Human Factors Certification in the Development of Future Air Traffic Control Systems

Alyson E. Evans

United Kingdom Civil Aviation Authority

Introduction

If human factors certification of aviation technologies aims to encompass the wide range of issues which need to be addressed for any new system, then human factors involvement must be present throughout the whole design process in a manner which relates to final certification. A certification process cannot simply be applied to the final product of design. Standards and guidelines will be required by designers at the outset of design for reference in preparing for certification.

The most effective use of human factors principles, methods, and measures is made as part of an iterative design process, leading to a system which reflects these as far as possible. This particularly applies where the technology is complex and may be represented by a number of components or sub-systems. Some aspects of the system are best certified during early prototyping, when there is still scope to make changes to software or hardware. At this stage in design, financial and/or time pressures will not rule out the possibility of necessary changes, as may be the case later. Other aspects of the system will be best certified during the final phases of design, when the system is in a more complete form and in a realistic environment.

Human Factors Input at System Conception

The need for any new aviation system is either generated by incumbent end users in the operational environment or by planners closely associated with the current system or the job to be done. The need for change arises because of failures or inefficiency in the current system or from a change in the future requirements for that system. In the United Kingdom, the very first conceptual stages of system design aimed at meeting a new requirement are usually carried out by end users or planners who will usually seek guidance from hardware and software engineers as the first step in design. It is rare that human factors specialists are involved at this stage in design, when ideas for designs are being generated and moulded.

It is, however, necessary that they are involved at this stage, before any firm requirement for the system has been put on paper. Their involvement as a member of the design team is necessary for a number of reasons.

Firstly, the human factors specialist views the system user as an integral component of the whole system. The human component, like others in the system, brings advantages and limitations. To make best use of the human component in the system, consideration must be given to generally accepted psychological strengths and weaknesses (Meister, 1971). A model of the current or future user can be constructed, or alternatively a survey of user needs, using questionnaires and/or interviews, may provide information which can help to outline user characteristics.

A survey of user needs can be helpful, even if the nature of the user in the new system is to change, as it gives a baseline of current needs against which future objectives can be planned. Optimisation of the role of the human component in the system is necessary at this stage of design to maximise efficiency and safety of the system and to provide the operator with a supportive usable system which allows job satisfaction rather than a system which is supported by the user and causes frustration.

The presence of a human factors specialist at this stage would ensure that the characteristics and psychology of the user are considered from the start. What is more often the case is that user/designers, along with software and hardware engineers, will look first to the available technology as a starting point in design. This tends to lead to an abuse of the flexibility of the human component within the design as the human is then 'worked around' the technology which is chosen, filling in functions which are not carried out by the technology. User/designers and engineers are not aware of how to best utilise the human component even though they may be very familiar with the system or the job to be done. Effective utilisation of the human component should be possible if the capabilities of both the human component and machines are bourne in mind and if sound human factors principles are applied. This is more likely to result in safer systems, with the human component having a minimal risk of failure, and which are more satisfying for the user to operate.

Secondly, the design of any new system is an iterative process. As design options are explored ideas are developed which need to be fed back to the design. A human factors practitioner is makes an essential contribution to this process by using human factors principles at appropriate phases in design and adapting them to the specific requirements of the system. Sequence and timing in the use of human factors principles are important. If sequence and timing are not appropriate, then benefit is lost and later certification will reflect this. When principles do not exist for some aspects of design or when a number of alternatives have been generated, then user opinion may be collected from design options which are tried out in a controlled fashion. Such information can, in turn, be fed back into the design process.

Finally, the human factors specialist can help to define the performance criteria necessary for the system to achieve its aim, including those necessary for the human component. Such issues are rarely addressed in any detail by user/designers at the start of design. Definition of system aims allows design to focus on supporting the human component and technology to achieve system output. Definition of performance criteria create standards against which the system can be evaluated or certified at a later date. Performance criteria used can be divided into three categories.

System Criteria. Overall system performance can be measured in terms of the output; i.e., "does the system achieve a specified level of output according to the standards set at the start of design?"

System output is an objective measure of performance, and if standards of output are not reached, questions are generated concerning the system design. Low system output may reflect poor equipment, procedures, or poor user performance. System output can be assessed in both quantitative and qualitative ways.

Task Performance Criteria. Levels of performance on individual tasks needed to achieve the output can be examined. Such tasks may or may not involve the user. Sub-tasks carried out by the machine may affect the user, however, so if deficiencies are corrected this can contribute to the improvement of the overall process. Again quantity and quality of individual tasks can be examined. Quality can be measured, for example, in terms of accuracy, efficiency, effectiveness, number of errors and timeliness of tasks and quantity in terms of number of aircraft processed by a sector in a specified time period.

Subjective Responses. Subjective responses allow the assessment of ease of use of the system by operators. Acceptable levels for ease of use are gauged at the start of design and can be measured throughout system design using questionnaires and interviews to cover many aspects of the system.

Workload measures also reflect ease of use of the system. Early in design, performance criteria may be defined in broad terms before the detail of functions and tasks of the system have been considered. The constraints on system development in terms of time, money, manpower, etc., should also be identified so that the limits, within which accomplishment of system goals must take place, are taken into account.

The performance of some criteria can be measured objectively and others can only be reached subjectively by asking the user to respond to direct or indirect questioning. Subjective measures of performance are obviously going to be subject to some bias from the respondent, but have proved successful and useful in highlighting problem areas in design of ATC systems for the Civil Aviation Authority (CAA).

The emphasis on involvement of a human factors specialist in such early stages of design has been made because human factors input at conception of system design is unfortunately rare. The argument for early involvement is common. The reasons why it is rarely the case is largely because the people who find themselves in the position of having to design new systems are so often users or ex-users who have a planning role. They are not aware of what human factors has to offer throughout the design cycle. Likewise, hardware and software engineers are largely unaware of human factors issues and of the reasons for fully considering the human component during the design process.

Human Factors Approach to System Design

A systematic approach to design of new systems is described by Bailey (1982). This outlines the human factors approach and can be used to illustrate where certification of various aspects of the system are best carried out.

Determination of Objectives and Performance Specifications. A broad statement of system objectives is the first requirement. For a new air traffic control (ATC) system, these may be

along the lines of: "provide an ATC system which increases the capacity of a major terminal maneuvering area (TMA)."

Following this, system performance specifications need to be developed which reflect in more detail what the system must do to meet its objectives; e.g., process air traffic at a faster rate using new routes and airspace divisions while maintaining specified separation standards and providing efficient flight profiles for aircraft. This must be achieved using the current number of air traffic personnel.

At this stage in design, it is important to have a thorough understanding of the end user. Consideration must be given to whether the end user will be the same as the current end user of whether the user will change. A change in characteristics of the end user may result from new demands of the job to be done or because of a demographic limitation. Interviews and questionnaires will yield information about the current end users from a fairly representative population. It is only when the user is understood that a future system can be designed to effectively include the user.

Likewise, the technology available to form the system should be fully understood in terms of its capabilities and limitations.

Definition of the System. Having started design with a high level statement of objectives followed by a description of performance requirements, the definition of functions which the system has to perform to meet its objectives and performance specifications takes description to a more detailed level again. Functions reflect the individual statements of work to be done in order for the system to meet its requirements; e.g.,. receive aircraft into sector, communicate with aircraft and other controllers, assimilate aircraft information from radar and from flight strips and maintain separation between aircraft. The functions should be defined whilst consideration is given to user needs which have been defined in the preceding phase.

Basic Design. A number of activities are carried out in this phase of design. The first is functional allocation, which involves division of functions between software, hardware and people. An example of such a consideration would be: should flight strips be updated by hand by controllers or automatically by machine and displayed on a screen? The relative capabilities of people and machines are well documented and may be referred to during the process of functional allocation. Such documented capabilities should only be used as guidelines, however, as the context within which the system will operate may influence decisions on allocation. Attention must also be paid to the technology available as continuing advances mean that capabilities are likely to change (Sanders & McCormick, 1987).

For those functions which are allocated to the human component in the system, the performance requirements need to be determined. Such performance requirements can be used later during testing and certification processes.

When human performance requirements are clear, then a task analysis is necessary to break down the human function into tasks which contribute to it. The sequence in which tasks are performed is listed and then each task is further broken down into the discrete actions required to carry it out. Diagrams representing the analysis are produced.

Task analysis allows the safety and efficiency of the system to be considered before it is constructed. It also forms the basis for designing human-machine interfaces, instruction manuals, job aids, determining personnel requirements, developing training programs and designing the evaluation of the system.

Interface Design. Following basic design, attention is turned to the design of workspace layout, controls, displays and human-computer interaction. Human Factors principles can be applied to all aspects of interface design and such principles are well documented. Summaries can be found in Sanders and McCormick (1987), Salvendy (1987) and Schneiderman (1987).

At this stage of design, it is important that principles are applied and that systems are certified as far as possible in terms of interface design, using a prototyping facility, before equipment for operation is purchased or before software becomes too costly or time consuming to repair.

Testing the Whole System

After basic design has been assessed during prototyping, it becomes necessary to test the whole integrated design in as realistic an environment as considered necessary and possible. This permits examination of the interaction of subsystems. It also allows investigation of the impact of realistic environmental variables on the whole system. The degree of realism introduced into the simulation should be decided upon with reference to the importance of intervening variables in the environment and also with reference to the safety criticality of the system.

A high fidelity simulator is appropriate for safety critical systems like air traffic control systems.

At the Air Traffic Control Evaluation Unit at Bournemouth, a simulation facility exists which is used for the final stages of development and then the evaluation of new air traffic control systems. It is a somewhat flexible facility which can be used to simulate a variety of ATC systems.

A description of this system can be used to illustrate how the Civil Aviation Authority has made steps towards human factors certification of new air traffic control systems.

Simulation Facilities at the Air Traffic Control Evaluation Unit (ATCEU)

The simulation facility at the ATCEU consists of two full replicas of air traffic operations rooms. The operations rooms are equipped to represent the two main ATC systems being developed for the UK at present. The operations rooms can also be configured to replicate various other ATC operations in terms of airspace, traffic and procedures.

In the text that follows, the central control function (CCF) development will be used to illustrate how the simulation facility is used to develop and evaluate future ATC systems. The CCF development is concerned with the airspace comprising the London Terminal Maneuvering Area (TMA).

The TMA airspace is made up of thirty-two control positions dealing with three airports – Heathrow, Gatwick and Stanstead. The development is being implemented in three phases. These discrete stages in the development of the overall system have been, and continue to be examined in a series of simulations which will span approximately ten years.

The operations room at the ATCEU is equipped with the new radar system, information display systems, flight strips and telephone systems which have been developed specifically for the CCF operation.

For each simulation the airspace and air traffic in question are computer generated along with flight strips and accompanying information content for the information display systems. Six to twelve traffic scenarios of 1 1/2 hours duration are generally used during each simulation.

Each of the computer generated aircraft is controlled by a 'pseudo pilot' who is usually an air traffic assistant who reacts to instructions and communicates as much like a real pilot as possible. During a typical simulation, about 21 of the sectors will be simulated and this requires 30 pseudo pilots to 'fly' the aircraft.

Each simulation typically lasts for 3 weeks during which current licensed controllers operate the control positions as they will be operated in the real world.

The objectives for each simulation are set by the designers of the system who are usually air traffic controllers. In development simulations objectives may reflect options in terms of airspace division, routes and procedures to be tested so that the most appropriate can be chosen for use. Equipment also undergoes final tailoring at this stage. In evaluations, the objectives reflect overall concerns about the operability of the system for real world implementation.

Experimental Method and Design for Simulations

During development simulations design options for airspace, routes and procedures are under examination. There may also be new pieces of equipment to examine as part of the whole system in realistic conditions.

A controlled experimental design is necessary which enables the air traffic controllers participating to see all options being examined from as many control positions for which they are valid, for as many of the traffic samples as possible. The time for which a simulation can be run is limited by cost and the limited amount of time for which operational controllers can be released from their work. This means that a completely balanced design is not possible and usually controllers do not see all the traffic samples from all control positions.

During an evaluation, the final system design is tested over one or two three to four week periods to examine operability for implementation. Because there are no design options to be tested, it is possible for air traffic controllers to experience the system from all control positions for which they are valid for all of the traffic samples which are produced. The number of exposures to the system which controllers experience during one or two evaluations makes results reasonably valid.

Measurement During Simulation

The measurements taken at the ATCEU currently fall into two of the categories defined at the start of this paper. These are system output and subjective responses.

There are three main aims behind the measurements taken during simulations. The first is to find out whether a new air traffic control system is acceptable and workable from the air traffic controllers perspective. The second is to discover what effect a new system has on the aircraft and whether it achieves what it set out to achieve in terms of aircraft movements. Thirdly, the relationships between aircraft movements and controller workload and opinion is examined. The information gathered from system output measures is used in conjunction with subjective responses to achieve these aims.

Subjective Responses

Subjective measurements involve the air traffic controllers in expressing opinions through questionnaires and interviews, and in rating workload states during and after simulation exercises.

Questionnaires. Questionnaires are tailored to the objectives of individual simulations. They are used to ensure that all the opinions of participants are captured and that all issues relevant to the objectives are considered by each participant.

Questionnaires usually cover topics such as new airspace, routes, procedures and coordination. Communication and workload may also be covered. When a new piece of equipment is under development or evaluation, a complete questionnaire will be devoted to addressing all aspects of that equipment in detail to ensure that human factors principles are applied as far as possible and that the end result is acceptable to the users.

Interviews. Interviews are conducted during simulations if a particular issue becomes of interest or concern. It may also be decided before a simulation that interviews are the most appropriate way of addressing an issue. Depending on the purpose of the interview, an individual or small group of participants may be interviewed informally or by using a structured checklist. Interviews are usually tape recorded and transcribed.

Debriefs. Debriefs are held at regular intervals during a simulation by the ATC system designers who are usually air traffic controllers. The ATC issues underlying the system under examination are discussed. Notes are taken during such debriefs and used to augment other recorded data.

Instantaneous Subjective Assessment (ISA) – Workload Measure

The Instantaneous Subjective Assessment (ISA) is a measure of workload which was developed at the ATCEU about six years ago. It provides a means by which workload states can be recorded from 20 – 30 controllers in a dynamic way during simulation exercises.

Workload is defined for controllers using the concept of spare capacity on a five point scale (see Table 1).

At each control position there is an 'ISA Panel' containing a vertical line of five colour coded buttons, each of which corresponds to one of the five levels of workload defined above. The panel also contains two small neon lights which flash for 30 seconds every 2 minutes during a simulation exercise to prompt controllers to input their workload state at that moment.

During a simulation exercise, the ISA inputs are displayed in real time on a PC screen, known as the Real Time ISA. A colour coded square is displayed beside names of all control positions simulated for each input made every two minutes. Thus the progress of workload during a simulation exercise can be monitored. Any incidences of prolonged high or excessive workload can be investigated as they happen by observing the controller concerned and discussing the situation with ATC system designers present.

Worklaod	Level		Spare Capacity	Description
Excessive	5		None	Behind on tasks. Losing track of the full picture
High	4		Very Little	Non essential tasks suffering. Could not work at this level for long
Comfortable	3		Some	All tasks well in hand. Busy but could keep going at this level.
Relaxed	2		Ample	More than enough time for all tasks. Active less than 50% of the time.
Under utilised	1		Lots	Not enough to do. Rather boring.

Table 1. Five-Point scale defining controller workload

If certain control positions show a pattern of high workload over a number of simulation exercises, then relevant participants are asked to assess their workload further, after a simulation exercise, using the NASA Task Load Index (see below). This may also be followed by an interview to discover what the participants felt the causes of high workload were. Hence the ISA can be used in a diagnostic fashion during a simulation. Such use often leads to changes in procedures or airspace division being worked out by designers and participants and tried out. If solutions do not work then the redesign process continues.

After the simulation is complete, ISA data is tabulated per exercise in terms of percentage of time spent at low, acceptable or high workload levels.

NASA Task Load Index (NASA TLX). The NASA TLX is a measure of workload used to compliment the ISA. It is a well documented measure (Hart & Staveland 1988) which breaks workload down into six components: mental demand; physical demand; temporal demand; frustration; effort and performance.

After a simulation exercise participating controllers make workload ratings according to each of the six scales. This is done using a personal computer. Controllers also weigh the relative importance of the scales so that an overall workload score can be calculated which takes into account the relative contribution of each dimension to the task. As mentioned above, this measure of workload is used to add detail to the overall workload scores collected by the ISA.

System Criteria

The main purpose of objective measurement during simulations is to collect information concerning the detailed movements of all simulated aircraft. This information is then used primarily to look at overall system performance in terms of output; i.e., aircraft movements, climb and descent profiles and landing rates. Trends in objective data may relate directly to subjective recordings so reasons behind subjective recordings can be explained clearly.

Alternatively, objective data may help to explain inconsistencies in subjective data recordings. The two types of data complement each other well in conveying a complete picture of how well the whole system, including the human component, is working.

The system performance data can be described as reflecting qualitative and quantitative aspects of system output.

Qualitative System Performance Measures

Aircraft Track Plots. The horizontal and vertical position of every aircraft is recorded every 12 seconds to build up the track history of all aircraft during one simulation exercise. The tracks are plotted as continuous lines on a chart marked with beacons relevant to the airspace under examination. Three colours are used to represent inbound, outbound and overflying aircraft. Such plots allow examination of route keeping, route layout in relation to stacks, military areas, etc. Where routes have been found to be too close to stacks due to the limited airspace given to airspace planners, the track plots have been used as hard evidence to argue for more airspace to be given to the system under development.

Conflict Plots. Separation criteria are set during a simulation at distances applicable to the airspace under examination. When such separation standards are broken, the information is recorded again on a plot which shows where the conflict occurred in relation to beacons in the airspace under examination. Conflicts are classified according to relative positions and headings of the aircraft involved. Plots are examined for clusters of conflicts which may be indicative of poor procedures, poor route or airspace design or of high workload.

Stack Analysis. The numbers of aircraft which hold at airport stacks are recorded, along with the levels occupied and the length of time for which they held. Stack usage gives an indication of how well the traffic is flowing through the system. Levels of stack usage within the system design are assessed for acceptability by the air traffic controllers responsible for designing the system.

Slice Analysis. To examine how well the system serves the air traffic, it is sometimes necessary to examine the heights achieved by aircraft at specific sector boundaries, within specific vertical and horizontal coordinates. To do this, a 'slice,' representing the two dimensional area, is placed at the sector boundary in question and the distribution of heights achieved by aircraft in that area is measured.

Profile Plots. The profile of aircraft tracks into and out of airports can be plotted to allow examination of the climb or descent profiles achieved by aircraft. If aircraft are held down or up due to inefficient airspace or procedures design, the proportion can be calculated. If this is unacceptable in Air Traffic movement terms then changes can be made.

Quantitative System Performance Measures

Aircraft on Frequency. The flow rates of aircraft are reflected in terms of total number of aircraft which were handled by a sector during a simulation exercise, peak number on frequency at one time, and the average number per hour through the sector. In this way a

check can be made of whether the new system increases the flow of aircraft by the number aimed for in the system objectives.

Landing and Takeoff Rates. Likewise the landing and takeoff rates reflect whether the system achieves that it set out to achieve in terms of aircraft movements at the airports under examination.

QSY Analysis. The location where each aircraft transfers frequency from one sector to the next can be plotted. These plots indicate numbers of aircraft transferring by location. The system designers examine this data to see that it fits in with their air traffic requirements for the system.

Speech Workload. The amount of time for which each controller is engaged in speech using the RT or the telephone or the intercom is recorded. When amount of time spent in coordination is under examination, the exact destinations of each telephone call is logged. Direct verbal coordinations may also be recorded. The aim of some new ATC systems is to reduce the number of coordinations necessary in a designated piece of airspace.

Use of Recorded Data

The range of measurements described are used to answer specific objectives which are derived during the planning phase for each simulation.

Tabulated output is aimed at directly answering the simulation objectives and for some recorded data comparisons can be made between airspace or route options simulated to judge increases or decreases in recordings, e.g., aircraft on frequency. For workload data, questionnaire data and loss of separation data experience of human factors practitioners and Air Traffic Control experts is used to judge whether the data reflects problems or not.

Trends and relationships between recordings are identified so that a picture can be built up of the way in which the whole system works. Such use of data goes some way towards what may be required in certification of an air traffic control system, but what is currently lacking is a set of approved standards against which measures can be taken. Standards would form a necessary and important part of certification and would need to be developed as a prerequisite of a certification process.

References

Bailey, R. (1982). *Human performance engineering*. New York: Prentice Hall.

Hart , S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp 139 - 183). North Holland.

Meister , D. (1971). *Human factors: Theory and practice*. New York: Wiley.

Sanders, M. S., & McCormick, E. J. (1987). *Human factors in engineering and design*. New York: McGraw-Hill.