the fourth level because cases of feedback and coupling between operator-performed functions are seldom identified at higher levels of analysis. This contrasts with some of the models used to predict manpower requirements, which use information generated by the second or third level of functional decomposition.

8.6.1 Deterministic, Stochastic, and Hybrid Models and Simulations

Many task network models are deterministic because they are based on a scripted input (mission analyses and scenarios) and a predetermined sequence of tasks. If human performance terms such as times and probabilities of completing each task are added, task network models can be used to run constructive simulations of operator tasks. Such simulations are hybrids; some aspects of the simulation are stochastic, but the inputs and sequences of tasks are predetermined.

Implemented as a Monte Carlo simulation, where a random-number technique draws values for task completion times and probabilities of task completion, the simulation becomes stochastic. Further elaboration of such models can make the inputs subject to variation, for example, by using probabilistic mission events. The model of the operator's response to such inputs or factors such as task load can also be made probabilistic. Stochastic simulations avoid the criticism that operator performance is not deterministic and cannot be properly represented by models; however, they require many replications, typically at least 100, to collect the necessary distribution of outcomes to properly represent the operation of a proposed system. In fact, when simulating complex enterprises such as command and control systems, it is important to examine the distribution of possible outcomes as the system is simulated (RSG.19, 1999) so many replications are required. This is not a significant problem for computer-based constructive simulations; however, the need for many replications is a limitation for complex human-in-the-loop simulations. Establishing clear start and end points for complex sequences of probabilistic events is another difficulty and can preclude drawing generalizations across a wide range of simulation runs.

Any human performance that affects the time or probability of completing a task can be incorporated in TNM simulations. This makes the technique useful for HSI investigations because personnel factors such as skill level can be expressed in the time-and-error distributions used in such simulations. Thus, TNM simulations can be used to investigate HSI trade-offs involving manning levels, training or skill levels, and human factors engineering (Knapp, 1996). One extensive simulation used task network models to reflect the impact on crew tasks of four technical upgrades to a maritime patrol aircraft: data fusion technology, multifunction workstations, voice interactive technology, and electronic library applications. Simulation results were interpreted in terms of changes to the operators' workload and effectiveness and changes to the system effectiveness [Canadian Marconi Co. (CMC), (1995)]. Simulating human performance using task network models is seen as an important tool in achieving significant HSI cost reductions in future systems. For example, such simulations have been used to explore decreased manning levels in future naval systems (Campbell and Laughery, 1999), and such simulations can link job and task skill demands with system design, manning levels, and system performance (Middlebrook et al., 1999).

The outputs from task network simulations predict some characteristic of human performance, such as time to perform a sequence of tasks. By applying suitable algorithms to the task simulations, operator workload can be predicted as the sequence of tasks unfolds (Hendy et al., 1990). The simulation outputs can also be used to generate measures of a system's effectiveness. For example, the number of occurrences of specific tasks, such as verbal communication with a particular unit, can be used in conjunction with workload predictions and figures of merit derived from subject matter experts to generate MOEs for mission segments. Other transformations of operator MOPs to system MOEs can be calculated from task performance data. For example, the rates of processing messages and message-processing delays or backlogs can be calculated to give a MOE a C2 system (Middlebrook et al., 1999).

Early attempts at modeling human factors issues used commercially available software such as the General Purpose Simulation System (GPSS) originally developed by IBM (Overmayer, 1975) or custom-written software (McCann and Sweeney, 1976). Software tools are now available to support TNM. The chief tools are SAINT, developed by the U.S. Air Force (Wortman et al., 1978) and Micro Saint developed by Micro Analysis and Design and available commercially (Laughery and Drews, 1985).

The IMPRINT tool was developed from Micro Saint to incorporate nine separate tools for HSI analysis that had been developed previously by the U.S. Army Research Laboratory. IMPRINT can be used to model both crew and individual soldier performance for operator and maintainer tasks. Detailed operator–machine interface designs can be evaluated through the effects on task performance. For some analyses, workload profiles are generated so that crew–workload distribution, operator–system task allocation, and workload coping strategies can be examined. Maintainer workload can be assessed along with the resulting system availability. Using embedded algorithms, IMPRINT also models the effects of personnel characteristics, training frequency, and environmental stressors on the overall system performance. Manpower requirements estimates produced by IMPRINT can be used as the basis for estimating manpower life-cycle costs. The predecessor to IMPRINT, the hardware versus manpower (HARDMAN) analysis and simulation tool was subjected to verification, validation, and accreditation (VV&A) (Allender et al., 1995), and the key analysis capabilities of IMPRINT have also been subjected to VV&A. (See Chapter 11 for more information on IMPRINT.)

8.6.2 Complex Simulations and Warfighter Exercises

Human interaction with advanced systems is complex and emergent. It cannot be fully predicted because it emerges from the interaction between what the operator does with the machine and what the machine imposes as tasks or constraints on operator behavior (Taylor, 1959; De Greene, 1991). Because of this, in the early years of human factors research there was a strong emphasis on making observations in real-world conditions (Green et al., 1995; Moroney, 1995). The disadvantage of this approach is that the real world is not controlled. Thus, it may not be possible to make the observations required to evaluate system or operator performance without an unrealistic amount of observation (arranging situations to provide the required observations is the basis of most experimentation).

Compared to highly controlled laboratory experiments that have a strong theoretical basis but no close relationship to the real world, complex simulations and exercises are loosely controlled and are near the opposite extreme of techniques for measuring performance (Chapanis and Van Cott, 1972). This is primarily because of differences in the numbers of independent and dependent variables involved. Laboratory experiments involve few independent and dependent variables—sometimes only one dependent variable. In contrast, exercises and field trials involve a large number of both types of variables. Complex human-in-the-loop simulations lie between these extremes as they take place in quasi-operational conditions, are close to real-world observations, but are arranged to permit useful observations to be made as required.

While complex simulations, field trials, and exercises cannot replicate the stress of actual combat environmental stressors, the stress of sleep loss can be included as a factor when measuring performance. Other stressors can and may need to be simulated. For example, Muir et al. (1989) reported the effectiveness of using financial incentives in studies of emergency evacuation times from aircraft. They also reported that without such incentives behavior was not realistic and did not provide realistic MOPs.

At the outset of a development program, complex simulations, field trials, and warfighter exercises can provide empirical data for use in models or can validate models used in OA (Bryson, 1989). However, the primary use of complex simulations and field trials is to validate predictions about performance made earlier in the program. These human-in-the-loop simulation efforts can extend over several years during the development of a major weapon system and require firm commitments of personnel from operational units over that period. They typically require operational units to commit both operational and support personnel, weapons, and logistics required for the trial and require from months to years of preparation. In addition, analyses of the results from such trials typically require four to six times the amount of time required to make the observations, so results are not available immediately after the completion of the trial. Because of the time and effort required, large-scale field trials are easier to manage when conducted separately from specific system development projects to explore new system concepts or concepts of operations that can lead to development projects. When the trials are used to complement other measures of performance and to validate systems concepts, they must be anticipated, planned, and budgeted well in advance.

The requirements for time and effort preclude repeating field trials and exercises. Therefore, a key factor in organizing field trials is to arrange them so that the required observations can be made. To achieve this, such trials may need to be scripted to evolve in a particular way or to have operators perform specific mission segments. The selection of relevant performance measures that can be observed reliably in a field trial is a highly skilled activity. All measures will be affected somehow by the way that a field trial evolves, from almost negligible effects to measures that are strongly associated with the final states of the trials teams and systems. This is particularly true of the evaluation of C2 systems where commanders' decisions can have a significant effect on outcome measures (RSG.19, 1999). Table 8.2 provides some examples of the range of field trials and the corresponding script requirements.

In an effort to reduce the organizational and logistic requirements, field trials are sometimes arranged to "piggyback" on planned military exercises or other trials. This approach affords much less control over the trial. One consistent problem is that insufficient time may be scheduled by the operational units conducting the exercises for the trial troops to train and establish repeatable standard operating procedures (SOPs) for the new systems being evaluated (Poisson and Beevis, 2000). Usually, performance with a crew-served system increases with training until a plateau is attained (Towill, 1989). When

TABLE 8.2 Measures and Scripting Required for Various Field Trials

Aim of Trial	Measurements	Script Requirements
Suitability of personal equipment and individual weapons	Operator evaluations of comfort or suitability when issued to operational personnel for use during regular duties	No script required (Webb et al., 1998)
Physical workload associated with new equipment or procedures	Physiological measures of thermal stress and physical performance in field conditions	Some scripting of physical tasks required (Tack, 1996)
Direct control of a destroyer during specific ship-handling maneuvers	Measurement of accuracy of ship's track-keeping; comparison with results of a mathematical control model	Affected by weather conditions and requires definition of the maneuvers (Lewis et al., 1966)
Low-level navigation in tactical aircraft	Measures of aircraft track, accuracy of navigation in normal and unusual circumstances	Affected by weather, day/night, and scenario; requires definition of scenario and reactions to becoming lost (Lewis et al., 1968)
Evaluation of new systems for an existing role, such as the use of a hovercraft for search and rescue	Evaluation against a given set of goals for the system; measures of completion of specific tasks	Affected by weather and scenario; requires definition of scenario and tasks (Lewis et al., 1967).
Digital battlefield equipment and procedures	Multiple measures of information flow, situation awareness, decision making	Affected by evolution of scenario; easy to lose "thread" between independent variables and outcome measures

the operators are familiar with a crew-served weapon system and have established SOPs, performance times can be expected to conform to a learning curve. The advantage of this is that the final plateau of performance can be predicted from three observations early in the trials. If SOPs are changed, performance does not improve in a consistent manner and the plateau of performance is impossible to predict. In such a case, it is possible that the performance measurements could result in the selection from among several candidates of a system that is not the most effective.

8.7 INTERACTIONS AMONG HSI DOMAINS

The general measurement model described above is intended to support HSI efforts during the development process by focusing on system-relevant aspects of human performance. When planning human performance measures for a system under development, it is important to remember that the various HSI domains are interrelated. Changes in design to improve one domain nearly always affect other domains [Office of the Deputy Chief of Staff for Personnel (ODCSP), 1997]. Such changes must be considered when conducting design trade-offs. For example, in many weapon systems the operator's performance is adversely affected by having to wear protective clothing. Typical problems are hindrance or inability to perform tasks when wearing cold-weather gloves, inability to use weapon sights or other displays when wearing respirators, and reduced reach envelopes due to the bulk of clothing (Poisson and Beevis, 2000). Training can overcome some of these problems, but the most restrictive combinations of protective clothing and equipment may not be routinely included in training. Thus, from the viewpoint of system acquisition, it is important that the operator–machine interface and the training system be designed to accommodate all necessary combinations of protective clothing and equipment. The human factors engineering, health hazards, and training domains interact, and operator performance must be measured under conditions that represent those interactions.

The human systems integration plan for the defense acquisition process must address design trade-offs (DoD, 1991a). Manuals for HSI do not provide much guidance on HSI trade-offs; Booher (1990), for example, includes only two references to “trade-off”. Domain interactions and HSI trade-offs that have been suggested include the following (ODCSP, 1997; Walters, 1992):

- increasing system costs through automation to reduce costs for manpower, or training, or reduced requirements for experience;
- increasing system costs through built-in test equipment to reduce the requirements for skilled personnel to avoid drawing from a classification that is projected to be under strength;
- increasing system costs through simplification of the user interface to reduce training time costs; and
- reducing costs for operator manpower by increasing support manpower and personnel requirements.

These trade-offs are binomial. However, it is clear that a change in one domain could interact with several other domains. Reducing manpower costs through automation or simplification of the user interface will increase system costs but might also increase the maintenance training requirement or lead to skill degradation. Also, automation entails its own set of performance issues and costs (Parasuraman et al., 2000).

Although the need is obviously great, there is no well-established body of knowledge on HSI domain trade-offs. Kennedy and Jones (1992) noted that weapon systems designers do not have the expertise nor the tools that are required to make MPTS trade-off decisions. For example, iso-performance curves have been recommended to support quantified trade-offs among personnel abilities and factors such as training time and training system effectiveness (Kennedy and Jones, 1992). The curves (see Fig. 8.5) show the relationship between personnel abilities measured on an aptitude scale and the time to train a given percentage of operators or maintainers up to an acceptable standard. An improvement in the ease of use of an item of equipment should result in a change in the iso-performance curve, because the new equipment requires less training than the predecessor to achieve the same level of proficiency. Unfortunately, Kennedy and Jones report that no data were archived by any of the armed services that would support the generation of iso-performance curves and that developing such curves must be done opportunistically.

8.7.1 The F-18 Example

To better understand interactions between the HSI domains and their effects on operator and system performance, Davidson et al. (1991) analyzed data from a review of human factors affecting flight safety and operational effectiveness of the F/A-18 Hornet aircraft operated by the Canadian Armed Forces. The data, derived from interviews with F-18 pilots, were categorized into the HSI domains and reviewed for interactions (Beevis, 1996). Figure 8.6 shows the interactions between 34 factors related to manpower, personnel, training, system safety, health hazards, and human factors engineering.

One thread of HSI domain interactions will be followed as an example. Through a landmark effort in human factors engineering (Merriman and Moore, 1984), the F-18 was designed and developed as a multirole aircraft that can be flown by one person. The F-18 replaced two-place interceptors on some squadrons, thereby reducing the manpower levels;

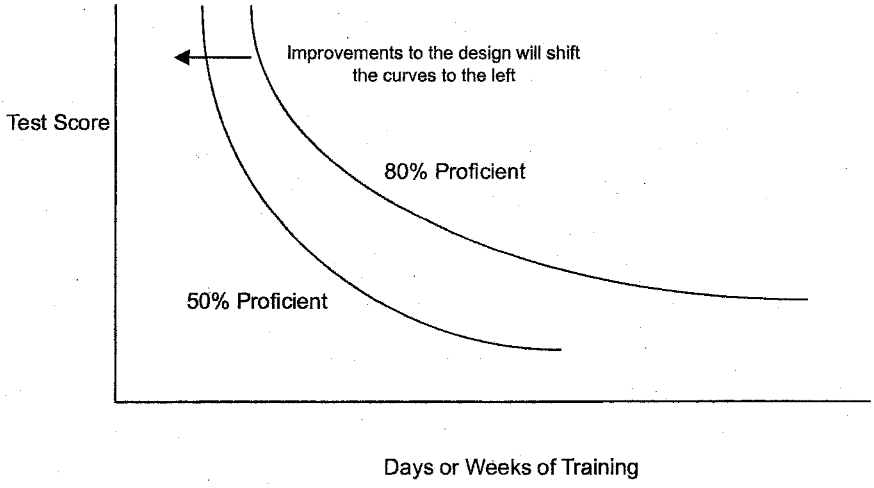

Figure 8.5 Iso-performance curves for ability and training time (after Kennedy and Jones, 1992).

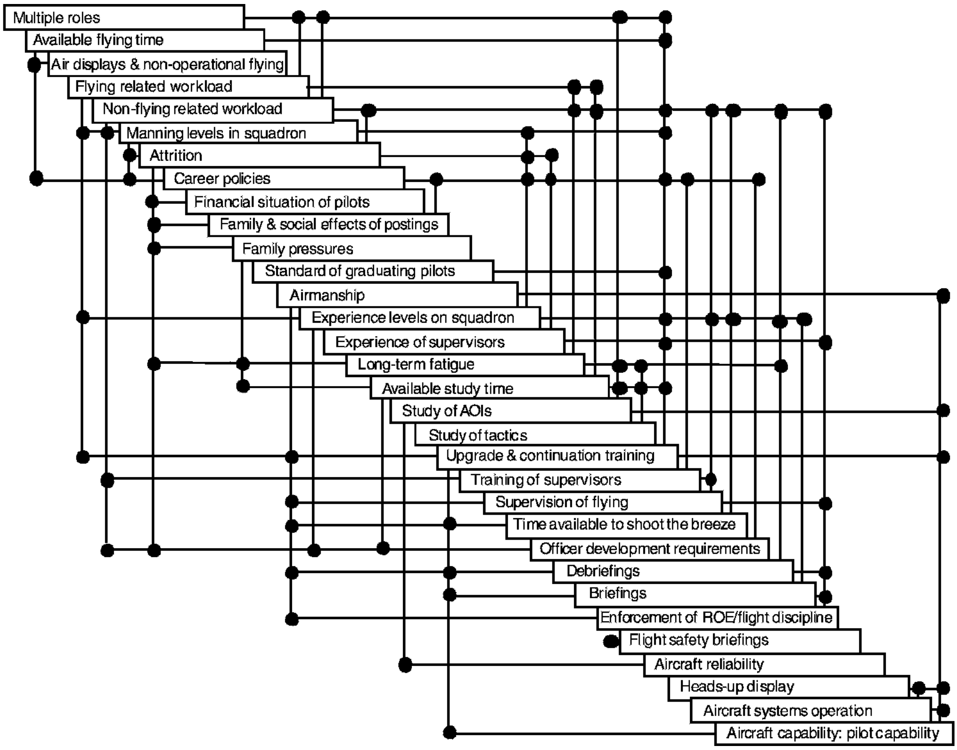

Figure 8.6[1] Factors affecting flight safety and operational effectiveness in F-18 aircraft (horizontal lines are outputs; vertical lines are inputs).

thus, the HFE domain interacts with manpower. The reduction in squadron manning levels increased the overall workload of nonflying activities on squadron personnel, including study time and classroom training; thus, the manning domain interacts with training. The capability of the F-18 made the study of tactics and aircraft systems more demanding, but the reduction in available study time made it more difficult for aircrew to do this; thus, training has internal interactions. In addition, lack of familiarity with aircraft operating instructions for emergency procedures affected flight safety. At the same time, on-squadron proficiency training and practice (upgrade and continuation training in Fig. 8.6) affected the standard of airmanship in squadron pilots. Airmanship affects the level of flight safety in the squadron as well as the quality of training once pilots are on-squadron; thus, training affects system safety in several ways. The U.S. Air Force has made similar observations. For example, an investigation of a fatal accident involving two combat aircraft concluded that the quality of crews had been hurt by too many deployments, which had precluded developmental training (Newsbreaks, 1994).

The interactions among the factors in the F-18 HSI study were analyzed using matrix algebra (the MICMAC method of Godet, 1991) to identify important direct and indirect interactions. Figure 8.7 shows the direct and indirect interactions among the HSI domains, derived from the relationships in Figure 8.6. It was concluded that even though the HSI domains of manpower, personnel, training, system safety, and human factors engineering do not seem to interact directly, they do have strong indirect interactions. Human factors engineering appeared to interact directly with only the training domain, whereas it interacted indirectly with manpower, personnel, and system safety factors. The results suggested that the interactions between these domains were probably more important than the interactions within the domains.

The personnel domain appeared to have the most interaction with other HSI domains. This seems reasonable given that personnel factors include the basic performance abilities of the humans in a system. Some important interactions between the 34 factors examined were a function of operational and organizational issues (e.g., career policy or operational commitments). This suggests that, once a well-designed system is introduced into service, operational requirements and policies may have a greater effect on operational effectiveness than the individual HSI domains. The importance of these factors is reflected in the conclusion of a NATO study group that the implementation of HSI would be facilitated by user descriptions that include information on individual units and organizational matters as well as the intended applications and environment (RSG.21, 1994).

Surprisingly, the health hazards domain had no direct or indirect interaction with any of the other domains. This may be a reflection of the particular application of the F-18, since there should be interaction between health hazards and system safety at a minimum. The seventh domain of HSI, personnel survivability, would also interact with both health hazards and system safety at a minimum, with desired interaction with HFE and probably training as well.

Overall, the analysis of HSI factors in F-18 squadrons showed that the pattern of the interactions is complex and does not lead to simple statements about trade-offs among the HSI domains. Rather than operating in isolation, operational practice, developmental training, manning levels, and the experience levels of personnel interact with the design resulting from the human factors engineering effort to affect the overall level of performance, effectiveness, and safety of an operational unit. For example, the thread of indirect domain interactions outlined above suggests that good human factors engineering can lead to deterioration in operational safety standards unless organizational measures are taken to avoid it.

8.7.2 Quantification of HSI Trade-offs

Many of the interactions examined in the F-18 case study provide qualitative information about trade-offs among the HSI domains (e.g., the need for protective gloves requires

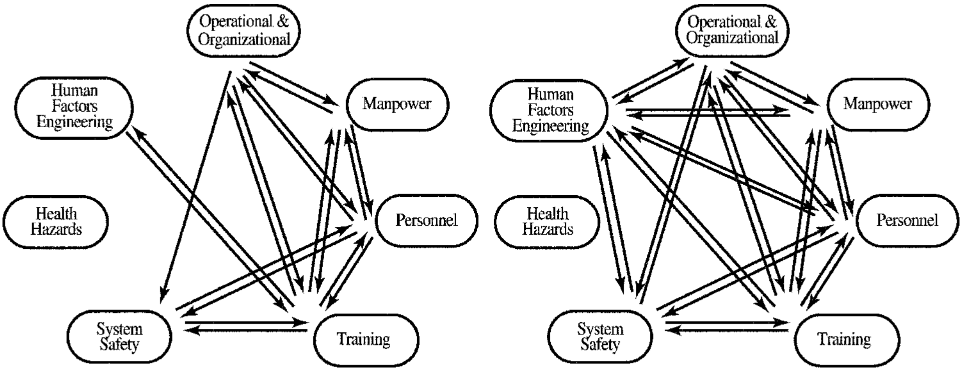

Figure 8.7[1] Direct (left) plus indirect (right) interactions among HSI domains for CF-18 aircraft.

additional human factors engineering effort to avoid decrements in performance). However, the design process requires quantitative information for trade-offs. Ideally, trade-off analyses should describe and compare either equal-cost or equal-capability options (DoD, 1991b). Thus, in a formal cost–benefit analysis all measurements must be transformed into an equivalent cost or performance delta.

Some costs associated with HSI issues are obvious, particularly those associated with manpower and personnel (DoD, 1991c). Several personnel cost models are available; for example, the Army Military–Civilian Cost System (AMCOS) is a database of active, reserve, and civilian manpower data that generates the manpower costs for the life cycle of a proposed system from "manpower-by-grade" information (Horne, 1987). System life-cycle cost models often include manpower and personnel costs. One model used for military system procurement is shown in Figure 8.8 (Kerzner and Bayne, 1991) with the most obvious personnel costs broken out. Reviewing its applicability to assist HSI trade-off analyses, the model developers concluded that it could be used to cost different system concepts, including ones with different manning levels or training costs. Procurement agencies have successfully used such cost models to compare HSI life-cycle costs across competing systems by requiring bidders to provide the data necessary to run the cost model.

In most cases, the life-cycle cost model approach does not help system developers and designers make the trade-offs required during the design process. First, during the design process, life-cycle costs are difficult to identify because the criteria are multivariate and elusive, including, for example, mission performance costs, safety costs, and logistic costs such as the supply, maintenance, and replacement of protective clothing and equipment. Both knowledge and tools are needed to identify such costs and to make appropriate trade-

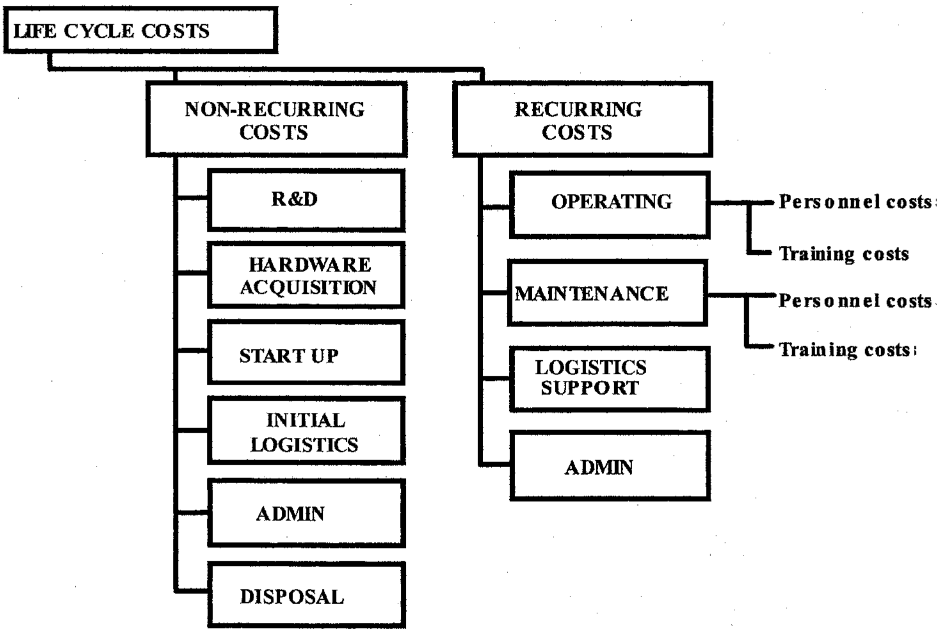

Figure 8.8 Life-cycle cost model.

offs. However, detailed analyses of the costs in HSI domains do not appear to have received much systematic study. Second, cost models that require a designer to produce a complete system concept to which costs can be assigned are unlikely to be used by designers at the desk level. The HSI manuals do not provide cost trade-off tools and typically approach cost control by identifying high-driver tasks (ODCSP, 1997). Due to these problems, HFE specialists rarely perform a formal cost–benefit analysis. Much too frequently, human factors (HF) specialists on design teams leave cost considerations to specialists and focus on the measurements related to the performance of the human–machine system. Unfortunately, this limits the supply of information that would support studies of the cost effectiveness of HSI. To address that problem, at least one large-scale program has been started.[2]

8.8 FUTURE TRENDS

When Erickson (1984) produced his recommendations for linking human performance measures to system effectiveness measures, the predominant human factors technique was laboratory experimentation. This is changing rapidly with the advent of large warfighter exercises and the complex simulations discussed previously.

The DoD is developing several large simulation systems that allow constructive instantiations of developing systems to be evaluated collaboratively at diverse locations literally all over the globe. Other engineering disciplines are responding to the need to maximize system effectiveness, minimize life-cycle costs, and reduce development costs and times needs with a *revolution in business affairs* (RBA). This revolution places much greater emphasis on iterative design, integrated program management teams, and the use of technologies such as synthetic environments, which expands the use of modeling and simulation and computer-aided design through an integrated approach known as *simulation-based acquisition* (SBA).

Unless HSI processes and techniques are able to link into and exploit these processes, it will become increasingly difficult for HSI specialists to influence the eventual design solution. Thus, it can be expected that HSI activities will become more closely associated with constructive, virtual, and live simulations. Given the limitations of knowledge about human behavior needed for constructive simulations and the costs and lead times associated with live simulations, much more use is likely to be made of virtual simulations and experimentation. The use of virtual caves and other virtual representations is still more of a laboratory phenomenon than an engineering tool, but the pendulum is definitely swinging toward the use of virtual and other realistic simulation environments. That will present its own measurement challenges, because evidence to date is that human behavior in a virtual environment has some significant differences from that in the real world (Wickens and Baker, 1995).

Perhaps even more disturbing is the dramatic change in the military environments that will challenge the new systems under development. Technology may prove counterproductive in many of the nonlinear and asynchronous environments that will constitute modern battlefields and peacekeeping missions (Barnes and Fichtl, 1999). This puts an added burden on the evaluation process to ensure that military flexibility and the ability to operate in diverse environments are considered as part of the HSI design process.

Finally, systems themselves are becoming more complex, and the concept of a system of systems is becoming an accepted part of military doctrine. Evaluating systems in isolation will become increasingly more difficult to justify. The cost of ignoring complexity outweighs the considerable cost of investigating developing systems using the full panoply of measurement techniques that we have discussed above.

8.9 SUMMARY AND CONCLUSION

In summary, a wide range of variables must be considered and measured if human factors are to be integrated effectively into the design of complex systems to improve total system performance and reduce costs of ownership. Thus, a prime goal of any HSI program should be the measurement of human performance that can generate MOEs related to top-level design requirements. Measurements must be made in the context of the growing use of a spiral concurrent engineering effort where early approximations of the system are modeled and field evaluated before any mature design concept exists. Since available measurement techniques are applicable at different points in the acquisition design and development cycle, the scale of measurement that can be used changes throughout the system design and development process.

Three general approaches to measurement have been found applicable to the development and evaluation of defense systems: (1) analysis and computer simulations, (2) laboratory experiments, and (3) complex human-in-the-loop simulations combined with large-scale field trials. Many activities in the early stages of system design and development involve the analysis and synthesis of a design solution. Task network simulations can predict a range of characteristics of human performance. The simulation outputs can also be used to generate MOEs. Experimental methods are most useful for addressing specific design issues and for investigating specific cognitive and human performance questions related to these issues. Validation methods that include complex simulations and field exercises are essential in allowing the designer to evaluate design concepts in a realistic military environment. The human factors associated with the different HSI domains may have important interactions, but these are hard to predict and there is little quantitative information available to support trade-offs between domains. It is important that more quantitative data to aid trade-offs be developed. Without such data, it is difficult for life-cycle cost models to help system developers and designers make the necessary trade-offs.

The general conclusion is that a careful blend of measurement tools can and should be used during the design and development process to uncover the performance benefits of various design options that are impacted by human systems considerations. If the derived benefits can be related directly to design requirements and overall system goals, the payoff for performance-centered HSI trade-off studies is significant.

NOTES

1. Figures 8.4, 8.6, and 8.7 are © HER MAJESTY THE QUEEN IN RIGHT OF CANADA (2003) as represented by the Minister of National Defence. Reproduced with permission.
2. See Defence Research and Development Canada. (2001). *Human Systems Integration Capability: Concept Description*. Ottawa, Canada: Defence Research and Development Canada, Director of Science and Technology for Human Performance.

REFERENCES

Allender, L., Kelley, T. D., Salvi, L., Lockett, J., Headley, D. B., Promisel, D., Mitchell, D., Richer, C., and Feng, T. (1995). Verification, Validation, and Accreditation of a Soldier-System Modeling

Tool. In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Barnes, M. J. (1978). *Display Size and Target Acquisition Performance*, Tech. Report No. 6006. China Lake, CA: U.S. Naval Weapons Center.

Barnes, M. J., and Fichtl, T. (1999). Cognitive Issues for the Intelligence Analyst of the Future. In *Proceedings of the 3rd Annual Federated Laboratory Symposium* (pp. 15–19). Adelphi, MD: U.S. Army Research Laboratory.

Barnes, M. J., Knapp, B. G., Tillman, B. W., Walters, B. A., and Velicki, D. (2000). *Crew Systems Analysis of Unmanned Aerial Vehicles Future Job and Tasking Environments*, ARL-TR-2081. Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Barnes, M. J., and Matz, M. (1998). Crew Simulations for Unmanned Aerial Vehicle Applications: Shift Factors, Interface Issues, and Crew Size. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp. 143–148). Santa Monica, CA: Human Factors and Ergonomics Society.

Barnes, M. J., McClean, M., and La Nasa, F. (1996). Empirical and Analytical Methods for User Centered Design: A Synergistic Approach for the Downsized Ground Control Station. In *Proceedings of the 64th Annual Military Operational Research Society Symposium* (p. 149). Ft. Leavenworth, KS.

Beevis, D. (1987). Experience in the Integration of Human Engineering Effort with Avionics Systems Development. In *Proceedings of AGARD Symposium on the Design, Development, and Testing of Complex Avionics Systems, CP-417*, Vol 27, pp. 1–9. Neuilly sur Seine, France: Advisory Group for Aerospace Research and Development.

Beevis, D. (1996). Human Functions and System Functions. In D. Beevis, P. Essens, and H. Schuffel (Eds.), *Improving Function Allocation for Integrated Systems Design* (pp. 63–79). Wright-Patterson Air Force Base, OH: Crew Systems Ergonomics Information Analysis Center.

Beevis, D. (Ed.). (1999). *Analysis Techniques for Human-Machine System Design: A Report Produced under the Auspices of NATO Defence Research Group 8*. Wright-Patterson Air Force Base, OH: Crew System, Ergonomics Information Analysis Center.

Beevis, D., and Essens, P. (1997). The NATO Defence Research Group Workshop on Function Allocation. In E. F. Fallon, L. Bannon, and J. McCarthy (Eds.), *ALLFN'97 Revisiting the Allocation of Functions Issue: New Perspectives*. Louisville, KY: University of Louisville, IEA Press.

Booher, H. R. (1990). *MANPRINT: An Approach to Systems Integration*. New York: Van Nostrand-Reinhold.

Box, G. E., Hunter, W. G., and Hunter, J. S. (1978). *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building*. New York: Wiley.

Burbank, N. S. (1994). *The Development of a Task Network Model of Operator Performance in Asimulated Air Traffic Control Task*, DCIEM-94-05. Toronto, ON: Defence and Civil Institute of Environmental Medicine.

Bryson, M. R. (1989). Experiments. In W. P. Hughes (Ed.), *Military Modeling*, 2nd ed. Washington, DC: Military Operations Research Society.

Card, S. K., Moran, T. P., and Newell, A. (1986). The Model Human Processor—An Engineering Model of Human Performance. In K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.), *Handbook of Perception and Human Performance*, Vol. II (pp. 45-1 to 45-35). New York: Wiley.

Cook, T. D., and Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis for Field Settings*. Chicago: Rand McNally.

Campbell, G., and Laughery, R. (1999). Modeling Human Performance on the Road to Manpower Optimization. *CSERIAC Gateway, X*(2), 12–13.

Canadian Marconi Company (CMC). (1995). *Aurora Life Extension Project, CP-140 Aurora Crew Efficiency Study: Report on the Preliminary Assessment of the Potential to Increase Crew*

Efficiency, Vol. 1, Department of Defence DSS Contract W8477-4-AC96/01-QC. Kanata, ON: CMC.

Chapanis, A. (1960). On Some Relations between Human Engineering, Operations Research, and Systems Engineering. In D. P. Eckman (Ed.), *Systems: Research and Design*. New York: Wiley.

Chapanis, A., and Van Cott, H. P. (1972). Human Engineering Tests and Evaluations. In H. P. Van Cott and R. G. Kincade (Eds.), *Human Engineering Guide to Equipment Design*. New York: Wiley.

Charlton, S. G. (1996). SITE: An Integrated Approach to Human Factors Testing. In T. G. O'Brien and S. G. Charlton (Eds.), *Handbook of Human Factors Testing and Evaluation*. Mahwah, NJ: Erlbaum.

Davidson, R. A., Beevis, D., Buick, F., Donati, A. L. M., Kantor, L., Bannister, S. R. H., Brook, E. A., Rochefort, J. A. P., and Turner, J. R. (1991). *Human Factors in the CF-18 Pilot Environment*, DCIEM Report No. 91-11. Toronto: Defence and Civil Institute for Environmental Medicine.

De Greene, K. B. (1991). Emergent Complexity and Person-Machine Systems. *International Journal of Man-Machine Studies, 3*, 219–234.

Erickson, R. A. (1984). *The Human Operator and System Effectiveness*, Tech. Publication 6541. China Lake, CA: Naval Weapons Center.

Flach, J. M., Vicente, K. J., Tanabe, F., Kazuo, K., and Rasmussen, J. (1998). An Ecological Approach to Interface Design. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp. 295–299). Santa Monica, CA: Human Factors and Ergonomics Society.

Godet, M. (1991). *Problemes at methodes de prospective—Boite a outils.* Paris: Edition Futuribles.

Gould, J. D. (1988). How to Design Usable Systems. In M. Helander (Ed.), *Handbook of Human–Computer Interaction* (pp. 757–785). North-Holland: Elsevier.

Green, R. J., Herschel, C. S., and Ellifritt, T. S. (1995). *50 Years of Human Engineering: History and Cumulative Bibliography of the Fitts Human Engineering Division*. Wright-Patterson Air Force Base, OH: Crew Systems Directorate, Armstrong Laboratory.

Greenley, M., Downs, G., and Espenant, M. (1998). Human Factors in an Armored Vehicle Simulation Project. In *Proceedings Interservice, Industry Training Simulation 48th Education Conference (IITSEC)*. Orlando, FL: IITSEC.

Hatley, D. J., and Pirbhai, I. A. (1987). *Strategies for Real-Time System Specification*. Chichester, UK: Wiley.

Hendy, K. C., Campbell, E. L., and Schuck, M. M. (1990). The Use of Human Information Processing Models for Workload Predictions Using Network Simulations. In E. J. Lovesey (Ed.), *Proceedings of the Ergonomics Society's 1990 Annual Conference, "Ergonomics-Setting Standards for the 90's"* (pp. 129–134). London: Taylor and Francis.

Horne, D. K. (1987). Army Manpower Cost System. *MANPRINT Bulletin, II*(4). (See also http://dticam.dtic.mil/hsi/index/hsi4.html.)

Kennedy, R. S., and Jones, M. B. (1992). *Simulating the Impact of MPTS Trade-off Decisions by Application of Isoperformance Methodology.* Orlando, FL: Naval Training Systems Center.

Kerzner, L. F., and Bayne, R. H. (1991). *A User's Guide to DND LCC 2.0*, ORAE Project Report No. PR 556. Ottawa: Department of National Defence, Operational Research, and Analysis Establishment.

Kirk, R. E. (1982). *Experimental Design: Procedures for the Behavioral Sciences*. Pacific Grove, CA: Brooks Cole.

Knapp, B. G. (1996). Task and Workload Analysis for Army Command, Control, Communications and Intelligence Systems. In D. Beevis, P. Essens, and H. Schuffel (Eds.), *Improving Function Allocation of Integrated Design Systems*. Dayton, OH: Crew Systems Ergonomic and Information Analysis Center.

Laughery, K. R., and Drews, C. (1985). Micro SAINT: A Computer Simulation System Designed for Human Factors Engineers. In *Proceedings of the Human Factors Society 29th Annual Meeting* (pp. 1061–1064). Santa Monica, CA: Human Factors and Ergonomics Society

Lewis, R. E. F., de la Riviere, W. D., and Logan, O. (1966). Sea Trials of Direct Control of Engines and Helm. In *Proceedings of the Ship Control Systems Symposium*, Vol. 1 (pp. VIII-A-1 to VIII-A-4). Annapolis, MD: U.S. Dept. of the Navy.

Lewis, R. E. F., de la Riviere, W. D., and Sweeney, D. M. (1968). Dual versus Solo Pilot Navigation in Helicopters and Low Level. *Ergonomics, 11*, 145–155.

Lewis, R. E. F., Storr, J. W., and Brewer, F. G. (1967). *Trials of an SK-5 Hovercraft for the Canadian Coast Guard*, DMRL Report 667. Toronto, ON: Defence and Civil Institute of Environmental Medicine.

McBurney, D. H. (1998). *Research Methods*. Pacific Grove, CA: Brooks Cole.

McCann, C., and Sweeney, D. M. C. (1976). *A Simulation to Study the Loading of Manpower and Facilities in the Data Interpretation and Analysis Centre for the LRPA, Part 3: A Guide to the Operation of the Mk.III Model*, DCIEM76-X-67. Toronto, ON: Defence and Civil Institute of Environmental Medicine.

McMillan, G. R., Beevis, D., Stein, W., Strub, M. H., Salas, E., Sutton, G., and Reynolds, K. C. (1991). *A Directory of Human Performance Models for System Design*, AC/243 TR/1. Brussels: NATO Defence Research Group Panel-8.

McMillan, G. R., and Martin, E. A. (1994). An Overview of Human Performance Models and Potential Applications to Combat Simulation. In C. A. Murtaugh (Ed.), *Human Behavior and Performance as Essential Ingredients in Realistic Modeling of Combat—MORIMOC III. Proceedings of the Military Operations Research Society Workshop, 27–29 May 1990*. Alexandria, VA: Military Operations Research Society.

Meister, D. (1985). *Behavioral Analysis and Measurement Methods*. New York: Wiley.

Meister, D. (1986). A Survey of Test and Evaluation Procedures. In *Proceedings of the Human Factors Society 30th Annual Meeting*, Vol. 9 (pp. 1239–1243). Santa Monica, CA: Human Factors Society.

Meister, D. (1987). Systems Effectiveness Testing. In G. Salvendy (Ed.), *Handbook of Human Factors*. New York: Wiley.

Merriman, S. C., and Moore, J. P. (1984). The F-18—A New Era for Human Factors. In *Proceedings of AGARD Symposium on Human Factors Considerations in High Performance Aircraft*, AGARD-CP-371, Vol. 19, pp. 1–5. Neuilly-sur-Seine, France: Advisory Group for Aerospace Research and Development.

Middlebrook, S. E., Knapp, B. G., Barnette, B. D., Bird, C. A., Johnson, J. M., Kilduff, P. W., Schipani, S. P., Swoboda, J. C., Wojciechowski, J. Q., Tillman, B. W., Ensing, A. R., Archer, S. G., Archer, R. D., and Plott, B. M. (1999). *Computer Simulation Models to Investigate Human Performance Task and Workload Conditions in a US Army Heavy Maneuver Battalion Tactical Operations Center*, ARL-TR-1994. Aberdeen Proving Ground, MD: Army Research Laboratory.

Moroney, W. F. (1995). The Evolution of Human Engineering: A Selected Review. In J. Weimer (Ed.), *Research Techniques in Human Engineering*. Englewood Cliffs, NJ: Prentice-Hall.

Muckler, F. A., Seven, S., and Akman, A. (1991). *Construction of a Military Occupation Specialty Taxonomy*, ARI Research Note 91–10. Alexandria, VA: U.S. Army Research Institute.

Muir, H., Marrison, C., and Evans, A. (1989). *Aircraft Evacuations: The Effect of Passenger Motivation and Cabin Configuration Adjacent to the Exit*, CAA Paper 89019. Cranfield, Beds, UK: Cranfield Institute of Technology Applied Psychology Unit.

Naikar, N., and Sanderson, P. M. (2000). Evaluating Design Proposals with Work Domain Analysis. In *Proceedings of the IEA 2000/HFES 2000 Congress* (pp. 202–205). Santa Monica, CA: Human Factors and Ergonomics Society.

Newsbreaks. (1994, October 17). *Aviation Week and Space Technology, 141*(16), 15.

Office of the Deputy Chief of Staff for Personnel (ODCSP). (1997). *A Handbook for MANPRINT in Acquisition*. Washington, DC: ODCSP.

Overmayer, R. W. (1975). *A Computer Model for Command and Control Analysis*, Report under Contract 00014-74-C-0324. Arlington, VA: Office of Naval Research, Naval Analysis Programs.

Parasuraman, R., Sherdian, T. B., and Wickens, C. D. (2000). A Model for Types and Levels of Interaction with Automation. *IEEE Transactions on Systems, Man, and Cybernetics*.

Poisson, R. M., and Beevis, D. (2000). *Human Factors Issues in Land Forces Weapons Systems Evaluations*, DCIEM Report TR 2000-032. Toronto, ON: Defence and Civil Institute of Environmental Medicine.

Rasmussen, J. (1986). *Information Processing and Human Machine Interaction*. New York: North-Holland.

Rasmussen, J., and Vicente, K. J. (1989). Coping with Human Error through Interface Design: Implications for Human Interface Designs. *International Journal of Man-Machine Studies, 31*, 517–534.

RSG.19. (1999). *Code of Best Practice (COBP) on the Assessment of C2*, RTO-TR-9. Neuilly-sur-Seine, France: NATO Research and Technology Organization.

RSG.21. (1994). *Liveware Integration—Final report*, NATO DRG Report No. AC/243 (Panel 8) TR/18. Brussels: NATO Defence Research Group Panel-8.

Tack, D. (1996). *Hot Weather CB Protective Clothing: Trial Concept and Outline Plan*, DCIEM-97-CR-36. Toronto, ON: Defence and Civil Institute of Environmental Medicine.

Taylor, F. V. (1959). *Human Engineering and Psychology.* Washington, DC: U.S. Naval Research Laboratory (unpublished preprint).

Towill, D. R. (1989). Selecting Learning Curve Models for Human Operator Performance. In G. R. McMillan, D. Beevis, E. Salas, M. H. Strub, R. Sutton, and L. van Breda (Eds.), *Applications of Human Performance Models to System Design*. New York: Plenum.

U.S. Department of Defense (DoD). (1991a). *Defense Acquisition. Part 7, Section B: Human Systems Integration*, Directive 5000.1. Washington, DC: DoD.

U.S. Department of Defense (DoD). (1991b). *Defense Acquisition. Management Documentation and Reports, Part 8: Cost and Operational Effectiveness Analysis*, Directive 5000.2. Washington, DC: DoD.

U.S. Department of Defense (DoD). (1991c). *Defense Acquisition. Management Documentation and Reports, Part 6: Manpower Estimate Report*, Directive 5000.2. Washington, DC: DoD.

U.S. Department of Defense (DoD). (1999, May). MIL-HDBK-46855A, Human Engineering Programs, Process, and Procedures. Washington, DC: DoD.

Walters, H. (1992). *Ministry of Defence: The MANPRINT Handbook*, 2nd ed. London: Controller, Her Majesty's Stationery Office.

Webb, R., Tack, D., and Gaughan, P. (1998). *DND Prototype Combat Helmet User Acceptance Trials*, HSI Report for DCIEM. Toronto, ON: Defence and Civil Institute of Environmental Medicine.

Whiteside, J., Bennett, J., and Holtzblatt, K. (1988). Usability Engineering: Our Progress and Evolution. In M. Helander (Ed.), *Handbook of Human Computer Interaction* (pp. 791–819). Amsterdam: North-Holland.

Wickens, C. D., and Baker, P. (1995). Cognitive Issues in Virtual Reality. In W. Barfield and T. A. Furness (Eds.), *Virtual Environments and Advanced Interface Design* (pp. 515–541). New York: Oxford University Press.

Wickens, C. D., and Hollands, J. G. (2000). *Engineering Psychology and Human Performance*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall.

Williges, R. C. (1995). Review of Experimental Design. In J. Weimer (Ed.), *Research Techniques in Human Engineering*. Englewood Cliffs, NJ: Prentice-Hall.

Wortman, D. B., Duket, S. D., Seifert, D. J., Hann, R. L., and Chubb, G. P. (1978). *Simulation Using SAINT: A User-Oriented Instructional Manual*, AMRL-TR-77-61. Wright-Patterson Air Force Base, OH: Aerospace Medical Research Laboratory.