Human Factors Certification of Advanced Aviation Technologies

Aviation Human Factors Series from
The Center for Applied Human Factors in Aviation (CAHFA):

- Verification and Validation of Complex Systems: Additional Human Factors Issues (1993)
- Human Factors and Advanced Aviation Technologies (1993)
- Situational Awareness in Complex Systems (1994)
- Human Factors Certification of Advanced Technologies (1994)

L. W. Hennessy
Series Editor

Human Factors Certification of Advanced Aviation Technologies

Edited by

John A. Wise
Center for Aviation/Aerospace Research
Embry-Riddle Aeronautical University
Daytona Beach, FL 32114-3900, USA

V. David Hopkin
United Kingdom Civil Aviation Authority
Farnborough, Hampshire GU14 6SZ, United Kingdom

Daniel J. Garland
Center for Aviation/Aerospace Research
Embry-Riddle Aeronautical University
Daytona Beach, FL 32114-3900, USA

Embry-Riddle Aeronautical University Press
Daytona Beach, Florida USA
Published in Cooperation with Eurocontrol

Proceedings of the Human Factors Certification of Advanced Aviation Technologies Conference held at the Chateaû de Bonas, near Toulouse, France, July 19-23, 1993.

Technical Editing by

L. W. Hennessy
Center for Aviation/Aerospace Research
Embry-Riddle Aeronautical University
Daytona Beach, FL 32114-3900, USA

ISBN 1-884099-03-3

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in other ways, and storage in data banks.

For information, write to:
Embry-Riddle Aeronautical University Press
600 South Clyde Morris Blvd
Daytona Beach FL 32114

Tel: (904) 226-6385
Fax: (904) 226-7050

©Embry-Riddle Aeronautical University Press 1994
Daytona Beach
Printed in the United States

Sponsored by:

- Eurocontrol
- U. S. Federal Aviation Administration
- Direction Générale de l'Aviation Civile
- National Aeronautics and Space Administration
- Embry-Riddle Aeronautical University
- Research Institute for Information, Science, & Engineering (RIISE)

Acknowledgments

The editors would like to acknowledge the work of those individuals whose dedication and untiring effort made possible the publication of this manuscript. In a way, each of them should have his or her name associated with this text because without any one of them, the text would not exist.

We must thank a large number of people and organizations for the success of the Institute, beginning with our sponsors. Only through the support of our sponsors was it possible to undertake the Institute and to bring representatives from so many countries together for a one week period. The sponsors for the conference included:

- Eurocontrol
- U. S. Federal Aviation Administration
- Direction Générale de l'Aviation Civile
- National Aeronautics and Space Administration
- Embry-Riddle Aeronautical University
- Research Institute for Information, Science, & Engineering (RIISE)

We are grateful to the Institute's staff who worked hard before, during, and after the meeting. The Editors owe a significant debt to Leonard Hennessy for his technical editing of the proceedings and in particular for his work with those authors who do not have English as their first language. The outstanding work in video and audio recording for the conference by Taylor Bracey directly contributed to the success of the meetings. Finally, we must thank the students who assisted in the preparation of the papers for publication: Jean-Philippe Cottenceau, Diane Farrow, Patrick Guide, Michel Masson, David McBride, Laurence Rognin, and Mark Wise.

The participants contributed actively in the exchange of their views and experiences which were drawn from a diversity of backgrounds and national origins. We would like to thank all of the participants for their contributions to the discussions throughout the conference and for their preparation of their individual position papers.

As Co-Directors of the Institute, it was our good fortune to have received the support of these many individuals and the sponsoring organizations in bring this volume to publication.

John A. Wise
V. David Hopkin
Daniel J. Garland

PRECEDING PAGE BLANK NOT FILMED

PAGE ii INTENTIONALLY BLANK

Table of Contents

Introduction

Philosophies of Human Factors Certification

Practical Approaches to Human Factors Certification

PRECEDING PAGE BLANK NOT FILMED

PAGE xii INTENTIONALLY BLANK

Selection and Training

Parallel Views and Topics

Reflections of Certification in Aviation

Issues in Future Design and Certification of Complex Systems

Conclusion

Appendices: The Group Papers and Participants

Preface

In several nations during recent years there have been independent and authoritative statements proposing the application of human factors principles and evidence to certification processes. Although a majority of these proposals have originated from aviation contexts, they have often envisaged that human factors should be applied wherever certification is currently conducted. The initial products of such broad application would not be system specific but could be used in many contexts, if necessary with some adaptation for particular systems.

The problems that the application of human factors to certification might help to resolve appear to be most serious in large human-machine systems that share certain definable characteristics. Such systems have a combination of complexity and integrality, are safety critical in that the penalties for any fallibility in certification are high, and require some revision or extension of any existing validation and certification procedures, all of which have evolved separately for different components or aspects of the system and none of which can deal with the whole system as a functioning entity.

This volume contains edited papers from a meeting held at the Chateau de Bonas in France in July, 1993 to launch the application of human factors to certification. Specialists with relevant knowledge and experience met for a few days in order to review appropriate subject matter, to define and evaluate feasible approaches, to consider applications and implications, to specify topic issues and topic boundaries, and to recommend productive courses of action. Some of the papers in this volume are tidied versions of texts prepared in advance of the meeting and they therefore emphasize ideas brought to the meeting, but many have been revised substantially since the meeting (a few have being completely re-written) to incorporate new or revised views formed then or since. Much of the time during the meeting was spent in small groups that discussed assigned themes and presented their conclusions in plenary sessions, and the Appendix records some of the ideas from these small group discussions. This volume contains much of the background material for the meeting and mentions many of its ideas and proposals, but it is not simply a record of the proceedings.

The editors encouraged those invited to prepare material for the meeting to choose their own approaches to the general theme. Among the wealth of ideas produced, some may initially seem idealistic or over-ambitious, but they represent aspirations and goals which would denote the successful application of human factors to certification if they could be achieved, and diluted versions of objectives do not serve long term interests well. The main concerns were to generate ideas and provide a forum for informed discussion and evaluation of them, and not to pre-judge their value or durability which would emerge from subsequent events. An objective of this initial meeting on human factors and certification was to start to define what the subject-matter, approaches and techniques should encompass, which implies some consideration of what should be ruled out as well as of what should be included. The tenuous structuring of the subject-matter devised for this volume will no doubt be revised as the needs become better defined.

The two introductory papers contribute to the framework for the meeting. Hopkin is concerned with ensuring that human factors contributions to certification would not only benefit certification but also reflect credit on human factors through their high quality. Endsley's paper poses the six basic questions which were discussed by the small groups, and thus provides the rationale for the proceedings of the meeting though not for these texts in advance of it.

The next six papers present a variety of philosophies and propositions about human factors certification. Wise and Wise compare a bottom-up approach with the more top-down approach which they equate with a systems approach and prefer. Wilson explores what is currently valuable in certification processes as a step towards realising further benefits, such as better designs and more professionalism within human factors. Hancock was stimulated by discussions at the meeting on the reasons for certification to draw parallels between certification and legislation as processes that constrain diverse and unpredictable entities by imposing frameworks upon them. Hanes approaches human factors certification by coupling the need to increase human factors participation in design processes with the practical means by which this could be achieved. Stein and Wagner, in considering certification in the context of system validaton, also emphasise the legal status and functions of certification procedures and processes. Koelman discusses the Operational Concept of future air traffic management systems in Europe in terms of their combination of tactics and strategies, and notes the intention to validate but not to certify the Operational Concept.

A series of five papers all adopt a more direct approach to certification activities. Jackson identifies some currently neglected human factors aspects of existing certification processes, singling out some social aspects of cognition as of particular significance. Evans believes that the successful application of certification processes implies adequate human factors involvement throughout the system design. Taylor and MacLeod draw on their experience of compliance with human engineering standards in the procurement of advanced aviation system, and emphasise that certification should require proof of process as well as proof of content and performance. Gilson and Abbott, in proposing that flight crews be certified for attaining mastery of sophisticated control systems, also propose criteria through which flight crews would be able to demonstrate that they had attained the requisite deeper levels of understanding. Small and Rouse consider system evaluation as a precursor to certification, and draw the distinction that whereas evaluation can show that the system behaves correctly, certification must show that it can only behave correctly.

Three papers then deal with aspects of human-machine integration. Gibson noted that all the models for the certification of flight training have to rely ultimately either on expert opinion or on the actual outcome of training. Haglund describes a project in Sweden that sought to improve the selection of air traffic controllers and reduce training costs but revealed the limited value of the selection procedure and a need to revise it. Macleod and Taylor argue that since the human's role in human-machine systems has evolved into being primarily cognitive, human factors certification must include a substantial understanding of and provision for human cognition in order to be effective.

A group of six papers approach certification issues by discussing topics that should feature in any established human factors certification of advanced aviation systems. Bernard is concerned lest certification has a deregulatory effect because the certification of a system could lead to loss of control over its subsequent evolution. Tattersall discusses the effects of workload, in terms of its methods of measurement and of alternative patterns of adjustment to variations in workload demands. Hancock notes that infallible certification must be inherently unattainable and suggests that attempts to devise systems that are not only generative and explorative but also skillful offer a more promising alternative. Day is concerned by the continuing entanglement of controllers and machines in relation to the attribution of blame for any failures, and foresees an extension into certification of the problems of this kind that are already encountered in verification. Westrum, drawing on the role of the test pilot, suggests that a test controller for air traffic control systems could bring comparable advantages, particularly with respect to user involvement in system evolution. Bukasa views the application of human factors to certification as a welcome sign of increased formalization and institutionalization of human factors contributions in regard to complex systems.

The next six papers all use ideas that already exist in aviation as a basis for discussing certification issues. Baldwin addresses the question of how human factors certification should be managed and organised, and notes that the reluctance of human factors specialists to commit themselves must be overcome. Harwood and Sanford point towards possible fresh approaches to certification by describing the field exposure of a forthcoming air traffic control automation system early in its development cycle instead of immediately prior to its implementation. Maurino and Galotti suggest how human factors certification requirements might be integrated into current certification processes, on the basis of existing ICAO regulatory requirements and guidance material. Paries reviews some aspects of current airworthiness regulations and certification processes related to human factors and cockpit design from the perspective of the occurrence of accidents with aircraft that have been properly certificated. McClumpha and Rudisill consider human factors aspects of civil flight deck certification with particular emphasis on the human-computer interfaces in automated cockpits.

Three papers address a variety of issues that arise in relation to the certification of complex future systems. Amalberti and Wibaux, on the basis of French experience with human factors in the certification of advanced automated cockpits, identify the three main difficulties as the relationship between human error and accident risk, the evolutionary nature of pilot expertise in contrast to the non-evolutionary certification requirements, and the status of human factors findings in relation to certification goals. Gaillard and Leroux contrast cognitive engineering as an approach to designing new tools with traditional methods of human-machine system evaluation and validation. Javaux, Masson and De Keyser emphasise that the combination of complexity and lack of transparency that characterises much current automation also limits the userís ability to cooperate with such systems.

In two concluding papers, Hopkin describes some current characteristics of human factors as a discipline that would influence its application to certification, and notes topics that seem pertinent to the theme of the meeting but were not discussed there. Debons and Horne make recommendations for further progress based on their retrospective analysis of proceedings in order to identify needs, and they recommend consideration of an independent agency for human factors certification.

In the Appendix, the notes are presented from the six discussion groups, respectively tasked to consider issues of what, why, who, when, where, and how, in regard to the application of human factors to certification. These notes are largely unedited but remain substantially in the forms in which the groups chose to present them. Their ideas, frameworks, and provocations should constitute steps towards further progress.

The success of a meeting of this kind depends on a great deal of work by many people, before, during, and after the meeting.

Introduction

PRECEDING PAGE BLANK NOT FILMED

PAGE xviii INTENTIONALLY BLANK

1.

Optimizing Human Factors Contributions

V. David Hopkin

Independent Human Factors Consultant

Introduction

Human factors as a discipline is concerned primarily, though not exclusively, with people in their work environments. Its objectives are to promote safety, efficiency, well-being, and productivity by ensuring a good match between the human and the job. This match seeks to optimize the relationship between human strengths and weaknesses on the one hand, and the tasks, equipment and demands of the job on the other. Insofar as human factors is based on a knowledge and understanding of human beings, it can be applied to every work environment and to every kind of human activity that constitutes human work. Accordingly, human factors can, and should, be applied to certification processes in general, and to certification processes in aviation in particular. Based on fundamental human factors knowledge about what humans are capable of and what they have difficulty with, the application of human factors should result in improvements and also in a better understanding of why the improvements have occurred, and of how further improvements might be made.

In principle there appear to be two different ways of relating human factors and certification processes. One starts from current certification objectives and practices, and suggests how human factors may be applied to them so that those objectives and practices are more effectively realized. This can be construed as a form of validation of existing certification objectives and methods (Wise, Hopkin, & Stager, 1993). Alternatively, human factors is applied primarily to certification as a process rather than as a product. The aim is to make that process an example of good human factors. A probable consequence of this approach is that because of human capabilities and limitations revealed, some of the objectives and practices of certification should be modified. This paper looks primarily at the second of these alternatives.

Examples of Alternative Approaches

All applications of human factors as a discipline should meet professional standards. They should constitute "good human factors" insofar as human factors is an independent discipline with its own criteria of what is and what is not safe, efficient, satisfactory, acceptable, optimal and compatible with good professional practice. There will occasionally be circumstances when

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

[redacted]

human factors requirements seem at least initially to be incompatible with those of other disciplines, and these requirements have to be reconciled to achieve practical agreement on what to do, yet without compromising basic human factors principles. It is possible to apply human factors evidence to existing certification processes, but it does not follow that the processes are acceptable in human factors terms, meet human factors requirements, or have satisfied some form of human factors audit.

A practical example can illustrate the difference. It is possible to certify a three pointer altimeter by applying human factors knowledge to it, and by making it as good in human factors terms as any three pointer altimeter can ever be. But as an altimeter it is still potentially dangerous under certain circumstances, with a history of human error because it remains too easy to misread by 10 thousand feet (Hawkins, 1993). What is wrong with the three pointer altimeter, in human factors terms, is that it has three pointers. A human factors approach which concentrates on good human factors would emphasize that no three pointer altimeter can ever meet basic human factors requirements satisfactorily because its well documented deficiencies are intrinsic to it, and changes to it are at best palliatives that cannot achieve the optimum human factors recommendation for the depiction of altitude in cockpits.

Human factors also specifies the information criteria to be satisfied. It must be possible not only to provide the information which pointers can give quite well, such as rate of movement and rate of change of movement, but also to incorporate a digital read-out of altitude. The way in which this digital read-out is portrayed must itself meet stringent human factors requirements in terms of character design, brightness contrast ratio, and size and legibility of numerals, taking account of the diversity of ambient lighting and pilots' minimum eyesight standards. Any residual error rates in readings must be demonstrably low, with the remaining sources of error either being readily noticed and correctable or erring towards safety rather than towards danger so that, for example, an aircraft may occasionally be too high but will never be too low and fly into the ground because of an altimeter misreading. This example can encapsulate the difference between the two possible human factors approaches to certification.

Another example, now of historical interest, concerns the attempts to introduce various peripheral vision directors into cockpits. The theoretical premise behind these was that they could visually replicate aspects of the expanding visual field on final approach to an airfield and thus provide the pilot with intuitive and non-distracting information very much as a streaming peripheral world does (Hopkin, 1962). Unfortunately, this did not work out in practice. The principles of a peripheral vision display depended upon human sensitivity to movement in the visual periphery which indeed is good, but also required the abilities to sense direction of movement, rate of movement, and rate of change of movement which are poor and fallible in the periphery. Thus the principles on which the instrument was based constituted an oversimplification and it never did provide additional information as a bonus without constituting a distraction or disrupting attention, a limitation which it was intended to circumvent. Certification for the purpose of meeting objectives which tried to optimize, in human factors terms, the "methods of portrayal" could be done, but good human factors practice would suggest that such a device cannot attain its objectives fully because of known human limitations.

2-3

PRECEDING PAGE BLANK NOT FILMED

Previous New Applications of Human Factors

Another argument for concentrating on good human factors in certification rather than simply the meeting of objectives is that the former has been the more customary practice when human factors has been introduced to a new application for the first time. The approach to the study of aviation maps exemplifies this (Hopkin & Taylor, 1979). A problem arose when it was found that the maps in use could not survive the photographic processing required to present them in a projected moving map display. One of the first steps was to commission a real map drawn by cartographic draftsmen, printed on cartographic presses and using standard cartographic paper and inks. The specification of this map was drawn up on the basis that it was not primarily treated as a map, but as an information display. Relevant evidence in human factors handbooks, standards and guidelines was applied to compile the map specification according to these perceptual principles for the portrayal of information, and coupled with a task analysis of how the map was used and what it was used for. A full description was derived of how and to what extent the photographic processing affected the cartographic information categories on the map.

The resulting product was received with some skepticism in the cartographic world because it violated some traditional cartographic principles. However, it proved better than expected, and it provided an excellent experimental tool to ascertain which of the existing human factors visual principles could be extrapolated to maps and which could not. Maps were far more complex visually than most of the previous applications of those perceptual principles at the time. The experimental map provided a kind of short cut (Taylor, 1976). It represented good human factors and it revealed where there were incipient incompatibilities or discrepancies between existing cartographic practices, objectives and forms of validation and those which would receive a human factors stamp of approval. In such circumstances it is vital not to presuppose that either discipline, whether cartography or human factors, has a monopoly of wisdom. Usually each discipline introduces further factors which the other has not hitherto considered to be relevant. Once identified, these explain disagreements and point towards optimum compromises.

Categorizations

If human factors as a discipline is applied solely to help existing certification objectives to be met, then the existing categorization of certification practices would be accepted. For example, the distinction between the certification of equipment, of procedures, and of personnel would be retained, although this distinction seems dubious in human factors terms, partly because of the apparent extent of their interdependence. An approach which tries to optimize the human factors however, would examine possible alternatives to existing classifications of certification processes for their compatibility with human factors objectives, with a view to identifying feasible alternatives and resolving any differences. The purpose here is not to be obstructive or stir contention and it certainly is not to call into question current certification processes. Such an approach would be foolish because the ultimate application of human factors to certification requires the collaboration of those concerned with certification, just as the introduction of human factors into cartography relied on the collaboration of cartographers. The objective is to ensure that human factors as a discipline is not initially compromised. The known requirements that have to be met in relation to certification are stated in their own right also from the outset, for they must not be compromised beforehand either. The best outcomes cannot be achieved if either of the disciplines, whether human factors or certification, has been compromised before the attempt to reach practical solutions has even begun.

Human Factors Professional Practices

Another reason for preferring the optimization of human factors in applications is that at some point the question arises of who has the authority to rule whether the human factors contributions to the certification process constitute good human factors practice. An authoritative view on this can scarcely come from outside the discipline of human factors itself. A procedure may be required that is somewhat analogous to submitting papers to journals for professional and peer review. Depth of knowledge of human factors and an understanding of certification are needed in order to pronounce authoritatively on the quality of the human factors work applied to it.

One of the issues that arises in applying human factors to certification is whether human factors practices and recommendations from other applications can be transferred to certification, or certification poses a significant proportion of different human factors problems that are specific to it, for which new solutions have to be found and proved. Maps again provide an example: the tendency was to treat their human factors problems as unique because maps had the visual complexity of their coding greatly exceeded that of most other displays to which human factors principles of visual information coding had been applied. It is also difficult to justify the transfer of existing human factors practices if they themselves do not constitute good human factors but are simply a set of empirical practices which have sufficed to meet objectives elsewhere, but for which no claim that they constitute ideal human factors can be proved. If they are simply a means of optimizing (in human factors terms) solutions adopted by others without reference to human factors implications, then they are scarcely suitable for transferring out of context to another application, whether it is certification, validation, maintenance or whatever.

One aspect where it seems essential to insist on good human factors concerns the issue of teachability. It is possible to devise certification or other processes which may be fulfilled by those select few who have devised them but are very difficult indeed for others to implement because they cannot be taught, although they can be done. An aspect of good human factors is to insist that teachability is examined so that a new option is not rendered impractical because of insuperable training problems.

The Introduction of Human Factors into Certification

A central issue is how to get started in any process of applying human factors to certification. How could it be applied? What would it be necessary to do? How would it be possible to recognize what kinds of functions are suitable for human factors certification? What procedures must be followed to achieve a human factors certification of equipment, of procedures, of personnel, or of all or some of them in combination? Commonly, one may start with a task analysis of the certification processes, but human factors also addresses those processes themselves and the extent to which the processes meet human factors requirements or could be improved or modified to do so. This is not to question the competence, skills, knowledge and experience of those currently concerned with certification, but to bring out how knowledge of human capabilities and limitations might enhance these processes, make them more efficient, more reliable, quicker, and more consistent. It follows if the emphasis is on good human factors that a human factors specialist must have the authority to pronounce whether the human factors is good (not whether the certification is good). It also follows that people with specialist human factors knowledge need to have a practical role in the certification processes. This does not mean, and is not intended to mean, that they should serve the functions of those concerned with certification, or that they need to have the whole range of skill, knowledge and experience of professional certification specialists. It is common to find in various certification procedures that it is necessary to call on other professional expertise to pronounce on particular aspects of what is being certified. Numerous disciplines may contribute in this way to check for example, that various kinds of engineered items meet the requirements, that the software is reliable, that there are no medical problems and so on. A human factors contribution would be to certify that there are no serious human factors problems either, and the role would be analogous to that of other professions whose knowledge may be needed to complete the certification process in certain circumstances.

Preferably, this should not be a random or apparently arbitrary process. Usually the human factors specialist is not only the best to say whether human factors problems have been satisfactorily resolved, but also to identify the human factors issues present. People in other disciplines do not compromise the fundamentals of their disciplines in order to provide a certification that may not be adequate. If the software is unsatisfactory, the software specialist must say so. Similarly, if the human factors aspects are unsatisfactory, the human factors specialist must say so.

This raises the question of the quality of existing human factors data, and whether it is suitable for such certification purposes (Boff & Lincoln, 1988). Obviously much of the evidence, for example, about portrayal of information, characteristics of input devices, and the criteria the communications channels must meet, is thoroughly reputable and well validated in most circumstances, but some cognitive recommendations, about the roles of human memory and understanding for example, may rely on fewer or less well established data. Every discipline contributing to the certification process can only use the best data available and those data become more suitable for certification as the discipline advances. Human factors is not an exception to this. One of the reasons however, for needing a human factors specialist, and one of the constituents of good human factors, is a knowledge of the evidence on which human factors guidelines and recommendations are based. Knowledge of the strength of that evidence indicates how far that evidence may be compromised or modified to meet certification requirements and where it must not be.

Résumé

It is contended that the application of human factors to certification should be regarded primarily as a process rather than a product, and that the human factors applications should exemplify good human factors practice even though they may prescribe changes in existing certification processes. Some human factors problems in certification can have no optimum solution unless the human limitations in which they originate can be traced. Certification will reveal relevant issues that are familiar to one discipline and unknown to another. These are a potential source of disagreement but also point towards compromises. Human factors specialists should make contributions to certification processes which guarantee recognition of human factors implications. The objectives are not to oust those whose profession is certification nor to undermine their authority, but to exercise human factors influences throughout certification and to demonstrate the acceptability of certification processes in human factors terms. The relationship between human factors and certification should be collaborative and mutually beneficial, since both share the same ultimate system objectives.

References

Boff, K.R. & Lincoln, J. (Eds.) (1988). *Engineering Data Compendium: Human Perception and Performance.* Ohio: Harry G. Armstrong Aerospace Medical Research Laboratory.

Hawkins, F.H. (1993). *Human Factors in Flight.* Aldershot: Ashgate.

Hopkin, V.D. (1962). Peripheral vision and flight information. In: Barbour, A.B. & Whittingham, H.E. (Eds). *Human Problems of Supersonic and Hypersonic Flight.* Oxford: Pergamon.

Hopkin, V.D. & Taylor, R.M. (1979). Human Factors in the Design and Evaluation of Aviation Maps. Paris, NATO AGARDograph No. 225.

Taylor, R.M. (1976). Human factors in the design and evaluation of an experimental 1:250,000 scale topographical map. Farnborough: RAF Institute of Aviation Medicine Report No. 545.

Wise, J.A., Hopkin, V.D., & Stager, P. (Eds) (1993). *Verification and Validation of Complex Systems: Human Factors Issues.* Berlin: Springer Verlag. NATO ASI Series F. Vol 110.

Aviation System Certification: Challenges and Opportunities

Mica R. Endsley

Texas Tech University

Purview

In dealing with the issue of aviation system certification six basic questions must be addressed: why, what, who, when. where, and how.

Why?

Perhaps the first question should be, "why certify?" The best argument for certification is to meet the desire of insuring the safety of the flying public. If we assume that adherence to human factors design principles and guidelines will achieve safer, less error prone systems, it makes sense to require that good human factors be incorporated into system designs and to implement a certification process to see to it that this requirement is met.

There are those who would argue that we cannot ensure the safety of any system, that any certification process is bound to be limited and constrained and, therefore, that certification is costly and ineffective and should not be bothered with. While these concerns are valid, I would argue that perhaps the only thing worse than certifying is not certifying. While there is no guarantee that certification itself will result in a completely safe system, the requirement for certification will go a long way towards encouraging system developers to incorporate human factors considerations into the design process. The political and economic realities of organizations dictate that little attention be paid to human factors unless there is a requirement for it and unless the success of the system (in terms of sales and acceptability to customers) is dependent on the degree to which the system meets these requirements. While there is no guarantee that this process will result in a truly, 100% safe system, it will bring us much closer than throwing up our hands in defeat over the inadequacies of our science.

What?

A more important question is, "what should be certified?" The most obvious answer is, of course, aircraft. But what kinds of aircraft – gliders, helicopters, hot air balloons, general

Human Factors Certification of Advanced Aviation Systems
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

aviation, commercial aircraft, and/or military aircraft? The greatest emphasis has been placed on commercial and military aircraft due to the greater complexity and potential impact of these systems. Yet this is not the whole aviation system, and each part ever increasingly relates to other parts.

I will put forth that increased attention needs to be paid to human factors certification of the whole aviation system: air traffic control consoles and systems, airway facilities, and maintenance systems. Too often these systems get ignored because after all, "all the real action takes place in the cockpit." Ignoring human factors considerations in the rest of the aviation system will create weak links in the chain because safe flying depends on all of these components. The controllers and maintenance personnel involved are as susceptible to error as anyone, despite their admirable dedication and perseverance in the face of often difficult circumstances.

Many aviation accidents can be linked to mechanical failures which either lead to an accident or add to other factors to create an accident. How many of those failures could have been prevented by more/better/different maintenance that may not be currently feasible or cost effective? How many light bulbs take an hour to change, leading to pilots rushing to make up for lost time? How many important bolts get left off because putting them in takes someone with the agility of Houdini, the strength of Schwarzenegger and the endurance of Spitz? How many navigation aids go out because of the challenges they pose to their maintainers? This is a greatly neglected area for which the human factors considerations are very real and which needs to be included in certification if it is ever to receive the attention that is needed. Certification needs to incorporate *aviation systems:* aircraft cockpits and maintenance interfaces, air traffic control interfaces, and supporting aviation facilities. Only by looking at human factors issues resident in each of these systems and the *interaction between them* will the goals of certification be met.

Who?

Who should perform certification? This is a critical issue. The bottom line is that a system's customer must ultimately certify that their requirements are met. But who is the customer? An airline, a regulating body, the public? As a society we often look to regulatory agencies to assume the role of "customer" on behalf of the large body of people and numerous organizations involved.

Perhaps the better question is "who within these organizations will certify?" The requirement for certification is an empty one if personnel performing the certification do not possess the necessary knowledge in human capabilities and limitations, design principles, testing procedures, experimental methods, etc. Currently, anyone can call themselves an ergonomist or human factors professional, with or without benefit of any of this knowledge (and many do). This is disturbing. Achieving a good design (and certifying a good design) requires a lot more than simply checking off fixed rules listed in some document. It is often dependent on the integration and interaction of various system components, of the tasks that are performed, of the characteristics of the specific user population, and of their experiences with other systems not even under consideration in the design at hand. Checklists fall far short of meeting these demands. Certification requires that the persons doing the certifying have a good understanding of how people function, of what works and what does not, and of how to go about evaluating a particular combination of components constructed in unique ways.

I will therefore assert that human factors certification of aviation systems needs to be performed by those who are certified to do so. The certification of human factors professionals is, in and of itself, difficult and troublesome and has been the subject of its own heated debate which shall not be repeated here. Even though certification will not ensure that the individuals so designated practice "good human factors" in their efforts, it will serve to ensure some minimum level of knowledge deemed important to the process. The ability of individuals to perform different tasks is highly variable, despite factors like training and experience, and this is no less so for human factors professionals. Although certifying the certifiers will not insure their competence, I again will argue that the only thing worse than certifying is not certifying.

The exact nature and content of this certification is debatable. Should the person have basic competence in human factors knowledge in general, or is a specialized body of knowledge in aviation system certification required? What is required for the certifier: testing, passing certain courses, experience, and/or demonstrated know-how in practice? This issue is its own barrel of monkeys, but one which we must open if the goals of aviation certification are to be met.

When?

A factor which will largely impinge on the success of the certification process is when it occurs. Certification is generally thought of as a final process: a test to be passed at the end of a system's development as the final hurdle before employment. This, however, is actually when the least amount of good can be done. Years of development work and often millions of dollars have been spent in creating the product that is put forth for certification. When a human factors evaluations reveals deficiencies – from things which are less then desirable to things which are major problems – there is a natural organizational tendency to balk at making what, at this point, will be extremely expensive changes. Certifiers find themselves as in the position of determining what is really bad, and what may be tolerable with a lot of training. Ultimately, we hope for better from certification, but when applied in this manner, better is difficult to get.

A far more preferable way to apply certification is as an ongoing process rather than a final test. First, the system's designers need to be fully cognizant of what will be required for certification: what design guidelines are expected, what tests will be performed, and what criteria will be used. Armed with this knowledge (and the sure reality that *they will be held to it*), the savvy organization can incorporate human factors as a part of the design process. This is probably the only way certification will come close to meeting its goals. To be successful, certification requirements and procedures need to be incorporated throughout the design process by the design organization with the cooperation/advice/interaction of the certifiers. The actual certification then becomes a far simpler culminating event, and less trial by fire.

Where?

Highly related to when is the question of where. As a process, the design organization needs to have adequate facilities for incorporating human factors considerations into the design – from prototyping tools to simulation facilities, where necessary. The certifying organization also must be prepared to perform the required final testing. The ultimate certification of any system is its performance in the field over an extended period of time with its designated user population. This of course may be not only unreasonably untimely but unsafe as well. While a certain degree of field testing is probably ultimately necessary, laboratory and/or simulator-based testing are probably indicated for most systems to meet this need.

How?

The hardest question - how – has been saved for last. Given that we want to certify and have reasonably competent people for doing so, how should we really go about certifying any system? Arriving at this answer is probably more than can be achieved, even in this workshop. At best, at this point, some issues that need to be incorporated can be discussed.

First, certification needs to require that basic human factors design principles have been adhered to. We have learned a few things in the past 50 years. Many concrete do's and don'ts have been established and can be readily examined in any system's design. Many of these are summarized in established standards such as MIL-STD-1472.

Next, certification needs to address the fact that many serious human factors issues can not be or are not addressed in these guidelines and standards. Such standards are almost always out of date by the time they are printed. The rapid pace of technological change presents a moving target. As a profession, we are constantly trying to keep pace with this change to develop standards for technologies which had never been considered or for which standards were not possible before. Even worse, each innovation (and specific implementation) creates a ripple effect in changing the way people interact with other parts of the system. For this reason, certification standards need to form a "living document" that can incorporate new knowledge as its generated (in itself this will create a serious challenge for certification).

Third, the system as a whole must be certified. Most design guidelines apply to components – a gauge, a lever, a chair. Many human factors issues, though, have to do with how those components are brought together and interact with each other in the context of particular tasks and with particular user populations. The acceptability of a given system design configuration cannot be adequately assessed in a vacuum. Certification must ultimately take into account how the system as a whole is used, how various components interact with each other in producing the system's performance, and who is using it. This requires that a complete understanding of the proposed (and possible) uses of the system be determined.

Finally, evaluating the implementation of complex system components, such as automation, is a tricky job. Certification needs to incorporate the application of established procedures to address issues like this that are not readily comparable to a handy guideline. Determining what these procedures should be and appropriate criteria to be used in the process is paramount.

Certification is neither simple nor straightforward. It is, however, a worthwhile objective which can provide benefits through the realization of well designed aviation systems that come as close as possible to the goals of safety and efficiency in the aviation system.

Philosophies of Human Factors Certification

N95- 34765

On the Use of the Systems Approach to Certify Advanced Aviation Technologies

51-54
55741
p. 9

Mark A. Wise[1] & John A. Wise[2]

[1]University of Central Florida, Orlando, FL USA
[2]Embry-Riddle Aeronautical University, Daytona Beach, FL USA

Introduction

The field of human factors is as varied and diverse as the human subject itself. But one of its most important applications is the facilitation of safety and efficiency in a particular working environment through the implementation of paradigms known about the human and their working relationship with machines and systems. During the period since World War II (which is often viewed as the birth of Human Factors) no area has been the subject of more human factors research than aviation. And in no time during that epoch is the influence of human factors more important, nor more imperative than it is today.

As technology driven designs have been finding their way into the national airspace system (NAS), there has been growing concern within the aviation industry itself, the Federal Aviation Administration (FAA), and the general public for a means by which to certify complex systems and the advanced aviation technologies that will be responsible for transporting, directing, and maintaining our airborne travel. While it is widely agreed human factors certification is desirable, the philosophy that will underlie the approach is debatable.

There are, in general, two different approaches to certification: 1) the top-down or systems approach; and, 2) the bottom-up or monadical approach. The top-down approach is characterized by the underlying assumption that certification can be best achieved by looking at the system as a whole, understanding its objectives and operating environment, *then* examining the constituent parts. In an aircraft cockpit, this would be accomplished by first examining what the aircraft is supposed to do (e.g., fighter, general aviation, passenger), identifying its operating environment (IFR, VMC, combat, etc.) and looking at the entire working system which includes the hardware, software, liveware and their interactions; then, evaluative measures can be applied to the subsystems (e.g., individual instruments, CRT displays, controls).

The bottom-up approach is founded on the philosophy that the whole can be best served by first examining it constituent elements. This approach would perform the above certification completely antithetically, by looking at the individual parts and certifying good human factors applications to those parts under the basic assumption that the*whole is equal to the sum of its parts.*

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PRECEDING PAGE BLANK NOT FILMED

PAGE 14 INTENTIONALLY BLANK

This paper will attempt to form an argument for the top-down (systems) approach, while addressing arguments against it, and pointing out the shortcomings and erroneous assumptions inherent within the bottom-up approach.

Certification

To develop a cogent argument outlining the advantages of the top-down approach to certification, it must first be established what the goals of the certification process are in general, and the certification problems that human factors will attempt to overcome. Certification, in a generic sense, is the process by which a product is declared appropriate for a particular task in that it matches or exceeds a previously defined set of "design to" criteria. In being "certified," it is implicitly understood that the product will safely and effectively perform the task for which it was designed.

Of Aviation Technologies

Certification of advanced aviation technologies involves, many times, the evaluation of products which are technologically new and previously unused or untested. This, in itself, poses an interesting problem because with uncharted equipment the standards by which previous, like, products had been evaluated now become obsolete and inapplicable.

Certification of aviation technologies is also unique from some other certification problems because of the extensive and unavoidable interplay between many systems, so that the systems themselves can be looked upon as subsystems of a larger system. For example, ATC, the fleet of aircraft, and maintenance could be seen as systems unto themselves, but on a larger, more universal scale, their boundaries are not so narrowly defined; each of the aforementioned "systems" are merely players in the entire NAS. Therefore, the certification of each of these interrelated entities by themselves falls, trapped, into the quagmire of "fuzzy" certification, which is neither desirable nor acceptable.

The challenge, then, is to overcome these obstacles. The means for doing so appears to be the implementation of the systems approach to certification. Because the foundation of this theory is built on the premise that the system as a whole is more important than the parts of that system, it could be argued that, by its nature, it avoids the aforementioned problems.

The Systems Approach

A system can be defined (in a broad sense) as the collaboration of functionally similar objects (humans, machines) working towards a common goal within its respective environment. The first and most important aspect in designing any system is to clearly define the goals and objectives of the system (Christensen, 1987; Meister, 1987). In an automobile, the goal is safe, efficient, land travel; in the government, the goal (at least hypothetically) is to serve and protect the citizens; and in the airspace system, the goal is to provide safe, expedient, air transport.

Since the first step in developing a system is the identify the goals, so should the first step in certifying that system be to identify the goals, then certify that system based on those goals.

What Came First: The product or the idea?

Thomas Edison once said that a new invention consisted of "one percent inspiration and 99 percent perspiration." But, the inspiration, the invention's objective/goal, comes first. The first step in developing any system is to first define what the goals will be; it is impossible, if not inconceivable, to begin to build a system without first deciding what it is going to do. The Wright Brothers did not just begin to assemble pieces of wood and paper, only to find out, to their amazement, that the "thing" they built could fly.

Just as the goals of the system are a prerequisite for its development, so should they be a prerequisite for the system's certification. The definition of a system's goals dictates the means by which the certification of the subsequent subsystems should be handled. It seems illogical and erroneous to attempt certification of products without first considering what the ultimate goal of the product is when it is placed in the context of the system.

Starting with the system's goal and working down provides a framework within which an evaluator can examine the parts as contributing factors towards said goal. This eliminates excessive redundancy among components and does not leave room for certain vital components to be left out of the system.

Working Environment

The systems approach looks at the system and the elements of the system in their working environment, and therefore can evaluate the system's ergonomic layout. By this we are referring to the positioning of controls, instruments, displays, etc., within the system and their functional relations to each other. This would be similar to a task analysis, where one wants to examine the physical relationship of functionally similar, and operationally dependent objects in the workplace. This certification is particularly applicable when introducing a new instrument or device into a workstation, in which it ends up being placed wherever an opening is available.

Going from a bottom-up approach, a yoke, for example, could be certified to be poses all of the characteristics of a sound human factors yoke, and it is certified on this criterion, but the evaluation ends there. The yoke, in all of its glory, is not optimally usable if it is placed behind the pilot's seat by the engineers. It is analogous to writing a book of poetry in Egyptian hieroglyphics in the twentieth century; while it may contain brilliant rhyme scheme and flowing poetic prose, no one can read it – it is merely wasted paper. Something is only as good as it is usable.

In contrast to the bottom-up approach, a top-down approach would look at a working simulation of the aforementioned cockpit, and ergonomic problems like the one mentioned above would be recognized. While it is not at all likely that such a huge error would ever come about, the point is still valid, and the problem is still real.

Not only does a system's ergonomic environment need to be considered in certification, but its operational environment needs to be considered as well. By operational environment we are referring to lighting, weather, temperature, etc. The minimum light emittance needed for a display or instrument is directly related to the environment that it will be use in. Therefore it is necessary to evaluate the instrument while in those conditions. This is not easily accomplished through a bottom-up technique, but is easily evaluated within the systems approach.

It could be argued, by the bottom-up proponents, that the product's environment could be replicated during the evaluation to take into consideration the lighting, for example. But, this only really takes care of half of the problem, because another consideration is the light being emitted from other displays, the glare, due to the angle of the display in the system, etc. All of these environmental factors would theoretically be observed in the systems approach through a simulation or mock-up.

Money, Money, Money

The major drivers in any system, whether it be in the developmental, evaluative or production stages, are cost and cost efficiency. The top-down approach is cost effective in two ways: 1) the certification personnel is not required to spend the same amount of time on every product in the system because not every part of the system is forced to meet the same criterion of human factors engineering; and 2) money is saved in the production of the products because of weighted criterion..

Toilet Seats and Tool Boxes. The United States Government was under a great deal of scrutiny in the mid 1980's for purchasing miscellaneous items for its fleet of C-130's (a military transport plane) which appeared to have greatly inflated prices attached to them; some examples were $15,000 toilet seats and $5,000 wrenches. The government justified the purchases by claiming that the equipment had to be "perfect" in order to be usable and safe in their operational environment. While the prices paid for those products were *probably* justifiable (because of the research and development costs for a few production items), it gives insight to problems that could be arise out of bottom-up certification: setting outrageously high standards for a product before its relative importance in the system and the system's environment are known.

If every product in the NAS had to be evaluated by the same standards, the prices for the products would be exorbitant. No one would argue that the toilet seats on a Boing 757 should have to comply with the same human factors standards as the plane's navigation system, to take it to the logical extreme. But, where is the line drawn? How can one judge which products need to pass strict human factors and ergonomic tests, without first looking at its role within the system. The point is that you cannot. As a result, every product, every display, control and widget would be subject to the same meticulous human factors standards. This would result in exorbitant prices for the products which would be felt directly by manufacturers and indirectly by the paying airline passengers.

The systems approach *would* look at the system and the parts that make up that system and do something that the bottom-up approach cannot do: decide the relative importance of each part, and be able to make a well informed decision as to what standards by which they *need* to be evaluated. So, a rarely used, unimportant product does not have as much time and money spent in its certification as a relatively vital, often used product.

In this way, the systems approach would require a reevaluation and alteration of current certification standards. Products should to be evaluated on their functional importance: first, and human factors standards; second, where depending upon the first cause human factors certification may or may not be necessary. For example, if a job required an employee to shovel three pounds of coal from the coal pile to the furnace each day, there would be no need to certify that shovel to optimize ergonomic standards. Any shovel from the local hardware store would satisfy the requirements sufficiently, and to require anything else would be superfluous, and cost inefficient. On the other hand, if the shovel were to be used eight hours a day, five days a week then it should be subject to more stringent encompassing human factors standards.

Workload

One of the main objectives in any human factors effort is to insure a good fit between the humans and the machines they operate. This is done several ways, including personnel selection, training, manning, etc.; underlying all of these processes is an evaluation of the workload incurred by the human while operating the system, whether it be psychological or physiological.

Workload is a very important aspect of any system design, and it is something that must be examined in the certification process. The top-down and the bottom-up approaches address the issue from different angles. The bottom-up approach would look at each part of the system and measure the workload involved with running that particular part. Then by summing all of the measurements, a gauge as to the amount of total workload that would be present in the entire system should be had. The problem with this is that, once again, there should be significant differences with the parts by themselves and with the human's management of those parts when they are incorporated into the system. This method does not take into account secondary tasks. There could be a significantly different amount of workload incurred by the human when they are operating the entire system, than the original guesstimation made; and this number could be either high or low. Since, neither a high nor low workload is desirable because of documented performance deficiencies (Rohmert, 1987), there needs to be a more accurate method: the systems approach.

The systems approach would have the advantage of observing the operator while managing the entire system. Subjective and objective tests could be run to determine the amount of workload, and appropriate measures could be taken to increase the crew, decrease the crew or leave it the same.

Bottom-Up Approach

Several of the problems which are inherent within the bottom-up approach have been described above. In addition to those problems there are others which not only show this evaluative philosophy to be incompatible with certification, but indeed prove it to be undesirable as well.

Inductive Conclusions

It has been argued by many well-respected modern philosophers, including David Hume and Immanuel Kant, that inductive arguments and assumptions can never be validated. The nature of an inductive assumption is that by observing past examples of a particular event, one concludes that: in the future, a similar cause will produce the same effect. For example, if we drop a penny, at time, t_1, and it falls towards earth, it will *necessarily* fall to earth at time, t_2. According to Hume there are no necessary causal connections, and any attempt to predict the future from the past is a fallacious one that it is built on a circular argument. Kant, not as critical, said that there is causality, but we can still never validate a future event based on similar past events.

The process of certification requires us employ inductive logic. We are essentially saying that "if X works now at t_1, then it will work in the future at t_2." Induction is a necessary part of certification and cannot be overcome or circumvented. But the number of times that an inductive conclusion must be drawn can be minimized. As with all necessary evils, the less the better.

Top-down certification must only use inductive logic once: i.e., in certifying that since the system works well during the evaluation, then it will work well in the future (in production use).

Bottom-up certification employs inductive reasoning early as well as later in the certification process, thereby making the probability of error more than twice as great. The human factors certification personnel must not only certify that part X will work when placed in the system – the first case of induction; but, they also must also assume that the system will work when fully implemented – the second occurrence of induction.

Whole ≠ Sum of the Parts

"...Because undermining the foundation will cause whatever has been built upon them to fall down of its own accord." - René Descartes

It does not take an advanced degree in engineering physics to deduct that a house that is made out of bricks could not be built on a foundation made out of straw, without collapsing under its own weight. Basic physics (and common sense) tells us that any physical structure is only as sturdy as the foundation upon which it is built. Similarly, in logic, an argument is only valid if the premises upon which it is "built" are true. If the foundation is weak or the premises are shown to be untrue, the argument crumbles under the weight of invalidity. Bottom-up certification is guilty of being built on a straw foundation.

Bottom-up certification uses as its foundation the premise that the whole is equal to the sum of its parts. While this statement may be true with a jigsaw puzzle, it is certainly not true in certification, nor any scientific endeavour. As far back as Aristotle – one of the first enquiring scientific minds and logicians – it has been recognized that:

> ... We often fall into error because our conclusion is not in fact primary and commensurate universally in the sense which we think we prove it so. We make this mistake when... the subject (element) which the demonstrator takes as a whole is really only part of a larger whole.

In certification terms, Aristotle would be saying that we often err when certifying a part as a entity in and of itself, when it is truly only a part of a larger whole – the system.

Later in *Posterior Analytics*, Aristotle's argument further repudiates the use of the bottom-up approach in certification. It says that while a part can be certified by itself the truth of that certification is only applicable to the part individually; it would not be true universally, because the part is in fact different when it is placed in the system.

With this weakness exposed the foundation upon which bottom-up certification rests is undermined – the theory is invalid.

A Bad Product with Good Certification Criteria

Another, less philosophical, more practical, problem has to do with certification criteria, and the role of the certification personnel in the process. It is conceivable for a product to be valid by human factors standards without being desirable by them. The three-pointer altimeter provides an excellent example (Hopkin, 1994). The altimeter could be certified on the grounds that it provides excellent contrast, brightness and font size. It is judged that it could be visible from every part of the cockpit, and from a performance standpoint it is accurate to +/- 0.5 feet. The problem with this instrument obviously does not within its design, but within the instrument itself.

Incident reports, and experiments analysis of the three-pointer altimeter have shown that is responsible for pilot-induced errors a dangerous amount of times. Misreading of 10,000 feet are not uncommon. Therefore, perfectly sound human factors instrument, is not a good instrument. This once again ties back to the problem of certifying a product, without looking at it in its working environment. In the bottom-up approach, this instrument, could be certified; not to say that it would not in the systems approach, but it is much less likely.

A Portrait of the Artist as a Certifier. To illustrate (quite literally) an error that can occur by using a bottom-up certification process, I will use an analogy: The Analogy of an Artist as Certification Personnel.

Imagine that you were hired out as a professional art certification consultant. This job required that you look at different pieces of art, then certify whether or not they represent what they are supposed to (e.g. an eye looks like an eye; a cow looks like a cow). One of your clients, a not to bright artist, comes to you and asks you to certify an eye for him which he has recently sketched (it looks like Figure 1). The eye looks good – good proportions, proper relationship between the pupil, iris, etc. – so you give it your stamp of approval: A good eye.

Over the course of the next two months the same dimwitted artist shows you another eye, then a nose, and then a mouth, all of which look like what they should represent; again the obligatory stamp of approval given for each feature. Finally, a couple of weeks later he shows you the whole thing, which looks like Figure 2.

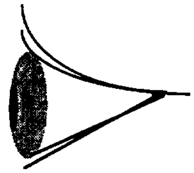


Figure 1. A good eye

The picture is poorly drawn, not because any of the parts themselves are poor, nor because they features are improperly aligned. The picture is poorly drawn because each feature was certified without knowing what the ultimate goal of the painting would be. Each feature, by itself is a good drawing (open for debate), and accurately represents the its respective object. But, when summed together, the whole is wrong.

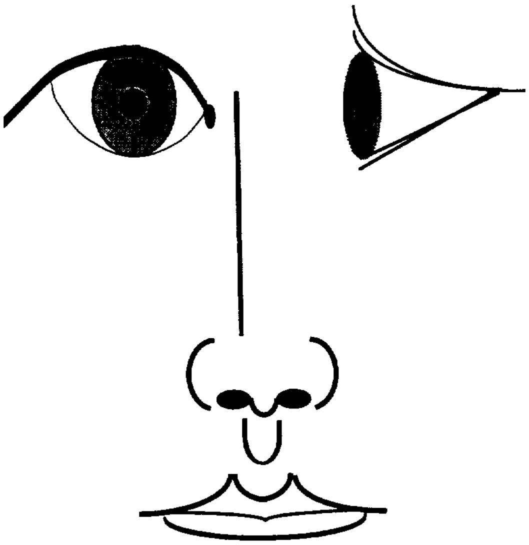


Figure 2. A bad face.

Conclusion

Certification of advanced aviation technologies should not only pose a unique professional challenge for the human factors expert as a scientist, but also as a consumer, who wants to have the safest air transportation for her and her family. To insure this safety, the best possible method of certification should be employed: the systems approach. This is not to say that the systems approach is infallible, but it certainly is superior to the bottom up-approach. The effectiveness of any certification is only as good as the individual(s) performing it. But, taking human error, or misjudgments out of the picture, the systems approach is more sound fundamentally, practically and philosophically.

It is because of its superiority and not it infallibility that the top-down approach is better suited to certification. David Hume (1977), philosophical empiricist, argued that human judgments and scientific decisions are always made after one entertains two or more opposing arguments, examining the possibility of each by weighing the their relative proofs, then believing the strongest case. "In all cases we must balance the opposite experiments, where they are opposite, and deduct the smaller number from the greater, in order to know the exact force of the superior evidence" (p. 74).

In this situation, the top-down approach provides the strongest case towards its cause; while agreeably it brandishes some problems, the positives highly outweigh the negatives. And from a Humian approach to decision making should rightly be chosen over its counterpart.

References

Aristotle. (1952). Posterior analytics. In: Hutchins, R.M *Great Books of the Western World* (G.R.G. Mure, Trans.). Chicago: Encyclopedia Britannica, Inc.

Christensen, J.M. (1987). The human factors profession. In: Salvendy, G. *Handbook of Human Factors*. New York: John Wiley & Sons.

Descartes, R. (1979). *Meditations on First Philosophy* (Donald A. Cress, Trans.). Indianapolis: Hackett Publishing Company. (Originally work published 1641).

Hopkin, V.D. (1994). Optimizing human factors contributions. In J. A. Wise, V. D. Hopkin, & D. J. Garland, (Eds.). *Human Factors Cerification of Advanced Aviation Technologies.* Daytona Beach: Embry-Riddle Aeronautical University Press.

Hume, D. (1977). *An Inquiry into Human Understanding.* Indianapolis: Hackett Publishing Company. (Original Work Published 1748).

Kant, I. (1926). *Critique of Pure Reason* (Norman Kemp Smith, Trans.). New York: St. Martin's Press. (Original work published 1787).

Meister, D. (1987). Systems design, development, and testing. In: Salvendy, G. *Handbook of Human Factors*. New York: John Wiley & Sons.

Rohmert. W. (1987). Physiological and psychological work load measurement and analysis. IIn: Salvendy, G. *Handbook of Human Factors*. New York: John Wiley & Sons.

N95- 34766

A Rose By Any Other Name: Certification Seen As Process Rather Than Content

53-03
[illegible]
p. 15

John R. Wilson

University of Nottingham

Introduction

It is perhaps indicative of perceptions and knowledge about certification that a small sample of colleagues, on hearing that I was writing this paper, all assumed it was to do with certification of human factors or ergonomics professionals. On the other hand, this may just illustrate the self-referential nature of professional groups and genuine concerns about professional certification, and these concerns are germane to the present argument since certification of systems may in the end be reduced to certification of professionals in the process (this is an issue returned to below). In fact, misunderstanding over the term "certification" may be in part responsible for misgivings and a lack of a wholehearted welcome for it, certainly amongst some ergonomists of the author's acquaintance. If it is seen as a formalisation and standardisation of their activities, then there is considerable opposition. When explained as a "design review and approvals" procedure, response seems more favourable.

Green (1990) believes that the two main factors safeguarding flying from human error are both related to certification and regulation. First is the increasingly proceduralised nature of flying whereby as much as possible is reduced to a rule-based activity. Second is the emphasis placed upon training and competency checking of aircrew in simulators and in the air, both generally and for all particular types of aircraft flown. This leaves, believes Green, other human factors that are relatively unaddressed as yet and which can give rise to human reliability problems. These include: hardware factors and especially the compatibility of control/display relationships and the way information is presented in relation to pilots' expectations; social factors and especially pilot/co-pilot relationships; and system factors including fatigue and cost/safety trade-offs. He also, importantly, identifies problems with the integration of the "electronic crew member" following increased automation. Human reliability failures with artificial intelligence and automation, due to over-reliance on the system fail-safe mechanisms, or to operator under-confidence in the integrity or self-regulating capacity of the system, or to out-of-loop effects, are widely accepted as being due to deficiencies in plant design, planning, management and maintenance more than to "operator error" – Reason's (1990) latent error or organisation pathogens argument. Reliability failures in complex systems are well enough documented to give cause for concern and at least promote a debate on the merits of a full certification programme.

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PAGE 24 INTENTIONALLY BLANK

PRECEDING PAGE BLANK NOT FILMED

The purpose of this short paper is to seek out and explore what is valuable in certification, at the least to show that the benefits outweigh the disadvantages and at best to identify positive outcomes perhaps not obtainable in other ways. On both sides of the debate on certification there is general agreement on the *need* for a better human factors perspective and effort in complex aviation systems design.

What is at issue is how this is to be promoted. It is incumbent upon opponents of certification to say how else such promotion be enabled! This is an exploratory and philosophical review, not a focused and specific one, and it will draw upon much that is not firmly in the domain of complex aviation systems.

Parallels

There used to be an unwritten "law" in work study – or motion and time study in the USA – that no one was so enthusiastic about work measurement and standardisation as those whose own jobs – they felt – precluded any possibility of such a process being applied to them. Equally, no-one was quicker than management to oppose, utterly, any attempt to assess their own work when this was suggested, on the grounds that this was inappropriate, represented an unnecessary effort, and was regardless impossible for any analyst to understand what they really do.

I saw first hand another salutary lesson within the past year or two. One of my students was attached to work with a "high flier" in a major UK consultancy; they were charged with applying quality assurance procedures to the activities of the consultancy itself. This involved vetting all areas of their operation for compliance with BS5750 and ISO9000, the relevant service quality assurance standards. The student, and the formerly popular and high achieving consultant, quickly became the villains of the group, pariahs to be avoided, because they were seeking to propose some formality for the consultants' work, some prescription of how they should operate. And what was the core activity of this consultancy group? Advising industry on the need, processes and procedures for quality assurance!!

What all this illustrates is the difference that perspective and standpoint make to opinions on formal systems, appraisals, standards and review processes. The reviewers or appraisers see them as bringing about order and rationality, and as ensuring that "the best" is retained and "the worst" is identified and eliminated. The reviewed or appraised, on the other hand, see formal systems as restrictive and petty, unnecessary interferences with their activities, and as leading to "throwing the baby out with the bathwater."

What other parallels can be drawn? First of course, and top of the agenda for many in the human factors community, is the notion of certification for human factors and ergonomics professionals. This is taking place in proposals for the Centre for Registration of Ergonomists in Europe (CREE) scheme, but has substance with the Board of Certification in Professional Ergonomics in the USA. A recent "Provocations" article in Ergonomics in Design (Senders & Harwood, 1993) took contributions from both sides of the argument. Senders and Harwood themselves point out that one danger is of formality driving out reality, in that any degree or certificate may become more important in itself than the individual's competence it implies. (Educators are, in fact, often faced with this from their students, when attempts to discuss and explore ideas are met by the students' desire to digest directive information tailored to an examination setting.) In the same piece, Schumacher and Dorst question certification in terms of need, process and impact. If we summarise and generalise some of their objections, these are:

- certification does not *ensure* quality (or integrity or ethics)
- what position will certification have in law?
- do we want the homogeneity that certification might bring?
- prescriptive criteria may stifle innovative designs.

Perhaps reflecting the difficulty of doing so, the response from Hendrick in the same paper makes little attempt to answer directly some of the questions about whether professional competence certification is needed at all, concentrating instead on defending the processes involved. He does, though, argue that certification can promote growth of a discipline and its image, although admitting that the process cannot guarantee "worthwhile job performance... competency...[or]...conformance to ethical, moral and professional standards" (Senders & Harwood, 1993).

What then can possibly be the arguments for certification, if it cannot even guarantee basic compliancies? To find some value of certification, perhaps we can look at other areas of ergonomics. In the domain of work organisation and job design, and in particular the implementation of changed work structures, it is often argued that the content of any change is of minor importance for successful outcomes compared to the change process. If the best process possible is put in place then we can afford to "change the change", iterating on the actual content as required. Translating this idea to certification in complex aviation systems, we could look upon certification as the means by which an improved development process is enabled, rather than as a limitation and detailed specification of the content of the development.

From the fields of product liability (in the 1970's and 1980's) and health and safety at work (in the 1990's), we can see some of the more systemic benefits possibly accruing from certification. Out of the imposition of regimes of strict liability have come better processes or systems of design amongst producers, and movement to a more beneficial standards regime, that of horizontal standards, amongst the lawmakers (see below). Consequences of the health and safety (ergonomics) legislation implemented in 1992/1993 across member states of the European Community as a result of EC Directives are even more marked. Although it would not be appropriate to be too starry-eyed and naive about beneficial outcomes, it certainly seems as if the need for employer conformance with ergonomics criteria and practices has stimulated the production of tools, techniques, instruments and methodologies for investigation and diagnosis that will be of value across a range of concerns. Not all of these developments are to be widely welcomed of course – I imagine we have all been shocked by some of the so-called "ergonomics aids" now on the market – but the overall effect in general has been one of dynamic growth in the discipline. The ergonomics community itself has had to produce new approaches and techniques, improve the validation of existing ones, and generally ensure greater justification for its guidelines and recommendations. Even colleagues who are dubious about the value or validity of specific requirements in the new regulations have been pleasantly surprised by the consequent pressures for quality in methods.

MANPRINT - Lessons Learned

Can we learn from very close parallels to human factors certification in complex aviation systems? Nuclear or military systems perhaps form comparable domains to an extent.

The well known MANPRINT developed for procurement in the U.S. Army has been adopted by both the British Army and, in modified form, the Royal Navy. Reasons claimed by the British Ministry of Defence for its adoption are as follows (MoD, 1992):

- The perceived success of MANPRINT in the USA, including improved maintainability of equipment and use of analytical techniques to ensure wider and better usability;
- A desire to identify and achieve the best balance between people and equipment. It is recognised that "high quality, multi-capable...better equipped motivated and properly trained" personnel will only result if manpower issues are considered as part of the equipment procurement process.
- Personnel costs now outweigh equipment costs and it is hoped to provide more control over these by being better able to anticipate, budget for or reduce costs through design improvements. It is also anticipated that MANPRINT would give a better specification generally and thus more cost-effective products, reducing overmanning, poor performance and errors.
- Better working conditions and reduced training costs;
- Greater requirements for cognitive skills rather than physical and the shrinking pool of skilled labour available, mean it is desirable to constrain designers to produce operable equipment for existing specified personnel. These reasons are labelled "skills drift" and "demographic trough or slide" respectively (Goom, 1993).
- New health and safety legislation applies to military as well as civilian systems and "covers areas which have traditionally been regarded as usability rather than safety issues", again an argument for the broad approach of MANPRINT.
- MANPRINT covers manpower, personnel, training, human factors engineering, health hazard assessment, and system safety (and habitability and environmental ergonomics for the navy). This ensures a single source of advice across all human factors issues, thus preventing or reducing sub-optimality in systems design.

Such support for MANPRINT and presumably for similar certification systems raises three questions. *First, are these claimed advantages real, secondly are they important, thirdly are they generalisable, especially to civil systems?* A sceptic might answer "no", "partly", and "no" to these questions. Certainly it is easy to be cynical about any claims on the part of the military establishment to be making efforts to reduce costs for instance. Nonetheless, a more reasonable view might be to answer: "the claimed advantages seem reasonable"; "yes, they are potentially very important"; "they might be generalised to other situations in other industries."

In fact, the strongest support for the certification process might derive from the fact that the claimed advantages are as much or more concerned with process as they are with content. If the key gains reported for MANPRINT are summarised and generalised, they look like:

- better human-machine systems designs and improved usability
- improved cost-effectiveness and cost control
- widening of the user base
- compliance with ergonomics/health and safety legislation
- more efficient design process.

Generally translated into systems design then, whilst certification might seem to be a cumbersome route to go down, looking at the potential benefits then the spin-offs might be decisive in judgements as to its value.

Standards for the Certification Process

Certification will require agreement on standards so that it may be useful and feasible. Debate over type and coverage of standards has a long history in the field of product regulation for instance. "The trade-off between voluntary and mandatory standards [concerns] acceptability, applicability and ease of formulation versus possible non-compliance ... even standards enshrined in legislation are of little value unless there is strict enforcement ...." (Wilson, 1984). Much support was given to the notion of performance standards rather than construction or dimensional standards. Problems with safety standards were further identified as: their inadequacy in scope and permissable levels of risk; their not addressing all foreseeable hazards or types of behaviour; specifications that tend to be generalised, partial and inadequate; and a general lack of a standardised format. Although these criticisms are still valid today for product ergonomics standards, there has been a major change in direction away from product or vertical standards and towards horizontal and generally hazard-oriented standards. Advantages of these are said to be faster development, easier updating, greater applicability, more consistency and better clarity and understanding about necessary safety levels (van Weperen, 1992).

Meister (1984) differentiates "attribute" standards, which describe how the product should appear or should function, and "performance" standards, which describe how the design product should perform (pp.215-217, 256-263). He sees the former as being general and applying mainly to the component or equipment level and the latter as particular at the subsystem/system level. He criticises the state of human factors standards in much the same way as consumer product standards have been criticised. As a consequence, Meister (1989) has subsequently stated that "... whether because the standard lacks substantive data support or because human factors is generally viewed ... as a constraint on ... freedom to design, MIL-STD 1472C [for instance] is honoured as much in the breach as in the observance."

In summary, if we are to have certification, then it must be related to some norms or standards or standardised procedures. In other words, a system might be certified if it can be shown to have attributes which meet certain recommended values or if its performance meets acceptable limits on certain recommended criteria. We can add to this a third form of certification, if it is shown that defined analysis or test methods have been applied to the design. In this last case, of course, the methods themselves will have to be certified first. Of relevance to complex aviation technology is a particular case of the last, "there is little doubt that a principal future use of simulators will be for licensing and certification" (Jones, Hennessy, & Deutsch, 1985). Principally used now for pilot training and proficiency approval, there seems no reason why artificial intelligence in the cockpit, and particularly its interaction with crew, cannot also be assessed in simulators. The interesting issue then is certification of the simulation system itself!

Currently under consideration is STANAG 3994 AI – the NATO Standardisation Agreement on the Application of Human Engineering to Advanced Aircrew Systems. The intention of this is to "standardise methods for the integration of human engineering procedures with the design and development of advanced aircrew systems." The purpose also is to provide a basis for agreements between contractors and procuring agencies on human factors scope, context, techniques and criteria. The draft STANAG 3994 AI defines a "general human engineering program model for military systems". The core requirements include those to do with analyses:

- **system analysis** – mission analysis, function analysis, potential operator capability analysis, potential equipment identification, function allocation
- **analysis of operator/maintainer tasks** – timeline analysis, task analysis, critical task analysis, decision analysis, error analysis, loading analysis
- **preliminary system and subsystem design** – information requirements analysis, control requirements analysis, workspace requirements analysis, and environmental analysis.

In addition, a number of other issues are also defined as requiring agreement between agency and contractor. These are: programmes of research involving experiments, testing and dynamic simulations with human subjects; detail on application of relevant human factors standards; development of software and hardware procedures in operation and maintenance; production of mock-ups and models for conformance testing; and development and planning for test and evaluation, including criteria justification and test interpretation details.

Perhaps most critical in terms of any sustainable argument for certification are the requirements to prepare a "human engineering programme plan." This must identify standards of relevance to the system and must identify what human engineering activities will be involved, time-scales and criteria, and should indicate how formal interaction between human factors specialists and other relevant design specialists will be achieved. Finally, provision is made for a tailoring of details in the standard to meet any specific requirements of the system under development.

Taking a domain outsider's viewpoint, several things seem apparent about this STANAG 3994. First, it appears at first sight to be complex and unwieldy, with potential for great overlap in particular amongst the many different analyses. Even co-ordinating these, making sure they are complementary but not too duplicatory (or even contradictory) will be a considerable project management task. Secondly, on the other hand, it is to be welcomed that so much emphasis is placed upon analytical activities and not on prescriptions of design detail. However, within some of these analyses are implied checklist comparison procedures, for instance "... multi-function control modes, and display menu selections shall be analysed ... and the resulting ... structure shall be plotted and analysed for ease and effectiveness of use", or "workspace [requirements] shall be analysed in terms of their access, vision, reach, egress, and emergency requirements, for the range of body sizes, clothing and protective equipment ..." (p5). This will presumably bring into play a plethora of other standards, guidelines and recommendations, the effect of which *may* be to impose a degree of complexity and restriction on design which is not commensurate with finding innovative solutions. Thirdly, again on the positive side, the document does allow for flexibility in its provisions according to circumstances. Finally, and thinking again about systemic gains, it may be that the most important benefit is the integration and collaboration required between ergonomists and engineers.

Any undue complexity, as suggested above, may have excessive cost implications. For instance, even under present regulation systems: "

> It is relatively easy for the profitable airline...but the airline operating in a more competitive area of the aviation system, where economic margins are extremely constrained, may simply be unable to undertake all of the desirable training and

> standardization of equipment without going out of business. The regulatory authority may have considerable difficulties in compelling such airlines to undertake costly procedures as the airlines may accurately point out that by doing so they will be made less cost efficient vis a vis foreign operators (possibly operating in a less regulated environment) with whom they compete directly...The temptation for operator and regulator alike, when faced with an acknowledged but intractable problem, is to undertake some unconscious dissonance resolution by regarding the problem as less serious than they might if it were readily soluble. (Green, 1990, p510)

If this concern applies to aviation systems developers and suppliers as well as to the operating companies, then the impact on the workability of any certification process may be serious.

To certify or not to certify?

So, why is someone who typically dislikes regulation, systematisation and quality assurance generally writing here to support certification in complex systems? The answer lies in what has been stressed above, namely that the systemic outcomes of having a certification programme in place may be advantageous enough to overcome any drawbacks of the regulatory regime and the content of any standards.

As a start, if human factors certification is to work in any domain we need to consider why it has not been in place there previously and address all potential reasons very seriously. The author is currently working with a very large, reputable transnational company. The design engineers pass their designs through every conceivable review and approval process – HAZOP, P & I Approval, Environmental Impact Assessment, Engineering Audit etc. – but, at the moment, there is not any human factors approval. We asked ourselves why such an absence of any formal human factors standards approval system exists. Possible reasons include:

- Historically, ergonomics has not been seen as important by engineers, managers etc.
- Ergonomics has been assumed to be included in all the other types of approvals and standards.
- Certification of human factors is not seen as cost-effective (in terms of there being few gains); there will still be problems afterwards (since people are seen as fallible), but much time and energy will have been expended meanwhile.
- Certification is genuinely seen, by engineers and/or ergonomists as not required
- Human factors certification may be resisted by ergonomists themselves, perhaps because of the requirements or restrictions it may put on them.
- Certification is seen as impossible to do, or at least impossible to do well.

Before even beginning to introduce ergonomics design review to this company, before even planning what might be included, we need to examine these reasons, see which are apparent in this case, and address the organisation issues involved.

The Case Against

There are two ways to mount a defence of certification: to counter or at least downplay the criticisms, and to promote advantages and benefits. Taking first the criticisms, a number of arguments against certification have been rehearsed implicitly or explicitly in this paper.

Certification may firstly be seen as *unnecessary*; presumably in this view either there is little or no room for improvement in ergonomic design of systems (hardly a sustainable argument) or else that the aviation world is self-correcting. That is, up to a point human factors deficiencies will be rectified anyway during development and commissioning and where they are too large or deep seated then the system itself will not remain in operation. I find this argument unconvincing unless we allow major failures in operation to also be a part of this self correcting process. However, what must be accepted is that, currently, we are talking about remarkably reliable and safe systems of hardware, software, procedures, communications and people.

An extension of the first objection is that certification is unnecessary for simple and/or relatively stable systems and is *impossible* to do adequately for complex systems. There may be some validity to this argument but the advantages of an improved quality of development process discussed below might counteract it to some extent. What must be recognised, though, is that any system of approvals must allow for trade-offs between human factors and between human and technical factors in its operation.

Further criticism might be that a certification regime is *restrictive* and *cumbersome*. If we replace "is" by "can be" then I would agree here. However, if we aim certification at performance and at assessments as against at design specifications, allow tailoring of standards to meet circumstances, and – most importantly – make ergonomics design review and approvals an intrinsic part of development rather than an extraneous add-on, this will all help dilute this complaint. Similarly, we can meet objections that certification might stifle innovation and lead to homogeneity in design.

Certification might also be criticised on the grounds of its being *misdirected*, with standards aimed only at reducing the incidence or consequences of active errors. In this view, standards may be much less help with latent failures or resident pathogens in the system; these are the system problems which may have lain dormant in the system for a long time and which are spawned by the activities of designers, managers and, indeed, regulators themselves (Reason, 1990). One could take a positive view though, that in fact it is these "violations", giving rise to latent errors, which are best attacked through a process of certification, due to the process itself being a good discipline upon all involved in high level planning and decision making.

As for a fourth set of criticisms, that certification is *untestable*, widely *unacceptable* and thus *unworkable* or *unenforceable*, this will largely be a function of the particular regulatory regime. To repeat again an earlier point, if a system of certification can be constructed such that it is seen to improve and streamline the development process and time as well as increase systems integrity, then acceptance will be more widespread.

The Case For

The case for certification can be made positively, as well as by minimising the validity or consequences of criticisms as above. Three areas of benefit may be defined, all systemic in nature in that they emanate from the fact that human factors certification of complex systems will have effects beyond defining and ensuring compliance with human factors standards.

First, we have the improvements in the design process that might be expected. Knowledge that a system must be certified in terms of human factors may not ensure all correct detail design alternatives are chosen – indeed, there is no such thing as a perfect design given all the tradeoffs that must be made. However, it will mean a greater likelihood that all relevant issues are addressed and their consequences assessed much earlier in development. Costly changes after prototyping or even during commissioning trials can be reduced in frequency and extent. A related benefit is that for any suppliers to be able to meet future certification requirements, technical and financial decision makers will have to coordinate much earlier and better with those responsible for the human factors.

A second benefit is predicted, based upon experience in other domains where ergonomics standards have been introduced or toughened. In the act of formulating, specifying and testing the process and procedures necessary to allow systems to be certified, the human factors community will have to respond to pressures for better methods, techniques and criteria, and will have to validate, justify and report them better.

Finally, this improved "professionalism" in human factors, the perceived benefits to the design process, and – if experience in industrial health and safety and ergonomics is any guide – increased interest of engineers in the resulting need for increased problem solving, will all act to produce a more human-centred design approach. Thus, through both the fact and process of certification, as much or more than by its content, the design of complex aviation systems will be improved.

Conclusions

We should not talk of certification only as a choice of two options – to certify or not to certify. If we draw an analogy in politics, like saying that electorates have a choice between the authoritarian right (prescription, control, punitive consequences of non-compliance) and the libertarian right (the individual has an absolute right to do as he/she pleases, and the 'market' will ensure instability is kept in bounds), such debates see anarchy as the only outcome if a choice between the two options is not made. There *are* other paths in government however, whereby individuals have rights or freedoms but also responsibilities towards society, and where society attempts to redress imbalances in power. Thus, a regime of certification *can* be implemented such that it provides a framework for complex systems design, a benchmark to aim for, and a bulwark against very poor design, whilst still allowing for innovation and not imposing too costly or cumbersome a design regime.

What must be addressed before instituting a formal certification system are the needs of such a programme. Benefits will be realised and disadvantages or problems minimised only when decisions are made on:

- Distinctions to be made and balance to be found between attribute, performance, personnel and process certification
- Desirable degree of prescription or latitude for design
- Identification, definition, agreement and validation of test methods, measures, criteria etc.

- Provision for flexibility and updating of requirements
- Examination of trade-offs between value and cost/time of the certification process
- Systems to certify the certifiers/certification systems
- Communication of outcomes of the certification process in more useful terms than just a pass/fail, yes/no judgement
- Consideration of implications of non-conformance, and thus enforcement

It must be stressed once again that the process of certification can be of value even if we are unsure of or unhappy about the content when it is first instituted. More than this, if we get the process right, then content problems – in appropriate requirements or missing tools or data for instance – will be rectified as part of the process being put into operation. We must remember, though, that the individuals who produce certification processes or who test and approve systems are themselves fallible, as also will be any intelligent systems built to help with certification. Perhaps this is the key issue for acceptance of certification – Quis custodiet ipsos custodes?

References

Goom, M.K. (1993), An industrial view of MANPRINT. Proceedings of the Annual Conference of the Ergonomics Society (ed E J Lovesey). Heriot Watt University. London: Taylor & Francis.

Green, R. (1990), Human error on the flight deck. *Phil. Trans. R. Soc. Land.B.*, *327*, 503-512.

Jones, E.R., Hennessy, R.T. & Deutsch, S. (eds), (1985). *Human Factors Aspects of Simulation*. Washington D.C.: National Academy Press.

Meister, D. (1985). *Behavioural Analysis and Measurement Methods*. New York: J. Wiley & Sons.

Meister, D., (1989). *Conceptual Aspects of Human Factors*. Baltimore: John Hopkins University Press.

Ministry of Defence. (1992). *The MANPRINT Handbook (2nd ed.)*. London: HMSO.

Reason, J. (1990). The contribution of latent human failures to the breakdown of complex systems. *Phil. Trans. R. Soc. Land.B.*, *327*, 475-484.

Senders, J.W. & Harwood, K. (1993). Provocations: To certify or not to certify. *Ergonomics in Design*, *1*, April, 8-11.

STANAG 3994 AI, *The Application of Human Engineering to Advanced Aircrew Systems.* Draft NATO Standardization Agreement.

van Weperen, W. (1992). A hazard-oriented approach to product safety criteria. *Proceedings of ECOSA Workshop on Product Safety Research in Europe*, Amsterdam, July, 10-19.

Wilson, J.R. (1984). Standards for product safety design: A framework for their production. *Applied Ergonomics*, *15*, 203-210.

Certification and Legislation

P. A. Hancock

University of Minnesota

Preamble

One of the mandates of each of the series of meetings on complex technical systems has been to review, rewrite, edit, amend, and elaborate upon initial position papers based upon insights and interactions that occurred at the meeting itself. I think this is a most laudable aim and one I was able to follow in previous papers (Hancock, 1991, 1993). However, on this occasion, I found that my understanding of certification in a socio-technical context was extended to such a degree that I felt constrained to develop a different paper from my other submission (Hancock, this volume) based upon the metaphorical, analogical, and literal relation between certification and legislation. I offer it with sincere thanks to Tony Debons and the other members of the 'why' group whose comments were so stimulating.

Introduction

I wish to compare the process of certification with how we develop and apply legislation, especially as it relates to individual behavior. Embedded in this parallel are analogical, metaphorical and literal relations as, at heart, certification must in some form intersect with legislation at a national, international, and global level (for complex aerospace systems at least). Behind legislation, particularly that which applies to individual action, lie assumptions about moral and social behavior. These are of course under continuous review and critique with such arguments taking more or less violent characteristics. However, we accept certain assumptions about behavior superimposed upon which are human created laws. We recognize that law in this sense is not strictly synonymous with law in a scientific sense (although the respective comparison is most instructive). There is continuous dispute about interpretation in legal circles and indeed judges, solicitors, lawyers, attorneys, barrister, and juries would be redundant if legal concepts were completely determined (this does not of course imply that scientific laws are without challenge). At a certain level, laws are empowered to *control* society. In a similar manner, certification is viewed by some as an effort toward control over certain events and processes. If, in the face of an undetermined environment we cannot guarantee control, at least certification is a representation for a *desire for control.* While systems are faced with the

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

vagaries of an uncertain environment, humans can be constrained more by law in their behavior since legislation is essentially arbitrary (and therefore is perceived as differing from scientific laws). However, we can imagine situations where individuals engage in unintended actions which result in transgressions of the law and also imagine circumstances in which we ask whether it is justifiable to transgress the moral basis of law (e.g., killing dictators). Therefore, the simile is partial since we seek certainty as an adjunct of control, but recognize that we cannot ubiquitously or even frequently achieve such an aim.

When control and certainty apply they are looking forward in time. That is, control is for a purpose and certainty statements always look to the future (since we believe the past to be determined). However, the function of *compliance* is one predominantly of enforcement. In legal terms, compliance to behavior is achieved by the force of power and punitive action. The majority of legal compliance comes from self-reference and appropriate laws which recognize enlightened self-interest (e.g., traffic laws). Compliance in certification terms is similarly framed. There is much hope that self compliance by professionals throughout the process means that best self-interest will achieve the best possible result. However, certification can be used as a tool of penalization. In essence, we *hope* professionals in system design and certification behave like law-abiding citizens, not needing policemen running around after them to ensure they behave properly. We *fear* that certification will be used as a bludgeon to cower individuals into compliance. The analogy is not complete since designers and system developers have to explore boundaries of what is known (in advanced aerospace systems), however most individuals do not 'press the envelope' of the law. The fallacious criticism of G. B. Shaw is appropriate here.

> *Only fools and idiots seek to change society. Therefore all societal change is affected by fools and idiots.*

At the heart of this fallacy is the enlightened designer who seeks constructive ways to facilitate change. If compliance is an insurance function that oscillates between the past and the future, *accountability* is a historical process that seeks to attach individuals to the decisions and actions that they take. In law, we can happily talk of "diminished responsibility" when individuals for differing reasons are unable to recognize the consequences of their actions. Similarly for certification we cannot indemnify completely a design as we cannot have complete knowledge of all potential interactive conditions. Accountability then can only be used when in the process of design test and evaluation something was neglected which *could have been known at the time.*

One of the most important facets of certification in complex systems is the problem of technical evolution and non-stationarity. That is, we are trying to hit a moving target with something which is inherently time-locked. Systems evolve and change quickly, and it is dauntingly difficult to keep certification going at the same pace. There is, of course, a parallel in law at present, where for technologies like DNA and computer systems, legislation literally cannot keep up. Hence, the old question emerges. How to provide stability against the background of instability. For some, the greater the instability the greater the need for stability; for others the greater the need for flexibility. These are not mutually exclusive aims in adaptive systems.

I believe there are strong parallels between certification and legislation. They each represent human attempts to place arbitrary frames upon diverse and unpredictable entities. They should each recognize that such frameworks are by constraint, often arbitrary and therefore cannot provide perfect fits. However, their presence appears preferable to their absence. In a real sense, science is also a member of this movement which is part of the human appeal for comprehensibility in the face of the incomprehensible. In that way, certification seeks nothing new but is part of a time-honored tradition which begins when the child must first make some sense of the "blooming, buzzing confusion." I have not explored the full spectrum of links and relations between certification and legislation. Also, I have not stated the obvious literal links where certification is, in actuality, a legal process. However, I hope the brief comments can help frame my final summary which asks the reader about their central beliefs of human society.

Summary

It might appear that there are inherent differences between my two observations on certification. I submit that there are none. In drawing parallels between certification and legislation, I imply a strong advocation of neither. I am still naive enough to be an optimist. In the end we can come down to one's opinion of people. Do we have to generate minimum standards of behavior and conduct against which to hold the lowest common denominator, or can we aspire to self-generated standards which inspire the application of a highest common factor? The aged, the jaundiced, the wordly-wise will sadly shake their head and quietly admit to the former pessimism. This is so especially in light of the apparently immovable bureaucratic behemoth of social institutions. However, we have seen massive social change in the past decade and if survival is a requirement, pluralism is essential. It is all, I submit, a matter of education and a matter of social self-enlightenment. Sadly, to many we seem to be headed, as a global society, in the wrong direction. While life has become more comfortable, at least to those of us who rely on and benefit so much from the higher realms of technology (but we should note, still a small minority of the planet's population), it does not seem that life has become progressively more fulfilling. Indeed, many of the ex turn-on, tune-in, drop-out generation still hanker after the elysian 'escape' (Wooley, 1993). However, I shall end this essay with a phrase that my father was most fond of: 'it only needs a small candle to change a large darkness.' I think the nature of how we deal with our technology and the promises and indemnities with which we vest it directly relate to our view of the future. I like Kennedy's observation:

> *Some see the world the way it is and ask, why? I see the world the way it can be and ask, why not?*

Let us hope that seeing the world the way it is does not blind us to the world that can be before it is too late.

References

Hancock, P.A. (1991). The aims of human factors and their application to issues in automation and air traffic control. In: J.A. Wise and V.D. Hopkin (Eds.). *Automation and Systems Issues in Air Traffic Control,* NATO. New York: Springer-Verlag.

Hancock, P.A. (1993). On the future of hybrid human-machine systems. In: J.A. Wise, V.D. Hopkin., and P. Stager (Eds.). *Verification and validation of complex systems.* Martinus Nijhoff: The Netherlands.

Hancock, P.A. (1993). Certifying life. In J. A. Wise, D. J. Garland & V. D. Hopkin, *Human Factors Certification of Advanced Aviation Systems.* Daytona Beach: Embry-Riddle Aeronautical University Press.

Wooley, B. (1993). *Virtual Worlds.* Blackwell: Oxford.

Human Factors Requirements in Commercial Nuclear Power Plant Control Rooms

Lewis F. Hanes

Independent Consultant

Introduction

The development of a human factors certification program for advanced aviation technologies presents many interesting questions. A key question is as follows: Is there a *need* for additional human factors participation in the design, development and implementation of advanced aviation technologies? Some evidence suggests that the answer is yes.

- The FAA National Aerospace Plan discusses the need for human factors activities.

- ICAO statements regarding human factors (Maurino & Galotti, 1993 this volume);

- The results of investigations which document the high percentage of accidents and incidents in airplane and air traffic control systems in which human errors are the cause or significant contributors.

A second key question is: What are the *ways* by which human factors knowledge and methods can be introduced into the design, development, and implementation of advanced aviation technologies? There are at least three ways, as described below.

- Research and development. Modern aviation technologies (e.g., "intelligent flight management systems") provide functional capabilities that human factors standards and evaluation methods do not address very well. Standards, methods, etc., are needed to support the design team that is designing, developing and testing the modern technologies incorporated into the aviation system.

- System development process. The United States military organizations (e.g., United States Air Force) require the application of human factors knowledge and methods during design, development and testing of major systems in which personnel play an important role. Typically, MIL SPEC H46855 is applied, which requires that a process be developed for applying human factors throughout system design, development and testing. In addition, various standards (e.g., MIL STD 1472D) are invoked, and as part of the human factors process it must be demonstrated that these standards have been

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

satisfied. Organizations acquiring airplane, air traffic control, etc., systems that include advanced aviation technologies may require that human factors as described above be a formal part of the design, development and testing activity.

- Certification. Certification of the human-machine interface system, procedures, training program, system functions, personnel, design team, etc., may be required. This topic of certification is the main theme of the workshop of which this book is a record.

Of course, all of the ways identified above may be implemented. If certification is given serious consideration, then it is useful to learn from the experiences of organizations that have a human factors program involving certification already in place. The Nuclear Regulatory Commission (NRC) has such a program. The purpose of this paper is to present a brief description of this activity. The reader should be cautioned that the information presented in this paper is based on the author's reading and interpretation of published reports. The NRC should be consulted for an accurate and current description of the human factors evaluation process in actual use.

Initial Human Factors Requirements

NRC interest in human factors issues associated with nuclear power plant (NPP) control rooms came into existence following the Three Mile Island–Unit 2 accident in 1979. The NRC Action Plan developed in response to the incident required NPP licensees and license applicants to perform detailed control room design reviews (DCRDRs) to identify and correct design deficiencies. These reviews included the assessment of control room layout, the adequacy of the information provided, the arrangement and identification of the important controls and displays, the usefulness of the alarm system, information recording and recall capability, lighting, and other human factors considerations that have an impact on operator effectiveness and plant safety (Ramey-Smith, 1985).

In 1981, the NRC issued guidelines for control room design reviews, NUREG-0700 (NRC, 1981), for use by utilities in conducting DCRDRs. NUREG-0700 consists of human factors guidelines adapted to NPP control rooms, and additional guidelines as required (Ramey-Smith, 1985). The formal NRC requirements for the DCRDR are contained in NUREG-0737, Supplement 1 (NRC, 1982), which was issued in 1982. This document contains technical and documentation requirements. The NRC required that licensees and applicants perform a DCRDR on their designs. The review had to consist of the following (Ramey-Smith, 1985):

- establishment of a qualified multi-disciplinary review team

- function and task analysis to identify control room operator tasks, and information and control requirements during emergency operations

- comparison of display and control requirements with control room inventory

- a control room survey to identify deviations from accepted human factors principles

- assessment of human engineering discrepancies (HEDs) to determine which HEDs are significant and should be corrected

- selection of design improvements

- verification that the design improvements will provide the necessary correction, and will not introduce new HEDs

- coordination of control room improvements with changes from other programs, such as operator training and upgraded emergency operating procedures.

The NRC pointed out that NUREG-0700 and similar NRC documents did not have to be used. It is believed, however, that every utility used these NRC reports in performing the DCRDRs.

Each utility prepared a DCRDR report, which included proposed control room changes and the schedule for change implementation. The NRC human factors staff reviewed each DCRDR submitted, and approved, approved with comments, or disapproved the document (Ramey-Smith, 1985).

The DCRDRs are nearly all completed, and many changes to control rooms have been implemented.

Requirements for New Design

The NRC and United States utility industry have begun a program to improve and standardize future commercial NPP designs (O'Hara, Higgins, Goodman, Gallett, & Eckenrode, 1993). The NRC has issued 10 CFR 52 to streamline the plant licensing process. The licensing process of Part 52 consists of a Final Design Approval by the NRC followed by a standard design certification that is issued as an NRC rule. This involves formal rule-making and the opportunity for a public hearing before the Atomic Safety and Licensing Board. This certification would be valid for 15 years (renewable). Neither the NRC nor the plant designer can change or impose new requirements on the standard design certification without new rule-making. Utilities would have the opportunity of purchasing the standard design and utilizing it as already approved by the NRC. To ensure that an as-built plant conforms to the standard design certification, inspections, tests, analyses and acceptance criteria (ITAAC) must be specified as part of the standard design certification. After certification, the NRC will ensure that the design has met the ITAAC (O'Hara, Higgins, Goodman, Gallett, & Eckenrode, 1993). To obtain a standard design certification under Part 52, a plant designer must submit a Standard Safety Analysis Report (SSAR) to the NRC for review. The NRC's review of the SSAR is issued as a Final Safety Evaluation Report (FSER), which will form the basis for the Final Design Approval (O'Hara, Higgins, Goodman, Gallett, & Eckenrode, 1993). Chapter 18 of the SSAR addresses Human Factors Engineering.

The Human Factors Assessment Branch of the NRC evaluates the human factors engineering (HFE) material (chapter 18 of the SSAR) submitted as part of the certification process for new plant designs. The review process is very different from the DCRDRs evaluated in the past. A major reason is that detailed control room and instrument design information is not available on which to make a safety determination. Due to changing technology, much of the detailed design will not be completed prior to the issuance of design certification (O'Hara, Higgins, Goodman, Gallett, & Eckenrode, 1993). Therefore, the NRC is performing the design certification evaluation based on an implementation process which describes the HFE program elements required to develop an acceptable detailed design specification. In addition, the applicant must submit ITAAC/Design Acceptance Criteria that will ensure that the design process is properly executed (O'Hara, Higgins, Goodman, Gallett, & Eckenrode, 1993).

Guidance for the HFE review of Chapter 18 of the SSAR submitted by the designer is provided in regulatory requirements (e.g., 10CFR 52.47, 10CFR 50.34(g) and 10CFR 50.34(f)) and guidance contained in such documents as NUREG-0700 (NRC, 1981) and NUREG-0800 (NRC, 1984). These documents, however, provide limited information on which to base a review (O'Hara, Higgins, Goodman, Gallett, & Eckenrode, 1993). Therefore, a HFE Program Review Model is being developed to provide needed guidance. This model is described in O'Hara, Higgins, Goodman, Gallett, & Eckenrode, (1993). A summary extracted from this reference is presented below.

The model is intended to provide a programmatic approach to achieving a design commitment to HFE. The commitment and scope of the HFE is as follows: human-systems interfaces (HSI) should be provided for the operation, maintenance, test, and inspection of the NPP that reflect state-of-the-art human factors principles. State-of-the-art human factors principles include principles currently accepted by human factors practitioners. "Current" is defined as a practice, method, or guide documented in the human factors literature within a standard or guidance document that has undergone a peer-review process, and/or is justified through scientific/industry research practices.

The model contains eight elements, each of which consists of an objective and factors that must be considered in the review process. The elements are listed below.

1. *HFE Program Management.* An HFE Design Team and an HFE Program Plan should be established.
2. *Operating Experience Review.* Problems and issues encountered in similar systems of previous designs should be analyzed so that they are avoided in the development effort, or retained in the case of positive features.
3. *System Functional Requirements Analysis.* Identify those functions which must be performed to satisfy objectives.
4. *Allocation of Functions.* A structured methodology should be used to allocate functions.
5. *Task Analysis.* This analysis should (a) provide a basis for specifying requirements for displays, data processing and controls, (b) provide a basis for design decisions, (c) assure that human performance requirements do not exceed human capabilities, and (d) be used as a basis for developing procedures, manning, skill, training and communications requirements.
6. *HSI Design.* HFE principles and criteria should be applied in design.
7. *Plant and Emergency Operating Procedure Development.* HFE principles and criteria should be applied in the development of procedures.
8. *Human Factors Verification and Validation (V &V).* V & V should be used to assure that the performance of the HSI achieves all HFE design goals as established in the program plan, all system functional requirements, and all requirements to support task accomplishment.

Concluding Remarks

The NRC has developed processes to review HFE associated with NPP control rooms and instruments. One process was developed and used to evaluate existing detailed designs following the accident at the Three Mile Island plant. A more recent process was developed in response to the NRC-nuclear industry effort to obtain certification for improved and standardized future commercial NPP designs. Because of changing technology, the detailed control room and instrument designs are not available for evaluation. Therefore, the NRC has developed a method to assess the HFE process used to develop the detailed designs.

An HFE group is located within the NRC that reviews and evaluates submittals by design groups. It approves, approves with comments, or disapproves the submittals. The results of this effort are considered by the NRC in deciding whether to certify the overall plant design.

The NRC has many years of experience conducting reviews toward certifying designs. The NRC has also learned a great deal on how to implement a design certification program. The aviation industry should study in detail the NRC lessons learned as it evaluates the need for and desirability of the certification of advanced aviation technologies.

References

NRC. (1981). Guidelines for control room design reviews (NUREG-0700). Washington, D.C.: NRC.

NRC. (1982). Clarification of TMI action plan requirements - requirements for emergency response capability (NUREG-0737, Supplement 1). Washington D.C.: NRC.

NRC. (1984). Standard review plan (NUREG-0800, Revision 1). Washington D.C.: NRC.

O'Hara, J., Higgins, J., Goodman, C., Galletti,G., and Eckenrode, R. (1983). Human factors engineering program review model for advanced nuclear power reactors. In Proceedings: *Topical Meeting on Nuclear Plant Instrumentation, Control and Man-Machine Interface Technologies*, La Grange Park, IL: American Nuclear Society, pp. 383-388

Ramey-Smith, Ann. (1985). Nuclear power plant control room design reviews: A look at progress. In *1985 IEEE Third Conference on Human Factors and Power Plants*. New York: IEEE, pp. 117-118.

N95- 34767

A Psychologist's View of Validating Aviation Systems

53-54
39143
p. 6

Earl S. Stein & Dan Wagner

United States Federal Aviation Administration

Introduction

All systems, no matter what they are designed to do, have shortcomings that may make them less productive than was hoped during the initial development. Such shortcomings can arise at any stage of development: from conception to the end of the implementation life cycle. While systems failure and errors of a lesser magnitude can occur as a function of mechanical or software breakdown, the majority of such problems in aviation are usually laid on the shoulders of the human operator and, to a lesser extent, on human factors (Nagel, 1988). The operator bears the responsibility and blame even though, from a human factors perspective, error may have been designed into the system.

Human factors is not a new concept in aviation. The name may be new, but the issues related to operators in the loop date back to the industrial revolution of the nineteenth century and certainly to the aviation build-up for World War I. During this first global confrontation, military services from all sides discovered rather quickly that poor selection and training led to drastically increased personnel losses. While hardware design became an issue later, the early efforts were primarily focused on increased care in pilot selection and on their training. This actually involved early labor-intensive simulation, using such devices as sticks and chairs mounted on rope networks which could be manually moved in response to control inputs.

The use of selection criteria and improved training led to more viable person-machine systems. More pilots survived training and their first ten missions in the air, a rule of thumb arrived at by experience which predicted ultimate survival better than any other. This rule was to hold though World War II. At that time, personnel selection and training became very sophisticated based on previous standards. Also, many psychologists were drafted into Army Air Corps programs which were geared towards refining the human factor. However, despite the talent involved in these programs and the tremendous build-up of aviation during the war, there were still aircraft designs that were man killers (no sexism implied since all combat pilots were men). One classic design error that was identified fifty years ago was the multipointer altimeter, which could easily be misread especially by a pilot under considerable task load. It has led to flying fully operational aircraft into the terrain (Fitts and Jones, 1947). The authors of the research which formally identified this problem put "Human Errors" in quotes to express *their* dissatisfaction with the traditional approach to accident investigation. It traditionally places the burden of guilt on the operator. Some of these altimeters still exist in older aircraft to this day.

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PRECEDING PAGE BLANK NOT FILMED PAGE 44 INTENTIONALLY BLANK

Human Factors in Complex Systems

The airspace system has become increasingly more complicated since the Second World War, and an emphasis on aircraft issues alone would not do it service. The potential for chains of events leading to system breakdown has increased with the volume of traffic and the complexity of both air and ground subsystems. While the concept of human factors is not new, it is continually being rediscovered or ignored by systems developers. In addition to cockpit crew operations, the modern civil and military airspace system must include airport ground operations, conducted by air traffic control, and maintenance of both air and ground hardware/software. This latter area is handled by the airway facilities personnel in the FAA, while the former is a function of airlines, the FAA, fixed base operators, and other airframe and power maintenance resources.

All of these systems and subsystems have people working in them, and in cases like the multipointer altimeter, the hardware and software are not user friendly. Systems have often been created in which the operator was the last to know what was going on, and human factors professionals only became involved by exception when the designers knew they had a problem. In the latter case, it is not unusual for the designers to come *to* human factors people with a request to find a better way to select and train because the hardware and software designs are already frozen. This is a step back in time where human factors by whatever name was only viewed as useful from the limited perspectives of training and selection.

Warm and Dember (1986) described their concerns over systems that are designed in such a way that there may be attentional lapses to the point that operators are no longer "awake at the switch." In such a situation, whether in aviation or not, the system is operating in "free flight" without any supervisory control by human hand. The fact that this can and does happen became clear several years ago when the Nuclear Regulatory Commission closed the Peach Bottom power plant in Pennsylvania. Inspectors found operators literally asleep at their stations. Obviously this problem was not identified and addressed during the development of the plant.

An example of a situation in which systems designers decided to evaluate their product before it was too late was reported by Kantowitz and Sorkin (1983). A consumer electronics company was designing a new answering machine that would have been the answer to all needs of business and industry. The designers were convinced that they had a very marketable product which would be quite profitable. However, someone decided that it would be reasonable to test the product using a sample of people who were the intended users: secretaries. They gathered a group of ten and provided the documentation and the equipment and told them to go forth and use it. Not one could figure out how to operate the system. The designers concluded that the problem had to be training, so they rewrote the documentation and developed a basic training program. Another group of ten secretaries was mobilized. Of these ten only two could learn to operate the system. They both had previous backgrounds in computer programming. The designers had created a system for themselves and not for the users. Fortunately, they had chosen to test. Had they attempted to market the product it would have been a financial disaster.

Are there parallels between the answering machine example and the evolution of aviation systems? Of course! Thomas (1985) described the development of air traffic control over a nine year period in the 1950's and 60's. He noted the difficulty in transitioning from the older broad band radars to the more modern narrow band digital systems. Part of the problem was not so much a matter of system design but rather a function of preparation of the users to accept the new equipment. Further, there were capabilities in the newer equipment that the operators tended to use or not based on their preferences and experience. At the time, very little consideration of the users was apparent in system development. This is changing slowly, and modern ATC systems often have controller panels involved in the design effort.

The Concept of Certification

Certification is a legalistic term. It implies an organizational standing such that the certifier or certifying agency has the power to determine if a system can be used and under what conditions. Certification implies sound methods that have met the tests of time, validity, and reliability. It suggests protection of the public from hazards generated by systems that have been poorly designed and/or from operators that are unqualified or unable. Certification is a novel concept in human factors, at least from the viewpoint of some stand alone process separate and distinct from the engineering aspects of system development. Human factors professionals must ask themselves if they really want to become involved in certification and if so why?

In the United States, the Federal Aviation Administration already carries the responsibility for certifying aircraft and personnel in aviation. It also certifies its own equipment and the people who control and maintain the airspace system. Most of this legal requirement is handled by subject matter experts who are flight examiners, controllers, and hardware systems specialists. In the past, the role of human factors personnel has always been in the background and has been principally advisory to those who would actually sign off on an aircraft or other system. While human factors as a science may have been viewed as an information source, it has been the exception rather than rule for anyone in the business of certification to ask the advice and council of human factors specialists.

One partial exception is recalled by these authors. This was an effort to provide the designers of the Automated En Route Radar Air Traffic Control (AERA) with empirical validation of a construct called "workload probe." The probe was a computer algorithm that would theoretically predict controller workload up to 20 minutes in advance based on weather and anticipated traffic. It was tested in simulation at the FAA Technical Center. Measures of real time controller workload were collected using a Cooper-Harper type scale called the air traffic workload input technique (ATWIT). The results of the workload predictions from the probe were significantly correlated with participants self ratings using ATWIT and with ratings by over-the-shoulder observers. This proved that the concept of workload probe was viable, but it hardly qualified as certification since AERA as a system is not yet ready for certification testing. However, the process of empirical testing could be viewed as model for future systems evaluation. Human factors are here to stay in one form or another and has the support of Congress based on law.

The Aviation Safety Research Act of 1988 led to an increased awareness of the possibilities of human factors in the air space system. It required the FAA in particular to expend a finite portion of its annual budget in human factors related to new systems under development. One of the most visible products resulting from the Act has been the *National Plan for Human Factors* (FAA, 1991). This is a very comprehensive document which theoretically defines the human factors research needs for the present and the foreseeable future. One is struck by the magnitude of the document and the implication/admission that there is a great deal which is not currently in the corporate body of knowledge concerning person-machine systems and subsystems in aviation.

Within the domain of aircraft certification, the plan states the following: " The FAA is responsible for the human factors evaluation and certification of aircraft. The personnel most responsible for this job are FAA certification pilots. These pilots are finding it very difficult to keep up with the human factors implications of the latest developments in flight deck automation and advanced technology aircraft." The suggested solution is the development of handbooks, checklists, and special courses.

While "certification" is cited in the chapter on aircraft from an aircraft perspective, conceptually it does not appear within the aircraft maintenance section of the plan. There are, however, a variety of research related issues to include the old standbys of selection and training as well as an expressed need to define the task structure more thoroughly.

The plan also notes that "... airway facilities personnel problems and needs have remained largely ignored" (FAA, 1991). These men and women work unseen behind the more glamorous positions of pilots and controllers. They maintain equipment, some of which retains vacuum tubes and most of which was designed without any reference to even common standards of ergonomic considerations. There are FAA standards of system maintainability, but it is notable that the plan does not address any issues related to certification of those standards or for that matter to the test equipment that personnel will use to accomplish the maintenance task. It does imply, however, that additional work will be necessary to evaluate such factors as maintenance documentation (an often ignored area of system operations) and the approaches to diagnostic support requirements.

It should be noted that airway facilities currently use a certification process. Before a new piece of equipment or a piece of equipment that was taken out of service for maintenance or repair is placed in service in the national airspace, it must be certified. That is, a piece of equipment, such as a radar, must be "certified" to be performing its intended function acceptably and within specified tolerances before it can be placed in service in the national airspace system. Of course, this does not mean the equipment is well designed, user-friendly, or even maintainable. It simply means that the required function is being accomplished.

The issues this process raises from an airways facilities perspective is that if a piece of equipment, a facility, or a system is accomplishing its intended function, how is human factors certification going to enhance this process? What are the criteria to be applied that say, "This piece of equipment or system is working better now (i.e., safer) than it would have if it were not human factors certified." Even if we are invited to certify newer items/systems designed to meet acceptable criteria, how can we certify older pieces of equipment that were not designed to human factors principles or standards?

These practical issues suggest that if certification is to become a reality, we should be prepared to support it for a long time, since existing equipment may still be in the field 10-20 years from now; and that a certification process must provide a value-added dimension (e.g. safety, reliability, maintainability, etc.), or else why do it at all. Additionally, certification of the individual cannot mean that the person has expertise in all areas of human factors, since the field has simply become too diverse to maintain proficiency in all domains.

Certification of air traffic control systems is traditionally accomplished by air traffic controllers and operational test and evaluation personnel. In the past, human factors personnel have been involved in research and development of new systems and to some extent in OT &E. However, human factors has not been considered a major element of the testing process but rather as a necessary adjunct or from a program managers perspective, a less than necessary evil. This view has been justified in the eyes of the system developers because when systems were developed in the past without adequate human factors support, any analysis of prototypes was bound to identify unforeseen person-machine issues late in the development cycle. The human factors plan discusses many areas of needed research for evolving systems, but addresses certification primarily from the perspective of controller personnel issues.

Introducing human factors into the certification process beyond what is cited in aircraft systems already will require a cultural change of the magnitude invoked by the Department of the Army MANPRINT program (Booher, 1990). In the introduction of his book, Harold Booher writes:

> People are both the cause and the solution. People are both the benefactors and the victims. Through human error in design, operation, or repair of machines, others are hurt killed or made unhappy or, at the least, inconvenienced. On the other hand, it is through human intelligence and unique human skills that equipment, organizations and knowledge enhancing products are designed and operated effectively, efficiently, and safely. (P. 2)

Booher sees the solution to these issues as a reorientation from hardware to people and conceives the mechanism to achieve this organizational change as "Total Quality Management."

In MANPRINT, the goal is to integrate human factors into every level of material development. This requires a complete systems view of each new piece of technology or hardware. It is mandated though Army regulations. This is both a program and a philosophy of system development. While the program is formally limited to the U.S. Army, the philosophy could go well beyond to other high reliability organizations in which small errors can lead to big problems. The MANPRINT philosophy suggests that there is a long term payback for good human factors in the initial development of a system. This is a life cycle approach to new technology. One of the problems identified by Booher (1990) is that the benefits of early investment in good design for long term system reliability do not usually accrue to the program managers personally because they move on to other developmental efforts. The key to success may be in educating all the personnel involved the process. This sounds reasonable in principle but is not easy to implement in practice. While there have been discussions within the Federal Aviation Administration of developing something similar to Army Regulation 602-2 (U.S. Army, 1990), to date there is no such document in place.

Human Factors Certification of Aviation Systems

There are two central questions that are recurrent. First, do we want to be involved in the certification process to an extent beyond what exists now – which is principally an advisory role – when asked? Second, if we were able to see that increased participation was desired, could we live up to the challenge? Do we have the methods and measures to go beyond our episodic advisory role?

As stated earlier, aircraft systems certification is well institutionalized and has a history going back before World War II. However, as indicated by the National Plan, both engineering and human factors are theoretically accomplished by flight test pilots. This places a great deal of weight on their shoulders which they are very willing to bear. One advantage they have over other aviation systems, such as air traffic control and airway facilities, is that because of their long history there are fairly well recognized standards of systems and individual performance in the cockpit. Aircraft systems are usually, although not always, designed with clear goals for what they are intended to do; this makes evaluation easier. Despite this, it will not be easy to increase the role of human factors in the overall certification process because this would involve a change in thinking on the part of the personnel currently doing the work. It could be viewed as an additional impediment to the process.

In ground systems, the situation is probably even more complicated. First, not all systems are designed with integrated goals other than the global desire to improve safety and performance while reducing operator workload. These are admirable yet non-specific goals and lend to the complexity of defining both system and individual performance. Can we define performance and use it as criterion for evaluation during development and operational testing? Past experience indicates that this is often a moving target which seems to progress as the system evolves. It is further complicated by moving beyond a general definition of performance with a qualitative description of what constitutes "good performance." This involves subjective decisions by subject matter experts who may well have differences of opinion. One can only evaluate the impact of a new system if it can be determined that it somehow is worth the investment in time, money, and safety. This improvement can only be measured if there is technical agreement on what "improvement" means.

From a human factors standpoint, it is not adequate to conclude that a system simply reduces workload because it may in fact reduce the load to the detriment of situational awareness or general alertness. It is the system performance that is the key, and the human is a critical component of the system that can and should not be equated with a piece of hardware. There are issues which machines do not become involved in such as motivation, esprit de corps, fatigue, and human information processing. All of these factors can have an impact on both human and systems performance.

In the process of evaluating a system, there are very few standards which apply consistently. Simply meeting the minimum requirements under MIL-STD 1472D (DOD, 1981) may not be nearly enough to adequately certify a specific application. For example, 1472D is purposefully vague when it comes to human workload. It says that we should not overload the operator. What constitutes an overload would vary from one application to another and for that matter from one operator to another. There have been discussions concerning rewriting this document for civilian aviation applications but no document has been circulated yet, although an FAA airways facilities human factors design standard has been written and is presently under internal review. Even when it is completed, it is likely that systems evaluation and subsequent certification will have to be accomplished on a case by case basis focusing on the design goals of each. For ground systems in particular, this will mean that considerable effort will be required in order to identify and validate suitable metrics that not only meet the criteria necessary for good measurement but have obvious relationships to performance of the systems being evaluated.

Falling back on human factors handbooks and data from days past is useful during design and early development of systems, but these authors believe that the bottom line of any system should be how it really performs either in actual prototype or preferably in high fidelity simulation before prototype testing is begun in the field.

To do either type of testing requires empirical and high quality measurement, and to date there is no general agreement on either the measures themselves or what constitutes acceptable performance. This is especially true when a system is under development to replace one which is currently operating. It can be anticipated that system change may be finite but not meteoric in magnitude. The more subtle the anticipated differences, the more difficult it will be to demonstrate them using conventional measurement and statistical tools. While empirical testing offers the evaluators an opportunity to reach out for what could be a better estimate of operational reality, it does have its drawbacks: time and cost. It takes longer and it costs more to follow a philosophy like MANPRINT in which human factors is integrated into the developmental cycle and empirical evaluation is accomplished whenever possible (at least in the ideal model of the philosophy). What this all means is that considerable effort must be expended to develop the measurement tools that would provide adequate credibility for human factors in the certification process. While there is a history in air side human factors, there is a very limited but fortunately growing level of expertise in ground side operations.

Criteria for Human Factors in Certification

Once again the questions are: Do we in the civil aviation human factors community want to become involved in the certification process and are we able to produce?

The answer the first question is: Of course we do! Even if we do not have all the answers today and all the tools of tomorrow, we can still be of more help than we were in the past if we are invited into the development cycle sooner. The answer to the second question is a qualified maybe. We need to develop ground side and improve air side measurement. This will not happen overnight despite the belief by some that it could be done very rapidly. If it has not happened to everyone's satisfaction in the past forty years, then even with a climate change in the engineering and system development disciplines it will still take time and resources to build adequate measures and methods.

There are some key criteria which need to be met if human factors is ever to become a full partner in certification of aviation systems. The first criteria to be met involves the pursuit of organizational change. Through a Total Quality Management emphasis or though an alternative educational process as suggested by Booher (1990), we need to produce organizational change and recognition that human factors is needed and is valuable. This may not be something you can mandate as the Army has tried but rather should involve attitude change out of self interest on the part of system developers. While human factors has come along way in the aviation community, there are still many who see it as a soft science, if they view in as a science at all.

Along with the organizational change regarding human factors, there needs to be some change in attitude in terms of accepting that there may be better ways of doing things in the development of new technology. The doctrine behind systems engineering is workable if the systems engineers remember to include the operators in their designs and if they seek out technical help related to person-machine issues. Human factors should be a system life cycle issue and not something that is only considered for the short run until the first prototypes are fielded. This type of thinking will show the rewards and costs of system use throughout their life expectancy and not merely the here and now. In many cases such costs are not considered; we live with systems that are user unfriendly and less efficient than they could be.

As human factors professionals, a critical criteria should be to never promise more than we can deliver and to deliver all that we promise. In order to do this, we have to be in the dual role of helping systems developers today avoid the obvious errors of design while constantly trying to develop new and better measurement tools, which are empirical and tailorable to specific applications. Measurement tools must be reliable and valid against systems goals. They also must meet the test of face validity if they are to ever be employed.

References

U. S. Army (1990). *AR 602-2 Manpower and personnel integration (MANPRINT).* Washington, DC: Department of the Army.

Boorer, H. R. (1990). *MANPRINT an approach to systems integration.* New York: Van Nostrand Rheinhold.

FAA (1991). *The national plan for human factors (draft).* Washington, DC: Federal Aviation Administration.

Fitts, P. M. & Jones, R. H. (1947). Analysis contributing to 460 "pilot error" experiences in operating aircraft controls. In H. W. Sinaiki (Ed.). *Selected papers on human factors in the use and design of control systems.* New York: Dover.

Kantowitz, B. H. & Sorkin, R. D.(1983). *Human factors.* New York:Wiley.

DOD (1981). *MIL-STD 1472D Military standard human engineering design criteria for military systems, equipment and facilities,* Washington, DC: Department of Defense.

Nagel, D. C. (1988). Human error in aviation operations. In E. L. Weiner and D. C. Nagel (Eds.). *Human factors in aviation.* San Diego: Academic Press.

Stein, E. S. (1985). *Air traffic controller workload: An evaluation of workload probe* (DOT/FAA/CT-TN84/24). Atlantic City, NJ: DOT/FAA Technical Center.

Thomas, D. D. (1985). ATC in transition, 1956-1963. *Journal of ATC,* 30-38.

Warm, D. D. & Dember, W. N. (1986). Awake at the switch. *Psychology Today, 20*(4), 46-53.

N95- 34768

Certification of Tactics and Strategies in Aviation

54-54
51104
p. 13

Hartmut Koelman

EUROCONTROL

The Need for Operational Concepts

The author's interest in Operational Concepts stems from his work at Eurocontrol, which is related to Air Traffic Management (ATM). Further details can be found in the annex of this paper (the European need for ATM Operational Concepts). These show that the "tactics and strategies" subject is not just to be seen as an academic research issue, but that there is a need for tangible results in the coming years to support the system innovation efforts which will see operational and industrial application after the year 2000.

But what is meant by the terms "tactics and strategies"? A story can help to clarify this. The author is a system analyst at Eurocontrol, but he is also a flight instructor at the University of Leuven (KU Leuven, Belgium). Although the prime purpose of the latter activity is to teach student pilots how to fly sail planes in flatland and mountain environments, it is also a platform for observing and analyzing human behavior.

As a flight instructor, one observes a variety of student shortcomings: insufficient perception of events, lack of situational awareness, lack of anticipation, inability to keep up with the rapid succession of events, faulty judgements, improper setting of priorities, lack of insight in cause-effect relationships, loss of concentration, etc. One also observes the learning curve in student pilots: from when they master attitude control and think that they can fly (except that the instructor has to tell the student what to see, what to do, and when/where/why/how to do it), and through the successive increase in skill and workload capabilities. Whenever a task is mastered, it becomes an automatism.

Automatism causes a workload reduction and the student is ready to tackle a new aspect of flight management. It ends with the situation where the instructor primarily monitors the judgement of the student (with associated problems of underload and boredom for the instructor). Meanwhile, as the instructor, one must constantly keep track of the student's (changing) performance limitations and decide when the student may be exposed to the expanding range of operating conditions. Last but not least, one must know when to take control of the situation so that the safety of the flight is not compromised but the student still learns from his or her mistakes.

Over the years, student after student completes training, and the instructor starts recognising patterns of cognitive behaviour. The conclusion is: everything can be considered a tactic or a strategy. The student masters the tactics needed to deal with a situation. Subsequently, he/she learns strategies which will control the application of those tactics. After that, super-strategies are needed to control the application of the strategies, and so on. The student learns a hierarchy

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

of techniques (i.e., an Operational Concept), starting from short-term anticipation and simple situation tactics to long-term anticipation and complex situation strategies. Many of these techniques are not written down as formal operational procedures, but have to be mentally acquired through personal experience. Finally, the student may develop into the experienced glider pilot who spends months planning and preparing for a world record attempt, waits years for the day with those unique meteorological conditions that the plan requires, and then executes that plan. Thousands and thousands of strategical and tactical decisions are needed in the ten or fourteen hours it takes to complete the flight.

The point is, (for certification, specification or any other purpose) to separate a system just into its design (the static part) and its operation (the dynamic part) is a far too simplistic view of the world, as is the separation of system operation just into strategies (the static part of operations) and tactics (the dynamic part of operations). The stability-volatility spectrum in system characteristics is usually not a black and white picture. It has many shades of grey. For the sake of the argument, it is unwise to allow oneself to become handicapped by a lack of conceptual richness in the English (or any other language's) vocabulary. It may be the case that the understanding of system operation or human cognitive behaviour (i.e., the task analysis) requires a whole hierarchy of planning techniques, with varying degrees of stability (from very tactical to very strategic and design-like). The same is true for the information which describes designs, strategies, tactics and the situational awareness which is associated with the planning techniques on these different levels. The terms "information" and "data" are typically erroneously used to convey the idea that every piece of information is equally significant (a typical engineering mistake). Information should be categorized in accordance with the different planning layers to preserve the operational significance of the different degrees of volatility of the information.

Typical Operational Concepts

An Operational Concept is a model which describes the dynamics of managing an operation. It is expressed in terms of planning and execution responsibilities, control loops and operational procedures. The operation is described for the composite human-machine system.

Traditionally in a composite human-machine system, the human still carries ultimate responsibility for the satisfactory performance of the composite system. Hence, the planning and execution responsibilities, control loops and operational procedures mentioned in the glider example, although supported by machines (automation), are primarily a human factors concern.

Operational Concept for Air Traffic Management (ATM)

If we look at the way ATM is organised today, and most likely still will be organised in the year 2015, we see the use of planning "layers" with different planning horizons (Eurocontrol, June 1992; EATCHIP Task Force on ATC System Integration, June 1992). The general objective for each layer is to deliver an acceptable situation to the next layer. Each layer works as a "filter" for the following one. This *filtering strategy* defines specific roles for each layer, with the higher layers addressing a general scope and a long planning horizon, and the lower layers concentrating on specifics and short planning horizons. A typical layering for ATM is as follows.

Development of Operational Concepts, standards and recommended practices on worldwide and regional scale (10 – 15 years lookahead): serves to provide guidance for system procurement actions and aircraft mandatory carriage requirements, items which have lead times of 5 – 10 years typically.

Planning the renewal and upgrade of infrastructure on regional and sub regional scale (5 – 10 years lookahead): deals with site construction, development and procurement of hardware/software, and strategic human resource planning.

Strategic Airspace Management (ASM) on a regional scale (several years lookahead): defines basic strategies for ATS Route Networks (ARN), airspace use (segregation/flexible), and traffic segregation.

Strategic Air Traffic Flow Management (ATFM) on regional scale (up to one year lookahead): strategic ATFM activities are intended to resolve major demand/capacity imbalance problems and generally concern summer traffic flow. They result in a Traffic Orientation Scheme based on published flight schedules, airspace structure and system structure and capacity. After processing, estimates of traffic loads over any navigation point or ATC sector can be provided. Discussions are initiated with all the partners concerned (states and aircraft operators), starting each year in October and ending in February. The outcome of this planning is the Traffic Orientation Scheme, which dictates the routes to be used by operators when planning flights from specific departure areas to specific destination areas during the coming summer (Martin, 1993).

Pre-tactical ATFM on regional scale (1 day lookahead): pre-tactical activity is directed at the specific situation one day ahead. On the basis of updated demand data (incorporating Repetitive Flight Plans (RPLs) and last minute changes notified by aircraft operators), archived data of traffic situations on a similar day in the recent past, and taking into account the latest information about capacity in the Area Control Centers (ACCs), a pre-tactical ATFM plan for the coming day is developed. This plan, which defines the restrictions to be applied to traffic flows on the day concerned, is published every day around noon in the form of an ATFM Notification message (ANM) and is dispatched to more that 1000 addressees (Air Traffic Services and Aircraft Operators). The ANM describes in a single message all the tactical ATFM measures which will be in force on the following day (Martin, 1993).

Tactical ATFM on regional and Flight Information Region (FIR) scale (several hours lookahead): on the day of operation itself the Flow Management Units (FMUs) will apply the measures announced in the ANM and will monitor whether the pre-tactical plan is having the desired effect. At the present time, tactical ATFM is based on the application of acceptance rates, expressed in terms of the number of flights per unit time that will be allowed to enter a specific congested area from a particular area of origin. Aircraft which plan to fly through a congested area – as detailed in the ANM – are expected to request a slot from the appropriate FMU. Slots are allocated mainly in the form of a revised departure time but sometimes as a time of arrival at an en-route point (Martin, 1993).

Air Traffic Control (ATC) Area Management on sub-FIR scale *(1 – 2 hours lookahead)* corresponds to operational supervision and tactical Airspace Management (ASM). This activity is responsible for dealing with events having a significant effect on traffic handling and throughput (such as NOTAMs, system failures, changes of airspace, airport and runway availability, meteorological hazard or traffic overflow). It adapts the sectors, selects the overall strategy for dynamic routing, airspace use, traffic segregation and runway usage in order to meet the strategic regulation plan (tactical ATFM plan) and the required capacity of the involved sectors. This includes assessing and smoothing the center's workload for the coming hours.

Planning ATC on single or multi-sector scale (20 – 30 minutes lookahead): serves to organise the traffic entering and leaving the planning area so as to avoid unmanageable situations inside the planning area (from a flight safety, flight economy and an ATC workload point of view). This involves aircraft sequencing, allocation of runways, routes, levels, delays and coordination with adjacent planning areas in order to establish agreed transfer conditions. The resulting plan must include a certain contingency to give Executive ATC the "manoeuvring liberties" needed to resolve unexpected problems.

Executive (tactical) ATC on single-sector scale (5 – 10 minutes lookahead): is responsible for implementing the plan established by planning ATC while maintaining satisfactory levels of safety (through separation assurance and aircraft guidance). The provided separation and aircraft guidance has to meet certain legal requirements (minimum separation values which depend on geographical, technical and institutional circumstances). Executive ATC has to monitor a highly dynamic system where the nature of the problems at hand can significantly change in the course of a few minutes. The resources available for problem solution are limited, and there are hard real-time constraints for the control loop which must detect these problems, develop solutions, and implement those solutions.

ATC Safety Net layer on single aircraft scale (2 minutes lookahead): complements the Executive ATC with functions such as Short Term Conflict Alert (STCA) and Minimum Safe Altitude Warning (MSAW). The purpose of this layer is to catch those safety threatening situations which were not resolved by the Executive ATC layer.

This hierarchy of layers can be continued to shorter and even negative time horizons (in this case the term "planning layer" is not appropriate any more).

Real-time operations layer on single aircraft scale (real-time): the physical act of communicating clearances, instructions, advice and requests to individual aircraft, executing procedures and changing the internal state variables of the ATM system as a reaction to the occurrence of triggering events (this progresses the chain of events). Real-time operations are what an incidental visitor sees happening when he/she observes a controller on duty. The incidental visitor would not see the meaning (all the tactical and strategical considerations) which are behind the observed real-time operations.

Forecasting and extrapolation layer *(past situations extrapolated to a target time, i.e. to real-time or to the future)*: serves to produce assumptions about the state vector of the relevant objects (individual aircraft, weather phenomena, the traffic flow, etc.) based on a description of the situation in the past (obtained via the history data collection layer). The purpose of forecasting and extrapolation is to close the feed-forward loops by transforming history data into a state which matches the time horizons used by the various higher planning layers. Forecasts and extrapolations may be deterministic (with reduced accuracy of extrapolated state variables) or probabilistic (if the target time is too far ahead of the recording time). A track extrapolation 10 seconds ahead is an example of deterministic extrapolation, and a two-day weather forecast an example of probabilistic forecasting. Note that forecasting and extrapolation are to be distinguished from prediction based on an object's intentions (such as the planned trajectory or the clearance of a flight).

History data collection layer (negative time horizon, i.e. delayed): this is called surveillance data acquisition in the ATC context. It serves to create an accurate recording of the air traffic situation (or weather situation, etc.) over time. The obtained accuracy depends on the sensors used. The average age of the most recent data (the delay) depends on the sampling rate and the processing/communication delay. The data may be used in real-time for planning and control purposes (after extrapolation), or for off-line applications (route charges, accident/incident investigations, statistics, etc.).

Operational Concept for Flight Operations Management

The Flight Operations Management Operational Concept is a layered model quite similar to the ATM Operational Concept. Aircraft Operators are faced with managing the three main phases of aircraft operations: flight time, taxi time and turnaround time in preparation of the next flight. Each of these phases is managed in a number of planning layers and properly coordinated with the other phases. To illustrate the similarity with the ATM planning layers, here is a typical list of planning layers for the flight phase, as applicable to a scheduled airline:

- longterm strategic planning, to determine business opportunities and decide between fundamental options (10 – 15 years lookahead)
- aircraft fleet planning (5 – 10 years lookahead)
- acquisition and planning of commercial routes and destinations (several years lookahead)
- development of timetables for the coming season (6 months – 1 year lookahead)
- negotiation of routes with strategic ATFM (6 months – 1 year lookahead)
- filing of Repetitive Flight Plans (RPLs) (approx. 6 months lookahead): strategic flight planning: determination of aircraft maintenance schedules and initial allocation of resources to individual flights (aircraft, crew, logistics etc.) (approx. 3 months lookahead)
- tactical flight planning for individual flights, based on pre-tactical ATFM (1 day lookahead)

- Direct flight planning (off-board crew activities), based on latest NOTAMS, actual weather forecasts, cabin briefing etc. (45 – 60 minutes before scheduled departure time, lookahead until a few hours after scheduled arrival time).

- Direct flight planning (on-board crew activities), based on load sheet, ATIS etc. This results in completed fuel order, computed take-off data, planned 4-D trajectory, estimated time en-route (ETE), etc. (15 – 45 minutes before scheduled departure time, lookahead until scheduled arrival time).

- Strategic flight management is the coordination with ATC to obtain pre-departure clearance (PDC), including a departure slot, and other flow restrictions. Adjust planned 4-D trajectory accordingly. This corresponds to decisions made by tactical ATFM. Subsequently obtain start-up approval (5 - 15 minutes before "Off Blocks").

- Pre-tactical flight management: obtain expected clearances for departure, en-route, climb, descent and approach, and adjust planned 4-D trajectory accordingly. This represents coordination with decisions made by the Planning ATC layer (during taxi and flight, 30 minutes lookahead).

- Tactical flight Management: obtain actual clearances for push-back, taxi, take-off, departure, en-route, climb, descent, approach, landing and taxi, and adjust planned 4-D trajectory accordingly. This represents coordination with decisions made by the Executive ATC layer (from 5 minutes before "off-blocks" to docking and engine shut down, 5 – 10 minutes lookahead).

- Prepare execution of tactical manoeuvres, based on the latest 4-D trajectory plan and tactical ATC instructions (vectoring). This represents execution of cockpit procedures and coordination with decisions made by the Executive ATC layer and the ATC Safety Net layer (during taxi and flight, 30 seconds to 2 minutes lookahead).

- Prepare execution of collision avoidance manoeuvres, based on visual observations and/or ACAS/TCAS (during taxi and flight, 5 – 30 seconds lookahead).

- Real-time operations layer (real-time): the physical act of operating (controlling) the aircraft and manoeuvring it in accordance with the most up-to-date 4-D trajectory plan, and the act of communicating with various partners such as ATC.

- Forecasting and extrapolation layer (past situations extrapolated to a target time, i.e., to real-time or to the future): for example, dead-reckoning techniques, fuel burn prediction, and other estimations.

- History data collection layer (negative time horizon, i.e. delayed): for example, flight data recording and navigation data acquisition.

Planning Theory For Operations

The foregoing describes the functioning of "layered" ATM and Flight Operations Management in specific (operational) terms. The advantage of this approach is that it is pragmatic and permits the reader to relate the story to his or her own operational experience as opposed to being theoretical. The disadvantage is that the same basic operational problems are unknowingly solved over and over again, for different lookahead time scales, in different terminology, and by people with different backgrounds.

Now this section of the paper addresses issues such as:

- the different goals that may be set in an automation strategy (these goals are heavily influenced by human factors)
- the development and assessment of certification strategies for operations management

Thus for certification purposes, we need a kind of theoretical insight into this "layered" planning technique. What is called "planning theory" in this paper represents an attempt to identify some of the basic underlying principles in the operation of an Air Transport System. The subject of planning theory is now presented in the following.

The Players in Operations Management

The Air Transport System. The Air Transport System is the domain of interest of this paper. The generic term *Air Transport System* refers to the aggregate of weather, airspace, aerodromes, aircraft, aircraft operators (commercial, military, general aviation and aerial work) and ATM/CNS (Air Traffic Management/Communication Navigation Surveillance) Systems, all operating together in a particular geographical region.

Actors. In accordance with the above definition, the Air Transport System consists of a number of interacting elements, such as airlines, aircraft, pilots, airports, runways, airspace, routes, ATC units, controllers, systems, weather phenomena, aircraft separation etc. For the sake of generalisation, I will call these elements the *Actors* of the Air Transport System.

Relationships between Actors. The operation of the Air Transport System is far more than the sum of the operation of the individual Actors. Indeed, these Actors are in constantly changing interaction with each other. Managing the operation of the Air Transport System means managing these interactions. Some interactions are to be promoted because they are necessary ingredients of the proper operation of the Air Transport System. Others represent problems, and system management efforts are directed at avoiding or removing such interactions. An example of the latter is the separation conflict between two aircraft. Some of the most important general types of interactions (relationships) are:

- *User/resource relationship:* Actors in a resource role exist in limited supply with a variable number of Actors in a user role competing to use that supply (establish a user/resource relationship). Examples are: aircraft operators with respect to aircraft, aircraft with respect to runways, clients with respect to database servers, aircraft with respect to route capacity, aircraft with respect to mutual separation, flights with respect to ATM, etc.

- *Competitor relationship:* users competing for the same resource are in a competitor relationship. Examples are: two aircraft approaching the same airport at the same time, two VHF radios attempting to transmit on the same frequency at the same time, areas of severe weather with respect to aircraft wishing to use that same airspace, etc.

- *Collaborator relationship:* resources able to distribute the user load between them (i.e. to reduce bottlenecks) are said to be in a collaborator relationship. Examples include: parallel runways, different flight levels, parallel routes, main vs. reliever airports, etc.

- *Buffer relationship:* resources able to temporarily absorb the user load of another resource are called buffers. Examples are holding patterns, route extensions, queues, contingency measures, etc.

- *Control relationships:* sometimes Actors are responsible for the operation of another Actor. This responsibility may range from defining an operating envelope (providing policy, guidance, operating constraints, or allocating workload), to assuming detailed control. Examples: pilots with respect to aircraft, controllers with respect to controlled flights, Air Traffic Flow Management (ATFM) with respect to Air Traffic Control (ATC), etc.

- *Target relationships:* when Actors are planning to reach a goal, actor and goal are said to be engaged in a target relationship. Example: a flight with respect to its destination airport.

- *Part-of relationship:* indicates the (permanent or temporary) assembly of individual actors into a composite system, having a certain state vector in common (such as a runway is part of an airport, a pilot is part of an aircraft in flight).

The Environment. In simple terms, the environment of an actor is everything that surrounds that actor. However, only that part of the environment is relevant with which the actor interacts in one way or another.

This leads to the following natural definition of environment: the total set of existing, expected and planned relationships with other actors. The environment is dynamic because the membership of this set is subject to change as time elapses: relationships disappear and new ones come into existence.

The Conduct of Operations

Operations in a World Without Planning. Object-Oriented Analysis (OOA) techniques such as Coad and Yourdon (1991) and Shlaer and Mellor (1992) describe the world in a mechanistic manner as a set of objects (actors), with predefined relationships, predefined life cycles (state transition diagrams), predefined event chains and communication capabilities.

This may perfectly suit the needs of system analysis for the purpose of developing static (non-adaptive) pieces of software or analysing rigid organisations (mechanistic systems), but it seems a bit inadequate for documenting the operation of complex goal oriented systems such as the Air Transport System.

OOA may correctly capture such a system on a syntactical, real-time operations level (existence of Actors, possible relationships, state transitions, etc.), but it misses out on the semantics (the whole layered planning process preceding the physical conduct of operations).

Operations in a Goal Oriented World. In contrast, the operation of the Air Transport System is highly adaptive (i.e., it is governed by a set of constantly *modified* and *(re)created* scenarios, scripts, procedures, project plans, flight plans, story boards, event models, rules, regulations, strategies, tactics, philosophies, etc.). This modification and (re)creation is the above mentioned layered planning process which precedes the physical conduct of operations. It happens this way because most of the actors in the Air Transport System are goal oriented entities. In order to reach a goal in a world full of uncertainties and conflicting requirements, one needs to plan the future and reduce the amount of improvisation. In fact, all the actors of the Air Transport System spend a considerable part of their energy on such planning activities.

So what are these planning activities all about, in a nutshell? They are about developing a scenario, refining and finally executing it, in an environment of external and self-induced perturbations.

The external perturbations are the unforeseen events, interactions and timing in an actor's environment. External perturbations occur because the environment may be inherently unpredictable, but also due to lack of overall coordination in the Air Transport System. The self-induced perturbations come from an actor's inability to accurately execute his or her own operational scenario. These are cases of mismanagement, in the operational sense. On top of that, the Actor may simply be following a bad scenario (e.g. with lack of feasibility and full of inconsistencies). This type of problem and the perturbations give rise to the need for constant situation assessment and revision of the scenario.

Fuzziness in the Planning Process. Actors deal with two types of scenarios: probabilistic and deterministic. The countdown towards the moment of physical execution of a particular operation is normally spent in different *uncertainty phases:*

- *PHASE 0*: the need for the operation has not yet been identified

- *PHASE 1*: the need for the operation is identified, including an approximate target time, but no plan or scenario is available

- *PHASE 2*: the phase of fuzzy and probabilistic scenarios

- *PHASE 3*: the reduction of uncertainty, to transition from fuzzy and probabilistic scenarios to a very limited number of candidate scenarios (scripts)

- *PHASE 4*: the phase in which one of these candidate scenarios has achieved a very high probability of occurrence, and has become a structurally stable script for the operation ("structurally stable" means that sequence of events is stabilised, and partners for various types of relationships are known, e.g. "contractual" status of relationships are established)

- *PHASE 5*: the phase in which the script does not structurally change, but the accuracy of its various parameters (such as timing, planned value of state vector, details of interactions with other Actors) is improved

- *PHASE 6*: the actual execution of that script, resulting in a physical operation (the chain of events is progressed)

- *PHASE 7*: the phase where factual data on what has happened is not yet available

- *PHASE 8*: the availability of history data describing what actually happened

To visualise the relationship between these uncertainty phases and time, an uncertainty-time diagram is used in this paper. The time axis is to be seen as absolute time; and the uncertainty axis lists the above phases to give a qualitative idea of the accuracy of a given scenario.

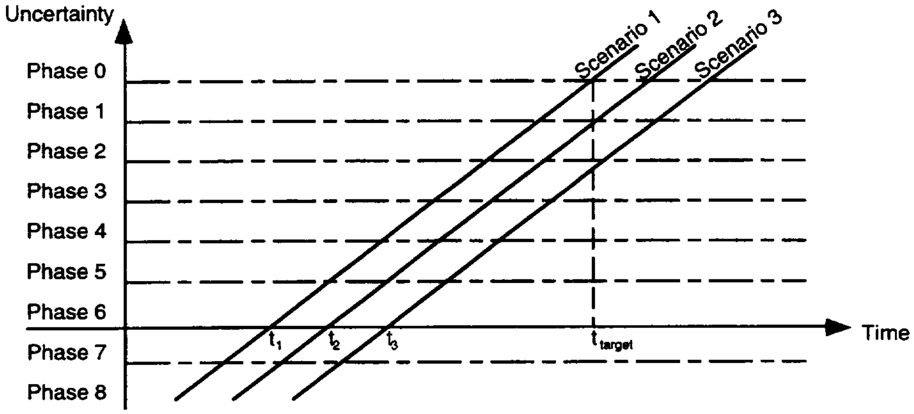


Figure 1.

Figure 1 illustrates the production of different versions of a scenario. Scenario 1 is associated with time t_1, scenario 2 with time t_2, etc. These reference times correspond to the real-time execution of what is spelled out in a particular scenario version. In the example of scenario 1, the part before t_1 is history, and the part after t_1 is the plan for the future. An example will clarify this. A flight plan is to be seen as a scenario. At time t_1, part of the flight has been completed already, the next leg is quite accurately planned, but the details of arrival are uncertain. For example, because of the unpredictability of the weather it is uncertain whether the alternate destination will have to be used or not.

This story can also be looked at from the perspective which is so well known from space vehicle launches: the countdown view. While counting down, a particular time target in the future (such as the time of arrival of a flight) is associated with different scenario versions as time elapses. Each new version is more accurate with respect to time target because prediction is less of a factor.

Figure 1 also illustrates that each scenario version is structured in a particular way: it is accurate in the short run and exhibits properties of the earlier fuzzy phases the further it looks into the future. It is important to note that Actors do quite a lot of planning and reasoning while still in the phase of fuzziness.

All this explains why operations management works as a "rolling program" of scenario development. In this approach, a particular situation at time target seems to be planned in an iterative fashion because it is associated with a number of different scenario versions. On the other hand, the planning process chases a moving time horizon because the subsequent scenario versions are time stamped differently, but must maintain the same outlook with respect to their time stamp.

This scenario revision process is implemented by a control loop. In project management, the control loop is called the *PDMA-cycle* (plan-do-monitor-adjust). In ATM, one usually distinguishes the following phases (Eurocontrol, June 1992; EATCHIP Task Force on ATC System Integration, June 1992):

- acquire information
- monitor current situation
- predict evolution
- identify problems
- propose and evaluate solutions
- choose solution
- communicate and implement solution.

Each cycling through these control loop phases produces a new version of the scenario. But let us return to the previously mentioned hierarchy of uncertainty phases. In system terms, this hierarchy could be expressed as the following strategy:

- determine the desired start state and end state (goal/target) of the operation
- plan the time of occurrence for start state and end state
- select intermediate states (sub-goals)
- determine the order of intermediate states (temporal organisation of activity)
- elaborate synchronization and coordination requirements for the operation (type and sequence of the relationships needed or expected during the operation)
- determine the partners for these synchronization and coordination relationships (production of a structurally stable script, establishment of "contractual" relationships with partners)
- work out the timing (start to end) of the synchronization/coordination with each of the above partners
- determine the detailed timing (accuracy) for the intermediate states
- operate in real-time, i.e. perform state transitions and interact (exchange events) with various synchronization/coordination partners
- collect history data.

Take any type of operation, say ATFM, planning ATC, flight planning, flight management, project planning, and it is possible to express the operations planning in the above terms. A third way of expressing this strategy is reflected in the traditional *WHAT-HOW-WHERE-WHEN* sequence:

- the *WHAT* phase identifies the operation
- the *HOW* phase is responsible for identifying the needed interactions
- the *WHERE* phase produces the stable script which identifies the partners for the individual interactions
- the *WHEN* phase puts on the accuracy by refining the timing.

One can continue repeating this strategy in different terminology disguises. In a systems development context this "scenario development strategy" is called a life cycle:

- user requirements definition
- operational concept definition (sometimes termed requirements analysis)
- operational requirements definition (also called system requirements definition)
- architectural design (alias technical concept definition)
- detailed design
- system procurement and installation
- system operation.

In project management terms, the hierarchy looks as follows:

- state the overall mission of the project
- determine the completion date of the project
- develop the work break-down structure (WBS)
- identify the interdependencies between work packages (production of Pert Chart)
- perform rough allocation of the total project duration to individual work packages (production of initial Gantt Chart)
- allocate resources to work packages
- refine timing of work packages by eliminating resource over-allocation (production of Gantt Baseline Chart)
- adjust the project plan based on plan deviations
- execute the project plan
- do progress tracking.

All these strategies are nothing more than variations on the same basic theme. Depending on the complexity of the operation and the expected number and magnitude of perturbations, the length of this countdown process – alias planning strategy – may take just a few seconds, and on the other extreme several years or even decades.

The Feasibility of a Scenario. The objective of each layer (or countdown/anticipation phase) is to deliver an acceptable situation to the next lower layer. This means: maintaining a set of conditions (i.e., a "solution space" or operational performance envelope) in which one or more feasible action plans exist. If the higher layer fails to maintain those conditions, the lower layer might not be successfully completed. The expression "to pass the point of no return" emphasizes the timing and state transition aspects of this feasibility collapse.

Consequently, one of the responsibilities of a higher layer is to maintain a constant awareness of the operational performance envelope of the next lower layer. The ATC concept of "minimum legal aircraft separation requirement" is an example of such an operational performance envelope. The above mentioned feasibility collapse may have internal and external (environmental) causes. A mistake in the calculation of aircraft endurance during flight planning is an example of an internal cause. So is the failure of a pilot to initiate the landing flare at the right moment, or the failure of a controller to detect a loss of separation between two aircraft. An unexpected weather change to IMC (Instrument Meteorological Conditions) during a VFR (Visual Flight Rules) flight is an example of an external cause.

The Impact of Perturbations. As mentioned, there is a constant need for situation assessment and revision of the scenario due to internal and external perturbations. In addition, the notion of "operational performance envelope" has been introduced.

The impact of perturbations depends on the magnitude of those perturbations. In this context, "magnitude" can refer to size as well as duration. If the magnitude of the perturbation exceeds the operational performance envelope of the planning layer under consideration, then there is nothing this planning layer can do to solve the problem. It is up to a higher layer to take care of the situation. That, of course, cannot be done in a reactive way after the problem occurred. By virtue of its longer planning horizon, the higher layer is supposed to have prevented the problem.

If the magnitude of the perturbation does *not* exceed the operational performance envelope of the planning layer under consideration but it exceeds the envelope of the next lower layer, then this planning layer is responsible. It has to modify the scenario within the possibilities of that particular planning horizon.

If the magnitude of the perturbation does not even exceed the operational performance envelope of the next lower layer, then this planning layer does *not* have to change the scenario with respect to that particular planning horizon.

Whatever layer is responsible, in a properly functioning system the problem is solved in anticipation (a certain time before it would actually occur). This can be seconds, minutes, hours, days or even years in advance. Additionally, this revision of the scenario by a particular layer invalidates all the plans under the responsibility of lower layers. This imposes certain time constraints on the lower layers which have to recreate their part of the scenario from scratch. Indeed, imagine the situation in which there exists a sufficient number of possible solutions on the shorter planning horizons, but the responsible human or machine is unable to produce these solutions in the available time. An example of this is the situation where the pilot "doesn't keep up with the airplane": he or she is overtaken by events rather than staying abreast of them.

The Impact of the Environment. In order to plan the scenario of an Actor with a certain accuracy to a certain time horizon, the predictability of the environment must be equal or better than the fuzziness/accuracy of the desired scenario.

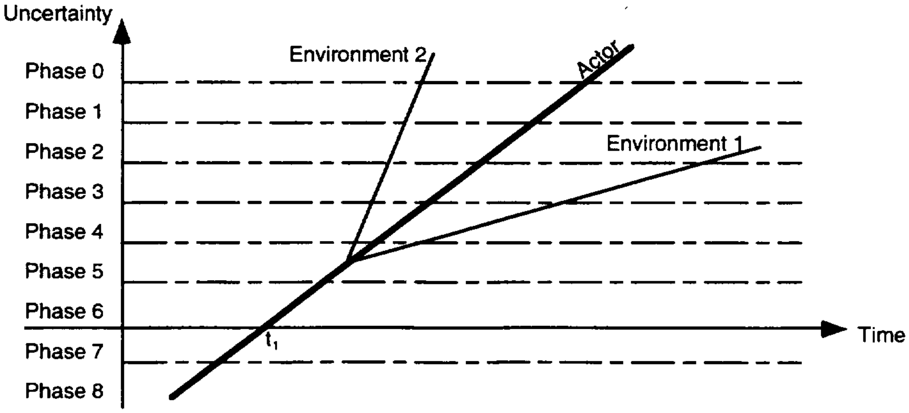


Figure 2.

This is illustrated in Figure 2. The scenario labeled "Actor" can be implemented in environment 1 but not in environment 2.

Let us illustrate this with the example of planning a conflict free 4-D tube clearance (the "Actor" scenario) in a given environment of other aircraft trajectories. The bottleneck of the planning process is the part with the greatest uncertainty. Uncertainty translates in this example into planning horizons, time windows (departure, overflight, climb, descent, arrival), positional accuracy and confidence levels.

Assume the following operational goal for the Actor: touchdown within 30 seconds of exactly 2 hours. Of course the Actor needs to have the capability to execute this scenario with the required accuracy. It is intuitively clear that it is feasible to plan this in an environment number 1 where the landing times of the other aircraft will occur with an accuracy of 10 seconds. It is equally obvious that such planning is pointless in an environment number 2 where the landing times of the other aircraft will occur with an uncertainty of 5 minutes, unless the minimum separation values (safety margins) are greatly increased, with a resulting reduction of control capacity.

The Impact of a Lack of Knowledge. The role of knowledge is quite similar to what was said about the environment. Note the choice of words in the previous example: "will occur with an accuracy of". If I replace this with the words "is known with an accuracy of", we see the impact of a lack of information.

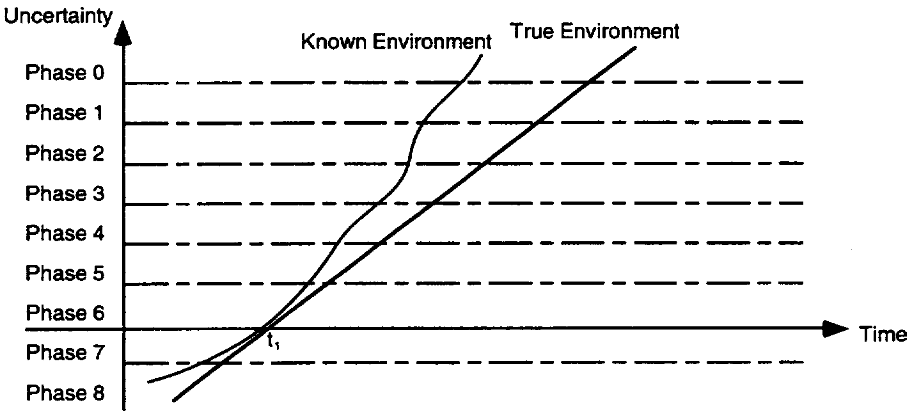


Figure 3

Figure 3 illustrates that the environment is assumed less deterministic than it really is, due to insufficient information. The Actors in the system have a lack of situational awareness, which translates into reduced planning horizons (the horizontal delta on the diagram) and reduced certainty at a particular outlook time scale (the vertical delta on the diagram). In plain words, the planning can only be as accurate as the knowledge on which it is based.

The attempt to approximate the inherent unpredictability as much as possible is the reason why future Operational Concepts strive to use better surveillance, better coordination and higher levels of system integration – air/ground integration and use of data link in particular.

Lack of Correlation with Reality. The other extreme is a lack of correlation with reality: sophisticated models of the future which make the Actors believe that the situation is very much under control.

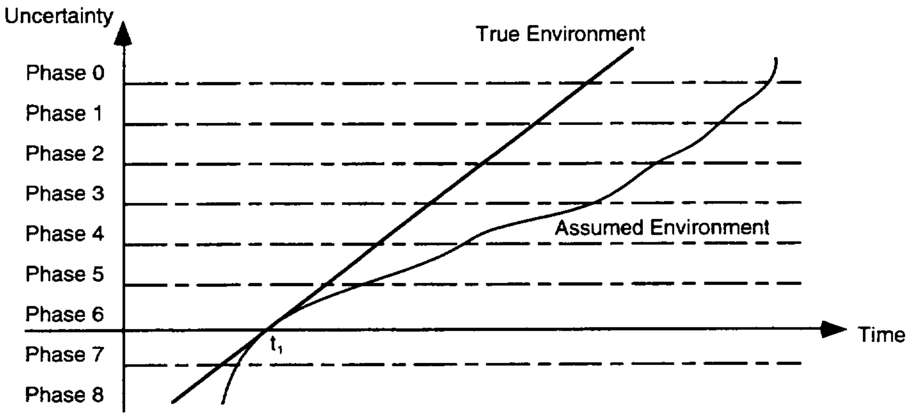


Figure 4.

That situation is represented in Figure 4. The description of the future is too precise: it is a collection of unfounded assumptions which will probably turn out to be false. For example, an ATC system which "sells" the idea that the next separation conflict of aircraft X will be with aircraft Y, whereas at the given moment it is inherently impossible to say whether it will be with aircraft Y or Z. A system acting like this will either fail or at least exhibit very poor performance because each such case of "over-confidence" probably creates a mistake.

The Need for Planning. Orville and Wilbur Wright did not need air traffic control in 1903. They certainly did not have any use for flow management. There seems to be a tendency that systems evolve to loose their simplicity over time (people call this "more advanced"): the environment becomes more complex, the system's internal complexity increases, the operation needs to be more optimized, uncertainty is less and less acceptable, and the operational performance envelopes are extended to enable the previously impossible. What used to be simple now requires advanced and accurate planning: no more "flying by the seat of the pants". This is now "follow the procedures" and "fly by the numbers". There was a time when aircraft flew but ATC did not exist, then a time with ATC but without planning controllers, and finally there was a need for ATFM (which is of course fairly recent).

In the future, one might see the beauty of Air Traffic Management: the completely deterministic Operational Concept with nearly 100% safety and nearly unlimited control capacity by virtue of highly accurate pre-planned (booked) 4-D conflict free flight trajectories from take-off to landing. You negotiate (book) a flight plan (all previously booked flight trajectories remain unchanged), and from then on everything unfolds like clockwork. Unfortunately, there are many uncertainties due to external events. Besides that, even if unexpected external events did not exist, operations managers on all planning levels might *want* to change their mind once in a while, instead of having to stick to plans which were cast in concrete long ago. A few things need to be remembered from the above:

- the need for planning tends to increase as systems, organizations and technology become more mature
- the stack of planning layers builds from the bottom up
- those higher planning layers always address the need for more global optimization of operations (the need for bird's eye views and crystal balls)
- the need for planning depends on the complexity of the operating environment (e.g. traffic density and geographical context)
- the need for planning depends on the complexity of the system (organisation and/or equipment) and the type of operation.

Optimization Strategies for Operations Management

What do goal oriented entities (humans, machines or composite systems) usually do to optimize their operations? A number of general strategies always return:

- have a plan, as early as possible
- have a plan, as feasible as possible (maximize contingency)
- have a plan which includes a strategy for dealing with probabilistic situations
- re-assess and revise the plan as often as possible
- have a plan, as stable as possible (i.e., the revisions should be as small as possible)
- have the ability to (re)create a plan quickly (improvise if necessary), to minimize the time delay between the last situation assessment and the availability of the new plan
- have the ability to stick to the plan (a minimum of internal perturbations)
- solve problems (external perturbations) as early as possible
- solve problems (external perturbations) as thoroughly as possible (full impact analysis of solution)
- if there is a choice, operate in as stable an environment as possible
- if there is a choice, operate in as predictable an environment as possible
- if there is a choice, operate in an environment with the least number of interdependencies (low complexity environment)
- avoid the unknown, i.e. operate in as well known an environment as possible (maximize the available information)
- split the planning process into different concurrently operating layers with different responsibilities, based on the dynamics of the possible perturbations (work with a hierarchy of plans)
- know the true extent of all the performance envelopes (how far can you go on each layer without compromising safety)
- devise a proper filtering strategy to dispatch perturbations to the responsible planning layer (full impact analysis of perturbation): a dispatch to a layer which is too high leads to unnecessary re-planning; a dispatch to a layer which is too low leads to safety problems.

Certification Issues

How should planning theory (i.e., the above considerations of planning layers), control loop phases and performance envelopes be seen within the context of certification? Rather than trying to give a complete answer, this paper attempts to give a number of useful indications. Before going into details, however, let's clarify the used terminology.

Definitions

For the purpose of this paper, I assume the following definitions:

- *Verification* is a review process, to check the system requirements against their source. The validity of the requirements themselves is verified, to ensure that a system which would be built to satisfy these specifications would also be suitable.

- *Validation* is the checking of a system design against the requirements. It is the production of (formal or experimental) proof, serving to establish a measure of confidence in the correctness and effectiveness of important system features. Validation is performed "after the fact", as for example during acceptance tests.

- *Feasibility Study* is similar to validation, but different in the sense that it is done in the exploratory phase ("before the fact"), in order to select a suitable solution amongst different possible alternatives. A feasibility study never replaces validation.

- *Certification* is the administrative "rubber-stamping" of a validation, an endorsement to give it an official status and level of authority.

The Limitations of Certification

There is one catch in regard certification: the correctness and effectiveness of the certified system features (the fitness for operation) are not endorsed under unlimited operational circumstances. Every certified system or person "carries" a piece of paper which lists these circumstances and/or limits of authority: system or person such-and-such is certified to deliver operational performance X during a period Y under operational circumstances Z. In fact, these limitations can be equated to the "operational performance envelopes" which were introduced earlier in this paper.

Certification of the Planning Process

Normally, Operational Concepts are not certified in their totality, probably because that is beyond today's state of the art. However, there is a need to develop, validate and certify the standards and recommended practices which are used in support of Operational Concepts.

Traditionally, certification was focused on systems (equipment or humans) in order to qualify their functional capabilities and operational performance. Emphasis always seemed to be on the *real-time* operational performance (uncertainty phase 6 in the diagrams of this paper), because that is the easiest to observe and, more importantly, because it represents the ultimate judgment on proper operation of the system. It is the stuff the whole operations planning process is finally all about.

Now, systems become more complex and rely more and more on planning (automated or human). This means that it is no longer sufficient to certify that "the system works", regardless of how it achieves this goal. The planning process itself needs to be certified because it determines those situations in which a system will and will not work. As mentioned earlier, the documentation of these limitations is a crucial element in certification.

In other words, the time has come to consider planning layers and their individual control loop phases as objects for certification, instead of just physical people, equipment, functions and procedures.

After having said this, it must be clearly stated that planning layers and control loop phases already exist in today's systems because all kinds of functions and responsibilities fulfill these roles. But these functions and responsibilities have not been consciously designed based on sound planning theory. Instead, they historically evolved in a bottom-up fashion, over many system generations, just as language is a product of history rather than a "careful design". Thus the certification problem is considered to be twofold:

- certify the operational principle (i.e. the effectiveness of a certain combination of planning strategies)

- certify the implementation (i.e. the functions, responsibilities, etc.) of these planning strategies to certain performance standards.

Let us rephrase this in a bit more detail:

- Operations management uses a layered planning process which includes strategies for the reduction of uncertainty and for dealing with perturbations. The performance of this planning process can be expressed as specified under "Optimization Strategies for Operations Management" and in terms of the uncertainty phases defined under "Fuzziness in the Planning Process" in this paper.

- The interoperability of these layers and phases depends on mutual awareness of operational performance envelopes (internal operation and expected range of external perturbations) plus proper matching of these layers and phases.

- Certification needs to concentrate on the effectiveness of these strategies (the quality of the produced scenarios) and on the interoperability of layers and control loop phases. The exact documentation of performance envelopes is a *key issue* in the certification process. This is the framework for certification of the *implementation components* of Operational Concepts, i.e., the functions, responsibilities and operational procedures.

- Large systems have many players: groups of people and automated systems operate together in teams to make collective tactics and strategies (planning layers and control loops) happen. This task sharing defines the information flows between people, on the human-machine interfaces, and between automated systems. Different tactics and strategies require different information flows.

- Functions, responsibilities and operational procedures are the "bricks" for building the implementation of an Operational Concept. The above mentioned information flows are the cement which keeps the "bricks" together. The existence of these "bricks" is to be justified in terms of the certified planning strategy, and the role of existing "bricks" is to be mapped on the planning layers and their control loop phases. New "bricks" should be designed to fit specific niches in that framework of layers and phases.

- To make the implementation of an Operational Concept perform as intended, the individual "bricks" need to be certified to meet the requirements imposed by the previously certified interface and performance specifications of planning layers and control loop phases.

Human Factors

All the above considerations about planning layers, control loop phases, scenarios, performance envelopes apply to any goal oriented system. That, of course, includes composite human-machine systems.

In such a system, the planning responsibilities outlined in the Operational Concept are allocated to humans and machines. This can be done in a top-down fashion (in an arbitrary manner), or bottom-up, built around the human capabilities and the state-of-the-art of technology.

Many studies have investigated the role of the human. The human factors field aims at determining the automation environment in which the human performs best. This paper does not attempt to draw specific conclusions from the existing literature, but in the end various strategies are possible to integrate the human into an automated system, or to support the human with automated functions. One strategy may be to give the complete responsibility for some planning layers to humans, and automate the remaining layers. For example, in ATC, automate safety nets but keep executive control largely manual.

Another approach is to take the control loop of a planning layer, automate some phases and leave responsibility for others to the human. To use again an ATC example: automate the monitoring and problem detection phases, but keep the proposal and evaluation of solutions, and the decision making phases manual.

No matter whether a layered or phased automation strategy is chosen, it is necessary to correctly use the strengths and weaknesses of humans and machines. In other words: within the framework of layers and phases, find out whether the human or the machine fits the requirements best, in terms of monitoring capabilities, problem solving capabilities, memory, speed, assimilation, pattern recognition, span of attention, reliability, etc.

Thus, the human gets certain modular chunks (layers, phases) of the overall Operational Concept (the planning strategy as described above). The problems of human factors certification can then be seen as the certification of interoperability of these chunks with the overall Operational Concept. In that sense, the human role is no different from an automated function taking the same responsibilities. The performance needs to meet the requirements as foreseen for that particular responsibility within the context of the total planning strategy.

Conclusions

General

The paper suggests that the "tactics and strategies" notion is a highly suitable paradigm to describe the cognitive involvement of human operators in advanced aviation systems (far more suitable than classical function analysis), and that the workload and situational awareness of operators are intimately associated with the planning and execution of their tactics and strategies. If system designers have muddled views about the collective tactics and strategies to be used during operation, they will produce sub-optimum designs. If operators use unproved and/or inappropriate tactics and strategies, the system may fail.

The author wants to make the point that, beyond certification of people or system designs, there may be a need to go into more detail and examine (certify?) the set of tactics and strategies (i.e., the Operational Concept) which makes the people and systems perform as expected.

The collective tactics and strategies determine the information flows and situational awareness which exist in organizations and composite human-machine systems.

The available infrastructure and equipment (automation) enable these information flows and situational awareness, but are at the same time the constraining factor. Frequently, the tactics and strategies are driven by technology, whereas we would rather like to see a system designed to support an optimized Operational Concept, i.e., to support a sufficiently *coherent*, *cooperative* and *modular* set of anticipation and planning mechanisms.

Again, in line with the view of MacLeod and Taylor (1993), this technology driven situation may be caused by the system designer's and operator job designer's over-emphasis on functional analysis (a mechanistic engineering concept), at the expense of a subject which does not seem to be well understood today: the role of the (human cognitive and/or automated) tactics and strategies which are embedded in composite human-machine systems. Research would be needed to arrive at a generally accepted "planning theory" which can elevate the analysis, description and design of tactics and strategies from today's cottage industry methods to an engineering discipline.

Planning Theory

A theory based on planning layers, control loop phases, uncertainty phases and performance envelopes would provide a modular framework to the task of designing and documenting Operational Concepts (i.e., sets of tactics and strategies). The second half of this paper represents an initial attempt to highlight the key issues of such a theory. When such a framework is used, the benefits may spin off to the certification task. In addition, it will put the role and contribution of human factors into clear perspective.

OOA

A few references to OOA (Object-Oriented Analysis) techniques have been made in this paper. It is felt that OOA is too mechanistic; i.e., it misses some expressiveness when used to analyze and document systems consisting of goal oriented entities. Planning theory could be a suitable candidate to remedy that problem.

Annex: The European Need For ATM Operational Concepts

Currently, a number of organizations around the world wish to bring Air Traffic Management (ATM) into the next century with significantly improved capacity, productivity and economy. To that effect some are conducting R&D programs and others are planning and procuring new systems which will still be in service after the year 2000.

Their vision of the future varies: from revolutionary to a more evolutionary approach. But even in the conservative case, everyone expects that due to the accelerated pace of technological innovation, Air Traffic Management will change significantly *more* in the coming ten to fifteen years than it did in the past half century.

The above explains today's interest in the development of the ATM Operational Concepts that are suitable for application in the period 2005-2015. To some degree, this paper has been written with the European ATM context in mind. Therefore it is useful to include a short overview of the activities going on in Europe (European Civil Aviation Conference, April 1990 and March 1992; Eurocontrol, June 1992).

The Transport Ministers of the European Civil Aviation Conference (ECAC) Member States, meeting in Paris on 24 April 1990 and in London on 17 March 1992, have noted the substantial growth which is forecast in air traffic demand in the ECAC area to the end of the century and beyond, and the considerable efforts which are being deployed to expand the system accordingly and to reduce air traffic congestion in Europe.

In order to unite and accelerate those efforts, the ECAC Ministers have adopted:

- the ECAC En-Route Strategy and action program to harmonize and integrate the operations of their air traffic control systems in the 1990's; and

- the ECAC Airports Strategy which will provide a concerted systems approach to the airport / air traffic system interface.

For these strategies, commonly known as the *ECAC Strategy*, the ECAC Ministers have adopted the following overall objectives:

- to urgently provide increasing airspace and control capacity, in order to handle the traffic expeditiously while maintaining a high level of safety;

- to improve the potential throughput of European airports and their surrounding airspace while maintaining safety and respecting the environment.

These initiatives will prepare the way for the introduction of a new generation of air navigation technology on the eve of the 21st century. To this end, the ECAC Ministers:

- have committed themselves to complete the phased action program for air traffic control which is the basis of the ECAC En-Route Strategy within a challenging but realistic time scale; and

- have resolved that the current program of research and demonstrations undertaken by Eurocontrol, the Commission of the European Communities (CEC) and ECAC Member States should be extended to cover new procedures and equipment required for air traffic management in and around airports.

Within the framework of the European Air Traffic Control Harmonization and Integration Program (EATCHIP), detailed planning is now well under way within Eurocontrol to give effect to the ECAC Strategy.

Meanwhile, the concept of a European Air Traffic Management System (EATMS) is seen as the 21st Century goal towards which the energies of Eurocontrol, the participating National Administrations and the European Industry should be focused. It takes into consideration those concepts that have already been accepted, FEATS, FANS, ECAC Strategy etc., and offers a method of *how* the future system would operate, given that the technology will be available. This concept will be developed as EATCHIP Phase IV evolution and implementation. Addressing the time scale 2005 - 2015, it will provide the basis for Phase IV of the ECAC Strategy and in particular for:

- adoption of a common functional model integrating the airborne and ground based components of the future EATMS;

- definition and implementation of advanced systems supported by extensive automation and enhanced data communications available via the Aeronautical Telecommunications Network (ATN).

But a common functional model and advanced systems are not enough. Before those can be defined, an agreement will have to be reached on the innovations which will be applied to the underlying Operational Concept (i.e., the tactics and strategies) of the EATMS. That requires proper attention to the Human Factors aspects of these innovations.

For the time being, no plans exist to *certify* the EATMS Operational Concept in its totality. However, various *validation* activities are foreseen, to be concluded by an Operational Concept demonstration program, scheduled near the end of the EATMS System definition and planning phase (which is planned to complete around the year 2000).

Disclaimer

The content of this paper expresses the opinion of the author and does not necessarily reflect the official views or policy of the EUROCONTROL Agency.

References

Coad, P., and Yourdon, E. (1991). Object-Oriented Analysis, Second Edition. Englewood Cliffs, NJ: Yourdon Press.

EATCHIP Task Force on ATC System Integration (1992). Open ATM System Integration Strategy (OASIS). EUROCONTROL Document 922011.

EUROCONTROL (June 1992). European Air Traffic Management System (EATMS) – Concept Document, Issue 1.1.

European Civil Aviation Conference (ECAC) (April 1990). Air Traffic Control in Europe – ECAC Strategy for the 1990s.

European Civil Aviation Conference (ECAC) (March 1992). Relieving Congestion in & around Airports - ECAC Strategy for the 1990s.

MacLeod, I. S., and Taylor, R. M. (1993). Does Human Cognition Allow Human Factors (HF) Certification of Advanced Aircrew Systems? In J.A. Wise and V. D. Hopkin (Eds.), *Human Factors Certification of Advanced Aviation Technologies*. Berlin: Springer-Verlag. [in press].

Martin, B., Central Flow Management Unit (CFMU), EUROCONTROL (February 1993). Progress through better Air Traffic Flow Management. ATC '93 Conference, Maastricht.

Shlaer, S., and Mellor, S. J. (1992). Object Lifecycles – Modelling the World in States, First Edition. Englewood Cliffs, NJ: Yourdon Press.

Practical Approaches to Human Factors Certification

PRECEDING PAGE BLANK NOT FILMED

PAGE 96 INTENTIONALLY BLANK

N95-34769

Human Factors Certification: A Useful Concept?

52-54 [illegible]

Alistair Jackson

EUROCONTROL Experimental Centre

Introduction: What this paper is about

This paper considers what is involved in certification processes and their relation to human factors aspects of systems. It derives from recognition of a lack of understanding of the processes and purposes of certification. This was encountered when attempting to address the workshop topic by integrating an understanding of human factors with the observed processes of certification. The paper considers what human factors (HF) certification might be and then develops a simple model of the elements of a certification process. It then tries to relate these elements to the needs of the aviation communities and other parties with an interest in the certification of advance aviation technologies.

What Could the Term "Human Factors Certification" Mean?

Consider the ways in which we might 'conventionally' interpret the phrase "human factors certification of advanced aviation technologies".

a) As *Human Factors Certification,* effectively a new concept that has derived from aspects of two existing areas of endeavour. While related to them, it has emergent properties that distinguish it from either.

b) As *the Certification of Human Factors,* i.e., as the certification activities and processes that are required to deal with human factors topics arising from advanced aviation technologies.

c) As *the Human Factors of Certification,* i.e., as describing the human factors issues which are associated with the activity and processes of certification of advanced aviation technologies.

The current status of these three alternatives is summarised in the following paragraphs.

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PRECEDING PAGE BLANK NOT FILMED

PAGE 98 INTENTIONALLY BLANK

Human Factors Certification as a New Topic

The first alternative is potentially the most interesting. Are we dealing with a whole new topic which emerges from the fusion of the activities of human factors and certification?

Alternatively, are we simply juxtaposing two terms, each of which is meaningful on its own, but which does not specially inform the other? Sadly, it appears to be the latter. If there is truly a new theme, then there is, as yet, no clear exposition of the nature or extent of the emergent properties that give it its uniqueness.

The Certification of Human Factors Issues

This activity is much simpler to understand. It is also likely to be the most useful interpretation for the purposes of the workshop. Currently, human factors contributes in a number of ways to the development and use of advanced aviation technologies. It should probably contribute in many more. Application of certification processes to these contributions would be analogous to certification practices in any other domain of human activity and indeed, it can be argued that it already takes place in a limited way. This normally occurs as an extension of other, more general, certification processes. It is now fairly standard within the UK that anyone offering input as a psychologist in the development of aviation technologies should have 'Chartered' status. Similarly, there are moves in the US to provide certification of human factors professionals. There is steadily increasing acknowledgment of the need for human computer interfaces (HCIs) to be designed in a user-centred way and an associated pressure that they be subject to the some kind of 'usability' testing. This might be considered analogous to the certification processes found in other systems especially those where safety is a priority. In a broader context there are now programmes like MANPRINT (Boorer, 1991), which operates for the procurement of military systems. MANPRINT is a methodology which seeks to ensure total system quality in terms of the human and the human's integration into the system. It could also be observed that many of the issues addressed in quality assurance schemes relate to 'human factors' aspects of systems

In summary, the certification of the human factors aspects of systems seems to be a worthwhile and practical activity and one that is already underway as part of general systems certification processes. The main question this approach poses for the workshop is: is there anything special about the human factors aspects which would mean that they are not being adequately covered by the processes used for other aspects of system certification? If the answer to this is yes, then the community must attempt to identify the nature of the inadequacies and the action required to remedy the situation.

The Human Factors of Certification

This approach poses rather more difficulties than its immediate predecessor. These difficulties lie, not so much with the proposed relationship between human factors and certification, but with understanding the nature of certification itself. To elaborate, any activity involving a socio-technical system will have human factors aspects and in almost all cases some of these will be sub-optimal. This implies that there must be scope for HF input. It seems to be an assumption that certification of advanced aviation technologies is such a socio-technical system activity and therefore HF should have a contribution to make in supporting and improving the process of certification. What that contribution should be will depend to a considerable extent on the nature of certification itself as well as on the nature of the systems being certified.

This places considerable emphasis on the need to understand certification. Similarly, the question asked at the end of the previous sub-section as to whether the human factors issues of complex systems have special needs in certification processes, presupposes an understanding of what is involved in certification. The main body of this paper considers the nature and scope of certification processes as an aid to understanding both the ways in which HF might contribute to such processes and their implications for certifying systems with significant HF aspects.

What is Certification?

It is clear in looking beyond advanced aviation technologies, that 'certification' can be applied to an astonishingly wide variety of things. It can be established in a number of different ways and serves a multiplicity of purposes – some more obvious than others. An initial examination identified three core elements in any certification process, namely the authority which supervises and performs the certification, the thing which is certified and the frame of reference employed in the implied evaluation process. For convenience these three elements can be termed the *agent*, the *object* and the *criteria* respectively. However, further consideration suggested that it is potentially misleading to consider these elements without providing a context in order to account for the motivations behind the activity of certification. This approach led to the assertions below about the nature and function of certification and to the schema shown in Figure 1. The following paragraphs discuss the elements of this schema more fully. However, it is important to make one underlying assumption very clear, namely: *"In what follows it is assumed that certification is a social or socio-technical process which can only be meaningfully understood in terms of social, as well as technical, processes and requirements"*.

Description of the Entities and Relationships in Figure 1

The description begins with the three original elements before introducing the supporting entities.

Objects. A very wide variety of things are currently subjected to some form of certification process. For the sake of both simplicity and generality, these are referred to as the *objects* of certification. Examples of classes of objects, which can be subject to certification, relate to and include *people, hardware, software, procedures and systems,* (the last being considered as being an agglomeration of some or all the other classes). In many instances, the objects are not actual physical objects but abstract properties or attributes, possibly related to a physical object. Examples, relating to people embrace motor skills (e.g., driving), abstract knowledge (academic qualifications), the ability to apply knowledge in a professional manner (doctors, dentists, pilots, air traffic controllers) social and cultural attributes (passports, nationality) and physical integrity (medical examinations). Examples relating to artifacts could include the ability to resist environmental stress of various kinds (mechanical, temperature, etc.) or to produce performance of different types (minimum lifetimes of light bulbs, stability of power supplies, etc.).

An *object* becomes a *certified object* when it has successfully passed through a certification process by meeting the criteria.

Generally, objects are certified in order to establish some form of quality or to preserve some type of standards. This often takes the form of defining some minimum performance criterion which must be exceeded. (See Reference Criteria, etc. below.) Although the present description of certification is used to embrace a very wide range of activities, not all of which are customarily called by that title, the term is most frequently used explicitly under circumstances where the characteristics being ratified are widely acknowledged as being important but are very difficult to define in practical terms. (These are the occasions where 'everyone knows what you mean,' but you find it almost impossible to write down in a clear, unambiguous manner.)

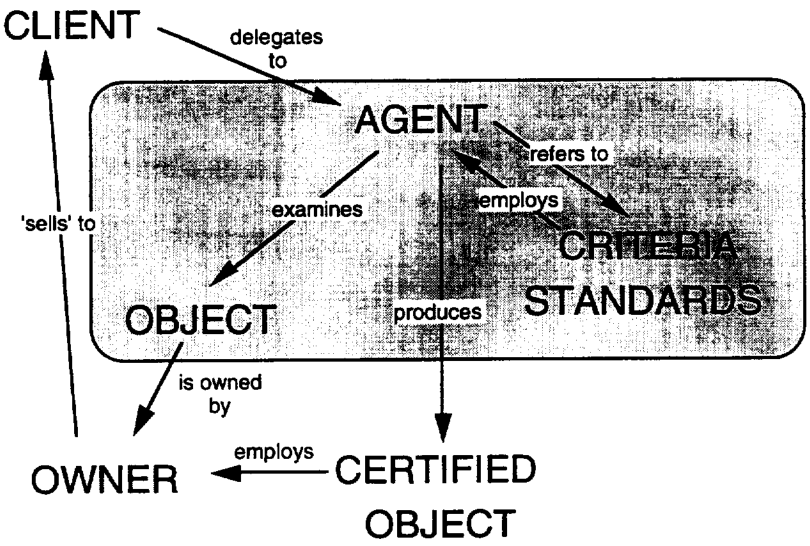


Figure 1. A simple SCHEMA for certification

Agents. Generally certification is carried out by some 'responsible' authority *(agent)* and consists of some process of examination of the object and its characteristics against some more or less well defined set of *criteria* (see below). In the world of certification, agents take many forms. In particular, they are frequently institutionalised as explicit authorities such as examination boards, standards organisations, inspectorates, etc. Under these circumstances, the authority of the agents and the process of certification are frequently supported by legislation.

The term authority is used intentionally, since a state of empowerment is implied by the existence of a process of certification. The certifying agent can grant or refuse the status of certification to the submitted object. The process of transforming an object into a certified object is important because there is a notion of added value associated with the process. The object has more utility, or is more marketable, to someone after it has been certified.

Reference Criteria and Standards. The *agent* subjects the *object* to assessment against some *standard or set of criteria* which by implication has been established beforehand for the purposes of certification. The nature of this reference system may be very precise (as is the case with standards for equipment performance and safety certification of hardware), they may be quite loose or they may even be a mixture of the two. This mixing of criteria quality is often to be found in the case of 'professional standards' where there may be quite precise requirements as to the holding of academic qualification but this frequently needs to complemented by less well defined, but equally important, peer review procedures. It is generally important that standards should be made explicit and observable. However, as has been mentioned, some types of knowledge and skill are very difficult to describe and define in a language based manner. As a consequence, it is important to note the dangers in assuming that imprecisely defined standards and criteria imply that the performance standards required by a certification activity are necessarily low. They may be very high indeed. For example, the final assessments for air traffic controllers validating on a sector, or for certain aspects of pilot licensing, are likely to involve assessment by specially nominated and experienced peers. In these cases, high standards are associated with implicit criteria embodied within the expertise of the assessors. Nevertheless, explicit criteria have considerable advantages. They provide a target, permitting the production of objects which are likely to meet certification standards. They can also help to establish the impartiality of certification processes.

Clients. Clients are the potential users or consumers of the products of certification. Clients delegate to *agents.* This delegation takes place for a variety of reasons. Sometimes it occurs to save the resources of clients by allowing them to assume certain standards with a minimum investment of time and money. Alternatively, certification can be undertaken by agents with a specialist technical expertise otherwise be unavailable to the client. The process of delegation is the origin of the agent's authority, i.e. the authority originates with the client and is delegated, along with responsibility, to the agent.

The unique identification of clients can be difficult because of the potential for hierarchies of delegation. To illustrate, consider the example of a ground air traffic control system. For a particular element such as an interface tool, the client might be the controller who will have to use the tool. However, it is also possible to consider that the client is the national administration which runs the ATC system and employs the controller. Even the administration may be considered as providing the ATC system on behalf of the airline operators and eventually on behalf of the fare paying passengers, a subset of the general public. Although for many certification processes it may suffice to identify the immediate client; for others, particularly those with legal implications, the design of adequate procedures may require a full understanding of the client hierarchy.

Very often the function of certification is to protect the interests of clients. For example, in the case of the institutional agencies cited earlier, the client is often the public and the certification processes are in place to ensure that services offered to the public meet 'adequate' standards of quality, safety, etc. Under these circumstances, what constitutes 'adequate' may bear a strong relationship to the public's perception and can be somewhat volatile depending on history and context. For example, after an accident or incident there could be a public pressure for higher standards etc., comparatively independently of the nature of the incident and the safety contribution of the suggested enhancement. The media can play a large part in shaping this perception. Amongst the requirements placed on both agents and certification standards may be the ability to be visible, to provide information and explanation, and to otherwise be the objects of media scrutiny

Owners. Owners, as the name suggests, 'own' or have control of the objects submitted to certification processes and normally thereafter for the certificated objects which result from the processes. They are very often the producers of an object, or they may wish to employ the object to provide some form of service. They are almost always the initiators for individual instances of certification activity, i.e., they submit objects for certification. As examples, consider that an airframe has a manufacturer, a skill is possessed by an individual, a car being road tested has a driver/owner, a software package was written by someone or for someone.

Owners generally wish to 'sell' or otherwise have their objects consumed by the clients. They tend to bear the initial costs of certification processes but would be expected to pass these on to the client. Owners are motivated to have objects certified since these are generally more salable after the certification process – clients 'buy' with more confidence. As in the case of clients, there may be a hierarchy of ownership.

Certification in the Context of Advanced Aviation Systems

In addition to the discussion above, at least two other contextual aspects of certification have to be considered in examining its potential role for advanced aviation technologies.

When does certification take place?

In the aviation community there seems to be a consensus that certification can be a comparatively global process applied to fairly large units, e.g. an airframe, a ground system, etc. This has significant implications. Firstly, it suggests that certification has to take place fairly late in the process of production the object of interest. Secondly, it suggests that certification of HF aspects is part of a larger process of certification addressing a number of relevant aspects of the object or system in question. (Effectively, this assumes the second interpretation of the term HF certification.)

Who might be seeking certification of advanced aviation technologies?

This is a question of some complexity. There are a number of potential beneficiaries with different motivations and requirements.

Potential suppliers of systems are likely to benefit from certification processes which employ explicit standards and criteria as these can be employed to set performance standards and targets during the design stages of production systems. This potentially reduces cost by improving the reliability of planning and reducing risk. In this context, HF certification would be expected to yield the same benefits as any other aspect of certification

Procurers of Systems could theoretically benefit indirectly from the cost and risk reductions for producers, but they also have a requirement for certification which is motivated by a professional interest in monitoring the successfulness of their own activity. They are seeking means to establish and assure that they are procuring systems of an appropriate character and quality.

Users/Participants in Systems. This term is employed to describe individuals involved within the system as contributor to the system's functioning, such as pilots in aircraft, air traffic controllers, etc. Like the procurers, this group has a professional interest in ensuring that system performance meets high standards, partly since it might be seen as reflecting on their own contribution. However, there are additional dimensions to their concern. Both pilots and air traffic controllers have legal responsibilities and liabilities. In an environment with increasing levels of automation and computer based assistance, it is especially important that:

- They should be able to have confidence in the system which they are using.
- They have no liabilities over which they cannot exercise their responsibility. For example, a controller should not be expected to have liability as a consequence of using erroneous information from an aid which he has no reason to recognise as unreliable.

These groups may view certification in general as a means of establishing justifiable trust in a system and of protecting against unreasonable liabilities. Because issues associated with human performance and human error are especially sensitive in terms of responsibility and liability, certification of HF aspects assumes a potentially critical role.

Customers of Systems such as fare paying passengers, or airline operators in either of their roles, as users of air traffic control systems or as purchasers of airframes, would wish to establish that everything possible and reasonable is being done to ensure both safety and effective system performance. Here again, the role of certification is closely related to creating confidence in a system and its operation.

Human Factors Specialists. Aside from a professional interest in seeing certification as a means of emphasising system quality, some HF specialists might see HF certification, or any certification process which makes HF aspects explicit, as a means of ensuring the insertion of HF throughout the design and production phases of systems. This latter view would not be shared by all HF professionals. There are other approaches to achieving this objective and it can be argued that while human factors achieved by a prescriptive approach may be better than nothing, it falls far short of the quality which should be the objective.

Summary

In conclusion, many stakeholders might see certification as a potential solution to their proper needs. Although it is not clear that certification is necessarily the best means of meeting their differing requirements it is already a recognised mechanism in several aspects of aviation and it is likely to be employed more widely.

If a general process of certification is going to take place, then it is incumbent on the aviation HF community to ensure that HF aspects of the target systems are adequately addressed within the certification process. This emphasis on our second interpretation of the term "HF certification" demands that we examine carefully the nature and extent of any HF aspects which are currently being neglected in such processes.

The discussion in this paper suggests that there are a number of such aspects relating to the creation of confidence in systems, the recognition and allocation of responsibility, and the management of liability. These areas, with their emphasis on the more social aspects of cognition, are not only comparatively new to the aviation HF community but are also potentially sensitive for the organisations and users which employ advanced aviation technologies. This situation presents us with a considerable challenge.

References

Boorer, H. R. (1991). *MANPRINT an approach to systems integration.* New York: Van Nostrand Rheinhold

N95- 34770

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

56-04
59116

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.

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

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.

N95- 34771

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

Robert M. Taylor[1] & Iain S. MacLeod[2]

[1]RAF Institute of Aviation Medicine
[2]Aerosystems International

Summary

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

A new approach to the application of HE, recently adopted by NATO, provides more systematic ordering and control of HE processes and activities to meet the challenges of advanced aircrew systems design. This systematic approach to HE has been applied by MoD to the procurement of mission systems for the Royal Navy Merlin helicopter. In MoD procurement, certification is a judicial function, essentially independent of the service customer and industry contractor. Certification decisions are based on advice from MoD`s appointed Acceptance Agency. Test and evaluation (T&E) conducted by the contractor and by the Acceptance Agency provide evidence for certification. Certification identifies limitations of systems upon release to the service. Evidence of compliance with HE standards traditionally forms the main basis of HE certification and significant non-compliance could restrict release.

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

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

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

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

Background

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

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

The Human Engineering Approach to Systems Design

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

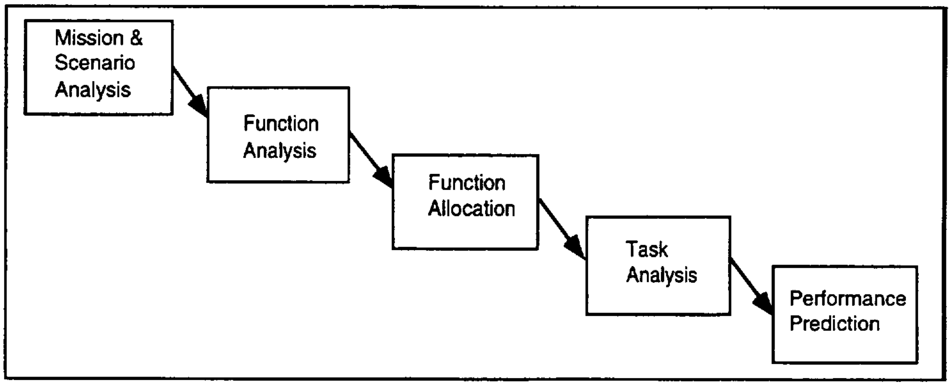


Figure 1. Stages of Human Engineering Analysis (From Beevis, 1992)

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

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

Test and Evaluation in Systems Human Engineering

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

Identification of Test Parameters

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

Test Plan

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

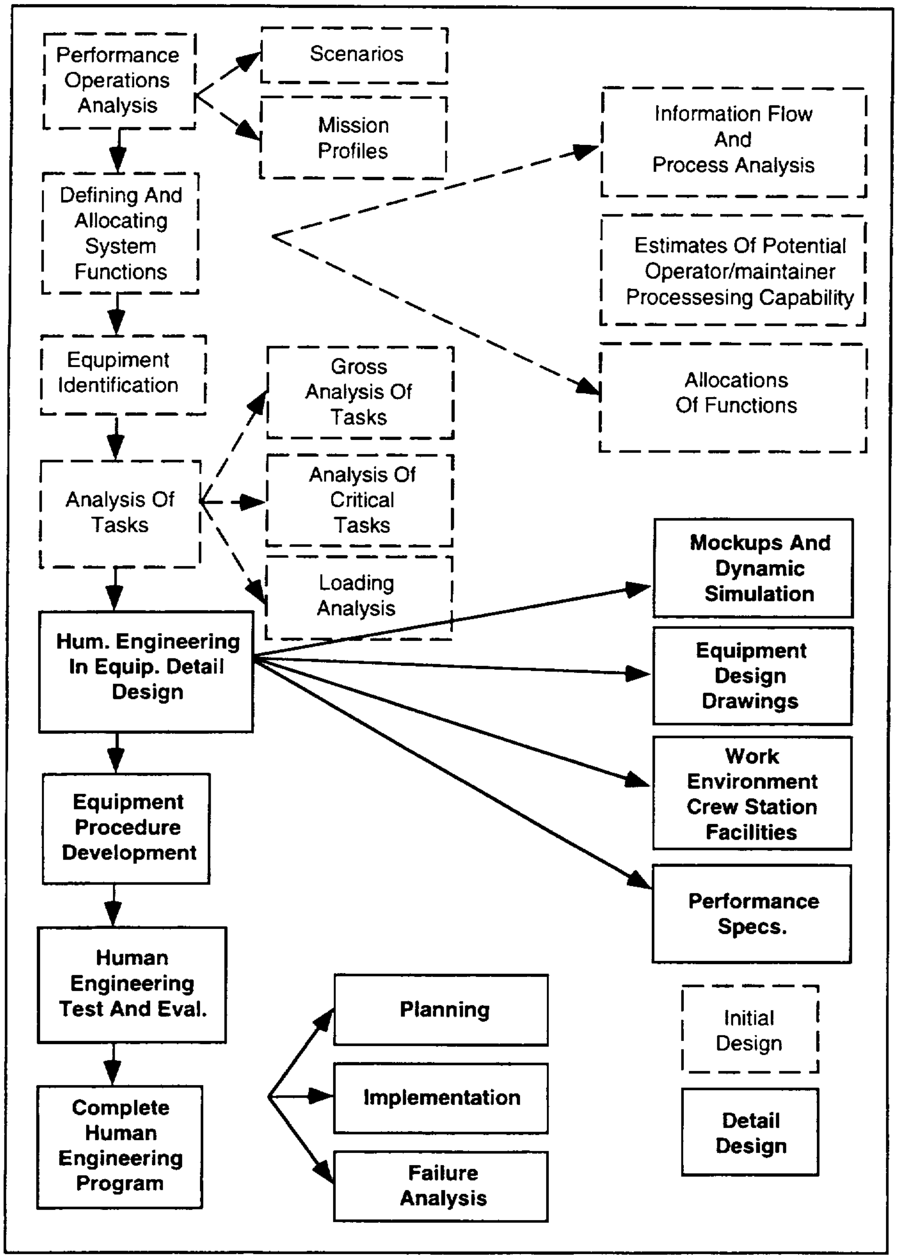

Performance Operations Analysis
Scenarios
Mission Profiles
Information Flow And Process Analysis
Defining And Allocating System Functions
Estimates Of Potential Operator/maintainer Processesing Capability
Equipment Identification
Gross Analysis Of Tasks
Allocations Of Functions
Analysis Of Tasks
Analysis Of Critical Tasks
Loading Analysis
Mockups And Dynamic Simulation
Hum. Engineering In Equip. Detail Design
Equipment Design Drawings
Work Environment Crew Station Facilities
Equipment Procedure Development
Performance Specs.
Human Engineering Test And Eval.
Planning
Initial Design
Implementation
Detail Design
Complete Human Engineering Program
Failure Analysis

Figure 2. STANAG 3994 Activities

Quality Assurance Compliance

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

NATO DRG Endorsement

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

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

Merlin Human Engineering

In the U.K. we have experience with applying MIL-H-46855 principles by citing STANAG 3994 as a mandatory reference on several air systems acquisition programs. We have been particularly keen on raising the profile and effectiveness of HE and emphasize shifting more HE risk in procurement to contractors while maintaining HE quality assurance. STANAG 3994 is perceived as a potentially valuable aid both for maintaining HE quality assurance and for managing HE risk in the procurement of complex mission systems. Also the risk for HE is perceived as particularly important during complex mission systems procurement. For complex systems, situation assessment and mission performance effectiveness are functions of the integration and interaction between the operator and the equipment's information processing and cognitive decision-making capabilities. The U.K. program which provides the most advanced example of STANAG 3994 application is the procurement of the Royal Navy Merlin (formerly EH101) Anti Submarine Warfare (ASW) helicopter. This project is known as the Merlin Prime Contract (MPC). The RAF Institute of Aviation Medicine (IAM), DRA Farnborough, and Aerosystems International have acted as HE technical advisors on the program. This paper is largely based on the HEPP acceptance/compliance assurance issues that have arisen on the MPC program.

Merlin Specification Rationale

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

Operational Performance and Acceptance Specification (OPAS)

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

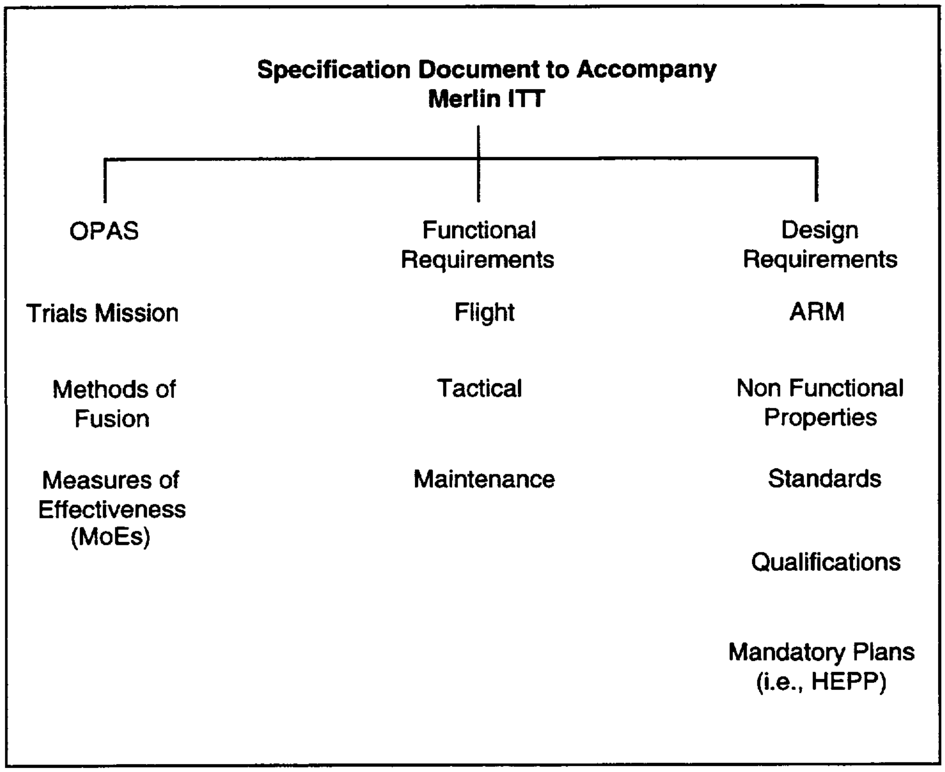

Specification Document to Accompany Merlin ITT
OPAS
Functional Requirements
Design Requirements
Trials Mission
Flight
ARM
Methods of Fusion
Tactical
Non Functional Properties
Measures of Effectiveness (MoEs)
Maintenance
Standards
Qualifications
Mandatory Plans (i.e., HEPP)

Figure 3. The Contents of the Merlin Specification

Merlin Human Engineering Program Plan

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

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

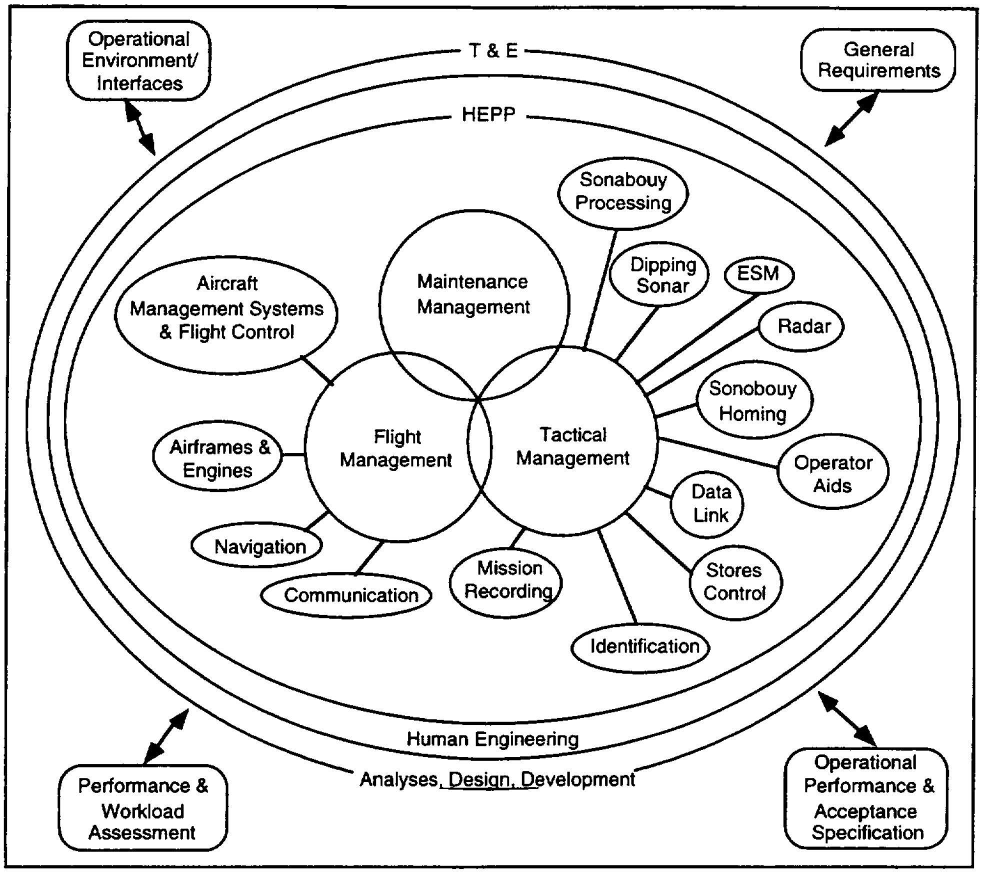


Figure 4. Merlin High Level Functionality

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

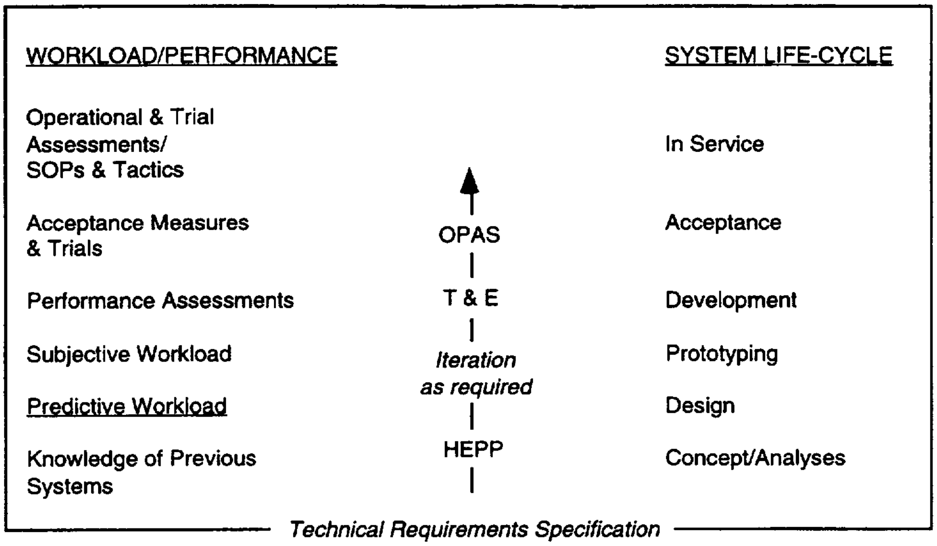


Figure 5. HE Testing Sequence in Merlin Life Cycle

Merlin Predictive Analysis

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

Workload Analysis

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

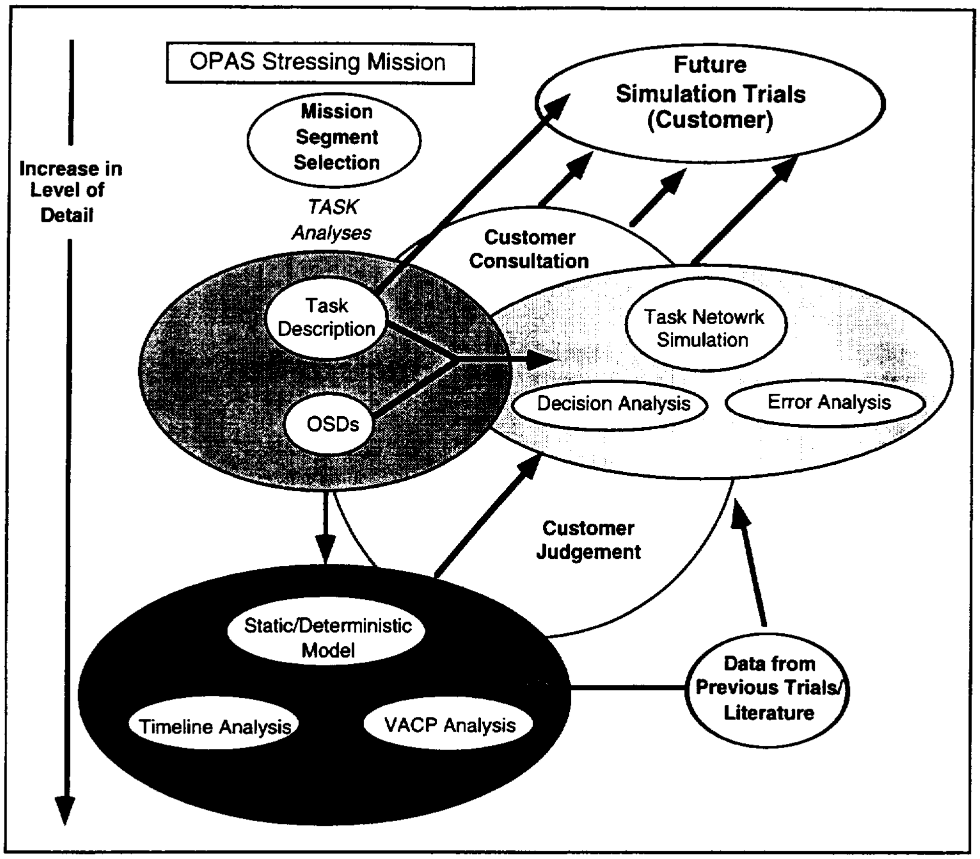


Figure 6. Relationship of Merlin HE Predictive Analyses

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

Decision Error Analysis

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

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

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

Certification

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

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

Government Functions

In government management of systems design the role of certification can be considered as a judicial function rather than a legislative or executive function. Certification is a judgment on the design standard of the system and carries with it major implications for program risk and cost. The following are further notions of how these functional distinctions can be applied:

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

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

Certification Authority

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

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

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

Certification Validity

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

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

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

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

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

Quality Assurance

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

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

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

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

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

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

Creative Evaluation

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

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

Progressive Acceptance

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

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

Certification of Human Behavior

The GFE Approach

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

Cognitive Functions

The traditional HE assumptions about human design requirements are at best limited in scope, and at worst invalid, if they are based on inappropriate models of human interaction in systems. They may lead to inaccurate, unrealistic, and optimistic assessments of overall system capability and effectiveness. Recent U.K. procurement experience indicates a tendency to be over optimistic with predictions of future operational performance of complex advanced systems under development. With the GFE approach, the risk for human functionality in total system performance is carried by the customer rather than by the contractor. Failure to achieve systems performance targets in T&E can be ascribed to human capability or performance variability. The problem then becomes one of the human not matching the machine rather than the converse, and it needs to be solved by improved customer-provided training or by enhanced customer selection standards, not by in-service system upgrades. This is increasingly untenable in a procurement climate seeking to minimize the risk to the customer. It is particularly inadmissible for procurement of complex advanced mission systems where system performance effectiveness is increasingly a function of operator-equipment integration and cognitive level interactions dealing with information processing, situation assessment, and decision-making. The RSG 14 report (Beevis, 1992) concludes that while it is generally assumed that new advanced systems place increasingly high demands on the cognitive aspects of operator/maintainer behavior, most HE techniques on the other hand lend themselves to the description of skilled behavior, not cognitive behavior. It seems that certification of HF in advanced future systems will require better resolution, analysis, and engineering of cognitive functions than presently available with HE techniques. Stiles and Hamilton (1987) describe how a cognitive engineering approach to functional analysis will be needed for identifying a pilot's intentions during his or her interface with the system, as well as for providing a design (information and/or control) to help achieve the intentions. The requirement for improved resolution of cognitive functionality is discussed further in the second position paper by the authors (MacLeod & Taylor, 1994).

Aircrew Certification

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

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

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

SRK Taxonomy

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

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

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

Conclusions

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

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

References

Barber, J. L., Jones, R. E., Ching, H. L., & Miles, J. L. (1987, September). *MANPRINT Handbook for RFP Development* (AMC-P 602-1). HQ U.S. Army Material Command.

Beevis, D. (1992, July). *Analysis Techniques for Man-Machine System Design.* NATO, AC/243 (Panel 8) TR/7.

Biggen, K. (1992, March). *EH101 Mission Workload Simulation Validation Trials Report* (Westlands Helicopters Report No. ER02Q002W).

Brown, H. M. (1992). Competency in the cockpit. In D. Saunders & P. Price (Eds.), *Developing and Measuring Competence. Aspects of Educational and Training Technology XXV.* London.

Brown, H. M., & Rolfe, J. M. (1993). *Training requirements or technical requirements.* Paper submitted for publication.

Lizza, C. S., Rouse, D. M., Small, R. L. & Zenyuth, J. P. (1992). Pilot's associate: An evolving philosophy. In T. E. Emerson, M. Rienecke, J. Riesing, & R. M. Taylor (Eds.), *The human electronic crew: Is the team maturing?* (U.S. Air Force Wright Laboratory Report No. WL -TR-92-3078). Wright-Patterson Air Force Base, OH: U.S. Air Force Wright Laboratory.

MacLeod, I. S., Biggen, K., Romans, J. & Kirby, K. (1993). Predictive workload analysis--RN EH101 helicopter. *Contemporary Ergonomics 1993.* London: Taylor & Francis.

MacLeod, I. S., & Taylor, R. M. (1994). Does human cognition allow human factors (HF) certification of advanced aircrew systems? In J. A. Wise, V. D. Hopkin, & D. J. Garland, (Eds.), *Human Factors Certification of Advanced Aviation Technologies* . Daytona Beach: Embry-Riddle Aeronautical University Press.

Metzler, T. R., & Lewis, H. V. (1989, June). *Making MANPRINT count in the acquisition process* (Army Research Institute Note 89-37). U.S. Army Research Institute.

NATO. *The application of human engineering to advanced aircrew systems* (STANAG 3994 AI).

Patton, M. Q. (1978). *Utilization-focused evaluation.* Beverley Hills: Sage.

Sherwood-Jones, B. (1987). Human-factors audits and fitness for purpose. *Proceedings of the CAP Scientific Conference.*

Stiles, L., & Hamilton, B. E. (1987). Cognitive engineering applied to new cockpit designs. *Proceedings of the American Helicopter Society National Specialists Meeting: Rotorcraft Flight Controls and Avionics.* Cherry Hill, PA.

Taylor, R. M. (1987). *Some thoughts on the future of engineering psychology in Defense.* Position Paper for the British Psychological Society Conference on the Future of the Psychological Sciences, Harrogate.

Taylor, R. M. (1990). *Merlin MPC Workload Acceptance Criteria* (IAM Letter Report 016/90). RAF Institute of Aviation Medicine.

Taylor, R. M. (1991). Human operator capability analysis for aircrew systems design. *Proceedings of a panel session at the British Psychological Society 1991 Occupational Psychology Conference* RAF Institute of Aviation Medicine Letter Report No. 004/91. RAF Institute of Aviation Medicine.

Taylor, R. M., & Selcon, S. J. (1993). Operator and automation capability analysis: Picking the right team. *Combat Automation for Aircraft Weapon Systems: Man/Machine Interface Trends and Technologies.* Neuilly Sur Seine: NATO AGARD CP 520.

U.K. Ministry of Defense. (1989). *Human factors for designers of equipment* (DEF-STAN-00-25).

U.K. Ministry of Defense. *Design and airworthiness requirements for service aircraft* (DEF-STAN-00-970).

U.K. Ministry of Defense. *Guide to quality assurance in design* (DEF-STAN-05-67).

U.S. Department of Defense. (1987). *Human engineering procedures guide* (DOD-HDBK-763).

U.S. Department of Defense. *Human engineering design criteria for military systems, equipment and facilities* (MIL-STD-1472).

U.S. Department of Defense. *Human engineering requirements for military systems, equipment and facilities* (MIL-H-46855).

U.S. General Accounting Office. (1981). *Guidelines for assessing whether human factors were considered in the weapon system acquisition process* (GAO FPCD-82-5).

N95- 34772

User "Type" Certification for Advanced Flight Control Systems

58-08
59148

Richard D. Gilson & David W. Abbott

University of Central Florida

Changing Flight Crew Roles

Advanced avionics through flight management systems (FMS) coupled with autopilots can now precisely control aircraft from takeoff to landing. Clearly, this has been the most important improvement in aircraft since the jet engine. Regardless of the eventual capabilities of this technology, it is doubtful that society will soon accept pilotless airliners with the same aplomb they accept driverless passenger trains. Flight crews are still needed to deal with inputing clearances, taxiing, in-flight rerouting, unexpected weather decisions, and emergencies; yet it is well known that the contribution of human errors far exceed those of current hardware or software systems. Thus human errors remain, and are even increasing in percentage as the largest contributor to total system error.

Currently, the flight crew is regulated by a layered system of certification: by operation, e.g., airline transport pilot versus private pilot; by category, e.g., airplane versus helicopter; by class, e.g., single engine land versus multi-engine land; and by type (for larger aircraft and jet powered aircraft), e.g., Boeing 767 or Airbus A320. Nothing in the certification process now requires an in-depth proficiency with specific types of avionics systems despite their prominent role in aircraft control and guidance.

New System Information

New systems now emerging will undoubtedly add safety to aircraft operating in the future airspace system, but the added information processing required to operate in that system will probably increase, not decrease, the crews' overall mental workload. For example, the capabilities of the Global Navigational Satellite System (GNSS) will eventually allow "uplinking" of the positions of individual aircraft via Automatic Dependent Surveillance (ADS), which could include data quantifying weather conditions surrounding each aircraft. The consolidation of specific traffic and weather into a national, even global, information network could be re-transmitted on demand to equipped aircraft. Moreover, air traffic control equipped with comprehensive and real time data about traffic and weather events in the airspace could dynamically manipulate airspace restrictions and uplink new configurations to cockpit electronic maps.

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

Even individual aircraft flight management systems could be automatically re-programmed from the ground for more efficient routing, then activated by acknowledgement from the crew. All this can provide crews with rich information about traffic, weather, and the airspace surrounding their aircraft, but it also increases monitoring responsibilities.

Monitoring

In the attempt to try to convey all information (without prioritization), the danger may be that there will be less knowledge conveyed (by overloading information). Design philosophy will play a role; some air carrier manufacturers (e.g., Airbus Industrie) prefer to display the status of many ongoing events and data, presumably with the objective to keep the crew better informed and aware. Other manufacturers (e.g., Boeing) choose to display principally abnormal status items, presumably protecting crews from lowering their attention by becoming accustomed to overlooking normal data. The implications of such designs are echoed in the ongoing debate about the effect of datalink cutting off "party line" monitoring of common radio frequencies. One side argues that crews are losing valuable situation awareness information by eliminating broadcasts from surrounding aircraft; the other side contends it is better to filter out all non-essential communications.

Supervision

Flight crews may have an even greater challenge supervising automated flight control systems than by actually performing tasks themselves. As Dr. Earl A. Wiener put it, "... bigger mistakes are made more often while supervising, than when in direct control" (Wiener, 1990). This may be due, in part, to less involvement by the crew; thus the potential for divided attention and larger casual errors. In addition, autopilots in the past were programmed to or from single ground reference points or altitude. These were easily remembered, and flight crews were kept involved as they made simple speed, time, and fuel calculations.

Today's flight management systems (flight-controlled autopilot coupled or not) have the programming and sequencing capabilities for dozens of navigational routes, flight profiles, and hundreds of waypoints far beyond the working memory of the crews. On-board computers (with databases) can processes information with ease and precision, leaving manual calculations passé and crews ever more dependent on system processing. Moreover, automatic sequencing is for the most part externally initiated at points of passage or interception, sometimes leaving the crews with the question "How in the world did I ever get into that mode?" (Sarter & Woods, 1994). The problem is that crew members may lose their awareness of the automated mode changes in their role shift from that of a supervisor anticipating future events to that of a monitor of lagging events. Often hardware and software are not designed to allow easy crew surveillance. Worse yet, mindless dependency on computer reliability, without mental supervision checking, produces the three most commonly asked questions on highly automated flight decks – "What is it [the flight management system] doing? Why is it doing it? What will it do next?" (Wiener, 1990).

Since current avionics are more likely to be integrated with aircraft controls, the consequences of errors are greater than even only a few years ago. The potential for errors extends across mistaken inputs (dumb use, i.e., non-checked use), misunderstood algorithms (misuse), and inadequate time for programming changes (rushed use). For example, transposition errors during the entry of latitude/longitude coordinates were implicated in the shoot down of Korean Airline 007 and in the 1977 collision of an Air New Zealand into Mount Erebus, Antarctica. Mode verification was apparently not done in the Airbus (A320) accident at Strasbourg, France, Air Inter Flight 148. The crew misinterpreted a desired flight path angle 3.3° for a vertical speed value of 3.3 x 1000 ft/min because both are entered into the same display. However, which one is active depends on the current mode (Monnier, 1992). As a final example, the slow response of a China Airlines crew in disengaging the autopilot during an engine failure in a Boeing 747 over the Pacific, resulted in a rollover and near terminal dive.

Thus, regardless of the level of automation, the pilot-in-command (PIC) remains the final authority. The FMS/FGS as an electronic crew member is not yet deemed capable of replacing the PIC's judgment. These systems currently perform exactly as programmed, yet they are not yet equipped with the artificial intelligence to correct PIC errors or suggest alternatives to a proposed PIC course of action. Moreover, in an informal survey we conducted, flight crews in both a Boeing 767 and an Airbus 320 did not know the manufacturer or the model of their "electronic" crew member, the FMS, admitting they were not fully aware of the full range of its capabilities, limitations, idiosyncracies, and underlying strategies.

Proposal

It is proposed that flight crews be certified for attaining mastery of sophisticated control systems. Perhaps the type of aircraft *plus* the type of FMS should become a new integrated type of aircraft/FMS certification, e.g., a B767/$A_{(FMS)}$ or B767/$B_{(FMS)}$. (Note, it is recognized that it becomes a judgment call as to when $A_{(FMS)}$ and $B_{(FMS)}$ are "significantly different" to require a different type rating.) Extending beyond automated flight control is the possible certification of crews to use software-specific navigational systems (FGS) which provide guidance during instrument flight operations. Given that manual control is often based on the same information displayed through the flight director that drives coupled systems, the argument for certification for use of uncoupled systems is just as compelling in IFR conditions.

An advantage to FMS/FGS certification beyond the demonstration of competency is that it provides hardware, software, and instructional system designers with human performance benchmarks to guide the design or training system process so as to best accommodate the human component. It also begins to insure that the total system (hardware, software, and people) will meet expected standards.

Justification

The basis for this proposal is that the traditional role of the flight crew is swiftly changing. Hands-on flight control is rapidly giving way to semi-automated or fully automated control even for the most routine operations. So much so that some long-haul crews that use automation for their infrequent takeoffs and landings fear losing their basic flight proficiency. Lower psychomotor effort without the need for constant attention to direct flight control has translated directly to a greater opportunity for monitoring and supervising flight systems, but only for crews who are highly FMS proficient. For low proficiency crews, the "mental" workload may be heavier. While certain mundane tasks such as navigational and altitude tracking are automated, additional capabilities and requirements are being added to the flight deck. This not only serves to supplant the "free time," but belies the mental effort involved in conducting the safest possible flight.

Testing

Testing should be done with the total system (i.e., the particular flight management system with a particular type of aircraft) to uncover any idiosyncratic aspects of system integration. The type certificate for aircrews should read, "(aircraft type)/FMS". Note that testing should consider both crew competency with the manufacturer's standard configuration as well as with their utilization of customized features. Unlike the past, the depth of testing should go well beyond just "how to use the system." In order to maintain mode awareness, the crew must know "how the system works" at a deeper structure – (based on a mental model) at least at a macro-level – and have the ability to track what the system is doing at any given time. For example, current FAA testing is for type ratings with the FMS fully functional and fully operational. In prior years, examiners turned *off* equipment to increase the difficulty level. A deeper understanding of the system(s) would be demonstrated with partial levels of automation and no automation – e.g., scenario manipulation to infer situation awareness (Sarter & Woods, 1994). Perhaps more importantly, crews should be able to accurately predict what the system will do next, allowing for anticipation of automated programming changes or hardware/software errors in the making; "thinking ahead of the aircraft" (Regal et.al., 1988).

References

Monnier, A. (1992). Rapport preliminaire de la d'enguete administrative sur l'accident du Mont Sainte Odile du 20 Janvier 1992.

Regal, D. M., Rodgers, W. H., & Boucek, G. P. (1988). Situational awareness in the commercial flight deck: Definition, measurement, and enhancement. Paper presented at the Aerospace Technology Conference and Exposition, Anaheim, California.

Rouse, W. B. & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin, 100*, 349-363.

Sarter, N. B. & Woods, D. D. (1992). Pilot interaction with cockpit automation: Operational experiences with the Flight Management System. *International Journal of Aviation Psychology*, 2(4), 303-322.

Sarter, N. B. & Woods, D. D. (1994). "How in the world did I ever get into that mode?", Mode error and awareness in supervisory control. *Proceedings of the First International Symposium on Situation Awareness* (in press). Orlando, FL: Center for Applied Human Factors in Aviation.

Tenney, Y. J., Jager Adams, M., Pew, R. W., Huggins, A. W. F., & Rodgers, W. H. (1992). A principled approach to the measurement of situation awareness in commercial aviation. (NASA Contractor Report No. NAS1-18788). Hampton, VA: NASA Langley Research Center.

Wiener, E. A. (personal communication, August 23, 1993).

N95- 34773

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

Ronald L. Small & William B. Rouse

Search Technology

Introduction

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

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

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

Naturalizing

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

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PRECEDING PAGE BLANK NOT FILMED

PAGE 124 INTENTIONALLY BLANK

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

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

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

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

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

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

Stakeholder Concerns

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

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

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

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

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

System Evaluation

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

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

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

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

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

Evaluation Methodology

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

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

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

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

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

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

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

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

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

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

Methodology Summary

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

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

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

Certification Issues

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

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

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

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

Certification Process

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

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

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

Other differences between evaluation and certification relate to rules of development that are designed to minimize the system's dynamic response to conditions. Certifiable systems should not have unpredictable failure conditions. For example, when we built a certifiable knowledge base development tool, we had to pay special attention to some specific software engineering issues, including:

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

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

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

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

Certification Team Composition

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

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

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

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

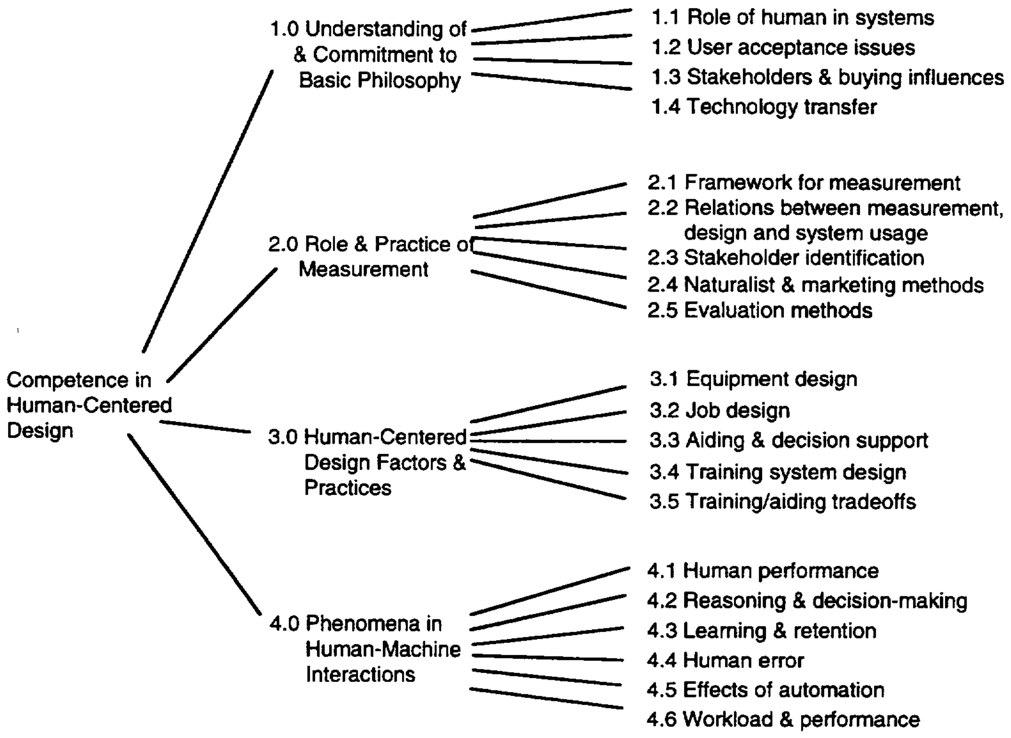

Competence in Human-Centered Design
1.0 Understanding of & Commitment to Basic Philosophy
1.1 Role of human in systems
1.2 User acceptance issues
1.3 Stakeholders & buying influences
1.4 Technology transfer
2.0 Role & Practice of Measurement
2.1 Framework for measurement
2.2 Relations between measurement, design and system usage
2.3 Stakeholder identification
2.4 Naturalist & marketing methods
2.5 Evaluation methods
3.0 Human-Centered Design Factors & Practices
3.1 Equipment design
3.2 Job design
3.3 Aiding & decision support
3.4 Training system design
3.5 Training/aiding tradeoffs
4.0 Phenomena in Human-Machine Interactions
4.1 Human performance
4.2 Reasoning & decision-making
4.3 Learning & retention
4.4 Human error
4.5 Effects of automation
4.6 Workload & performance

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

Conclusion

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

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

Acknowledgments

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

References

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

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

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

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

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

Selection and Training

PRECEDING PAGE BLANK NOT FILMED PAGE 134 INTENTIONALLY BLANK

N95- 34774

Certification of Training

Richard S. Gibson

Embry-Riddle Aeronautical University

Introduction

Training has been around as an informal process for countless years. Most higher order animals require some level of training in hunting, social skills, or other survival related skills to continue their existence beyond early infancy. Much of the training is accomplished through imitation, trial and error, and good luck. In some ways the essentials of training in aviation have not deviated from this original formula a great deal. One of the major changes in aviation and other technical areas is that more complex response chains based on a broader base of knowledge are now required.

"To certify" means many things according to the American Heritage Dictionary of the English Language (Morris, 1969). These meanings range from "to guarantee as meeting a standard" to "to declare legally insane." For this discussion, we will use the definition "an action taken by some authoritative body that essentially guarantees that the instruction meets some defined standard." In order to make this certification, the responsible body subjects the educational process, training, training device, or simulator to some type of examination to determine its adequacy or validity.

Academic Accreditation

In the academic community, the certification process is called accreditation. This refers to the granting of approval to an institution of learning by an official review board after the school has met specific requirements. In the United States, most universities and colleges are accredited through regional associations, which are voluntary associations of educational institutions. For example, Embry-Riddle Aeronautical University is accredited by the Southern Association of Colleges and Schools (SACS), which is the recognized accrediting body in the 11 U.S. Southern states (Alabama, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Texas and Virginia). SACS and other regional associations establish a set of criteria that the members must meet. These criteria address areas that are considered important to the effective operation of a college or school. In the case of SACS (SACS, 1989) this includes institutional purpose, institutional effectiveness, educational program, educational support services, and administrative processes. The accreditation process is a personnel intensive procedure involving an internal review conducted by the university's faculty followed up by a formal review by a visiting team from SACS composed of faculty from other universities and colleges. The process takes over

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PRECEDING PAGE BLANK NOT FILMED

PAGE 136 INTENTIONALLY BLANK

a year and consumes thousands of personnel hours and many thousands of dollars. Generally a satisfactory accreditation is valid for a period of ten years before the accreditation must be reaffirmed. The reward to the university is that other universities will recognize the credits awarded to their students, and also that the university qualifies for many government loan and grant programs. Failure to win or retain accreditation can have catastrophic consequences.

Traditionally this rather complex process has relied upon the expert judgment of subject matter experts for both the self-study and the visiting review team. More recently, as the result of pressures from state legislatures interested in proof of the value of various college programs, there has been an increasing emphasis in the use of more objective, verifiable measures, such as the pre- and post testing of students (Did they learn anything?), performance of graduates on licensure examinations (Did they learn enough?), to surveys of employers of graduates (Did they learn anything useful?), etc. And, as importantly, asking how the institution has used this information to improve its programs. As this process of using objective evaluations continues to grow, the accreditation process shifts from using construct validity, based upon a systematic review by experts, to using empirical validity based upon observable results.

In addition to the regional accrediting associations, there are many specialized accrediting bodies based on specific academic disciplines, such as engineering, business administration, computer science, and psychology. Their procedures are similar to the regional associations. While not as important as the regional accreditation, the specialized accreditations demonstrate that the programs accredited meet the specialized requirements of various professional associations. Since all of these accreditations are paid for by the requesting institution, the cost in both time and money is significant.

Since all of the accreditation processes are pass/fail procedures, the outcome is not to guarantee academic excellence but to set the level of minimally acceptable academic mediocrity. The primary effect is to bring the weaker institutions up to a level of defined acceptability. This assures the consumers (the students and their parents) that they will get some reasonable value for their investment. However, for our purposes, the use of independent associations to establish and regulate accreditation or certification criteria can serve as one type of possible model for the certification of training, training devices, or simulators.

Professional Licensure/Certification

Another approach to the certification problem can be found in the process of licensure for selected professions. The responsibility can be divided between a government regulatory body and a professional society. For example, for the licensure of clinical psychologists in the State of Virginia, the applicant must have graduated from an American Psychological Association (APA) approved graduate program, must have passed an APA national licensing examination, must pass a state written examination, and finally a state administered oral examination (Regulations of the Virginia Board of Behavioral Science). As in the case of accreditation, the full costs are borne by the applicant. Again, the license does not mean that high quality services will be provided by the licensed individual. It does mean that sufficient minimum standards have been met so that the licensee is not considered to be an undue risk to the public. This joint relationship between a professional society and a government regulatory body provides another type of possible model for the certification of training, training devices, and simulators.

Interestingly, the APA has another level of recognition called the Diplomate. An individual with a Ph.D. and appropriate experience can apply for Diplomate status, and after a favorable review of credentials and the passing of a special examination, be awarded Diplomate status. This means that the association is essentially certifying the individual for private practice of the profession. Unfortunately, state licensing procedures do not give special recognition for the Diplomate status: the licensing process is the same for an individual with or without the Diplomate.

This brings up additional issues with respect to certification. The issues are, "who will recognize the certification" and "what is its economic value." With both academic accreditation and professional licensing, there is significant economic value for being certified. However, with the case of the Diplomate, the certification may have intrinsic value to the recipient of the recognition or certification, but have little or no real economic value.

Training, Training Devices, and Simulation

Early aircraft simulators tended to look like miniature aircraft with stubby wings and tails. Their design gave them face validity. If they looked like airplanes and the instruments and controls appeared to be the same, they should be useful in teaching flying skills. Buyers of aircraft simulators have consistently had a strong bias toward purchasing devices that looked and acted like the real thing without actually becoming airborne. Researchers have tended to follow behind the development curve with questions such as: Does the training transfer? How much fidelity is enough? What is the cost effectiveness of simulator versus aircraft training?

A study by Provenmire and Roscoe (1973) used the Link Gat-1 simulators to train pilots to pass their final flight check in the Piper Cherokee Aircraft. Student pilots were given either 0, 3, 7, or 11 hours of training in the simulator before continuing their training in the aircraft. The results showed that larger amounts of time in the simulator led to larger amounts of time saved in the aircraft; however, the amounts of additional flight time saved diminished in the familiar shape of a learning curve. These results were important for two main reasons. First, they provided a basis for calculating the marginal utility of the simulator. Training in the simulator was cost-effective until the Incremental Transfer Effectiveness Ratio dropped below the simulator/aircraft operating cost ratio – in this case, about 4 hours in the GAT-1 for training student pilots to pass the final flight check for a private pilot's license. Second, the data also indicated that there was an upper limit to the transferability of simulator time to improved performance in the aircraft. Beyond a certain level of practice, about 8 hours in this case, the students were not showing increased benefits to their aircraft performance. This suggests that any within-simulator improvements were simulator peculiar without additional transfer value.

In an extensive review of the use of maintenance simulators for military training, Orlansky and String (1981) concluded that student achievement in courses that used maintenance simulators was the same as or better than that in comparable courses that used actual equipment trainers. In fact, not only was the training as good, it was cheaper. In one case that they cited, the total costs for the same student load over a 15-year period were estimated to be $1.5 million for the simulator and $3.9 million for the actual equipment trainer; that is, the simulator would cost 38 percent as much to buy and use as would the actual equipment trainer. In a subsequent study (Gibson and Orlansky, 1986), it was noted that student confidence and performance closely paralleled instructor ratings of simulator fidelity. They concluded that to make any generalizations about the effectiveness of simulator-based training without considering the fidelity of the simulators would be unwarranted.

Simulators that offer very high fidelity do not represent a serious problem for certification. The problem becomes more difficult as training devices depart in various ways from being faithful replicas of the aircraft and aircraft systems they represent. While initial students can benefit from a variety of relatively low fidelity training devices and simulators, experienced pilots receiving refresher training tend to need high fidelity simulators. The FAA Advisory Circular (AC) 120-45A specifies the evaluation and qualification requirements for six of a possible seven-level-of-flight-training devices. Level 1 is currently reserved and could possibly include PC-based training devices. A flight training device is defined by the FAA as:

> a full scale replica of an airplane's instruments, equipment, panels, land controls in an open flight deck area or an enclosed airplane cockpit, including the assemblage of equipment and programs necessary to represent the airplane in ground and flight conditions to the extent of the systems installed in the device does not require a force (motion) cueing or visual system; is found to meet the criteria outline in this Advisory Circular for a specific flight training device; and in which any flight training event or checking event is accomplished.

PC-based training devices do not meet these criteria, but many are offering some fairly impressive approximations. There will be a fairly steady pressure for some type of certification of some of these hardware/software combinations for currency or refresher training.

Trainer Certification and Training Verification

Assuming that this will eventually happen, there are two problems to be addressed: one is the extent that any simulator training transfers and the other is to have a system that will verify the amount of flight experience with the training device and the quality of the individuals performance. The Provenmire-Roscoe model provides one way to establish the transfer effectiveness and to establish a metric for the upper limits of substitution for using the PC-based devices. This may be too costly and it may be necessary to assess performance relative to accepted reference simulators.

This would be similar to the field practice of tests and measures in which paper and pencil intelligence tests are generally judged by how well they correlate with the individually administered intelligence tests, such as the WAIS.

The other problem will be that of verification of the actual amounts of practice. Regulators are wary of accepting unconfirmed self reporting. Emerging technology may offer some assistance. It is currently possible to log onto networked games and play other opponents interactively; and the network charges for the time used. In the future, it may be possible to access approved (certified) networked software provided that you have the right PC hardware configuration and practice flying. The network could keep the necessary records. Another option would be to use a "smart" card system that would use the PC and attached "smart" card hardware to provide a record of the training hours. Obviously, there would have to be periodic checks in higher fidelity systems to provide the training not available on the PC, and check for possible abuses of the system. A pilot who had high PC training time but who performed poorly on the "check rides" would lose PC privileges.

Conclusions

Numerous models for training certification exist. All models require either construct validation based on expert opinion, or some form of empirical validation that examines the results of the training. To be effective, the certification needs to be recognized by the appropriate regulatory agencies.

Techniques exist to assess the training effectiveness of training, training devices, and simulators. However, because of the cost and effort required, there is a need to examine the relationship between performance on low fidelity devices as a predictor of performance on higher fidelity intermediate devices that could be used as reference standards. If PC-based systems win certification, there will also be a need to establish a reporting verification system based either on network usage or some type of "smart" card.

References

Gibson, R. S., & Orlansky, J. (1986). *Performance measures for evaluating the effectiveness of maintenance training*. (IDA Paper P-1922). Alexandria, VA: Institute for Defense Analyses.

Morris, W. (Ed.). (1969). *The American heritage dictionary of the English language*. Boston: Houghton Mifflin Company.

Orlansky, J., and String, J. (1981). *Cost-effectiveness of maintenance simulators for military training*. (IDA Paper P-1568). Arlington, VA: Institute for Defense Analyses.

Provenmire, H. K., & Roscoe, S. N. (1973). Incremental transfer effectiveness of a ground-based aviation trainer. *Human Factors*, 15, 534-542.

Southern Association of Colleges and Schools. (1989). *Criteria for accreditation commission on colleges*. (1989-1990 ed.). Atlanta.

N95- 34775

Presentation of a Swedish Study Program Concerning Recruitment, Selection and Training of Student Air Traffic Controllers: The MRU Project Phase 1

11-53

Rune Haglund

Civil Aviation Administration, Sweden

Introduction

Background Phase 1

The Director of the ANS Department has set up an objective for the efficiency of screening and training procedures for air traffic controller students which implies that all students admitted "shall be considered to have the qualification for – and be given the means of – completing the training".

As a consequence, a study project has been established. It is run by the ANS Department with members from the Swedish CAA, in close cooperation with Uppsala University.

The task force of the MRU project consists of following members:

- Mr. Rune Haglund, Project Manager, Senior ATS Specialist, Swedish CAA
- Mr. Bertil Andersson, Air Traffic Controller, Swedish CAA
- Mr. Björn Backman, Industrial Psychologist, Swedish CAA
- Mr. Olle Sundin, Manager Arlanda ATS, Swedish CAA External expert
- Professor Berndt Brehmer, Ph.D., Department of Psychology, Uppsala University.

Graduation Rate

On the first of January, 1978 the military and civil ATS systems in Sweden were totally integrated into the Swedish Civil Aviation Administration. As a preparation for this alteration a new ATS Academy was created and a new integrated air traffic controller training programme implemented in 1974. One of the aims for this training programme was to decrease the failure rate to a maximum of 20 percent.

This objective has not been reached. However, since the start 1974, the average failure rate has been reduced by almost 20 percent. This improvement cannot be described as a steady curve. Instead, there is a great deal of unpredictible fluctuation around an average figure for successful training results:

• During the 1970's	Average 54%	Range 27% - 71%
• During the 1980's	Average 66%	Range 57% - 86%
• During the 1990's	Average 74%	Range 63% - 90%

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PAGE 142 INTENTIONALLY BLANK

PRECEDING PAGE BLANK NOT FILMED

Conclusions about the success rate and trends regarding the present recruitment, selection and training procedures are based on simple Analysis of Variance. Each decade was considered a group and the success rate of every class in that decade is the dependent variable. McNemar's (1969) formula for groups of unequal size has been used to test the significance of differences between the means.

$$F=s^2_b/s^2_w = .082/.015=5.467 \; p < 0.05$$

The conclusion of the task force is that there are systematic differences between the decades and that they are due to the greater experience of the people involved, instead of the systematic changes in recruitment, selection and training procedures. Interestingly enough, the failure rate has decreased by one percentage unit per year since the start of the new integrated ATS Academy. The mean for the years 1990–1993 has been calculated on the basis of 8 completed classes with a total of 190 accepted students and from which 140 graduated (74 percent). This outcome can be compared with the rate numbers of 80 percent graduating from the FAA Academy that FAA reports (MRU Delrapport 3, 1993).

Economic Review

As a key figure for reviewing the costs of the recruitment, selection and training system, one can calculate the total costs per graduated student. The total cost to the Swedish CAA for providing a new licenced TWR/TMC air traffic controller is 205,000 USD. For an ACC controller, the cost increases to a total of 255,000 USD.

This total cost can be divided with the total amount of weeks in training as shown below (currency in SEK). Figure 1 shows the costs accumulated over weeks.

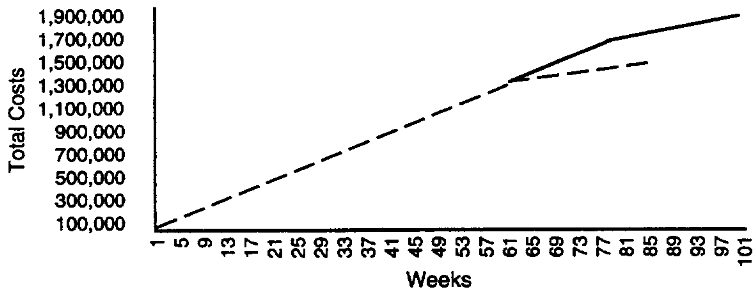


Figure 1. Total Costs over Weeks.

Today, the Swedish CAA has achieved a balance between supply and demand with respect to air traffic controllers. This leads to an acute planning problem that can be described as follows.

Today CAA has to employ 27 students to be able to deliver 20 air traffic controllers into the ATS system. This is due to the unpredictible span between accepted and graduated students in the present screening and training system: the current system will provide an outcome of qualified licenced air traffic controllers that varies, by chance, from 17 up to 24. This uncertainty has great negative effects on both planning and economics (lack or surplus of personnel).

The outcome of today's system of recruitment, selection and training of controller students is not satisfactory because it generates both a costly fluctuation around the mean and allows students who do not have the necessary abilities to remain in training for too long before they are expelled. An efficient system with a more predictable outcome and a higher success rate, i.e., a deviation of not more than 10 percent around the mean and a 90 percent success rate, would save the Swedish CAA at least 520,000 USD per class of 20 students.

Goal Setting

It is the opinion of the MRU task force that money and other resources invested in developing procedures for recruitment, selection and training of controller students, so that the outcome is less affected by random errors, will be a good investment.

The Research

In order to pursue the causes of today's random errors, the MRU task force issues followed a two step procedure. Step one involved a job analysis, and step two a study of the correlations between tests and training results. These two steps were taken in order to validate current recruitment, selection and training methods in use by the Swedish CAA.

Job Analysis in Order to Determine the Job Criteria

A number of acceptable procedures exist for conducting a job analysis. One way is to interview observers who are aware of the aims and objectives of the air traffic controller's profession and who see the controllers perform their profession on a duties frequent basis. Thus supervisors, peers and instructors may be interviewed about their observations of the critical requirements of the air traffic controller profession.

Current international research and analyses of the controller's job show that the air traffic controller profession is a very complex occupation where the tasks are performed in a very special work environment.

The Selection Procedure

Brehmer (1993) notes that the current selection procedure is based on a series of tests and interviews. The tests have been chosen by ABAR, a consulting company specialising in psychology of work and organisation. The choice of tests seems to have been made on the basis of a general analysis of the air traffic controller's job. But there has been no standardisation or statistical evaluation of the effectiveness of the selection procedure, except for later addition: the use of percept-genetic techniques.

The paper-and-pencil tests are described in terms of four factors (ABAR, 1978):

- *Flexibility and ability to find new solutions.* The aspect is measured by means of three tests: "Skeppsdestination" (Ship's destination), "Instruktionsprov II" (Instruction test II) and "Kravatt" (Neck tie) and concerns the ability to improvise and make decisions in unexpected situations.

- *Logical ability.* Logical ability is measured by means of two tests: Raven's matrices and Number series, which are designed to measure logical ability.
- *Spatial ability.* The aim is to measure the ability to construct a three dimensional picture of the air space from two dimensional information. Three tests are used for this: "Klossar" (Blocks), "Pl–tmodeller" (Metal Sheet Models) and "WIT Puzzles."
- *Attention to detail, carefulness, and short term memory.* This factor is measured with five tests: "Korrektur ABAR" ("Proof-reading ABAR"), "Sifferkorrektur" (Proof-reading of numbers), "Namnminne" (Memory for names), "Sifferminne" (Memory for numbers) and "Figuridentifikation" (Identification of figures).

In the final test battery, memory is treated as a separate factor and in addition to the tests mentioned above, two additional tests are used: "Uppskattning" (Estimation) and "Felletning" (Error search). The motives for including these are not given, and it is not clear what they are supposed to measure. In addition, a percept-genetic (PG) test is included together with an interview which aims to assess the applicant's motivation for the job. A test of capacity to process different information simultaneously is also included, and an interview by personnel from ABAR which assesses ability to cope with stress, ability to cooperate, ability to take initiative, and motivation for the air traffic controller job, i.e., many of the factors also covered by the interview by the consultants in charge of the PG test. Finally, there is an interview by personnel from the Civil Aviation Administration.

The MRU Hypothesis 1. International research and job analyses regarding the air traffic controller's profession show that a majority of the work behaviours can be described in terms of cognitive skills. The first hypothesis is that a job analysis in the Swedish work environment will replicate these results.

The MRU Hypothesis 2. The second hypothesis is that self-confidence plays an important role in coping with the critical job factors.

The MRU Hypothesis 3. The third hypothesis is that interpersonal skills play a significant part in being a skilled air traffic controller or student.

The MRU Hypothesis 4. The fourth hypothesis is that there is a significant difference in test results between those who successfully complete their controller training and those who fail.

The MRU Hypothesis 5. The fifth and final hypothesis is that training based on cognitive skills training, coaching and mentoring the students will be more effective than traditional training methods (e.g. on the job training, OJT).

Methods and Research Procedures

Job Analysis Procedures

The interviews were conducted as focused group interviews with a representative sample of ATS units. A total of 11 ATS units and 2 training units were visited. 127 air traffic controllers participated in the focused group interviews. The interviewers who were experienced in using this method worked in pairs. Each ATS unit had been contacted in advance about the purpose of the interviews. The interviewers used interview guides prepared in advance.

The interviewees were asked about how skillful air traffic controllers coped with stressful situations or events. Both the stressful situations, or events, and the effective work behaviours were recorded. The interviewers compared notes afterwards and only notes which agreed were accepted. Almost 400 different measures to cope with stressful situations were recorded.

All responses noted were thereafter recorded and tabulated according to frequency. Thus the content of the job analysis was a frequency table of stressful events and corresponding key behaviours (the effective way how to cope with stressful event).

Control of Validity

The second step of the job analysis was to transform the responses of the interviewees into different questionnaires for different types of ATS units (i.e. TWR, TMC and ACC). The same procedures as that described above were applied to the students' working situation.

The different questionnaires were distributed to a representative sample of 158 air traffic controllers and instructors working at TWR, TMC or ACC (radar and non radar). Their task were to list, on a 7-point scale, the importance of the behaviours and how often the related situations occurred in the daily work life. Step number two was taken as a measure of the relevance or content validity of the results.

The recorded events and behaviours were compared with the causes of failures for students undergoing training. This as a test of the predictive validity of the job analysis.

The final step was to compare the results of the job analysis with data from the air traffic incident and information report system which exists at the ANS department. This step was taken to check the construct validity of the job analysis.

Analysis of the Test Battery Material

The material (Brehmer, 1993, MRU Report 7) consists of 145 students who have been admitted to air traffic controller training 1990-91. The students come from the courses starting 9007 (26 students), 9008 (24 students), 9009 (1 student), 9011 (26 students), 9107 (28 students), 9108 (24 students), and 9111 (16 students). There were 58 women and 87 men. Thirty-seven had failed and 104 succeeded, or at least not failed at the time when the evaluation of the test battery was done. Data with respect to success was missing for four students, who had taken a leave of absence from the training.

Complete data are available for only 134 of the 145 students. The number of students for which data are available varies from 134 for the selection variable with the lowest number of students to 141 for that with the highest number. It is not likely that this will have had any important effect on the conclusions.

Analysis of Relations Among Variables

Two different kinds of analyses have been performed: regression analyses and discriminant analyses. Both of these aim at assessing the extent to which it is possible to predict success in training from the various predictor variables.

Regression analysis, which shows how well success can be predicted from the predictor variables, as well as the relative importance of different predictor variables, is the standard method for this purpose. However, some objections can be directed at this method in the present case where the outcome variable is binary. Therefore, we also made discriminant analyses which show the extent to which it is possible classify the students into two groups: those who pass and those who fail the training. As we shall see, these two methods give the same results.

In these analysis, the sex of the applicant has been entered as a predictor variable in addition to the test and interview variables. The analyses are based on the 134 cases for which complete data were available.

Results

Introduction

The demanding tasks accounted for in the charts below have been compiled into five categories describing the nature of those stressful tasks. The terms used can be explained as follows.

Traffic Processing

This term is used to describe the actions taken, and the decisions made, to establish a safe and well organized flow of traffic by use of clearances, separations, applicable working methods and planning.

Coordination

This term describes the communication between air traffic controllers used to exchange information, obtain clearances, revise previous information or hand over the control of an aircraft to establish a safe and well organized flow of traffic.

Disturbances and Irregularities

This term is used to describe situations and duties when normal working methods cannot be used (e.g. technical malfunction, irregular behaviour of an aircraft etc.).

Fluctuating Workload

Description of the events and situations connected with uneven flow of traffic (e.g. high traffic intensity with a variety of performance characteristics, followed by low traffic intensity, different flight status and a mix of military and civil aviation).

Personalities and Social Skills

These terms describe how the persons interviewed perceive air traffic controllers, their personalities and social behaviours.

The results of the studies show that the reported actions and behaviours are involved with Information Gathering, Decision Making and Communication in connection with traffic processing and coordination. Actions and behaviours caused by (high) level of Ambition and (high) demands on Performance account for a large portion of the strain appearing with irregular flow and varying traffic.

ACC

The most significant behaviours in the categories "Information Gathering and Decision Making are found in the ACC function. Five of the ten most common behaviours involve Information Gathering. The problem area "Social Relations" was not awarded the same significance as behaviours more closely connected to traffic processing, which is shown by the number of behaviours that came up in the group interviews. In the ACC function the highest importance was given to behaviours dealing with coordination and traffic processing:

- Accurate, short and precise coordination with proper prioritization
- Identifying conflicts early and following up on traffic.

TMC

In the TMC function, the majority of the behaviours that are ranked high in importance or frequency appear in the areas "Decision Making and Communication." The reasons for these behaviours can mostly be found in the straining tasks in connection with coordination and traffic processing.

A similar division of work behaviours as in the ACC function appears also in the TMC function, with the difference that by comparison, it is more important to:

- Dare to say "no"
- To be, and be perceived as being, determined.

TWR

In the TWR function, the most significant behaviours are found in the category "Decision Making, followed by "Information Gathering" and demands on "Ambition and Performance." The category "Communication" was awarded lower significance than in the ACC and TMC functions.

In the TWR function, the importance of behaviours categorized under coordination decreases in favor of behaviours in the area of traffic processing. In the TWR function, the highest importance is awarded to:

- Making decisions, looking out and following up on traffic
- Working with confidence in one's ability and maintaining concentration also during periods of low traffic intensity.

Conclusions

As a description of the air traffic controller's profession and of air traffic control services, the survey largely corresponds with what the persons interviewed reported as significant behaviours in maintaining a safe and well organized flow of traffic. We can therefore conclude that the job analysis is valid as well as reliable.

Movements and changes occur in the air and/or on the ground, and the air traffic controller is expected to handle these processes in a safe and orderly manner from her or his position in the tower or control central. The air traffic controller is thus not physically situated in the surroundings where these changes occur and cannot experience the movements and changes with her/his senses. Instead, the air traffic controller must create a mental picture of the present situation, or of what the situation will be like within a limited time frame using the fragmentary information provided. As a support in constructing this mental picture, the air traffic controller has a number of technical aids: radio systems, direction finder, radar systems, data displays and monitors, telephones, telefax, telex etc.

Attribution	**Ranking**
Slow starter, unprogressive learning curve.	1
Rigid and uniform working methods.	2
Passivity, lack of initiative, inactivity, late decisions.	3
Low stress tolerance, makes mistakes in complex situations.	3
Lack of theoretical knowledge.	3
Slow worker, slow in decision making.	4
Inequality of performance.	4
Tense and nervous personality.	4
Inadequate coordination.	5
Insecure when working, doubts own decisions.	5
Insufficiant understanding of the ATC system.	5
Excessively dependant on instructor.	5
Inability to switch from low to high workload.	5
Lack of concentration.	5
Constant inability to maintain separation.	5
Insufficiant planning skills.	5
Careless, not following instructions.	5
Lack of motivation, discontinuance due to other education.	5
Total number of attributions	**33**
Average number of attributions per student	**5**

Table 1. Attribution and rank, i.e., reasons for failures during basic training between 1990 and 1993. The attributions come from 17 randomly selected students who failed to complete the training program.

The predictive validity of the job analysis. A useful basis for studying success and failure is Heider's (Hastorf, Schneider and Polefka, 1970) Theory of Attribution. The central issue of Heider's theory is viewing behaviours as caused either by environmental factors or by the individual herself. The conception of reasons also leads to predictions about future behaviours and concequences. The attribution itself becomes a deterministic prediction of the future: chance factors are not considered.

Problem area	Controller	Failed trainee
Decision making	1	1
Ambition	2	2
Information gathering	3	5
Relations	4	3
Communication	5	5
Irregularities	6	4
Technical environment	7	6
Theoretical facts	8	5

Table 2. A comparison between air traffic controllers and failed trainees, regarding ranking of problem areas.

Spearman's correlation of rank (Renyon-Haber, 1971) describes the statistical connection in ranking of problem areas between air traffic controllers and failed students. The correlation is .78.

The Construct Validity of the Analysis

One noticeable discrepancy between the findings of Mattson (1979) and the MRU project is that Mattson only found two separate activities in air traffic control services: Decision Making and Communication. The MRU project has found that, in traffic processing and coordination, Information Gathering and Processing are highly important as a preparatory stage and that Self Confidence is an important characteristic of the controller.

The interpretation made by ANS/HQ of the irregularity reports, taken from the ANS department's air traffic incident and information report system from 1991, is: "The air traffic controller assumes or expects, often as a result of indistinct or incomplete phraseology, that a pilot will act in a certain manner. The controller therefore neglects to take measures that would ensure the pilot to perform in the manner assumed by the controller."

The executives in charge of the ANS/HQ judged that all of those incidents could have been avoided if the controller had taken action to ensure that the pilots performed in the way intended by the controller.

The importance of following up on one's decisions and measures is regarded as a key behaviour in the annual analysis, published by the ANS/HQ, and also in the MRU project's job analysis.

One interpretation in the job analysis is that important behaviours in the TWR function are Decision Making, to look out and follow up on traffic, to work with confidence in one's ability and to maintain concentration even during periods of lesser workload.

This corresponds with the summary presented in the annual statistics, published by the ANS/HQ, concerning irregularity reports.

The interpretation of those reports show that the most serious incidents occur in the immediate vicinity of the airport. The final evaluation of the incidents is that operators and supervisors have not sufficiently emphazised methods of working and phraseology.

The importance of a distinct and fixed phraseology, in order to verbally express one's decisions and measures to the party or parties concerned, constitutes a key behaviour in all sources accounted for.

Relationship Between Test-Results and Training Outcome

Table 3 shows means and standard deviations for the different predictor variables. For each variable, these computations are based on data for those students for whom the result in that variable was known; i.e., the number varies from 134 to 141. In these analyses, the results for the PG test, which are reported only in terms of two categories, + (a positive value) and +/- (doubtful) have been dummy coded with 1 for the +/- category and 2 for the + category. No students with a pure value on PG (i.e., students for which the prognosis according to this test was clearly negative) had been admitted.

It is important to note that the means in the predictor variables in terms of standard scores (stanine scores in this case, with a mean of 5 and a standard deviation of 2) are generally about one unit above the mean, and that the standard deviations do not differ very much from 2. This is likely to result from the compensatory effect of summing the scores for the individual tests into scales, as is done in the selection procedure. A result of such a summation is that high scores in one variable will compensate for low scores in another variable in the scale. It is therefore possible to find minimum scores of 2, and even 1, in most of the tests. This means that the scores for the tests for those who have been admitted to air traffic controller training will not deviate too much from the scores of those who apply, as is also shown by the fact that the standard deviations do not differ very much from 2. This means that correlational analyses, the results of which are affected by the standard deviations (but not the means) of the variables, will be meaningful, and that the effects of the possible restriction of range as a result of the fact that the students have been selected on the basis of the tests being evaluated, will not be too serious. We certainly do not have to expect that the restriction is so serious that it will be impossible to detect the relations that might exist between success in training and the predictor variables.

For some of the variables, the restriction is, however, considerable. This is especially true of the interview variables and the PG test. For these variables, it is clear that the whole range of scores is nor represented in the present sample. Concerning PG, only two categories are found in that no student with a clearly negative prognosis has been admitted, and the distribution of PG scores is quite skewed, with very few +/- values. In the ABAR interviews all of the scores are between 5 and 7, and in the interviews by the Civil Aviation Authority personnel all scores are between 5 and 8. This means that it is difficult to say very much about these variables from the results of the regression and discriminant analysis.

Regression Analyses

In the regression analyses, training outcome, sex and PG have been dummy coded. The results of a regression analysis with all predictor variables are shown in Table 4.

The multiple correlation, R, is 0.413. This represents an overestimation of the strength of the relations between the predictors and the outcome in that it capitalises on sampling error. Moreover, the ratio of predictors to observations is high (19 to 134). After correction for such errors, the adjusted squared multiple correlation is 0.032 ($p < 0.25$). That is, there is no significant relation between the predictors and the training outcome and the whole set of predictors explain only 3 percent of the variance in the training outcome.

	INST (R)	INST (F)	INST (S)	SERIER (R)	SERIER (F)
M	33.46	4.48	6.43	19.32	3.02
S	2.83	2.47	1.62	3.19	2.04
	SERIER(S)	KLOSS (R)	KLOSS (F)	KLOSS (S)	KORR (R)
M	6.39	39.41	8.11	6.51	21.49
S	1.56	3.39	3.74	1.65	3.20
	KORR (F)	KORR (S)	SKEPP (R)	SKEPP (F)	SKEPP (S)
M	2.44	6.18	41.70	5.14	6.12
S	1.76	1.70	5.21	4.47	1.61
	UPPSK (R)	UPPSK (F)	UPSK (S)	WIT (R)	WIT (F)
M	5.97	127.28	2.38	5.55	12.13
S	1.29	20.86	2.26	1.82	2.20
	KRAVAT (F)	KRAVAT (S)	PLÅTMO (R)	PLÅTMO (F)	PLÅTMO (S)
M	2.18	6.26	28.11	2.64	5.59
S	1.71	1.58	4.11	2.53	1.63
	PS IF (R)	PS IF (F)	PS IF (S)	FELLET (R)	FELLET (F)
M	51.24	1.44	6.22	16.39	1.08
S	5.61	1.44	1.74	2.92	1.20
	FELLET (S)	MATRIS (R)	MATRIS (F)	MATRIS (S)	VAR 1 (R)
M	6.77	38.13	8.09	6.31	18.86
S	1.45	4.91	3.67	1.43	3.24
	VAR 1 (S)	VAR 2 (R)	VAR 2 (S)	VAR 3 (R)	VAR 3 (S)
M	6.59	12.68	6.60	18.06	6.32
S	1.36	2.39	1.46	3.32	1.40
	VAR 4 (R)	VAR 4 (S)	VAR 5 (R)	VAR 5 (S)	SUMMA (R)
M	17.97	6.26	13.32	6.99	32.76
S	3.78	1.54	2.37	1.44	4.28
	SUMMA (S)	SIMULTAN	MINNE	ABAR	LFV
M	7.07	5.71	6.04	5.55	6.01
S	1.15	1.29	1.51	0.55	0.67
	PG				
M	1.92				
S	0.27				

Table 3. Showing means (M) and standard deviations (S) for the different predictor variables.

Variable	Coefficient	*p*
Intercept	0.125	0.869
ABAR-intervju	0.061	0.430
Skeppsdestination	0.057	0.032
Felletning	0.054	0.073
Instruktionsprov	0.007	0.791
Klossar	-0.026	0.328
Korrektur ABAR	0.055	0.050
Kravatt S	0.005	0.849
Kön	-0.019	0.844
LFV-intervju	-0.121	0.073
Matriser	-0.013	0.687
Minne	0.002	0.930
PG	0.124	0.416
Plåtmodeller	-0.011	0.694
PS IF figurer	0.026	0.291
Serier	0.019	0.504
Sifferkorrektur	-0.027	0.302
Simultankapacitet	0.016	0.645
Uppskattning	-0.033	0.307
WIT Pussel	-0.009	0.806

Table 4. The results of the regression analysis with all predictor variables.

The best predictors are "Skeppsdestination" (B=0.057, *p*=0.032), "Korrektur ABAR" (B=0.055, *p*=0.05), "Felletning" (B=0.054, *p*=0.073) and the interview made by the personnel from the Civil Aviation Authority (B=-0.121, *p*=0.073). The latter variable has a negative weight; however, applicants given a high rating on the basis of this interview are less likely to succeed. Further support for the conclusion that these are the most powerful variables for predicting the training outcome is given by the results of a stepwise regression analysis which selected three of these variables ("Skeppsdestination", "Korrektur ABAR" and the interview by the Civil Aviation Authority personnel). The multiple correlation for this stepwise regression was R = 0.334, R2 adjusted = 0.091, F 3/130 = 5.437, $p < 0.01$).

Discriminant Analysis

The discriminant analysis was performed with the same predictor variables as the regression analysis above. The results from the initial F-tests for these variables agreed with those from the regression analysis (as would be expected) in that significant F-values were obtained for "Skeppsdestination" (F 1/132 = 18.885, $p < 0.01$) and "Korrektur ABAR" (F 1/132 = 4.248, $p < 0.05$). As in the regression analysis, the results for the interview by personnel from the Civil Aviation Authority (F 1/132 = 3.429, $p < 0.07$) and "Felletning" (F 1/132 = 3.586, $p < 0.07$) were close to significance. The discriminant function correctly identified 12 out of the 36 who failed and 89 out of the 98 who succeeded in the training; i.e., 101 (75 percent) applicants were correctly identified. This should be compared with the number to expected if the predictor variables are ignored and only the base rates are considered; i.e., the number of correct classifications that would be expected randomly. This yields an expected rate of correct classifications of 62 percent or 83 students. Thus, the discriminant function improves the selection by 18 cases, compared to no selection procedure at all. Thirty-three students are incorrectly classified, compared to 47 that would be expected on a random basis. This agrees with what would be expected on the basis of the uncorrected multiple correlation between the predictors and the training outcome of R = 0.413. As noted in the discussion of the regression results, the sample estimate represents an overestimation of the possibilities of predicting the outcome; this is true also of the results of the discriminant analysis. Unfortunately, there is no procedure for estimating a discriminant function corrected for sampling errors comparable to the procedure for the multiple correlation. However, in the present case with a binary outcome variable, multiple regression and discriminant analysis are basically the same, and we should therefore expect that after correction for sampling errors, we should have the same decrease in effectiveness; i.e., we should expect that the ability to make correct classifications using the discriminant function should decrease by about 80 percent after sampling errors have been taken into account. A reasonable estimate of the improvement in the number of correct classifications of training outcomes from using the current set of tests and interviews is 3-4 cases (about 3 percent). This is the same estimate as that which we obtained from the regression equation.

Discussion

The Main Task

The main task for this phase of the MRU project was to evaluate current screening and training procedures and create recommendations aiming at a reduction of the present span between intake of students and output of examined new air traffic controllers.

The Analysis

The task force chose to conduct a job analysis based on the critical incident technique. The result shows that 300 reported key behaviours could be catalogued into 5 groups:

- Decision making
- Self confidence
- Information gathering and processing
- Social relations
- Communication

This result verifies the three hypothesis stated about the air traffic controller profession: Behaviours which are related to self confidence are mostly reported in connection with unexpected events and variabilities. The results from the job analysis have been compared to attributions for failure in the basic controller training. It has also been compared to incidents that have occured in actual operations according to the current official report system. Both students and controllers fail to perform the key behaviours at a sufficient level.

The training process requires the students to practise key behaviours from the very first day, aiming at minimizing the number of errors to reach a full performance level, and finally to reach a mastery level. Today an uneven learning curve is the most frequent cause of failure during the basic controller training.

At Stockholm ATS and Arlanda ATS units in Sweden, attempts have been made to improve on current methods of basic training. The results from this attempt to apply modern training techniques (for example to use programmed skill training and to transform the instructor into a mentor and a coach), is now the most promising measure taken to improve the outcome of basic air traffic controller training. To quote one of the members of the task force, professor Berndt Brehmer: "It is astonishing how little effort is made in general to train and develop an operator in a high tech environment by modern training technology, and how much one still relies on an old fashioned on the job training provided by a more experienced fellow-worker."

Present rate and variation in span in the outcome of the Swedish basic air traffic controller training can only partly be explained by inadequate psychological tests and screening procedures. To reduce the uncertainty in the outcome, it is important to improve training and learning of key behaviours for the air traffic controller work, as well as to develop screening methods with high reliability and validity. This will give a prompt and positive result. Efforts must also be made to create a continuing job analysis in order to keep up with a changing technology and maintain screening and training methods with the highest possible effectiveness. An important prerequisite for a successful training result is an efficient selection procedure based on a sophisticated chain consisting of information/introduction/skill tests assessing the substance of the most important groups of key behaviours.

The Relationship Between Test-Results and Training Outcome

The results of the analyses presented above show that it is not possible to predict the outcome of the training on the basis of the variables used in the selection procedure.

One possible reason for these depressing results, and this is true both for the regression results and those from the discriminant analysis, is that these results are based on data only for those who were admitted to the training; that is, we have a classical case of restriction of range. To ascertain the effects of this, we need to look at the standard deviations of the various predictors for the sample used in the calculations. The relevant results are shown in Table 3. As already noted in the discussion of these results, the restriction is not as severe as might have been expected. The standard deviations for the predictor variables are between 71 percent and 91 percent of those for the unselected sample used to determine the stanine scores, and for most of the variables, the standard deviations for our sample are about 80 percent of those in the unselected sample. Moreover, we have the full range of the predictor variables for many of the variables; the lowest values are 1 and 2 for many of the variables. There is therefore little doubt that we would have been able to detect the relations that might exist between the predictor variables and the training outcome. The fact that we find very few significant relations, and that the correlations that we have found are very low, can therefore hardly be explained in terms of restriction of range. Instead, it seems more reasonable to assume that the results express real deficiencies in the selection procedure. That is, the predictor variables are not very powerful predictors of the training outcome. This is hardly surprising in view of the fact that these variables have been selected on the basis of a very general job analysis without real standardisation and statistical evaluation; i.e., the tests have not been chosen on the basis of an empirical evaluation of the actual predictive validity of the tests. This means that there was no reason to expect that the selection procedure would be very effective.

One could, of course, argue that the selection procedure concerns the job as an air traffic controller and not the training. It may well be that the training makes demands that differ from those of the job and that an evaluation of the selection procedure in terms of the training outcomes is not quite relevant. To answer this question, we need a more penetrating analysis of the demands that the training courses actually make compared to those that the job makes. At the present time, we do not have the data required for such an analysis.

Another objection is that the analysis may rely on the wrong model. The present analysis is based on a model where the probability of success in training is assumed to be a monotone function of performance in the selection variables (see Figure 2). That is, this model makes the reasonable assumption that if some ability is required, more of that ability leads to a higher probability of success than less of the ability.

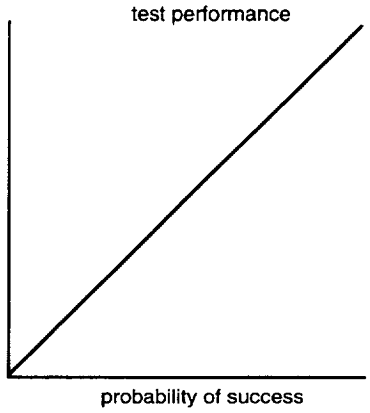


Figure 2. Basic model for the analysis in this study. This model assumes that there is a monotone relation between training outcome and test performance.

An alternative model is illustrated in Figure 3. This alternative model assumes that the training only requires some minimum ability, and that all students having at least this minimum ability will have the same probability of succeeding in the training course.

If this model is true, the possibilities of detecting relations between the training outcome and the selection variables would be limited, especially if the students in the training course had been selected so that all of them had values exceeding the critical value. In the present case, this does not seem to be a very serious problem, however, since the full range of values is represented for most variables in the present sample. Thus, it should have been possible to detect whatever relations might have existed between the test variables and the training outcome, even if model 2, rather than model 1, would have been valid for the present data. Moreover, when measurement error is added, model 2 will generally be impossible to distinguish from model 1.

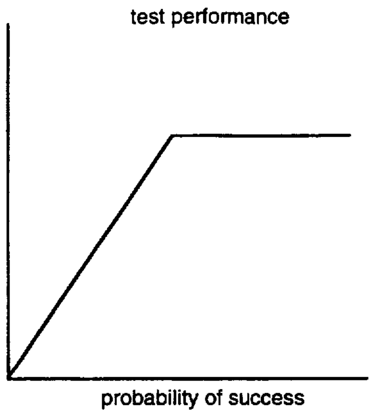


Figure 3. Alternative model. This model assumes that the probability of succeeding in training increases up to some critical values and that it then stays constant at the same level.

The analyses have been based on the individual tests rather than the scales used by ABAR in the actual selection. The reason for that is that our analyses yielded no support for these scales in that we found that these scales were intercorrelated, while the tests included in the scales were not intercorrelated as they should have been. Moreover, a principal components analysis failed to yield the scales as components. Thus, there was little support for the usefulness of these factors. Additionally, regression and discriminant analyses based on the scales used by ABAR did not give better results than the analyses based on the individual predictors.

The results with respect to PG deserves special comment. This is the only variable included in the selection procedure on the basis of an empirical evaluation procedure. In this procedure, the PG test was given to the applicants, but not used for the selection. That is, the evaluation concerned the extent to which this test could improve the selection over above what could be achieved with the original test battery. The results were quite encouraging, but we must now conclude that the conclusions from the original evaluation were overly optimistic. Thus, Svensson and Trygg (1991) concluded that it should be possible to decrease the proportion of students failing the air traffic controller training to less than 10 percent if the PG test was used. As shown in the present analysis, this has not been the case. Even when the PG procedure is used, the proportion of students who fail is 26 percent.

It was, however, not realistic to expect that one would have as good results with the present sample as with the standardisation sample used to determine what PG-variables should be used for the selection. First, the initial evaluation did not take into account the total effectiveness of the selection procedure with PG as one of many selection variables. The value of PG in such a procedure is dependent, not on the correlation between this variable and the training outcome (which is what was reported in Svensson and Trygg, 1991), but upon its unique contribution, which is dependent on the partial correlation between PG and training outcome, after its correlations with other selection variables have been taken into account. That such intercorrelations exist is demonstrated in the present sample, despite the severely restricted variation in the PG scores. Such intercorrelations decrease the weight that the PG results will receive in the final selection.

Second, one must expect a certain shrinkage in correlation when the test is used for a new sample because the values obtained for the first sample capitalised on sampling errors. In the present sample, the unique contribution from PG is far from significant. However, the extremely skewed distribution of PG values makes this correlation suspect.

A possibility of evaluating the effectiveness of the PG procedure is to compare the failure rates before and after the introduction of this test. The relevant comparison here should be with the failure rate for the 1980s, when the mean failure rate was 33.9 percent (with considerable variation among courses). In the 1990s, after the introduction of the PG procedure, the mean failure rate so far has been 26 percent, although this may well be an underestimation because not all students have completed their courses. That is, not all students have yet had a chance to fail. With this in mind, the maximum estimate of the improvement from PG would be 7.9 percentage units, but this would assume that all of the decrease in the failure rate from the 1980s to the 1990s can be attributed to the introduction of the PG procedure. This seems unlikely, especially in view of the fact that the decrease in the failure rate from the 1970s to the 1980s was about 6 percentage units (from a failure rate of 40 percent in the 1970s to a rate of about 34 percent in the 1980s) without any new selection procedures.

In the regression and discriminant analyses, three variables stand out. One of these, the results from the interview conducted by personnel from the Civil Aviation Authority, receive a negative weight. That is, they are systematically wrong: students with a high rating in these interviews perform systematically worse than those with a low rating. This suggests that this procedure must be improved.

Only two of the selection variables have systematic relations with training outcome: "Skeppsdestination" and "Korrektur ABAR". The former of these is supposed to measure flexibility, and the latter is to measure carefulness. The correlations are low, however, and they may well have been produced by chance. Therefore, one should not rely too much on these results until they have proved valid also for other samples.

Conclusions

The present evaluation of the selection procedure is clearly limited, first because it is based on a limited sample, and second because it is based on the results for a group that has been admitted to the training on the basis of the selection procedure that is being evaluated. The restriction of range problem does not seem to be as severe as one might have suspected, however. It should therefore have been possible to detect whatever relations might exist between the selection variables and the training outcome. We must therefore conclude that the fact that it has been hard to find such relations probably means that they do not exist. There are therefore good reasons to reconsider the present selection procedure. It is not possible to decide whether it is possible to design a better selection procedure on the basis of the data we have today. For this, we need a careful analysis of the air traffic controller's job to determine what demands that this job makes, and how these demands can be met by means of selection and training.

Future ATS Systems

Coming automated ATS systems cannot replace the human controller. But manual repetetive work can be eliminated and in that way facilitate information seeking and information collection.

If a new technology or a new system is to be introduced, it is fundamentally important to be assured that the operators accept the new technique, and that the new technique will create opportunities for them to improve their performance. The controllers must also be informed in advance in what way they will be trained to achieve this new standard of performance.

In our view, a continued automation of the air traffic controller's work will only further emphasize the importance of adequate training to execute "new" key behaviours.

References

ABAR. (1978). Förslag till bedömningsvariabler vid psykologisk lämplighetsprövning av flygledarspiranter, våren (Proposal for assessment of variables concerning psychological aptitude tosts for air traffic controller applicants).

Andersson, B. (1993) Sammanställning av enkätsvar som inkommit från flygledare. MRU-projektet (Synthesis of questionnaire answers from operational air traffic controllers). Delrapport nr 4. Luftfartsverket.

Brehmer, B. (1993) Prestation vid urvalet och utbildningsresultat i flygledarutbildning. MRU-projektet (Performance in selection and training results in air traffic controller training). Delrapport nr 7. Psykologiska Institutionen, Uppsala universitet.

Cambell, J. P. (1988). Training design for performance improvement. In J. P. Cambell., R. J. Cambell. *Productivity In Organizations*. Jossey-Bass. USA.

Cullen, J., & Hollingum, J. (1987). *Infr total kvalitet* (Introduction of total quallity). Konsultfölaget AB.

Dolgin, D. L., & Gibb, G. D. (1989). Personality Assessment in Aviator Selection. In Richard S. Jensen (Ed.). *Aviation psychology*. Gower Publishing Company.

Flannagan, J. C. (1954). The critical incident technique. *Psychological Bulletin. Vol 51*. Sid 327- 358.

Haglund, R., Andersson, B., Backman, B., & Sundin, O. (1993). Presentation av bakgrund till MRU-Projektet samt planering för dess genomförande (Work analysis concerning air traffic controllers and air traffic contller students). MRU-projektet. Delrapport nr 1. Luftfartsverket.

Haglund, R., Andersson, B., Backman, B., Sundin, & O. Rapport. (1993). Arbetsanalys flygledare och flygledarelever. MRU-projektet. Delrapport nr 2. Luftfartsverket.

Haglund, R., & Backman, B. (1993). Studiebesök på FAA Huvudkontor och Mike Monroney Aeronautical Center (Study visits at FAA Washington Headquarters and Mike Monroney Aeronautical Center). MRU-projektet. Delrapport nr 3. Luftfartsverket.

Haglund, R., Andersson, B., Backman, B., & Sundin, O. (1993). Resultatredovisning, urval av och utbildning av flygledarelever (Results concerning selection and training procedures for air traffic controller students). MRU-projektet. Delrapport nr 5. Luftfartsverket.

Haglund, R., & Backman, B. Analys av samstämmighet mellan arbetsanalyser flygledare och operativa systemets avvikelserapportering (Analysis of concordance between work analysed of the air traffic controller's profession and data from the air traffic incident and information report system at the ANS Department). MRU-projektet. Delrapport nr 6.

Hastorf, A. H., Schneider, D. J., & Polefka, J. (1970). *Person perception.* Addison Wesley.

Hopkin, V. D. (1989). Cognitive Demands of Automation in Aviation. In Richard S. Jensen ed: *Aviation psychology.* Gower Publishing Company.

Hopkin, V. D. (1980). The measurement of the Air Traffic Controller. *Human Factors.*. 22(5), 547-560.

Josefsson, B. (1992) Information - processing in air-traffic controllers. 3-uppsats. Institutionen för pedagogik och psykologi. Universitetet i Linköping.

Latham, G. P., Wexley, K. N. (1981). *Increasing productivity through performance appraisal.* Addison-Wesley.

Manning, C. A., Kegg, P. S., & Collins, W. E. (1989). Selection and screening programs for air traffic control. In Richard S. Jensen (Ed.). *Aviation psychology.* Gower Publishing Company.

Mattson, J. (1979). *Kognitiva Funktioner vid flygtrafikledning* (Cognitive functions in air traffic control). Lund University Sweden. Department of Psychology. Dissertion Series.

McNemar, Q. (1969). *Psychological Statistics.* John Wiley and sons.

Runyon, R., & Haber, A. (1971). *Fundamentals of behavioral statistics.* Addison-Wesley.

Smith, R. C. (1973). Job attitudes of air traffic controllers: a comparison of three air traffic control specialties. Washington, DC: U.S. Department of Transportation. Federal Aviation Administration.

Sundin, O., Blyckert, J., Hansson, B., & Landerstedt, B. (1993). Utbildning av flygledare på ATS Arlanda och ATS Stockholm (Training of air traffic controller students at ATS Arlanda and ATS Stockholm). MRU-projektet. Delrapport nr 9. Luftfartsverket.

Svensson, B & Trygg L. (1991). Personlighetskarakteristika hos elever i flygledarutbildningen (Personality characteristics of candidates for air traffic controller training). Rapport från Lunds universitet, Institutionen för tillämpad psykologi.

Sörli, N. (1992). Frafall under flygelederutdanning (Loss of students during air traffic controller training). Examensarbete. Universitetet i Oslo 1992.

Wanous, J. (1980). *Organizational Entry. Recruitment, Selection and Socialization of Newcomers.* Addison-Wesley Publishing Company.

Wing, H., & Manning, C. A. (1991). Selection of air traffic controllers: complexity, requirements and public interest. Washington, DC: U.S. Department of Transportation, Federal Aviation Administration.

N95- 34776

Does Human Cognition Allow Human Factors (HF) Certification of Advanced Aircrew Systems?

512-54

Iain S. MacLeod[1] & Robert M. Taylor[2]

[1]Aerosystems International
[2]RAF Institute of Aviation Medicine

Introduction

A system may be defined as a set of parts with the output of the whole greater than the sum of the output from the individual parts. The systems approach is considered to be a formal and systematic set of procedures for systems development. Within the systems approach, systems certification is defined as the result of an applied examination process devised to formally test and affirm that the system being inspected satisfies certain accepted criteria. If certification is achieved by a system, it indicates that the system should fit its intended purpose and that it meets specific requirements of reliability, safety and performance.

Within the definition of a system, it is obvious that a system may contain a human component. However, it should be noted that only recently have there been formal acknowledgements that a system is made up of human and equipment components (e.g. in the military – U.S.A. DoDI 5000.2 February 1991[1]; UK Defence Standard 00-25 dated July 1989[2] , NATO STANAG 3994AI dated 1990-91[3]).

Human Factors (HF) certification should be an integral part of systems certification. HF certification implies that HF specification, testing and evaluation have a secure foundation; therefore, the process of certification follows as a matter of carefully progressed HF appraisal of the system. The final HF certification tests should be the culmination of a planned process of certification allowing orderly and conditional HF certification to be progressed throughout the duration of the project. HF certification is ultimately concerned with how efficiently the human

[1] DoDI 5000.2 defines a total system as including the humans that will operate and maintain the equipment.

[2] Defence Standard 00-25 states that " This Standard should be viewed as a permissive guideline, rather than a mandatory piece of technological law". A *System* is defined as:

> 'A purposeful organisation of equipment (hardware and software), personnel and procedures all of which interact and thus influence each other to produce some specific result or goal.'

[3] STANAG 3994AI. This STANAG does not define a system though its list of definitions strongly implies that the human is an integral part of an 'advanced aircrew system'.

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PAGE 162 INTENTIONALLY BLANK

PRECEDING PAGE BLANK NOT FILMED

element(s) of a system can perform through their use of the system, and how human performance affects that system's performance capabilities and the safe achievement of system related goals.

Specification and certification of engineered systems can be conducted under any of several well documented and accepted methods. Further, HF analysis and measurement has attracted a great deal of attention since WWII and can also be reliably performed[4]. However, the human role in complex human-machine systems is recognised as becoming predominately one of supervision, understanding of problems, judgement, choice and decision making. Thus, the main emphasis of the contribution of the human to human-machine systems has changed in nature from physical to cognitive.[5]

If the human contribution is largely cognitive, the HF specification and certification of human complex interactions through complex systems should require a sound knowledge of human cognition. Applied knowledge of cognitive processes to the required quality does not currently exist in the realms of HF, engineering or psychology. Therefore, with such a gap in knowledge, any HF certification of complex or advanced aircrew systems must be carefully qualified.

Cognition and Operation of Advanced Aircraft Systems

Conceptual issues

Engineering ideas are about as far removed conceptually from the ideas of human psychology as any ideas can be. This is because engineering ideas are mechanistic/physicalistic and based in the natural sciences[6], whereas psychology can be termed the 'Science of Mental Life, both of its phenomena and its conditions'[7]· Disciplines such as HF, Cognitive Psychology and Industrial Psychology attempt to bridge the conceptual gap. In addition, disciplines such as Engineering Psychology and Cognitive Engineering are attempting to adopt more teleological approaches to the appreciation of human work.[8] However, with the rapid advances in computer based systems and automation, and their burgeoning complexity, it appears that the conceptual gap still remains wide. To give examples from diverse HF viewpoints over the last decade, the following four quotes show a general HF/Psychological concern on engineering approaches to system design and automation, approaches that are biased to only considering observable manifestations of human behaviour.

[4] A good description of the methods generally used by 'behavioural specialists', including HF engineers, is contained in Meister (1985).

[5] Cognition is normally taken to refer to the part processes of the mind involved in human knowing such as reasoned thought, understanding and judgement. In contrast, mental or psychological processes are more global and cover the total remit of mind functions and processes including such as emotion and long term memory.

[6] *Physicalistic* taken as pertaining to the physical, mechanistic or the observable. A great deal of the tenets of HF and Psychology are founded on physicalistic approaches i.e. anthropometrics, biomechanics, human manual control of systems, many approaches to HCI design, behaviourism to name but a few.

[7] Definition from James (1890).

[8] *Teleology* - as applied to the human refers to human mental goal seeking and purposeful behaviour. For an interesting and teleological associated discourse on paradigm shifts in science see Ackoff (1972).

i. ...Physicalistic descriptions can only capture those aspects of man which submit to the metaphor of the machine, and must fail to account for the rest. This inadequacy of the physicalistic approach becomes gradually more clear, as the complexity of man-machine systems increases. (Hollnagel, 1983)

ii. ...we are still making the same seemingly contradictory statement: a human being is a poor monitor, but that is what he or she ought to be doing. (Wiener, 1985)

iii. ...the designer who tries to eliminate the operator still leaves the operator to do the tasks which the designer cannot think how to automate. (Bainbridge, 1987)

iv. Is it possible that our advanced command and control systems will require cognitive human performance that defies our ability to measure and predict? ... What none of the existing models are much good at is analysis of cognitive behaviour. (Miles, 1993)[9]

Problems in the Appreciation of Unobservable 'Mental Life'

The greater the complexity of a human-machine system, the greater the problems in efficiently integrating the human component into the system. With a complex system, the human may have difficulties in maintaining a concept of system performance and fitting that concept to the human role within the system. Such difficulties may not only exacerbate problems that the human finds in system control or supervision but may also encourage the human to enter incorrect or inappropriate inputs into the system. Therefore, the human performance at the human-machine interface (HMI) of a complex system must be assisted in an attempt to ensure that situational awareness[10] is sustained, the human is aided in the obviation of human system related 'errors',[11] is helped to skillfully maintain a necessary defined role within the system, and is neither overworked nor bored.

There are innumerable HF standards and guidelines on how to define and certify simple systems. These standards and guidelines are normally advisory and invariably stress the physical aspects of systems and/or the use of empirical evidence. They mimic the form of system specification in that components of a system are specified by their physical or manifest functions and the logical interrelationships of these functions.

However, the human catalogue of skills transcends the physical,[12] especially when the human has to cope with complexity and uncertainty. Cognition is hidden and may be either abstract or have manifestations in observable human activity. Thus, the processes of human

[9] Miles, J.L. (1993). *TASK Analysis - Foundation for Modern Technology* presented at MRC Workshop on Task Analysis, University of Warwick, U.K.

[10] *Situational awareness* refers to an understanding of all the factors affecting mission performance, including the status of the aircraft and its mission system, and the tactical, spatial and the geographical environment external to the aircraft. See Taylor (1993). An older, higher level concept used by UK aviators was termed 'Airmanship'.

[11] Human system related errors could be operator, maintainer or designer based or be based on a combination of all the 3. For consideration on error forms see (Rasmussen (1986), Reason (1990)).

[12] As an example, Welford (1976) identified three types of human skill as:

Perceptual - The skills that code and interpret incoming sensory information;

Motor - The skills associated with skilled movement but controlled by perception and cognition;

Intellectual or Cognitive - skills considered by Welford to be the most important as they link perception and action through decision processes.

judgement are also hidden. For example, the human assessment and judgement of the quality of equipment related information might be a continual background task with no directly associated manifest actions on the part of the equipment observer. However, if the observer's judgement leads to choice, and the choice requires action, an observable human activity will result. Therefore, physicalistic system functions may or may not have an equivalence within the cognitive functions of the human component of the system.

Moreover, physical systems are constructed by logical rules of engineering while human logic is dependent on gleaned knowledge, human mental functions and heuristics that are based on aeons of human evolution. Training and experience can tune human abilities into skills with respect to the human role within a system. Training cannot mould a human into a metaphor of an engineered system component. Thus, the overall system must be built to consider the possible contributions of human and machine, to allow one to complement and appreciate the capabilities and system inputs of the other.

Traditional functionality. Traditional systems engineering stresses that the concept of design must be based upon a detailed understanding of the functionality of the system.

> Functional analysis requires specifying function in the abstract. The very fact that a design is undertaken presumes an engineering concept and in turn a fairly limited range of engineering solutions. (Price, 1988)

Thus, the transposition of engineering functions into required equipment performance is brought about by design based on a choice from a limited range of solutions. However, the manifestation of performance by the human depends on the individual human's innate mental abilities, developed physical and cognitive skills, the existing level of fatigue, and personal and organisational mores and ambitions, to name but a few influences. Underlying that human performance, human cognitive functions rely on human mental processes and are related to human progress towards goals[13] through the use of tools. They may or may not be associated to the parallel performance of certain equipment functions within an engineered system as suggested by the traditional approach. They may however be part of system processes that encompass both engineered and human system functions. Two examples are given as indicators of the problems inherent in using traditional physicalistic approaches to the conception of human machine systems:

Example One. An aircraft is flying from one airfield to another. During the flight a system subset might be performing an automatic navigation calculation to update aircraft positional information in parallel to a pilot's radio communication to inquire on the weather at the destination airfield. The equipment will be working to a fixed schedule whereas the pilot inquiry might be prompted by an observation that present position weather is different from that forecast or by his detection of a similar inquiry from another aircraft with respect to an airfield adjacent to the destination.

The two activities are related by the overall purpose of the flight but are not necessarily performed in parallel. Indeed, the subject communication activity might not take place on every flight. The automatic equipment calculation is deterministic whereas the pilot's communication mainly depends on the vagaries of the weather and the performance of the pilot.

[13] *Goal* can be defined as the end result towards which a mental or physical effort is directed. Goal can also be descibed as an objective towards which the individual consciously or unconsciously strives (Adler (1929)).

The pilot's human function allows him to decide on the form of his work with relation to prevailing circumstances regardless of his assigned system function. However, the machine is unaware of the flying environment apart from that environment's influences on the navigation calculations. Here a system related problem may arise from both the differences and possible divergences in the short term goals of the human and machine. Traditionally the human inputs information and direction to the aircraft systems to bridge the differences.

From this example there are unlikely to be repercussions critical to the safe completion of the flight. However, it is only a *simple* example. With a complex human-machine system the question should be raised on how much and when man/machine compatibility in roles is to be ensured, both from the machine and human standpoints.

Example Two. Task analysis is a form of predictive analysis used for the consideration of human tasks with relation to the operation of a system towards system related goals. However, this predictive modelling and analyses is usually based on engineering related functionality and is thus biased towards the observable aspects of human performance ignoring such as judgement, understanding and choice (two examples of such analyses are Operational Sequence Diagrams (OSDs) and Hierarchical Task Analysis (HTA)).

However, task analysis is concerned with the analysis of human tasks. Human tasks require the human to apply both cognitive and physical effort. This effort is needed to direct a system towards the achievement of preconceived goals both tactical and strategic (Tactical and strategic goals will be discussed in more detail in Section 4 below). The problem aired here is that not all human tasks can be considered and analysed if only mechanistic based or observable tasks are considered.

Traditional approach to HMI. The study of HMI, this encompassing Human Computer Interfaces (HCI), is also supposed to consider the amalgam of human and machine system components. However, what an HMI normally shows is an interface design tuned to foreseen needs for the human to equate to the engineered functionality of the equipment. Only recently have there been any signs of a consideration under certain applications to the cognitive needs and abilities of the human operator/maintainer (e.g. Macintosh and Windows WIMP in 'Desktop Metaphore'). However, the concept still appears one way, that the human has to appreciate the machine.

There are many HMI and HCI design paradigms in existence but they will not be further considered here. As an aside, prototyping is meant to be an exercise where an HMI or HCI can be demonstrated and tuned to obtain the optimum interface between the machine and the human for a particular application. In reality, HMI prototyping with advanced aircrew systems is frequently used only an exercise of demonstration and not as an analysis of the man machine system performance capability allowed by the HMI.

Consolidation. To reiterate, the human may be considered a complementary part of an engineered system, but not as a piece of equipment that can be easily specified. The traditional partitioning of system functionality to allocate functions to humans and machine so that each performed the most appropriate (machines are better at/human are better at)[14] *assumed that physicalistic system functionality could be directly transposed to either human or machine.* In defence of the traditional approach, developments on the theme considered the complementary

[14] Originating in the work by Colonel Fitts.

nature of human and machine in a system and acknowledged that some functions could be performed by either with equal efficiency.

Indeed, the simple physicalistic transposition of functions to human or machine may have held true in situations where human used machines as tools to be directly applied to work performed under immediate human attention. However, the subject transposition is much less likely to be true where human work through complex systems towards mission goals[15], where direct allocation of numerous system functions to the humans or the machine becomes more difficult to determine at anything but the highest level of consideration and, finally, when human work is based more on human cognitive performance than on psychomotor performance.

There appears to be a gradual realisation, especially in the U.S.A., that the standards and guidelines produced in the early 1980s are set in the physicalistic engineering mores of the 1970s. In the 1970s and early 1980s systems were less complex and the human was generally closely involved in operating directly *with* systems to achieve goals (often in a one to one relationship as in the use of a computer based word processor application) rather than the current push to operate *through* systems to meet goals (as with an airborne mission system where the tactical performance of the aircraft is directed by the operator through an HCI updated from advanced navigation, communication and sensor equipment).

As already suggested, one of the problems with the drive to automation with airborne systems is that the human has been forced, in many instances, out of the primary role of a system operator and into the primary role of system supervisor, this without the development of tools to assist in the performance of the new role (or even a determination that the new role is suited). Many studies have shown that automation may have decreased the occurrence of certain error forms but has introduced new categories of man-machine system error that have yet to be fully understood (Weiner op cit, Woods & Roth (1988)).

Advanced systems are being designed forgetting the underlying tenet of systems design – that the whole is greater than the sum of the parts. The human and machine components of a system must complement and assist each other within the system. It is not fruitful to enhance the speed in which a system can operate if the quality of system support to the human decision processes is impaired.

Thus, it follows that an underlying contention of this paper is that the new forms of complex system operating problems and errors tend to be cognitive rather than psychomotor based. Also that new forms of errors should be considered to be mainly dependent on the achieved efficacy of human machine system design rather than as mainly resident with the human operator of the system.

Problems Inherent in HF Specification with Traditional Systems Design

The problems inherent in HF specification with traditional systems design will be illustrated using examples from two well known standards.

Traditional design by Def Stan 00-25. Systems design and development rely on a traditional series of analyses through system planning and preliminary design to detailed design and development. The initial system requirement analysis is usually conducted by the customer and considers such needs as system purpose, sphere of operations, types of system components,

[15] For a discussion of goals in aircraft missions see Taylor (1993).

system reliability to name but a few. This initial requirement is stated at a high level. The system requirements analysis is the basis of the specification that initiates the system process.

Figure 1 shows the UK Def Stan 00-25 model of HF activities conducted during system design. In Figure 1, it can be seen that it is presumed that an allocation of system functions can be performed in the traditional manner. Of interest, the particular UK concept of Task Synthesis entails:

> ...the design team, using their judgement and expertise, proposing a combination or sequence of tasks appropriate to the function. (Def Stan 00-25, Part 12, p14)

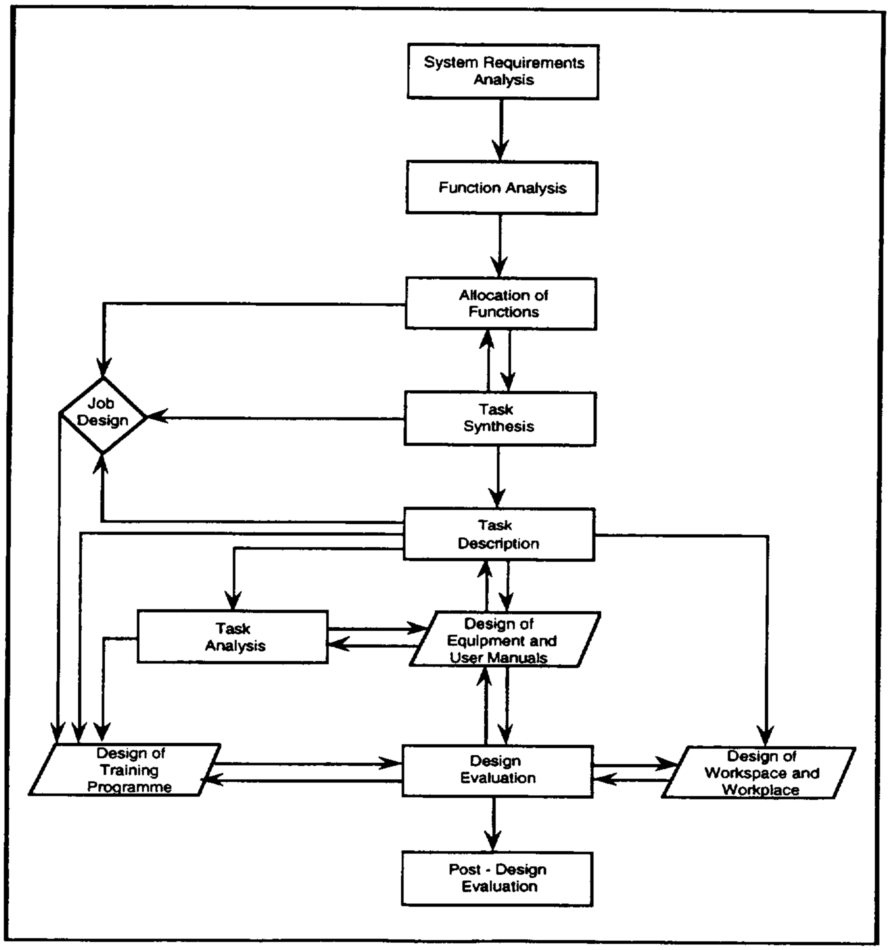


Figure 1. Human Factor Activities Conducted During System Design

Note that the previous forms of analyses on which the Task Synthesis is based are wholly physicalistic and conceive that:

> "The major system requirements are physical...but there are always explicit behavioral requirements." (op. cit. p11)

Design by STANAG 3994. In STANAG 3994AI the problems inherent in straight physicalistic function allocation to the human are recognised in that a 'Potential Operator Capability Analysis' is mooted alongside analyses of human 'Decision', 'Error', 'Information Requirements' and 'Control Requirements'. Indeed, the Task Analysis mooted explicitly covers human cognition as it states that the analyses:

> ...shall show the sequential and simultaneous manual and cognitive activities of the operators/maintainers, and include those aspects of their tasks which involve planning and maintaining situational awareness, as well as decision making and control activities. (p4)

However, the task analysis is to be based on preceding analysis. The emphasis of the preceding analyses is still seen as placing an over reliance on physicalistic functionality. From the descriptions of the preceding analyses, there appears to be an underlying assumption that a form of mapping can be made from the systems functionality of 'Advanced Aircrew Systems' onto human functionality within the system and the associated human cognitive processes. To give an example of difficulty in such mapping, the STANAG example of Function[16] (e.g. control air-vehicle) is decomposed under 'Function Analysis' through –

> ...successive levels of detail to a point where individual functions can be unambiguously identified, prior to allocation to human, hardware, or software system components.

It has already been argued above that 'Function' is an engineering concept within systems engineering and that human and engineering functions differ. Therefore, from that conceptual base it can be conceived that human mental facets such as human understanding, judgement and choice cannot be easily mapped onto a system function such as 'control air-vehicle' This is true especially as many complex system control functions must be hidden to the operator and much of human cognitive processes must be governed by factors such as previous human training, experience, the effects of the flying environment and immutable human heuristics.[17]

Nevertheless, though the STANAG concept of 'Function Allocation' still mainly relies on the assumption that physicalistic functions can be mapped into human functions, it is strongly influenced by "...the review of potential operator capabilities" (p3). It is suggested that, in reality, some of the 'human' human-machine functions requiring consideration might emanate *solely* from the 'Potential Operator Capability Analysis' area, especially through a review of operator tasks in similar systems. Importantly, it needs to be recognised that some essential

[16] *Function* - A broad category of activity performed by a system, usually expressed as a verb + noun phrase e.g. control air-vehicle, update way-point.

[17] Some basic fundamental characteristics of human cognition that appear to be common to all humans. See Tversky, A. and Kahneman, D. (1974).

human-machine system functions can be purely cognitive, albeit open to influence from human understanding on the significance of information available from the pertinent man machine system or other sources.

Figure 2 below gives an indication of the initial analyses required by the STANAG 3994.

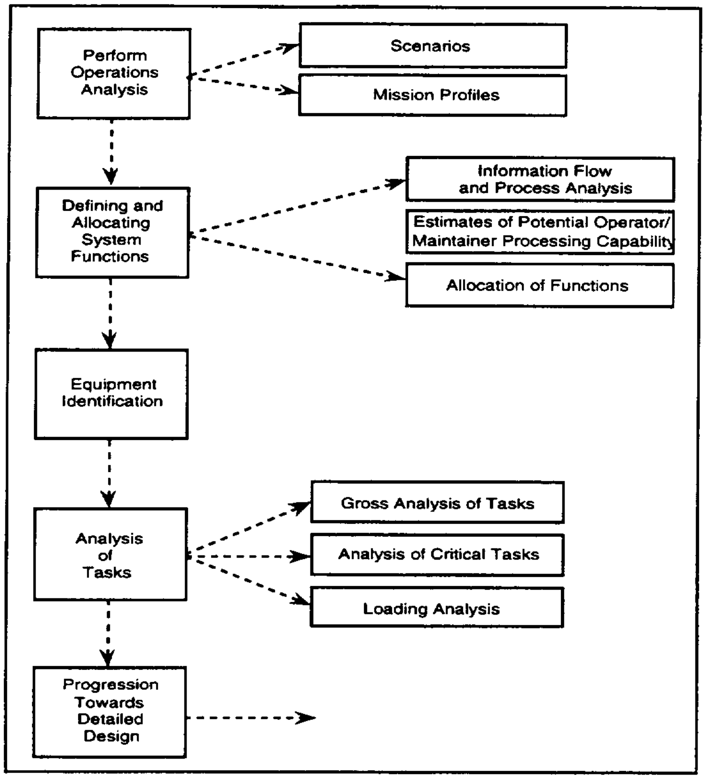

Perform Operations Analysis
Scenarios
Mission Profiles
Defining and Allocating System Functions
Information Flow and Process Analysis
Estimates of Potential Operator/ Maintainer Processing Capability
Allocation of Functions
Equipment Identification
Analysis of Tasks
Gross Analysis of Tasks
Analysis of Critical Tasks
Loading Analysis
Progression Towards Detailed Design

Figure 2. STANAG 3994 General Model for Early Human Engineering Programme

From the considerations above, an improved high level definition of a human-machine system can be presented:

> A man machine system is a complex system that works towards the achievement of specified goals using the dynamic application of the diverse capabilities of that system, assisted by a cognitively directed human and equipment effort, to create an expected system performance.

Fully automated systems beyond the realms of human ken[18] are not considered by this paper. The design of complex systems require that an adequate degree of human work related understanding, judgement and choice exists if optimum human decision processes are to be considered at a system working level. Indeed, effective human-machine systems design

[18] *Ken* - an old Scottish word meaning understanding within current knowledge or sight.

requires a recognition of both optimum and sub optimum human work processes in order to appreciate and conceive safe and efficient system operation throughout the range of possible human abilities and skills.

To reiterate,

> *It needs to be recognised that some essential human-machine system functions can be purely cognitive.*

Thus, it is necessary that better consideration is given to the important role of human cognition within human-machine systems during the specification of advanced aircrew systems. Before attempting to further advance such a consideration, it is first sensible to have a more detailed appraisal of some of the extant components of the system design process, and methods of their classification, starting with system functionality.[19]

System functionality. The specification of system functionality delineates the span of system capability and is one of the fundamentals of the traditional system approach to design. Functionality is based on a knowledge of intended system purpose, usage, the technology available, previous like systems (if any), and the level of human behavioral involvement with system operation.

Regardless, there must be a method of classifying the importance of functions. One method is to classify functions as either as *Necessary* or *Accessory*[20]. Whilst this classification method should be useful regardless of the nature of the functionality examined, traditionally, systems design only considers tangible physical[21] functions.

Of interest, the Necessary/Accessory functional classification may appear to map conveniently onto the standard UK MoD specification of system features as 'Essential' or 'Desirable'. However, "Essential/Desirable" are indications of the MoD priority on requirements and may or may not be associated with considerations on the criticality of the feature with respect to the achievement of mission success.

It has already been argued that the physicalistic functionality required by a system cannot be simply or easily mapped across to the functionality pertinent to the operation of the human component of the system. For advanced and complex aircraft systems, it is important that a method be devised of classifying functionality from several standpoints: the physicalistic or equipment standpoint; the standpoint of the human cognitive component; the standpoint of amalgamated equipment and cognitive system components, etc.

However, one improved approach to functionality classification verges on the recognition of the prime importance of human knowledge and cognition was mooted by Price (1985) in his

[19] It has been argued to this point that the abstract concept of functionality used for systems design has a physicalistic, engineering or empirical bias. This is historic in origin and is still the usual concept of functionality (See Ackoff op. cit.). However, functionality can be categorised in many alternative ways including 'Material or Informational', 'Necessary or Accessory' (Price (1985)). It is argued here that for an advanced or complex system it can also be categorised as 'Cognitive, Equipment or a combination of both'.

[20] *Necessary* functions are functions that are deemed to be essential to allow a system to successfully meet its goal(s). Absence or failure of a necessary function will result in a failure to meet the system goal(s). *Accessory* functions provide system redundancy, allow alternative paths to task completion or add capabilities that enhance the system. The failure of an accessory function is not critical to the successful performance of a system.

[21] It could be argued that if anything is tangible it can be considered under material functionality. However, this is not the case in the consideration of material functionality as it attends to things both tangible and physical with relation to a system. For example 'fear' is sometimes described by individuals as tangible, may be related to a system operation but is not physically part of a system.

classification of *Material* and *Informational* functionality. Material functionality is seen as purely physicalistic and refers to traditional system engineered or equipment functionality. Depending on the intent of the particular study this functionality can be generic or specific to an adopted equipment.

Informational functionality. Informational functionality concerns information that is associated with the system usage of physicalistic functionality. Thus, informational functionality is closely associated with material functionality. As an example, an informational function might be 'plan drop weapon' and has a direct material equivalent of 'drop weapon'.

Nevertheless, there may not necessarily be an obvious direct link between informational and material function. For example, the informational function of 'consider weather effects on Radar' has no direct material equivalent as humans cannot as yet dictate the weather and the material Radar functionality that could be related is diverse (e.g. adjust Radar picture, adjust Radar scanner tilt, switch off Radar, inform crew of Radar effects, etc.). Any departure from the basic 'verb+noun phrase' defining informational functionality leads to difficulties in determining the equivalent material functionality. Therefore, it should be questioned as to whether this direct matching is of any use to determining the total functionality needed for an advanced or complex aircraft system.

The traditional answer to the preceding question would be to the affirmative as the high level function 'plan drop weapon' might be decomposed into several sub functions such as 'plan select weapon', 'plan open bomb doors', 'determine time to warn crew of attack' etc. Again, however, the traditional approach would be to only consider the informational functionality directly associated with a material function or functions of 'drop weapon'.

Consideration of more complex functions highlights some of the application problems with the 'Material/Informational' classification. More complex functions such as 'choose best weapon and attack tactic' are obviously an amalgam of material, informational and cognitive functions with the cognitive being paramount. Such functions can also suggest processes at work: the threads, information flows and controls that tie functions together and give them meaning within system operation and performance.

However, a function such as 'question evidence' is essentially cognitive and may be prompted by human knowledge and experience rather than the by the physical evidence presented by a system. This latter function cannot be described as informational as there is no obvious association with a material function. Thus, the traditional informational approach must be questioned as only in the case of simple material functionality (such as 'lower seat height') is there likely to be near direct associations of informational functionality.

Still considering the example of 'drop weapon', in *reality* the associated human-machine related functionality is likely to be vast and could involve such as a cognitive association of information considering aircraft performance (height, speed and attitude), an assessment of target performance (speed, aspect height, manoeuvre), a recollection of given 'Rules of Engagement', an awareness of positions of 'friendly' forces that may be at risk, an appreciation of aircraft stores remaining, to name but a few possibilities. Therefore, it is argued that the above attempt to break the physicalistic description of functionality into its material and associated informational components is still engineering associated and would require a redefinition to allow it to fully consider human cognition within a human-machine system.

Some attempts have been made through Knowledge Engineering to encapsulate human cognitive functions within the materialistic functionality of advanced aircrew computer based systems. It is beyond the remit of this paper to consider whether Knowledge Engineering can successfully capture complex human cognitive based expertise, and then usefully incorporate that expertise as system related functionality within a dynamic and advanced aircrew system. Nevertheless, to be truly successful any such attempts must explicitly recognise human expertise and the importance of human cognitive functions to the operation of a man machine system.

Summary of the assessment of functionality. It has already been argued that much of the physicalistic functionality associated with a complex system has no obviously associated human equivalent as the processes of cognition are hidden and may result in no visible human action or system input. Further, in aircraft systems there are important sustainers of human situational awareness that are not specific to system design but that are important to system performance. Such sustainers may be forgotten to the detriment of aircraft operation if an advanced aviation system design was to be purely based on the physicalistic or material approach. These include, but are not limited to, environmental/system associations such as arise from changes to such as ambient noise, vibration, or visual sightings. Indeed, sometimes the meaningful indicator is an absence of system derived information when the information has been determined by other means and should be present. Finally:

> Much is yet to be done, especially in analysing human cognitive requirements in working with automated machines and in putting a methodology into effect that will bring humans and machines systematically together to do those things that each can do best and that they can accomplish jointly to improve system performance. (Price, 1985)

Performance. Performance is the manifest result of the work undertaken by a system. A system is designed to achieve a particular quality/level of performance and in reality achieves a standard of performance that is seldom truly equivalent to that aimed for by the design. Predicted system standard of performance is determined by the designed amalgam of desired system functionality and, sometimes, a planned capability of the human component within the system. In reality, the achieved standard of performance is mediated by the environment, the achievable performance of the designed system, system reliability and actual human standard of performance as allowed through the use of the system and by influences external to the system. The human ultimately directs advanced aircrew system performance. System performance cannot be fully addressed unless the functionality and expected performance of the human component of the system is considered, this alongside equipment performance during the process of system analysis and design.

Traditions of Enforced Compromise. Traditionally, the performance of complex airborne systems has been inferior to that predicted, a compromise has been accepted and the system has only reached the desired standards after an introduction of system enhancements introduced some time after the introduction of the system into operation. Extrapolation of past practices suggests that the more complex the system the longer it will take to bring an inferior system performance up to the level of expected or acceptable performance. In the military there are several advanced systems where the latter point has been borne out (they are known but should not be aired too much in public).

The solutions applied to the introduction of systems that are obviously inferior to the requirement appear to be selected from one or more of the following:

1) Compromise and accept the system as better than previous systems. ('do what you are told route').

2) Compromise and accept the system limitations until time/finances can be found to improve the system. ('traditional route often involving an expected mid-life update even before the initial system delivery').
3) Compromise and increase the number of personnel operating the system. ('throw manpower at the problem – the serfdom route').
4) Compromise by expanding the training programme and improving the quality/experience of personnel employed with the system (expensive, but can be blamed on the quality of the personnel available in the past, or past mistakes in recruitment or on the 'needed' complexity of the system).

Methods outside the traditional compromises.

5) Reappraise the design method. Ensure that the next system is designed using the 'lessons learned' (a form of reappraisal of design method should always happen after any system design).
6) Immediately cancel the production of the system (a final and shameful resort with blame attributed where and when possible and probably an obfuscation of any lessons to be learned).
7) Develop and apply suitable methods for specifying and certifying the design and build of advanced aircrew systems (this depends on a realisation that current methodologies are inadequate and the existence of a will to improve).

Of the mooted seven solutions given above, the first four compromises have been in existence for decades, the fifth solution is what is needed now and in the future to prevent the recurrence of expensive failures and the sixth has certainly been evoked recently (several programmes in the U.S.A. and UK spring to mind). The seventh is obviously the ideal. Of course, certain necessary compromises or 'trade-offs' should be inherent in design and should not be decried.

Design

Comprehensive and efficient design is the key to the achievement of required performance from complex systems but depends on the standard of system specification. Design is the bridge between system specification and the achieved system performance. This paper has argued that complex system design have traditionally considered the human in the physicalistic/ mechanicalistic sense.

Until recently, the problems associated with the parallel needs of promoting human understanding alongside system operation and direction were generally equated by the natural flexibility and adaptability of human skills. However, the information rates and complexity of modern systems often place system processes beyond the supervisory or manipulative capabilities of the human, because human cognitive attributes and performance have not been properly considered within the design of the system. Where human cognition has been considered it is normally only where the concept of human cognition has parallels into the physicalistic mould of determining material functionality.

The next section will consider the actual specification produced for the RN Merlin helicopter and discuss some of the lessons learned from that specification's consideration on system functionality.

Functionality specification for the RN Merlin helicopter

The emphasis of this paper is on the consideration of system functionality and its influence on the processes of HF certification. The overall specification process for the Merlin helicopter is presented in greater detail in the paper of Taylor & MacLeod (1993).

The definition of Merlin functionality constituted one of the three parts created for the Merlin specification exercise. The functionality document was termed the Functional Requirements Definition (FRD).

Functional Requirements Definition of Merlin Specification

The (FRD) considered system functionality for the Merlin, as

> Functionality is not solely derived from the definition of the requirements for the individual systems and their interaction, ...those systems interact with the crew and systems outside the remit of Merlin, and the operational environment in which they are to operate. (MPC Specification 1990)

Primary and secondary objectives were used to consider functionality.

Primary Objective – To define the minimum acceptable functionality for the Merlin. This involved specifying the functionality for existing and new systems. The functionality considered here was material functionality. Thus, considering the existing equipment:

> ...a set of 'Major Functions' were identified. To allow the current documentation set to be as effective as possible, the Major Functions are chosen to be approximately equivalent to the systems fitted to the current helicopter [EH101] and are therefore not intended to be a 'pure' functional breakdown from operational requirements. (MPC 1990)

Secondary Objective – System management functions fundamental to the successful integration of all systems on board Merlin. System management is defined as:

> The usage of the Merlin's System through the tools devised from the amalgam of Human Engineering and other engineering approaches to the system design. (MPC 1990)

The management areas were considered as Flight, Tactical and Maintenance. The three management functions, and specific Sensor functions, define the parameters that are to be displayed to the crew and the controls necessary to influence the operation of the Merlin. *However, the interdependency between the management areas was not addressed in any detail.* Flight and Tactical management were split into the following subsets

Mission Management
Those functions necessary to permit procedures and equipment to be employed that assist the crew in conducting the tasks required of Merlin.

o

Information Management
Those functions necessary for the collation and processing of data to determine a future flight path or for the collation and processing of a tactical picture.

Human-Machine Interface Management
Those functions necessary for display of the tactical picture and system/equipment status, and crew interaction with the tactical picture.

Sensor Management
Those functions necessary to control the operation of the various related sensor functions in a consistent manner.

It can be seen that under this form of functionality classification there was an effort to consider the functions that the crew would have to perform to manage the specified material functionality of the Merlin within, but not between, each of the three defined management areas. However, there is a potential pitfall with the focus on the human component of the system as a manager. Firstly, team work does not depend solely on good management but on a myriad of influences. Further, by definition, managers supervise and administer the resources available to them. *A manager does not necessarily lead but directs resources within certain specified rules.* In contrast, a leader does not necessarily need to manage, but guides others using such as foresight and tactics as well as resources. Indeed, the leader's used resources may be beyond those immediately available and may not necessarily be closely governed by rules.

In the Merlin case, the rules applied to the appreciation of system management failed to consider the system as a whole and thus restricted the scope of the given *Secondary Objective.* It is a matter of conjecture as to how efficiently the system might support leadership. However, the lack of a whole system appreciation on system management must raise a question on the efficacy of the aid that the resulting system will offer to the leader or to the system manager.

To summarise on the management of real time complex systems. Firstly, a human is basically a poor supervisor. If the human is not involved in the operation of the system, the human's attention and reactions to system cues are liable to suffer. Moreover, if the human functionality within a system is not fully specified it will be difficult to properly managed within the designed remit of the system. At the worst, the human resource is then managed in a procrustean manner to fit into the machine design and may become involved in an incessant combat with the machine in order to achieve system goals.

However, particularly with consideration on information management, there was some consideration on the human cognition requirements for the management of the Merlin systems. Examples are:

> The symbology used to represent information displayed to the crew shall be developed in accordance with human engineering principles given in chapter...to take account of the cabin and cockpit environment and human *understanding*.
>
> Classification also depends on human *understanding* of presented data. Aspects of this process shall be evaluated as part of the Human Engineering Programme Plan. (MPC 1990)

Nevertheless, the main source of management functions did reside in the material functionality of the aircraft equipment. This was partly because the definition of the 'Major Functions' was mainly determined by the existence of equipment already adopted for the helicopter during earlier development. It was also caused by a lack of in-depth consideration on extant maritime tactics or on future possible tactics for the aircraft. The particular reasons for this will not be discussed here. However, the result is that the requirements of possible tactical performance cannot be fully equated with the existing equipment functionality.

Thus the problems associated with a necessary task related judgements and direction through human cognition at the human machine system interface could not be fully addressed. For this reason, much of the functionality that could have been ascribed to human mental properties was instead placed for consideration within the responsibility of the Human Engineering Programme Plan (HEPP), a plan which had to be constructed under the mandate of the already discussed STANAG 3994.

HEPP and Design Requirements

The application of Human Engineering to the Merlin system is governed by a mandated and agreed HEPP. Human Engineering placed emphasis on the human component of the system and introduced a planned approach to this aspect of the Merlin's design, in order that important facets were recognised and addressed.

Because of the already designed equipment, the HEPP and other design requirements were mandated too "late in the day" to be as effective as they might have been if they had been in existence from the onset of the system analysis and design process. Moreover, the Merlin specification was *not created* under the full remit of the STANAG 3994, though the STANAG was mandated by the completed specification. Therefore, some of the STANAG's recommended system analyses were not fully considered in the specification including 'Mission Analysis' and 'Potential Operator Capability Analysis'.

However, for the new aircraft equipment, and hopefully for consideration of the integration of human performance within systems operations, the HEPP serves as a valuable aid for indicating areas of the system where improvements may have to be made or new procedures devised. Of course, before the system can be accepted into RN service it must pass a formal Operational Performance and Acceptance phase (OPAS) without a need for too much compromise. The HEPP is one of many ways of assessing the risk of successfully completing the OPAS.

Moreover, the HEPP is based on the FRD and Design Requirements. The problems with the method of the FRD specification of requirements have already been discussed. Therefore, with some system analyses already conducted, the biases of the analyses are bound to be reflected in the performance of the HEPP. As the HEPP is also concerned with HF acceptance of all forms of HF analysis, tests and trials, any HF certification of the Merlin system should bear in mind the initial problems of specification.

Discussion

The RN Merlin was respecified as a system many years after the onset of the initial design process for the helicopter. However, the respecification process allowed the remaining aircraft development process to be defined both with relation to the expected aircraft system performance and the possible risks associated with meeting the OPAS performance requirements. Throughout the HEPP, the HF input to design and the aircraft certification process has been stated with appropriate qualifications. This makes clear the process and requirements of HF certification.

The important HF lessons learnt from the specification work were:

- To be fully effective, a HEPP should be produced and started early in the systems analysis process .
- It is obvious that more care has to be taken in the consideration of human cognition with respect to the design of a man machine system, especially where cognition may have specialised functions within the human-machine system.
- The omissions in the specification with relation to human cognition must represent a source of unspecified risk within the process of HF Certification.

Through the above arguments, it can be reiterated that appreciation of human cognition is an important facet of complex man machine specification and, ultimately, certification. A possible method for the consideration of cognition during system analysis and specification will be outlined next. In part, the method is a development on areas of STANAG 3994.

The Incorporation of Human Tactics and Strategies

Introduction

A tactic is defined as an arrangement or plan formed to achieve some short term goal. The goal may be an end in itself or serve as a stage in the progress towards a later objective. A strategy governs the use of tactics for the fulfilment of an overall or long term plan. Tactics can be formal written procedures or reside in human mental processes. Normally the human tactic selectively directs the formal system related procedural tactic.

Human tactics and strategies are not only physicalistic, they are mental.[22]. A tactic is procedural, may be mainly skill based and is flexible and adaptable to a changing environment. In a human-machine system, the performance of tactics and strategies is enhanced by system equipment designed to aid the human to interpret information contained in the working environment and, also, to survive in that environment. Strategies allow the human to be selective in the use of tactics, to choose the most effective or most expedient for the fulfilment of the foreseen plan.

The human perceives the world through information gleaned from the senses. This perception can be achieved through direct observation of the world or through the use of a

[22] Explanation of the human usage of a mental model of the world have been given by many researchers from various viewpoints and considering many possible constituent parts (cognitive maps, schemas, frames, scripts, goals, plans and schemes to name but a few). The use of 'tactic' and 'strategy' in this paper is for the sake of explanation and is not intended to supplant what has gone before but to aid in the current exposition. The terms are used here as they conveniently afford a mirror of cognitive activity onto related aircrew operating procedures (which can be broken down to tactics and strategies in the militaristic sense). Reference is given to Adler (1929), Bartlett (1932), Craik (1943), Schank et al (1975), Neisser (1976) and Card et al (1983) to name but a few.

human-machine system's interface with the world. The perceived information is interpreted through the use of knowledge, rule and skill based mental and cognitive processes[23] that may vary between humans. The use of the interpreted information is then governed by human tactics and strategies. These tactics and strategies are tuned through training and experience and are governed by the human role within a human machine system and the human's interpretation of that role. The results of human tactics and strategies are manifest in observable human performance and skills.

Therefore, tactics and strategies are continually mediated by both the information that the operator already possesses and information gleaned from the working environment. In the performance of work, the former information is purely human in origin and influences the latter whilst the latter is often human-machine in origin and can eventually contribute to human work experience. All human-machine systems perform under the directed influence of the human component's tactics and strategies.

Difficulties in Concept and Application

One of the main difficulties of concept with human tactics and strategies is how to translate the abstract into something concrete akin to a system function, and then in a form through which it can be applied to human-machine system analysis. The first stage is to make the abstract tangible with respect to stages of job performance.

STANAG 3994 mentions the use of '...a review of tasks in similar systems'. To be effective, such a review requires an in-depth examination of tasks, possibly using Subject Matter Experts (SMEs) and knowledge elicitation techniques[24]. The matters to be examined include:

- Common problem areas requiring cognitive effort; i.e. the interpretation of sensor data, the determination of rates of change of data, the understanding of particular types of information.
- The association of the common problem areas with jobs (i.e. tactics and strategies) or forms of task.
- Any timing data that may be available; i.e. with this form of problem and that form of task, why does the operator take a certain time to gather evidence and to resolve the problem?
- Any evidence that can be collected on how to ameliorate the operator's cognitive effort, if deemed to be excessive.

The two areas of operator applied cognition that should be investigated by the examination are:

i. The application of cognitive processes to the performance of a task or sub task.
ii. The application of cognitive processes to progress tactics or strategies associated with a group of tasks or parts of a mission.

[23] Rasmussen's SRK theory, the consideration of Skills, Rules and Knowledge based behaviour as determinants of human performance. See Rasmussen (1986).

[24] Knowledge elicitation techniques can be considered under 2 categories; Direct and Indirect methods. Direct methods are used where an expert can be asked directly to indicate their knowledge. Indirect methods are used to infer the experts knowledge from the experts performance at other or similar tasks to those on question. Direct methods include: Interviews, protocol analysis, Kelly grid, concept sorting. Indirect methods include '20 Questions', concept recall and listing. Further reading includes Kidd (1987).

An Avenue Towards a New Method of Systems Design

The efficacy of the common forms of the Fitts list must be questioned. If it is accepted that some essential human-machine system functions can be purely cognitive, then it is necessary to develop a method to identify these cognitive functions and their implications within the system design. Some considerations on an avenue to such a method are outlined in the remainder of this paragraph.

Examination of similar systems. The initial action is to look at what systems have been used before for similar aircraft roles. This does not necessarily involve the examination of near identical jobs to those envisaged for the new system, though such an examination is preferable. It may be that there are no near identical jobs and that equivalent jobs will have to be used. For example, the operation of a collision avoidance Radar screen might give indications of problems that may be encountered in the operation of a weather avoidance Radar.

From the initial action, a system (or systems) is chosen for further examination. Obviously, the greater the number of systems examined the better within the constraints of time and budget.

The next stage is to use knowledge elicitation techniques, as appropriate and with the cooperation of suitable SMEs, to determine where the SMEs assess the greatest cognitive loads reside. To assist this process, the concepts of human tactics and strategies should be explained as well as the equipment and tasks being considered.[25] It is important that a series of tasks is considered in order that both tactics and strategies can be properly addressed. If possible, the questioning of an SME during the operation of an aircraft system or simulator is preferable. If actual equipment of some form is not available, some form of task analytic simulation might suffice provided the SME is well acquainted with the form of task representation used. It must be accepted that there is no current method of ensuring that all system critical cognitive processes may be examined.

A novel method of assessing human problems and capabilities associated with flight is allowed by the 'MicroPat' tool as developed by Bartram et al at Hull University (bought by some UK military agencies and Cathy Pacific Airlines). This tool was designed to perform psychometric assessment of aircrew candidates using dynamic simulations of the standard functions of aircrew systems. It is suggested that this tool, or another of the same form, could be used alongside knowledge elicitation techniques as a means of determining and assessing broad categories of aircrew cognitive functions associated with generic equipment functions of aircrew systems (i.e. combination of the human usage of artificial horizon, altimeter and heading reference[26]).

It is suggested that the elicitation of knowledge in the task related area will be easier than the elicitation of knowledge on tactics and strategies; a task normally being performed under operator focused attention whereas an operator's consideration on a tactic or strategy may not necessarily be continuous and may be resident in non declarative memory. Considering verbal protocols as an example of a task elicitation technique, concurrent and retrospective protocols may both be suitable to gain a fair indication of a task related use of cognition. However, to

[25] The SMEs use of tactics and strategies, both mental and in the material / militaristic sense, was approached in Macleod et al (1993) during the process of the predictive examination of workload for the RN Merlin. This study also included an examination on the effects of operator errors and decision processes on system tactical performance.

[26] An example is a tool produced to assess candidate aircrew's ability to perform mental navigation calculations whilst flying a simple computer based aircraft simulator. See Bartram (1988).

elicitate information not in immediate memory (i.e. strategies) might require some form of prompting or the use of several knowledge elicitation methods.

The examination of the verbal protocols must be based on a model and some decided categories of cognition. The question on whether verbal reports are or are not epiphenomenal is a subject of continued debate that will not be addressed here.[27] The concern here must be the benefit that any method brings to the final application compared to the benefits possible from other methods. Considering the inadequacies of the traditional approach, it is suggested that even a classification of human-machine system tasks as requiring associated operator cognitive processes of a 'High/Medium/ Low' nature is better than no consideration at all.

Once an early understanding is obtained on the operators use of cognition or cognitive functionality, task related use of cognitive functionality should be considered alongside the chosen system's equipment functions embedded in that task. Where any association of cognitive and equipment functions is not possible, but the equipment function is understood, each form of function should be specifically labelled as a *task related equipment function* e.g. lower undercarriage. Where a task related function is deemed to be purely cognitive it should be labelled as a *task related cognitive function* (e.g. assess visibility). Where task related equipment and cognitive functions must be associated, their association should be labelled as a specific *task related associated function (e.*g. determine position on glide-slope of airfield approach). Any cognitive functionality that cannot be related to a task, but to a series of tasks or a tactic or a precess, should also be considered as a *specific system cognitive function* and labelled (e.g consider tactics to be applied to surveillance of manoeuvering target).

Appreciation within the new system design. The next step should be to apply the data on cognitive processes obtained from the examination of similar systems as outlined above. The difficulties in determining and incorporation values of cognition into the system life-cycle design processes are considerable. To strive for the synergy necessary for a complex human-machine system, Cognitive Task Analysis techniques will be applicable here.

The method of incorporation human cognition into system design requires a detailed system functional and process analysis process, probably using a form of dynamic modelling technique. This technique to examine and combine the required and refined functionality of the new system with that obtained from the examination of similar systems.

The knowledge gleaned in this fashion could then be checked and further refined through the use of flight/air mission prototypes, mockups and simulators. Of course, the ideal scenario would be to continue the refinement process using data collected from actual aircraft equipment and flight prototype trials.

Whatever method is used, a careful consideration of cognition will require an iterative process in the early stages of the system analysis process. This iterative examination is seen as essential to consider and amalgamate the information gleaned from old systems (i.e. cognitive tactics and manifest operating procedures) with the detailed functionality and expected performance of new equipment.

Valid basics for the understanding of cognitive functionality can only be determined through practice in investigation and application. What is eventually required for the system designer is a set of 'rules of thumb' through which the subject of human cognitive functionality can be effectively approached within the realms of system design as a whole.

It is easy to pay lip service to theory and say that equipment should be built to appreciate the human and the human trained to appreciate the machine. The basis of such mutual appreciation

[27] For detailed coverage of the debate see Ericsson & Simon (1984) and Nisbett & Wilson (1977).

must be a better practical understanding of human cognition as applicable to advanced aircrew systems. However, such an understanding will involve a great deal of experimentation, preferably in the field rather than the laboratory, and an education of system designers to convince them that such an effort is necessary.

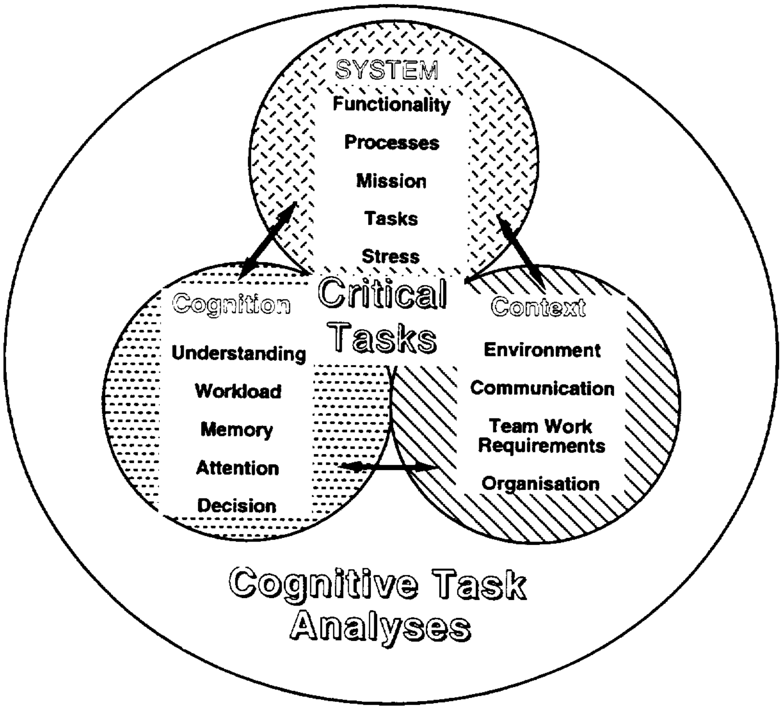


Figure 3. The Remit of Cognitive Task Analysis

It is suggested that a careful appreciation of cognition within the specification of the functionality for a new system should allow improvements in the following:

- The determination of Necessary and Accessory Functionality.
- The initial assessment of the numbers and quality of personnel required to operate the system.
- The assessment during system design of the form of operator training required.
- The production and progression of a HEPP for the new system.
- The design of the system HMI / HCI.
- The efficiency of trade-off process during system design.
- The creation of formal operating or tactical procedures.
- The assessment of achievable system performance and the risk inherent in the design.

Summary and Conclusion

This paper has examined the requirements of HF specification and certification within advanced or complex aircrew systems. It suggests reasons for current inadequacies in the use of HF in the design process, giving some examples in support, and suggest an avenue towards the improvement of the HF certification process. The importance of human cognition to the operation and performance of advanced aircrew systems has been stressed. Many of the shortfalls of advanced aircrew systems must be attributed to over automated designs that show little consideration on either the mental limits or the cognitive capabilities of the human system component.

Traditional approaches to system design and HF certification are set within an over physicalistic foundation. Also, traditionally it was assumed that physicalistic system functions could be attributed to either the human or the machine on a one to one basis. Moreover, any problems associated with the parallel needs, of promoting human understanding alongside system operation and direction, were generally equated in reality by the natural flexibility and adaptability of human skills.

The consideration of the human component of a complex system is seen as being primarily based on manifestations of human behaviour to the almost total exclusion of any appreciation of unobservable human mental and cognitive processes. The argument of this paper is that the considered functionality of any complex human-machine system must contain functions that are purely human and purely cognitive. Figure 3 indicates the place of Cognitive Task Analysis as an aid to this process. Human-machine system reliability ultimately depends on human reliability and dependability and, therefore, on the form and frequency of cognitive processes that have to be conducted to support system performance. The greater the demand placed by an advanced aircraft system on the human component's basic knowledge processes or cognition, rather than on their skill, the more insiduous the effects the human may have on that system.

This paper discussed one example of an attempt to devise an improved method of specification and certification with relation to the advanced aircrew system, that of the RN Merlin helicopter. The method is realised to have limitations in practice, these mainly associated with the late production of the system specification in relation to the system development process.

The need for a careful appreciation of the capabilities and support needs of human cognition within the design process of a complex man machine system has been argued, especially with relation to the concept of system functionality. Unlike the physicalistic Fitts list, a new classification of system functionality is proposed, namely:

- *Equipment.* System equipment related.
- *Cognitive.* Human cognition related.
- *Associated.* Necessary combination of equipment and cognitive.

This paper has not proposed a method for a fuller consideration of cognition within systems design, but has suggested the need for such a method and indicated an avenue towards its development. Finally, the HF certification of advanced aircrew systems is seen as only being possible in a qualified sense until the important functions of human cognition are considered within the system design process.

(This paper contains the opinions of its authors and does not necessarily reflect the standpoint of their respective organisations).

References

Ackoff, R.L. (1972) *Science in the systems age: Beyond IE, OR and MS.* Address to the Operations Research Society of America, Atlantic City, New Jersey.

Adler, A. (1929) *Individual Psychology.* London.

Bainbridge, L. (1987), Ironies of automation. In: *New Technology and Human Error,* Rasmussen, J., Duncan, K. & Leplat, J (Eds). John Wiley & Sons.

Bartlett, F.C., (1932) *Remembering: A study in experimental and social psychology.* Allen & Unwin.

Bartram, D. (1988). *Development of computerised test of aircrew navigation ability,* paper presented to the British Psychological Society Occupational Psychology Conference, University of Manchester.

Card, S.K., Moran, T.P. & Newell, A. (1983) *The psychology of human computer interaction* Lawrence Erlbaum Associates.

Craik, K.J.W. (1943). *The nature of explanation.* Cambridge.

Department of Defence. (1991). *Defence Acquisition Management Policies and Procedures.* Department of Defence Instruction 5000.2. Washington, DC: Office of the Secretary of Defence.

Ericsson, K.A. & Simon, H.A. (1984) *Protocol analysis. Verbal reports as data,* The MIT Press.

Hollnagel, E. (1983), What we do not know about man-machine systems. *International Journal of Man-Machine Studies,* 18, pps 135-143.

James, W (1890). *Principles of Psychology.* New York.

Kidd, A.L. (Ed) (1987). *Knowledge acquisition for expert systems: A practical handbook,* Plenum Press.

MacLeod, I.S., Biggin, K., Romans, J. & Kirby, K. (1993) Predictive Workload Analysis – RN EH101 Helicopter, *Contemporary Ergonomics.* Taylor and Francis.

Meister, D. (1985) *Behavioural analysis and measurement methods.* John Wiley and Sons.

Miles, J.L. (1993). *TASK Analysis – Foundation for Modern Technology.* Presented at MRC Workshop on Task Analysis, University of Warwick,, G.B.

NATO Standardisation Agreement (STANAG) – STANAG 3994AI First Draft. *The Application of Human Engineering to Advanced Aircrew Systems..*

Neisser, U. (1976) *Cognition and reality: principles and implications of cognitive psychology,* San Francisco.

Nisbett, R.E. & Wilson, T.D. (1977) Telling more than we can know: Verbal reports on mental processes, *Psychological Review,* 84, pps 231-259.

Price, H.E. (1985). The allocation of functions in systems. *Human Factors,* 27(1), pps 33-45.

Price, H.E. & Pulliam, R. (1988). Functional analysis and allocation of functions. In: Helander, M. (Ed) *Handbook of Human-Computer Interaction.* Elsevier, North-Holland, pp641.

Rasmussen, J. (1986) *Information Processing and Human Machine Interaction* Elsevier, New York.

Reason, J.T. (1990) *Human Error,* Cambridge University Press.

Schank, R.C. & Abelson, R.P. Scripts, plans and knowledge, *Proceedings of the fourth international joint conference on artificial intelligence.* Tbilisi.

Taylor, R.M. (1993) *Human Factors of Mission Planning Systems: Theory and Concepts*, NATO AGARD AVP Lecture Series 192, (In press).

Taylor, R.M. & MacLeod, I.S. (1993), *Quality Assurance and Risk management. Perspectives on Human Factors Certification of Advanced Aviation Systems*, Proceedings of the Workshop on Human Factors Certification of Advanced Aviation Technologies, Toulouse, France, 19-23 July (In Press).

Tversky, A. & Kahneman, D. (1974), Judgement under uncertainty: heuristics and biases, *Science,* 1 85, pps 1124 -1131.

UK Directorate of EH101, MoD. (PE) *Merlin Prime Contract*, MPC Specification Number: D/EH101/2/90, 1990.

UK Defence Standard 00-25. (1989), *Human Factors for Designers of Equipment.* Part 12: Systems, Issue 1, dated 15 July 1989.

Welford, A.T. (1976). *Skilled Performance: Perceptual and Motor Skills,* Scott Foresman

Wiener, E.L. (1985). Beyond the sterile cockpit, *Human Factors*, 27(1), pp 87.

Woods, D.D. & Roth, E.M. (1988). Cognitive Systems Engineering. In: Helander, M. (Ed) *Handbook of Human-Computer Interaction*. Elsevier, North-Holland, pp 16.

Parallel Views and Topics

Certification: Ballistic Blessing or Continuous Process?

Thomas E. Bernard

University of South Florida

The science and art of occupational health and safety has evolved rapidly during the twentieth century through the safety movement and worker compensation laws, to the point that it is widely recognized and accepted. Human factors is also a discipline dedicated to reducing adverse risk to life, limb and property that has grown in the last fifty years. My intention in this position paper is to view human factors certification of advanced aviation technology with the experience and baggage of an occupational health and safety professional.

In order to even begin, let's at least accept certification as a formalized method to reduce the risk to life, limb and property. To announce my bias in advance, I believe the industry and the public is best served by review and assessment. This is not achieved by design and performance standards, but by education and interdisciplinary approaches (e.g., ISO 9000). In response to workshop authority, I will briefly mention some of the interrogatories and their role.

Who

"Who" is either active or passive. In the active sense, it is the question, "who will certify?" The obvious answer is "we will be happy to certify your system." The "we" is you and me. But logically, "who" is the manufacturer, the customer, or some third party (independent agent).

In the passive sense, the question is "who will be certified?" For occupational health and safety professionals, certification is for the practitioners. The industrial hygienist and the safety professional are certified by their respective boards after the completion of appropriate education and experience and by passing an examination. A similar certification process for ergonomists/human factors professionals in the United States is in its infancy.

Within the aviation industry, plenty of people are "certified" in some fashion: from pilots to mechanics to air traffic controllers. While it may meet opposition for other aspects of operations or design, certification of people in the design process is not a wholly new idea.

My only fear about the certification of people is the degeneration of quality of personnel. Certification is generally believed to be a floor for the ability to practice, but it soon becomes the requirement for practice. When it becomes the requirement, it does nothing to promote superior practice and perhaps limits the superior person who is uncertified.

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PRECEDING PAGE BLANK NOT FILMED

PAGE 188 INTENTIONALLY BLANK

What, When, Where

What: Advanced aviation technology

When: At appropriate times

Where: Convenient locations

Why

The implicit given of the workshop is that human factors certification is an essential good and should be encouraged. (And I sometimes encourage a belief in God, baseball and Chevrolet. Note that each of these seems to drift in and out of favor.) Taking the given, then, dispenses of the interrogatory "why".

But just to carry the thought of "why" a little further, "why should we bother?" Unlike the history of occupational health and safety, the need is not driven by a long line of unacceptable risk and attendant consequences. I offer as evidence the fact that the public is not outraged (as it has been about the social costs of occupational disease, dismemberment and death). (Outrage is a construct proposed by Peter Sandman to describe the effects of perceived risk, as opposed to real risk.) Remember that "airplanes are not routinely dropping out of the sky" and the rates of mortality and morbidity do not approach that of motor vehicles. In summary, the current practice in the design of aviation technologies has served us well. The remaining answer to "why" is the belief that the technology is going to grow faster than our current practices will allow us to control the risk.

Another facet of the "why" deals with "why do I need human factors?" I view this as a marketing activity. The goal is to better market human factors as a discipline that is able to contribute significantly to risk reduction. It is this marketing element that health and safety professionals have over human factors professionals. We can point to federal standards for worker protection and cry wolf. This, however, is a trivial approach and it has limited benefits to the promotion of health and safety. The truly successful have demonstrated the value of their services and needs for investment through both cost-avoidance and improved productivity and quality. This latter approach establishes a longer-term, more positive relationship with decision-makers (those with the power and money).

It is clear that there is a kernel of interest in human factors in the aviation-related industries and organizations by the number and breadth of organizations represented at the workshop and in the literature. It is necessary to grow this kernel into a flourishing plant by demonstrating effectiveness at every opportunity until there is a wide-spread implementation of human factors.

On the other hand, manipulation is available. If manipulation is the preferred approach over the reasoned approach above, then it is worthwhile to explore the role of the public. Often public outrage is required to move an organization (most of which are characterized by substantial inertia). The commercial aviation system of the western world is quite safe, and the public as a whole recognizes this. To create outrage, the public must come to believe that the risk is much greater than it really is. Traditionally, this is accomplished through disaster. It is also caused by disproportionate fear. For example, the chemical and nuclear industries are quite capable of generating public outrage in the presence of insignificant risk. Using imagination, the human factors community can create a market through public outrage. (Some caution will be needed, and this is not a personal avocation.)

Certification vs. How

If we accept the why of certification, the next workshop implication is that certification is a dependent variable such that:

certification = function of who, what, when, where, how.

I would argue that the dependent variable should be an outcome variable and not an independent variable that contributes to risk reduction. The outcomes result in safer aviation as a whole. On a less grand level, the outcome of interest is reduced risk from some system, bounded by some flexible delineation. Taking the independent variable down one more layer perhaps can bring us to "how do we reduce risk?" (This is the case of a Debons knowledge word instead of information. In fact, most people probably look at certification as information and not as knowledge.) The model then becomes:

How = function of who, what, when, where, certification.

How

How is often thought of as being formulated in standards. Many of the statutory and consensus safety standards that govern general industry dictate the design. The desirability of design standards is that industry knows exactly what to do and auditing agencies know exactly what to look for. The problem is that it stifles the imagination; alternative approaches to the same goal are rarely considered. There are times, of course, when design standards are important. They are particularly useful to establish communications protocols. But they should be advocated in the extreme and not as the routine.

The first alternative to the design standard in health and safety is the performance standard. In essence, this specifies the maximum acceptable level of a hazard, and discretion is left to the individual plants on methods of control. Now, innovation is encouraged and may even be rewarded by increased profitability.

I equate certification with standards (either design or performance). Both require an explicit goal and some yardstick for measurement. The outcome, however, is invariably the just acceptable to meet the goal. In health and safety, a problem occurs when the standard changes with new knowledge. In most cases, the "just acceptable" becomes "unacceptable."

A new way of approaching worker health and safety is beginning to emerge from a number of directions. These include wellness programs, new management styles (e.g., Total Quality

Management, Product Liability Programs, ISO 9000), and some government health standards (e.g., process safety and cumulative trauma). The fundamental features of these are early recognition, prospective analysis and interdisciplinary teams. Because the promotion of reduced risk is a universal concern, even with advanced aviation technologies it makes sense to involve as many disciplines as possible.

The inclusion process then is one of bringing in human factors as well as risk management, law, accounting, etc. And the process is ongoing; it does not start with a system concept and stop with customer acceptance. It tracks field experience and brings the lessons into new concepts.

Conclusion

If certification marks the end-point or is limited to a design or performance standard, it will become a kind of ballistic blessing (with the system out-of-control once it is "certified"). To assure the continued protection of life, limb and property in the commercial aviation industry, I propose a program of continuous review and assessment. In this regard, certification can either be defined as "how" this program is accomplished, or it is an audit to indicate the program is active and effective.

N95- 34777

Practical Guidelines For Workload Assessment

Andrew J. Tattersall

University of Wales, Cardiff

Introduction

There are numerous factors that need to be taken into account in a comprehensive human factors certification process. This paper deals with one major factor which is a central area for human factors concern when humans are required to interact with complex systems, namely workload. If the process of human factors certification of systems is to succeed, then workload assessment must be incorporated into the evaluation and certification process. Proper and effective evaluation will ensure that the workload experienced by users of any system, in aviation or otherwise, is taken account of in system design and development.

There is now a vast amount of literature on workload assessment. This interest has been stimulated primarily because of the need to design complex task environments that do not place undue demands and requirements on the human operator. The principal applications have typically been in process control and aviation settings such as air traffic control and aircraft cockpit design. This paper will address a number of practical issues that need to be taken into account when any evaluation of existing systems or future systems is carried out.

There is little dispute that workload is a multidimensional concept (Damos, 1991; Gopher & Donchin, 1986; Moray, 1982; O'Donnell & Eggemeier, 1986), but one distinction that is not very often made explicit is between acute and chronic dimensions of the impact of workload. If workload is defined in terms of the costs that operators incur in performing tasks (Kramer, 1991), then the distinction is even more apposite. An understanding of this distinction will, to a certain extent, aid evaluations of the nature of workload in work settings.

First, if one is concerned with the acute effects of workload, then the main focus of the research will be on the interference between tasks in dual- or multiple-task situations. A principal question to be asked is whether the tasks are too demanding in terms of the human information processing requirements, in which case performance on one or more of the primary work tasks may be degraded. The logic of many of the approaches in this area is based upon multiple-resource theory (e.g. Wickens, 1984, 1991), in which it is proposed that there are a variety of processing resources that are limited in their capacity. The extent to which tasks will interfere with each other when carried out concurrently will depend upon the extent to which they compete for common resources. Evaluation of workload in this case may include primary and secondary task performance measures, subjective measures and certain psychophysiological measures. The ultimate intention of such approaches is to predict performance in multiple-task situations.

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

Second, if one is interested in chronic symptoms of heavy workload then the main concerns are with the effects of managing the demands of work over a day, a week or a prolonged period of time. The after-effects of work are an important consideration. In other words, to what extent is the physiological and emotional state of an individual affected over a period of time, and does subsequent performance at work show any decrement due to these changes in state? For example, individuals may become increasingly fatigued because of the sustained demands in the job that need to be managed effectively. This may result in breakdown of skills over the longer-term, and current models of workload are unable to predict these outcomes with any degree of certainty. Typical assessment techniques might again involve the use of performance measures, subjective measures and physiological measures of workload. In addition, subjective and physiological measures of individual state are likely to be useful as there may be implications of the longer-term effects of workload for well-being, both emotional and physiological, health, performance and safety at work. Cross-sectional studies of occupational stress and health provide a useful way to gain background information about possible sources of stress, and the coping strategies and behavioural styles which might be generally effective in moderating the effects of work demands (e.g. Farmer, Belyavin, Berry, Tattersall & Hockey, 1990). However, it is also important to investigate the effects of exposure to different job demands in the longer-term. The development of a model of stress and workload based on observations of individual patterns of response to various demands will lead to the more accurate prediction of states or situations in which a breakdown in skills might occur.

Questions That Should Be Asked

Before any evaluation is carried out, a number of questions need to be addressed that will enable the appropriate measures to be taken for the particular problem that is to be tackled. An approach that has been used to great effect in assessing the usability of human-computer systems (Ravden & Johnson, 1989) is to provide users or designers of systems with a checklist of items to consider. The questions below are based on part of that checklist, and they may be useful for workload assessors to identify the key areas of concern about workload in the initial stages of an evaluation.

1. What is the general question? Is the concern with the overall system or a specific piece of equipment, and is the focus primarily on performance or on the health and safety of the workforce?

- Is the primary concern with task design (the scheduling or allocation of tasks within or between individual operators' jobs)?

- Is the primary concern with equipment design (for example, to evaluate the effects of the introduction and integration of a new item of technology, or to evaluate the difficulties of working with one item of equipment)?

- Is the primary concern with health and safety (the outcomes of working within an existing or proposed system perhaps as a result of sustained task demands, underload or boredom, etc.)?

2. Is there a specific problem? That is, do operators complain about specific tasks or functions?

- What are the best aspects of the system?
- What are the worst aspects of the system?
- What parts of the system give the most difficulty when carrying out the task?
- Are there parts of the system that are confusing or difficult to understand?
- What are the most common mistakes made in using the system?
- What performance problems exist with the system?
- What changes do operators suggest might be made to the system to make it more effective and usable?

3. What is the aim of measuring workload?

These questions will help to define the problem and the goals to be set for the workload assessment exercise. By focusing on these issues, the task of choosing a set of workload assessment techniques and specifying the research environment and design should be more straightforward and the results of studies easier to interpret and act upon.

Factors To Consider In Choosing A Particular Technique

Many different measures of workload have been developed, but their effective use in particular situations will depend on various factors, including their sensitivity to changes in demand, their ability to distinguish different kinds of demand, and their suitability or relevance to that situation. Extensive reviews of these techniques have been produced before, many of which discuss criteria for application (Gopher & Donchin, 1986; O'Donnell & Eggemeier, 1986; Hancock & Meshkati, 1988). The main factors to consider are as follows.

Validity. Moray (1988) has argued in a recent review of the development of mental workload research that, because no clear, precise definition of workload exists, it is difficult to establish the validity of different techniques. He suggests that the reliability of measures has to be sufficient for practical purposes until such a definition is agreed upon.

Reliability. Measures should be accurate and provide similar values from different operators doing the same task. There should be a good correlation between the values produced by different techniques if they are used to assess the same dimensions of workload. They should also have test-retest reliability, although as yet few reliability studies have been carried out.

Sensitivity. The concern here is with the effectiveness of the technique to discriminate between different levels of primary task load. In other words, does the technique actually measure changes in task demands and identify conditions of extreme workload?

Diagnosticity. Diagnosticity refers to the extent to which the technique is able to distinguish between different types of task demands and to identify the particular components within complex tasks that result in difficulty. Some techniques provide a general measure of resource allocation or effort, whereas others, such as secondary-task methodology, may be more sensitive to variations in different domains of processing. They may, for example, distinguish between verbal processing requirements and spatial processing requirements of tasks. Primary task measures of performance provide a global measure and are not really suitable for this purpose. Subjective measures, unless used to assess different task components, systems or functions for example, are generally not terribly diagnostic. The need for diagnosticity really depends upon the aims of the study. If the aim is to assess the introduction of a new piece of equipment or change in working procedure, then diagnosticity may not be critically important. If, on the other hand, there is a need to assess the demands of different control actions (e.g. manual or spoken) in relation to different modes of information presentation (e.g. visual or auditory), then the diagnosticity of the technique will be an important factor.

Intrusiveness. This refers to the extent to which the workload assessment technique disrupts the performance of the primary work task. The disruption could result from the use of obtrusive equipment or the application of a technique, or, in the case of secondary-task methodology in particular, simply the requirement to perform a concurrent task (Ogden, Levine & Eisner, 1979). If safety is a major concern (in air traffic control, for example), then clearly workload assessment techniques that may degrade performance should not be used. Simulation exercises may provide useful data should techniques which may be intrusive also provide other useful attributes.

Generality, Acceptability and Applicability. On a pragmatic level, it may be useful to choose a technique that can be used in different situations, perhaps so that comparisons can be made between different conditions. Ease of use and special requirements that may restrict the application of a technique, such as the need for special, expensive equipment, can be important considerations, as can the extent to which operators are accepting of the particular technique.

Workload Assessment Techniques

The main concern here is with what is termed 'mental workload' rather than physical workload, for which there are reliable and established assessment techniques available (e.g., Rohmert, 1987). The concept of mental workload has proved to be more difficult to define and measure, which causes some concern when the focus is on the tasks of air traffic controllers and pilots as these tasks primarily involve cognitive processes rather than place great physical demands on personnel. It is not easy to estimate the demands of these tasks and therefore to predict the consequences of different levels of demand. We not only need valid, reliable and sensitive measures of workload, but also good methods based in sound theory to analyse the cognitive activity that is required to perform these types of complex tasks.

A further important factor to consider is that different types of measurement technique may relate to different dimensions of workload and therefore may provide a different perspective of the particular demands of the task. Indeed, some subjective rating techniques are designed to assess a number of different dimensions, such as time pressure, frustration or anxiety, and mental effort. Certain physiological measures may be most sensitive to one particular dimension, for example, heart rate variability may be more likely to be associated with changes in effort, whereas mean heart rate may reflect changes in anxiety or physical effort.

Furthermore, there needs to be consistency in the way that terminology is applied, and finally, the operational procedures should be standardised as much as possible.

Self Report Measures

Subjective measures are relatively easy to employ, and asking workers to rate the levels of demand they experience and their state of well-being and health at least has face validity. It is intuitively attractive simply to ask controllers or other workers about the levels of demands they are experiencing and the impact of work demands at different levels. A number of subjective workload assessment techniques have been developed. Among the validated scales that are widely used in aviation settings are the NASA Task Load Index (TLX) (Hart & Staveland, 1988), and the Subjective Workload Assessment Technique (SWAT) (Reid & Nygren, 1988). They both assess perceived workload on a number of dimensions, usually after the task has been performed. Nygren (1991) has suggested that they are both useful measures of workload and that SWAT is sensitive at both individual and group levels. Hill, Iavecchia, Byers, Bittner, Zaklad & Christ (1992) compared SWAT, TLX, the Modified Cooper-Harper scale (Wierwille & Casali, 1983), and the Overall Workload scale (Vidulich & Tsang, 1987) for sensitivity, operator acceptance, response requirements, and any special procedures they require. All were found acceptable and sensitive to different levels of workload. Nygren (1991) points out, however, that the psychometric properties of these scales need to be fully understood, in addition to their implications for task performance, before they are applied extensively.

A recent development in this area has been the attempt to design subjective techniques which provide ratings of workload during primary task performance rather than after carrying out tasks. One such example is the Instantaneous Self Assessment technique (ISA), which was initially designed for use with air traffic control tasks (Hulbert, 1989; Jordan, 1992). Few evaluation studies have been carried out, but it appears to be a relatively sensitive measure of workload (Tattersall & Foord, 1993). A similar technique was used by Rehmann, Stein and Rosenberg (1983), who suggested that gaining concurrent workload evaluations was more accurate than post-task ratings. In complex tasks which involve multiple elements or phases, the ratings may be more clearly related to changing task demands than retrospective ratings. Tattersall and Foord (1993) found, however, that ISA responses, although correlated with other subjective workload measures, interfered to a certain extent with the primary tracking task, whether responses were made by speech or manually.

One problem with subjective measures has been their diagnosticity and to a certain extent their reliability, whereas their validity and sensitivity have been fairly well established. A further worry is that they are not always found to correlate with measures of performance. Tasks that are performed better are sometimes found to have higher ratings of workload (e.g. Yeh & Wickens, 1988).

Self report scales of a different type have been developed to assess mood (e.g., Mackay, Cox, Burrows & Lazzerini, 1978; Warr, 1989), and longer-term health. Importantly, significant relationships have been found between subjective responses and specific physiological responses, such as that between cortisol and subjective distress (Frankenhaeuser and Johansson, 1986), and effort and heart rate variability (Aasman, Mulder & Mulder, 1987; Vicente, Thornton & Moray, 1987). If one is interested in the relationships between workload, performance and individual state, then repeated measurement of mood and other variables will be likely to be necessary.

Physiological Measures

These techniques include measures of cardiac function, brain function and other physiological processes. In terms of ECG measures, a number of studies now suggest that the power in the mid-frequency band of the heart rate variability (HRV) spectrum (0.07-0.14 Hz) is related to the level of mental effort invested in a task by an individual (Tattersall & Hockey, 1990; Aasman, Mulder & Mulder, 1987; Mulder, 1980; Vicente, Thornton & Moray, 1987). Such variability has been found to decrease as a function of task difficulty in a number of laboratory tasks such as tracking, memory search and classification tasks (Mulder, 1980). It is argued that bursts of suppressed vagal control correspond to periods of momentary effort or controlled processing. Mean heart rate may offer a more sensitive measure of response load, and is certainly influenced by physical activity and perhaps anxiety, which may limit its usefulness in relation to workload. The assessment of pupil dilation (Beatty, 1982) and EEG measures, including evoked potentials have also been used effectively, particularly in laboratory situations, but the advantage of measures of cardiac function are that they can be continuously and independently applied without intrusion to the primary task.

Other techniques, such as analyses of urine and blood provide measures of changes in physiological state through assessment of cortisol, adrenaline and noradrenaline excretion. Urine analysis allows measures of longer-term changes in state through assessment of cortisol and catecholamine concentration. Sustained stress states tend to show increased levels of these hormones. Blood or saliva samples may provide shorter-term measures of fluctuations in state.

There are potential problems with the diagnosticity of physiological measures (Wierwille & Casali, 1983), and a further problem to be aware of is that physiological processes are sensitive to the effects of physical activity and to emotional factors that may have an effect on physiological functions.

Performance

Two major approaches to performance assessment are primary task techniques and secondary task techniques. Primary-task measures are normally only useful for giving an indication of the impact of gross demands. These measures may be easy to obtain in some situations, but in others, such as air traffic control, it is difficult to generate a simple measure of a controller's level of performance that would meet the criteria outlined earlier. If one examined the safety record in air traffic control for example, one might conclude that workload is only a minor problem. However, such a measure may not be immediately sensitive to the effects of changes in task load or working procedures, and will only give a crude indication of the cumulative effect of sustained and high task demands over a long period. Task strategies may differ between skilled and inexperienced operators, and the health and state of the operator may determine the perceived difficulty of a task. The effects of these factors may only be detected by primary task measures once performance suffers or errors are made. From a safety persepective this may be already too late a stage to investigate the adequacy of human factors aspects of the system.

Secondary task measures may be more sensitive to changes in demand or working procedures but unless the allocation of information processing resources to the two tasks is controlled it can be difficult to interpret the observed secondary task performance decrements. Norman and Bobrow (1975) introduced the important concepts of resource-limited processes, which are limited by the effort invested in a task and the priority placed upon task performance, and data-limited processes which are constrained by the quality of information rather than by increases in effort. In work situations, operators may compensate for any increase in task demands by increasing the amount of effort invested in the task. Therefore observed performance levels may remain constant but the operator experiences increased workload. Conversely, a reduction in the level of performance may result either because operators cannot maintain the level of effort expenditure required, or because they lower their criteria for adequate performance. Therefore task performance in resource-limited tasks may be limited by the effort put into the task (related to the priority an individual places upon performance), as well as the difficulty of the task. Secondary tasks have to be chosen carefully in order not to introduce structural interference with the primary task, however secondary-task performance measures can provide a more systematic technique for analysing interference in multiple-task situations than many of the other workload measures.

Observations of error and slips of action may provide useful data concerning the demands of different tasks, but a sound theoretical model of errors must be used to categorise different actions and activities. The recent work by Empson (1991) and Stager (1991) highlights promising approaches to the study of errors in air traffic control.

Workload In Applied Settings

An important point to be made is that there is variation in the way that people do tasks and therefore in the effects and consequences of what is termed workload. Prolonged active management of the resources required to meet task demands may lead ultimately to a deterioration in performance, but there may also be implications for short-term well-being and longer-term health. The experience of workload is thus unlikely to depend simply on task load, but rather on the interaction of task demands, how these demands are dealt with by an operator, and the level of performance achieved. Task demands are important but are mediated by effort and the priority placed on the particular tasks.

The level of control that operators are able to exert in complex systems is an important factor in the relationship between task demands, performance and well-being. Studies typically show an advantage for active control over passive control (Rasmussen & Rouse, 1981; Sheridan & Johannsen, 1976). In other words, open-loop strategies, involving a greater degree of planning and broader understanding of the system as a whole, are seen to be more skilled and efficient than the closed-loop mode. However, Umbers (1979) found that even experienced operators resort to the closed-loop mode when under high levels of workload or when unfamiliar situations or problems occur. This could be a cause for concern as remedial action may be applied once critical events have occurred rather than the more desirable situation in which impending catastrophic events are predicted at a time when something can be done to prevent them occurring.

Air traffic controllers have been found to vary their strategies according to task demand. Sperandio (1978) suggested that controllers take fewer variables into account as the traffic load increases. Similarly, in a study of process control, Bainbridge (1974) found that subjects under pressure used quicker, less accurate methods of finding data values. Although discretion to use open-loop control may normally be preferred by operators and lead to enhanced performance and safety, this discretion could be seen as a demand in itself imposed by the structure of the task or job.

Steptoe (1983) discussed a number of studies that show how the effort required to exert control over situations may result in more pronounced physiological responses. Increased physiological activation, indicated by higher levels of blood pressure and heart rate, was observed when subjects were engaged in effortful problem-solving or activity in a controllable situation. Thus, performance may be maintained at a desirable level (determined by personal and perceived organisational goals) but the effort required to deal with the demands is observed as costs in other psychophysiological systems (e.g., Frankenhaeuser, 1986; Mulder, 1986). Frankenhaeuser has demonstrated various changes in catecholamine and cortisol excretion with increased work demands. The patterning of these changes reflects active management of work and opportunity for control over work. Lundberg and Frankenhauser (1978) found that lack of control was associated with elevated physiological arousal during noise stress, but ratings of effort and the particular pattern of endocrine activity were found to be different depending on levels of performance. Increased ratings of effort, and higher levels of both adrenaline and noradrenaline were found for subjects who performed well in noise, but no changes were observed in the group whose performance was impaired. Generally it has been found that increased catecholamine excretion and lower levels of cortisol excretion and lower levels of anxiety are associated with active processing strategies linked with increased control and effort investment. Distress and both increased catecholamines and cortisol levels are associated with passive conditions or strategies (Frankenhauser, 1979; 1986). Using different physiological measures, Tattersall and Hockey (1990) identified different activities in a simulated flight engineer task which resulted in different cardiovascular costs and subjective ratings of effort and concern. Heart rate appeared to be associated with concern, particularly during activities such as landing and take off, but suppressed heart rate variability and increased subjective ratings of effort were associated with the requirement for problem-solving activity in different activities.

Thus, in order to understand the relationship between demanding situations and changes in performance, well-being and health, it is necessary to investigate changes in different domains. This involves the short-term and long-term assessment of individual state, both physiological and affective state (in terms of mood and well-being), as well as cognitive activity (as implicated in performance). An example of such a study was carried out to investigate the impact of naturally varying workload in air traffic control on a range of measures including performance, and physiological and affective state (Farmer, Belyavin, Tattersall, Berry & Hockey, 1991; Tattersall & Farmer, 1993).

Data were collected from 66 air traffic controllers during two whole working shifts at different times of the year. One shift was during the busy summer period from June to August and the other during the relatively quieter periods, in terms of traffic load, in spring or autumn. Subjective ratings of mood (anxiety/tension, depression, and fatigue) were derived from a 20-item mood adjective checklist (Warr, 1989), which was completed at the beginning and the end of the shift and during each break period (usually every two hours during the shift). Subjective ratings of workload were collected at the same times throughout each day using the NASA Task Load Index (TLX) (Hart & Staveland, 1988). Saliva and urine samples were collected from which measures of cortisol and catecholamines were derived. ECG measures of heart rate and heart rate variability were collected continuously throughout the shift. The controllers also completed a set of cognitive tasks at the beginning and the end of each day. These tasks assessed visual vigilance and verbal reasoning and were used because of the difficulty in measuring overall performance in complex work systems such as air traffic control. Errors in performance may be infrequent and minor slips difficult to detect (Empson, 1991). Therefore tasks were chosen which were thought to simulate different aspects of air traffic control work, and to be sensitive to fatigue and levels of workload over the working day, for example in vigilance performance after prolonged periods of monitoring.

The analyses of these different measures show a number of interesting differences between and within high and low workload shifts indicating the negative consequences of dealing with sustained demands. Subjective ratings of workload were higher during the summer months. A simple measure of traffic count, however, did not appear to be an adequate predictor of subjective workload. Communications load was more clearly associated with the TLX dimensions of mental demand and effort, and traffic load with frustration.

The dimensions of mood were affected in different ways by increased workload. Anxiety showed a significantly greater increase during high workload days but pre-shift levels were not affected by workload. In contrast, levels of depression and fatigue were both higher at the start of the day under high workload conditions and were also elevated during the high workload shift. The sustained demands of the busy summer months appear to result in chronic after-effects of fatigue and depression, whereas anxiety was affected more transiently. However, repeated measurements are required at different times of day from off-duty days in addition to workdays in order to confirm this suggestion.

Salivary cortisol concentration was greater during high workload than low workload shifts, and greater before the shifts than later. It declined during the shift but not as clearly in the latter half of the shift under high workload. There were no reliable effects of workload in the analyses of urinary cortisol or urinary adrenaline but there was a significant interaction between workload and the two halves of the shift for urinary noradrenaline. Noradrenaline excretion decreased over the low workload shift but increased in the second half of the high workload shift. These findings perhaps reflect active coping with the quantity of demands during high workload shifts. The pattern of hormone excretion during high workload is consistent with the findings of Frankenhauser (1986) for example, in that cortisol and noradrenaline excretions are greater under conditions associated with lowered control and increased distress. This pattern may have long-term consequences for the health and well-being of controllers if sustained over long periods.

Heart rate did not differ reliably between high and low workload shifts but differences between the working periods, rest breaks and pre- and post-shift testing periods were found which may reflect differences in physical activity during the different periods. Increased heart rates were also found in the busier units (Heathrow Airport Control Centre and London Air Traffic Control Centre), but it is difficult to associate levels of workload or task variables with these differences as the factor of work location failed to interact with any of the other variables. Heart rate variability did not show a difference between high and low workload conditions, although there was a significant effect of activity, indicating increased mental effort investment during the performance testing periods, compared to the working periods and rest breaks. At the level of analysis carried out to date the heart rate results are not very clear, but they confirm earlier results that heart rate variability is sensitive to effort expenditure in laboratory tasks (e.g. Mulder, 1980). However, this measure may show different patterns of effort expenditure for different individuals and the gross level of analysis of group data could be obscuring the more complex effects of workload.

Visual vigilance performance was affected by workload but verbal reasoning performance appeared to be insensitive to the effects of work between testing times, or between high and low workload shifts. For visual vigilance performance the signal detection measure of d' (sensitivity) was lower before the shift than after the shift, and there was a significant interaction between workload and time of testing, revealing improved sensitivity over low workload days but not during high workload days. No significant effects of workload or time of testing were found for the criterion measure (beta). Verbal reasoning performance was superior at the end of the shift compared to pre-shift performance. Speed of response did not differ significantly between high and low workload shifts, but was faster after the shift than before. A comparable analysis of errors showed no differences between high and low workload but errors were less frequent at end of the shift. These results support the notion that controllers become more actively engaged in their task during the shift, which is consistent with findings that the speed of unpaced work seems to increase later in the day (Broadbent, 1971). The performance of air traffic controllers tends to show an improvement over the day with the important exception of visual vigilance sensitivity, the ability to detect signals in noise. This measure showed an improvement on low workload days but not on high workload days, and suggests that heavy work demands in air traffic control may have a detrimental effect on monitoring performance.

Further analyses will investigate the role of individual differences in workload management by examining differences in the trade-offs between performance, affective state and physiological state as a function of style of coping and locus of control. It is hypothesised that more active copers will be more likely to maintain performance under high workload conditions but will show greater physiological and psychological effects of this activity. It would be desirable for further research to focus on the effects of sustained demands over longer periods of time than was possible in this study, as there are indications that there may be chronic effects of sustained exposure to high workload. Other research suggests that the morning shifts investigated in this study were not associated with the highest ratings of workload and fatigue (Farmer et al., 1990), and therefore the reported effects may actually be an underestimate of the potential impact of workload on well-being and performance.

It is argued that multi-level measurement techniques can provide a broad assessment of the impact of different work demands. Further studies, both controlled laboratory-based studies and field-based studies, are necessary to refine the techniques, but a model of stress and workload management based upon findings from such studies should allow the more accurate prediction of states or situations in which a breakdown in skills might occur. Such a breakdown is referred to by air traffic controllers as 'losing the picture', when they experience difficulties in attending to, and remembering accurately, relevant information about aircraft under their control. It is precisely this kind of situation that should be avoided in work in which safety is critically dependent upon performance.

Conclusions

The practical problems that might be encountered in carrying out workload evaluations in work settings have been outlined. Different approaches have been distinguished that may determine the type of research design used and provide assistance in the difficult choice between workload assessment techniques. One approach to workload assessment is to examine the short-term consequences of combining various tasks. Theoretical models of attention allocation (e.g. Wickens, 1984) will underpin specific studies of interference and the consequences of task demand and task conflict for performance. A further approach with a different temporal orientation may lead us to a better understanding of the relationships between work demands and strain through the analysis of individual differences in cognitive control processes. The application of these processes may depend on individual differences in long term styles and short term strategies, but may be used to prevent decrements in work performance under difficult conditions. However, control may attract costs as well as benefits in terms of changes in affective state and physiological activity. Thus, strain associated with work demands may only be measurable in the form of tradeoffs between performance and other domains of individual activity. The methodological implications are to identify patterns of adjustment to workload variations using repeated measures and longitudinal sampling of performance as well as subjective and physiological measures.

Possible enhancements to workplace design must take into account these human factors considerations of workload in order to avoid potential decrements in individual performance and associated organisational problems.

References

Aasman, J., Mulder, G., & Mulder, L. J. M. (1987). Operator effort and the measurement of heart-rate variability. *Human Factors, 29,* 161-170.

Bainbridge, L. (1974). Analysis of verbal protocols from a process control task. In E. Edwards & F. P. Lees (Eds.), *The Human Operator in Process Control.* London: Taylor and Francis.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91,* 276-292.

Broadbent, D. E. (1971). *Decision and Stress.* London: Academic Press.

Damos, D. L. (1991). *Multiple-Task Performance.* London: Taylor and Francis.

Empson, J. (1991). Cognitive failure in military air traffic control. In J. A. Wise, V. D. Hopkin & M. L. Smith (Eds.), *Automation and Systems Issues in Air Traffic Control.* Berlin: Springer-Verlag.

Farmer, E. W., Belyavin, A. J., Berry, A., Tattersall, A. J., & Hockey, G. R. J. (1990). *Stress in Air Traffic Control I: Survey of NATS Controllers.* RAF Institute of Aviation Medicine Report No. 689.

Farmer, E. W., Belyavin, A. J., Tattersall, A. J., Berry, A., & Hockey, G. R. J. (1991). *Stress in Air Traffic Control II: Effects of Increased Workload.* RAF Institute of Aviation Medicine Report No. 701.

Frankenhaeuser, M. (1979). Psychoneuroendocrine approaches to the study of emotion as related to stress and coping. In H. E. Howe, & R. A. Dienstbier (Eds.), *Nebraska Symposium on Motivation.* University of Nebraska Press.

Frankenhaeuser, M. (1986). A psychobiological framework for research on human stress and coping. In M. H. Appley & R. Trumbull (Eds.), *Dynamics of Stress.* New York: Plenum.

Frankenhaeuser, M., & Johansson, G. (1986). Stress at work: psychobiological and psychosocial aspects. *International Review of Applied Psychology, 35,* 287-299.

Gopher, D., & Donchin, E. (1986). Workload – An examination of the concept. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance Volume II: Cognitive Processes and Performance.* New York: John Wiley & Sons.

Hancock, P. A., & Meshkati, N. (1988). *Human Mental Workload.* Amsterdam: Elsevier.

Hart, S. G., & Staveland, L. E. (1988). Development of a NASA TLX (Task Load Index): Results of empirical and theoretical research. In P. Hancock & N. Meshkati (Eds.), *Human Mental Workload.* Amsterdam: Elsevier.

Hill, S. G., Iavecchia, H. P., Byers, J. C., Bittner, A. C., Zaklad, A. L., & Christ, R. E. (1992). Comparison of four subjective workload rating scales. *Human Factors, 34,* 429-439.

Hulbert, T. (1989). A comparison of the 'NASA-TLX' and 'ISA' subjective workload rating techniques. Civil Aviation Authority Air Traffic Control Evaluation Unit, Bournemouth, UK. Internal Report.

Jordan, C. S. (1992). Experimental study of the effect of an instantaneous self assessment workload recorder on task performance. *Defence Research Agency Technical Memorandum DRA TM (CADS) 92011.* Portsdown, Hants: DRA.

Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. In D. L. Damos (Ed.), *Multiple-Task Performance.* London: Taylor & Francis.

Lundberg, U., & Frankenhauser, M. (1978). Psychophysiological reactions to noise as modified by personal control over noise intensity. *Biological Psychology, 6,* 55-59.

Mackay, C., Cox, T., Burrows, G., & Lazzerini, T. (1978). An inventory for the measurement of self-reported stress and arousal. *British Journal of Clinical Psychology, 17,* 283-284.

Moray, N. (Ed.)(1982). *Mental Workload: Its Theory and Measurement.* New York: Plenum.

Moray, N. (1988). Mental workload since 1979. *International Reviews of Ergonomics, 2,* 123-150.

Mulder, G. (1980). *The Heart of Mental Effort.* University of Groningen, The Netherlands.

Mulder, G. (1986). The concept and measurement of mental effort. In G. R. J. Hockey, A. W. K. Gaillard, & M. H. G. Coles (Eds.), *Energetics and Human Information Processing.* Dordrecht: Nijhoff.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7,* 44-64.

Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors, 33,* 17-33.

O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance Volume II: Cognitive Processes and Performance.* New York: John Wiley & Sons.

Ogden, G., Levine, J., & Eisner, E. (1979). Measurement of workload by secondary tasks. *Human Factors, 21,* 529-548.

Rasmussen, J., & Rouse, W. B. (Eds.). (1981). *Human Detection and Diagnosis of System Failures.* New York: Plenum.

Ravden, S., & Johnson, G. (1989). *Evaluating Usability of Human-Computer Interfaces: A Practical Method.* Chichester: Ellis Horwood.

Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload.* Amsterdam: North-Holland.

Rehmann, J. T., Stein, E. S., & Rosenberg, B. L. (1983). Subjective pilot workload assessment. *Human Factors, 25,* 297-307.

Rohmert, W. (1987). Physiological and psychological work load measurement and analysis. In G. Salvendy (Ed.), *Handbook of Human Factors.* New York: John Wiley & Sons.

Sheridan, T. B., & Johannsen, G. (Eds.). (1976). *Monitoring and Supervisory Control.* New York: Plenum.

Sperandio, J. (1978). The regulation of working methods as a function of workload among air traffic controllers. *Ergonomics, 21,* 195-202.

Stager, P. (1991). Error models of operating irregularities: Implications for automation. In J. A. Wise, V. D. Hopkin & M. L. Smith (Eds.), *Automation and Systems Issues in Air Traffic Control.* Berlin: Springer-Verlag.

Steptoe, A. (1983). Stress, helplessness and control: The implications of laboratory studies. *Journal of Psychosomatic Research, 27,* 361-367.

Tattersall, A. J., & Farmer, E. W. (1993). The regulation of work demands and strain. In G. P. Keita, & S. L. Sauter (Eds.), *Job Stress 2000: Emerging Issues.* Washington, D.C.: American Psychological Association.

Tattersall, A. J., & Foord, P. S. (1993). An experimental evaluation of instantaneous self assessment as a measure of workload. Submitted to Ergonomics.

Tattersall, A. J., & Hockey, G. R. J. (1990). The assessment of workload in a complex monitoring and fault diagnosis task. In D. Brogan (Ed.), *Visual Search.* London: Taylor & Francis.

Umbers, I. G. (1979). Models of the process operator. *International Journal of Man-Machine Studies, 11,* 263-284.

Vicente, K. J., Thornton, D. C., & Moray, N. (1987). Spectral analysis of sinus arrhythmia: A measure of mental effort. *Human Factors, 29,* 171-182.

Vidulich, M. A., & Tsang, P. S. (1987). Absolute magnitude estimation and relative judgement approaches to subjective workload assessment. In *Proceedings of the Thirty First Annual Meeting of the Human Factors Society.* Santa Monica, CA: Human Factors Society.

Warr, P. B. (1989). The measurement of well-being and other aspects of mental health. *Journal of Occupational Psychology, 63,* 193-210.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of Attention.* New York: Academic Press.

Wickens, C. D. (1991). Processing resources and attention. In D. L. Damos (Ed.), *Multiple-Task Performance.* London: Taylor & Francis.

Wierwille, W. W., & Casali, J. G. (1983). A validated rating scale for global mental workload measurement applications. In *Proceedings of the 27th Annual Meeting of the Human Factors Society.* Santa Monica, CA: Human Factors Society.

Yeh, Y. Y., & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors, 30,* 111-120.

[illegible]
P. 21

Certifying Life

P.A. Hancock

University of Minnesota

Preamble

Systems have three possible states: stable, transient, and failed. When a system is stable no certification is necessary. If a system is in transition no certification is possible. If a system has failed no certification is relevant. I argue, therefore, that certification is a palliative and an anodyne to societal concerns over the potential destruction that advanced systems can wreak. I further submit that the manifest 'need' for certification is part of an occidental view that nature must be tamed, constrained, and contolled. It is unlikely that our cultural myopia will be excised by the present polemic. But, within the fullness of time, mutual co-evolution and validation by nature itself will fulfill the argument for me.

Introduction

When my father died, I was some 40,000 ft. above Iceland. I will never be able to reconcile myself to the fact that I could not see him one more time before he died. So when I saw him in Cheltenham hospital's 'Chapel of Repose' much of what I felt was anger and frustration, diffusely directed.

I realize now, some years later, that part of that frustration had to do with life itself. As I stood in front of his body I could not help but feel that he was only asleep. After all, he had not changed substantively since I had last seen him. But physical appearance belied what we all know and all must eventually face ourselves; what made my father *my father* had gone.

As I started to write this diatribe on certification, I realized that some doctor had been asked to certify that my father was dead. Indeed, it became obvious that many agencies required this certified evidence to remove him from the 'lists of the living.' As you might imagine, and I hope you do not experience, the bureaucracy of dying is as obscene as the event is disturbing. (I exonerate the process only on the grounds that the individuals involved proved both sensitive and caring in a job where repetition must eventually dull the sensibilities.) In the Cheltenham office of the Department of Health and Social Services, I pondered on the comparison between my father's death and the demise of any system in society, biological or technical.

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PRECEDING PAGE BLANK NOT FILMED PAGE 206 INTENTIONALLY BLANK

Certifying Failure

My immediate response was, why bother? Nothing in the process of certification was going to bring him back and therefore, for me as an individual, the process was redundant. However, while my father was important to me as an individual, he was, as an entity, important to society in another manner. It was in this latter sense that he was being liquidated. As a result, one critical question on a societal *and* individual level, for both humans and machine systems, was why they failed. Whether it's a post-mortem or an accident investigation the process is the same; a post hoc analysis of what went wrong.

The fundamental assumption is that knowledge of what went wrong last time will help us to avoid the 'same' sequence of events leading to failure next time. With respect to my father such reasoning is vacuous. There can be no 'next' time. With respect to a theory of technical systems operation such reasoning is also becoming progressively more naive. That is, events of failure are becoming more idiosyncratic and less deterministic. Despite our continual attempts to extract or even impose pattern on failures, we are faced with the certainty that as nature explores the combinatorial explosion of interactive states that complex systems can achieve, no two failures will be exactly the same and are liable to become progressively divorced in type. Hence our search for pattern will devolve to ever higher meta-levels of description until we provide the single parameter, unified field theory of failure, namely; "It Broke!" The alternative to this is articulated in a recent chapter (Hancock & Chignell, 1993) and is directly dependent upon the resolution of an empirical proposition concerning the demise of natural ecological systems. This implies there is a power law relationship between the frequency of failure occurrence and the magnitude of that occurence. Should the power law, founded in the application of non-linear dynamics, apply to technical systems, it would provide strong evidence that technical systems are as "natural" as ecosystems and are subject to the same constraints. In particular, it would imply that catastrophic failures of systems are the magnification of events that propagate through systems and predominantly result in minor, frequently unnoticed, perturbations. Where and how intentionality might supress or deflect such propagations remains a moot issue until a veridical power-law relationship has been demonstrated. Had my father been murdered, as a technical system might be sabotaged, we use such knowledge to apportion blame. However, that is an outflow of certification of failure, a prime reason for it.

We *have* to believe in regularity, since we humans at a fundamental level *invent* it. Therefore, we have to, in the occidental world at least, subscribe to the notion that the future is, at least partly, predictable from the past and therefore controllable. To subscribe to a radically differing version of this belief is to risk being labeled, almost literally, 'insane.' Indeed, as Schrodinger (1944) observed about the self-consistency of these rules:

> It is well-nigh unthinkable that the laws and regularities thus discovered should happen to apply immediately to the behavior of systems which do not exhibit the structure on which those laws and regularities are based.

However, at the heart of the schizophrenia of this position is our manifest dissonance between a view of time as a linear dimension in which unique progression obviates exact repeatability versus time as cyclic phenomenon in which repetition and recurrence dominate (Toulmin & Goodfield, 1965). Our present zeitgeist is to believe that the future must be like the past in some way, but cannot be the past exactly. And the 'laws' that have structured the past should have a consistent influence in the future. This latter assumption is a belief[1], not an empirically supportable statement, as is the predicate of regularity and consistency in a more subtle way.

I also felt reasonably confident in asserting that the post-mortem cause of my father's death was about as accurate as the most cursory of accident investigations. In reality doctors deal with death in much the same way they deal with illness. They do not have the time for exhaustive diagnosis of particular problems, hence they frequently treat symptoms or provide palliative agents of widespread capability that will cover the source of the problem without ever necessarily identifying exactly what the problem is. Cause of death is even less liable to receive in detailed examination since the problem rarely proliferates. If death is likely or if there is some specific reason (e.g., homicide), they do what failure analysts do anyway: they pass the problem on to a specialist (e.g., a forensic pathologist). What is frequently not acknowledged is that because we do not fully understand the phenomenon of life, we cannot always specify why life is extinguished. With respect to complex systems, as they grow less determinate in their actions (indeed as many such systems already are), their 'cause of death' may become equally difficult to specify. Right now many professional medical personnel will acknowledge that some individuals die because 'they no longer wanted to go on living.' Can we expect an analog of this in our machine systems?

In summary, certifying death is fundamentally irrelevant. At a bureaucratic level, there are many boxes to be checked and some superficial reasons why we need a paper that records demise. But as with the munchkin doctor in 'The Wizard of Oz,' repeatedly asserting the absence of life is hardly an answer to the future of life. (I do not comment on the social function of leave-taking and grieving, but note that such processes occur when we let go of our possessions as well as our loved ones.) Certifying failure states in complex systems is similarly redundant. Post-mortems identify a concatenation of circumstances which connote progressively longer chains of interactive failures, where *a priori* prediction of such failures has not, and some would suggest, cannot be anticipated. The search for pattern in such failures will inevitably turn up some commonalties, since humans can turn up commonalties in the most diverse array of electro-magnetism. However, prevention based on post-mortem is inevitably a losing battle.

Certifying Stability

If certifying failure should prove irrelevant, shouldn't we certify systems for stable states of performance? That is, shouldn't we be able to assure ourselves that withinside the design

[1] The foundation of these beliefs has been most eruditely articulated by Sheldon Glashow in the New York Times (October 22, 1989) which stated: "We believe the world is knowable, that there are simple rules governing the behavior of matter and the evolution of the universe. We affirm that there are eternal, objective, extrahistorical, socially neutral, external and universal truths and that the assemblage of these truths is what we call physical science. Natural laws can be discovered that are universal, invoriable, inviolate, genderless, and verifiable. They may be found by men or by women or by miced collaborations of any obscene proporations. Any intelligent alien anywhere would have come upon the same logical system as we have to explain the structure of protons and the nature of supernovae. This statement I cannot prove, this statement I cannot justify. This is my faith."

operational envelope, the system reliably does all that we say it should? In part this depends upon what we mean by complex systems. Let us consider the nature of machines and consider indeterminancy in machines in the same manner we consider the potential for intelligence for machines. Some four decades ago, Scriven (1953) could be fairly unequivocal. He asserted that:

> "Machines are definite: anything which was indefinite or infinite we should not count as a machine."

Today we cannot be as certain. As a result, Scriven's (1953) subsequent argument about the incompleteness of Godel's theorem is not without problem. However, the process of reasoning is instructive.

> Godel's theorem must apply to cybernetical machines, because it is of the essence of being a machine, that it should be a concrete instantiation of a formal system. It follows that given any machine which is consistent and capable of doing simple arithmetic, there is a formula unprovable-in-the-system – but which we can see to be true. It follows that no machine can be a complete or adequate model of the mind, that minds are essentially different from machines.
>
> We understand by a cybernetical machine an apparatus which performs a set of operations according to a definite set of rules. Normally what it is to do in each eventuality; and we feed in the initial "information" on which the machine is to perform its calculations. When we consider the possibility that the mind might be a cybernetical mechanism we have such a model in view; we suppose that the brain is composed of complicated neural circuits, and that information fed in by senses is "processed" and acted upon or stored for future use if it is such a mechanism, then given the way in which it is programed – the way in which it is "wired up" – and the information which has been fed into it, the response – the "output" – is determmined, and could, granted sufficient time, be calculated. Our idea of a machine is just this, that its behavior is completely determined by the way it is made and the incoming "stimuli": there is no possibility of its acting on its own; given a certain form of consturction and a certain input of informaiton, then it must act in a certain specific way.

In arguing the mind cannot be like a machine, Scriven is limited in a number of ways. First, there is no rationale for suggesting that a mind can explore all possible states of a statement space. That is, as we cannot know everything it may well be the things we don't know that contain anomalies intrinsic to Godel's contention. Second, the argument about seeing what is true, but is improvable in the system can rapidly become a teutology in which we ask *how* the seeing or realization is done. Thus the theoretical difference between mind and machine may be obviated by practical exigency. For the purpose of the present argument, we cannot then state all possible conditions within an operational envelope with certainty. What certification devolves to in this case is an assessment of probability. As a consequence, the heart of certification would seem to represent a customer warranty. For small individual objects, this interaction may be appropriate, since the vendor and the customer are divorced in some spatio-temporal fashion. However, the complex systems about which the present discourse revolves are not the creation of one individual nor are they bought by one individual. In essence, society is at one and the same time, both vendor and client. Even within this global perspective, it is frequently the agency that operates a system that regulates and certifies a system.

We would like to think that if all individual parts of a system were certified then the overall system would be 'safe.' This is bottom-up, wishful thinking. It is the *sine qua non* of design, that objects and systems are created for stability of action and hence should be 'certifiable' with the design space. Yet, here it is the combinatorial explosion of potential interactions, as much as nature's own test and evaluation of those interactions which defeats the hoped for assertion. I should note here that combinatorial explosion of interaction alone does not connote instability as represented by the transient states of operation. This is examined below.

More critically, what are we designing such systems for? It is the frequent observation of the more experienced members of the design community that you never get the opportunity to create complex systems from the ground-up. Almost always they are evolutionary in that new elements are added to older system until the working environment is a palimpsest of overlaid versions. If this is the case, we will never be able to completely specify the parameters of a system that is itself 'underspecified.' More to the point, as we build systems that are beginning to cost in the billions of dollars (e.g., national airspace system, Intelligent Vehicle Highway Systems [IVHS]), we will want them to deal not only with existing conditions but also with future anticipated demands.

Hence, future complex systems must be generative and creative in exploring potential operational 'spaces' in order to be cost effective (Hancock, 1993). In consequence, such systems perforce will be underspecified, for not to do so would be to defeat their evolutionary purpose. Systems that are intentionally underspecified cannot be certified for all phases of operation. Thus we arrive at an impasse. That is, the very systems that we seek to certify should, by design, defy certification.

However one seeks to justify certification for stable states of system operation, one will devolve to this paradox. The paradox is that certification is a guarantee of future operation and implies a predictive determinism about that future state. If such deterministic foreknowledge could be achieved, the operation involved would be completely automatic and by definition not a complex system of the sort relevant here. However, as the future is conceived of as either partly deterministic or totally indeterminant, we want systems to adapt to unforeseen conditions and to explore 'strange new worlds' in order to justify their cost. Under neither circumstance is certification necessary or indeed feasible.

In his argument concerning the potentiality of machine intelligence, Turing (1950) examined the same issued from an inverted position and countered the argument that machines could not be intelligent because of the informality (or indeterminacy) of behavior. He indicated that:

> *It is not possible to produce a set of rules purporting to describe what a man should do in every set of circumstances. One might for instance have a rule that one is to stop when one sees a red traffic light, and to go if one sees a green one, but what if by some fault both appear together? One may perhaps decide that it is safest to stop. But some further difficulty may well arise from this decision later. To attempt to provide rules of conduct to cover every eventuality, even those arising from traffic lights, appears to be impossible.*

Given both the paradox of certification and the improbability of comprehensive future prediction, certification around stability appears a vacuous endeavor indeed.

Certifying Transition

If we do not need to or cannot certify failure and are excused from certifying stability, surely we have to explore certification in the intervening realm where systems fluctuate between stability and failure – the regions of transition. This appears most relevant, since it is during the process of incipient failure and recovery from potential failure that represents the most critical active phase of operation. The problem again is one of predictability.

Certification is an assurance of determined causality. That is, we undertake to state that if a sequence of conditions prevail and a sequence of processes are in operation, a series of outcomes are guaranteed. However, when we step into transitional states, we enter regions that by definition provide increasing uncertainty.

I noted above that the societal investment in large-scale complex technical systems implies that they should be generative and explorative. I shall extend this description to imply that such systems should also be 'skillful.' I use skillful in a specific context here. The context is one that has been used in examining adaptive systems (Holland, 1991). It has been posited that adaptive systems are so structured in response to their initially experienced environmental contingencies. That being that adaptive systems, of which life is the pre-eminent example, grew at 'the edge of chaos.' The latter condition is one where the phase plane of operation devolves from a stable condition toward a chaotic one. (A random regime does not allow sufficient consistency to allow responsive systems to develop, a system in energetic stasis cannot develop adaptive strategies.) It is at the edge of chaos that adaptation develops. 'Skill' in this context is the ability to explore the edge of chaos and the advantages intrinsic to residence in that region without fallback to immaleability or transition into chaos itself.

Systems in transition reside in the region between stability and chaos (not to be directly equated with complete failure). Hence, certification of skillful systems in transition is to suggest that we can 'predict' the response of an adaptive system whose primary function is to cope with unanticipated conditions. The imperilment of such a procedure is now surely laid bare. We cannot certify a system in such conditions, since to do so would be to constrain the very stages of response of a system that we want to be open and unconstrained in order to recover to a state of operational stability.

Certifying What?

I have presented a polemic which has used an analogy with life. Life is a successful adaptive complex system that is predicated upon the environment but is, we believe, not totally constrained by it in terms of its response. Within some bounds we can engage in a certification of life, but why would we? I have suggested a parallel between the failure of a system and death. By extension, the parallel holds for health (stable states) and disease and trauma (transition states), although I have not articulated these latter conditions in as much detail.

I have suggested that certification of stable and transient system states is a relatively futile exercise, since I posit that the very systems we are focusing on are ones which imply open, explorative, and non-deterministic functions. Certification of failure is a time honored societal endeavor to provide information on how to obviate failure in successive systems. In deterministic systems with high frequency of occurrence in the same fundamental state (e.g., DC-10's), this can be a useful function. For one-off large scale systems of progressive indeterminacy such certification serves a more social role in apportioning blame or accountability. I submit that the latter function is a societal palliative for the fears that such indeterminacy brings. I further submit that this is an occidental pre-occupation and one that stems from the notion of controlling and taming nature. As I have previously indicated (Hancock, 1991), the Titanic is the leitmotif of this 'world view.' I take all other aspects of certification to be 'lowest common denominator' insurance.

Hope For The Future

In reviewing the above, it might appear to be a rationale for doing nothing with respect to the design, test, and evaluation of systems and to fatalistically accept the uncertain outcome that nature 'chooses' to provide. I reject this fatalism wholeheartedly. What is objected to is an attitude of mind that proposes that we can 'know' all the states of complex systems we have already created and are creating by the moment.

Therefore:

i) I advocate a great exercise of humility, aspecially with respect to an understanding of the influence and effect of the technology we create.

ii) I advocate a societal change in attitude from the legalistic 'blame' we seemed destined to fix, to a recognition of societal responsibility for the things we collectively build.

iii) I advocate a recognition of the explorative and adaptive nature of ourselves and by extension the manufacturanda we create to extend ourselves.

iv) I advocate the need for the immediate integration of those whose innovative work is enlightening complex adaptive system operation with those who design, test, and evaluate such technical assemblies.

v) Finally, I advocate a strong thrust of research in the area of 'skillful' systems who possess an acknowledged degree of skill in recovering to stability.

In sum, I advocate the replacement of the procedures of certification with the exploration of training 'skill' in complex human-machine systems. I am not foolish enough to believe such recommendations are liable to actually enact change. I adhere more strongly to these statements even following the meeting and the interaction which occured. I also take as a cunard the notion of certification as process, since there is then no fundamental difference between certification and design, test, and evaluation. I take such an arguement to be without meaning.

Acknowledgements

I am happy to acknowledge the support and contribution of so many people who have made possible this and other of my work. First and foremost, I would like to thank the organizers, John Wise, David Hopkin, and Dan Garland (and express my regret that Dan was unable to attend the meeting he had done so much to promote). My thanks extend beyond the immediate organizers to Paul Stager and Marv Smith. They have been kind enough to invite me to a series of interesting and insightful meetings and my benefit and therefore debt is immeasurable. I am happy to acknowledge support from a number of agencies. In particular the Minnesota Department of Transportation(Mn/DOT) and the Center for Transport Studies continues to be a major supporter of our Laboratory. The issues of adaptive complex systems is supported by the Naval Air Warfare Center where Jeff Morrison is our enlightened contract monitor. I am especially grateful for the support of NASA through Sandy Hart who was the first to support my research efforts. Of course, the views expressed here are my own and do not represent the viewpoint of any of the agencies named.

References

Hancock, P.A. (1991). The aims of human factors and their application to issues in automation and air traffic control. In: J.A. Wise and V.D. Hopkin (Eds.). *Automation and Systems Issues in Air Traffic Control*, NATO. New York: Springer

Hancock, P.A. (1993). On the future of hybrid human-machine systems. In: J.A. Wise, V.D. Hopkin., and P.Stager (Eds.).*Verification and validation of complex systems*. Martinus Nijhoff: The Netherlands.

Hancock, P.A., & Chignell, M.H. (1993). On human factors. In: J. Flach, P.A. Hancock, J.K. Caird., and K. Vicente (Eds.). *The ecology of human-machine systems*. New Jersey: Erlbaum.

Holland. J.H. (1991). *Adaptation in natural and artificial systems*. Cambridge: MIT Press (University of Michigan Press, 1975)

Schrodinger, E. (1944). *What is life*? Cambridge: Cambridge University Press.

Scriven, M. (1953). The mechanical concept of mind. *Mind*, 62,

Toulmin, S., & Goodfield, J. (1965). *The discovery of time*. Harper & Row: New York.

Turing, A.M. (1950). Computing machinery and intelligence. *Mind*, 59, 433-460.

When It All Goes Wrong Who Will Be Blamed?

Victor Day

EUROCONTROL

Introduction

In the past, it was often the case that the bringer of bad news was killed by the king who received it. While we may no longer go to such extremes, the principle nevertheless lingers on in ATC in another form. It is highly probable that, in the future automated world of ATC, when an incident (or accident) related to ATC occurs, the first recourse will be to blame the controller, after all he or she is in charge and ultimately responsible for his or her ATC sector. This has a direct parallel with the "pilot error" problem, where the pilot is ultimately responsible for his aircraft. Doubtless, much investigation will be pursued to determine whether or not the controller was indeed responsible. However, ask any controller for their opinion of the probable outcome of such an investigation and they will reply "they will still blame me, no matter what the reason."

Why should this be so? Simply because the controller is there, and is the easiest person to identify; others who may have contributed to the problem may no longer be identifiable or cannot be located. Following an incident, when the outraged public is hungry for blood, blaming the controller will be the easiest and most readily available solution for any organisation in order to extract itself from an awkward situation: it has happened before in other fields, and will certainly happen again.

Knowing this, the controller will take measures, consciously or not, in order to avoid such a situation. Such measures are likely to ensure that the level of traffic he or she is handling does not greatly exceed what he or she could handle without the extra automation, in which case the millions of pounds of expenditure in providing advanced ATC systems with greater capacity potential are to no real advantage.

The control staff have a very real concern about the pressure to handle more flights and reduce delays, and for the relentless introduction of automation to achieve this aim. This paper tries to highlight some points of concern which go beyond the aim of a providing a 100% available system; it attempts to address the issues of a complete ATC system which is a composite of many skills – design, technical and human – which span the complete lifetime of the system from its inception to implementation, operation and maintenance, and of which the human element, the controller, is a significant part.

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

A Survey of Today's Problems

Let us take an example of automation already widely in use today: the use of radar labels. Supposing a system error should allocate callsigns to wrong aircraft or even lose all the labels simultaneously, in a busy sector. Would the official view be "it was not the controllers fault", or would it be "the controller is responsible at all times to ensure the system has properly identified the flights", and "the controller is responsible at all times for maintaining proper identification of aircraft under his control." Considering that the controller may have 20 or more flights on the frequency climbing, descending and crossing, the controller is, even today, dependent upon the system to track the flights. Indeed, if the controller were not, then he or she would not be able to handle the 20 or more flights simultaneously.

Consider such a case, and that a fatal accident occurred. After a search for the "bug" which caused the problem, would the programmer be found? Would the controller be held responsible? If employed by an Administration, would that Administration be ultimately responsible? What if the system was a turnkey system supplied by a company, would they be responsible? If the programmer has left the Administration or company, is he or she still responsible? What if other programmers have altered the code, or a system designer has changed the interface, or a new piece of hardware has been attached later, which was not foreseen in the original program; who is to blame, who will be responsible? With all these potential problems, is it not easier to fall back on the premise "the controller is responsible at all times for maintaining proper identification of aircraft under his or her control", and blame the controller? After all, he or she is much more 'available' than any of the others "kill the messenger" said the king!

These problems are with us today, and so far in this part of the world, there has not been a fatal accident directly attributable to such system failure problems. Under pressure for higher productivity and reduction of delays, the controllers continue to make use of the system and, perhaps naively, trust in ultimate justice in the event of a future fatal accident. So far, no question of legal liability has been tested; it is therefore possible to hold the controller responsible and give a re-training and re-validation period, with some justifiable resentment on his or her part; sometimes a more understanding viewpoint may prevail, and the incident can been closed without further action being taken; such a philosophy can be pursued when financial restitution is not in question.

However, should there be an accident, and institutions are obliged to pay large sums of money in compensation, the pressure to find a culprit on whom to shift the responsibility and evade the financial burden will become irresistible, and the most obvious and available person will be the controller.

The Problems of the Future

One of the specific aims of increased automation in ATC is to relieve the controller of unnecessary and unproductive work, in order to permit the better use if his or her capacity for other more important problems. So far, automation has generally centred around the printing of the paper strips, transmission of basic data between systems, automatic radar identification, tracking and labeling of flights, and more recently, safety net features such as Short Term

Conflict Alert. However, the trend is to replace the paper strips by automated systems incorporating monitoring, conflict detection and resolution advisories into the ATC system. How will such trends affect the controller's responsibility in the future, and in the event of a failure, his or her ability to cope with the complex situation which occurs?

We must consider that "failure" does not necessarily mean a total system failure; it can be partial failure of a particular but essential function, or even a logic error within a function (such as not detecting a conflict) which only becomes apparent when the eventual loss of separation occurs.

The very essence of improved automation is that it off-loads work from the controller. This therefore means that, as numbers of aircraft under control increase, the work cannot be done without particular functions. As controllers become more familiar with the automated facilities, they will become more reliant on them, and subsequently less capable of working without them (as in the 'glass cockpit' revolution in the aircraft, where pilots are concerned about losing flying skills is an equivalent situation). Therefore, the failure of any automated function in the future ATC system may well be catastrophic. The provision of an alternative backup system, including paper strips, will not be suitable if it requires skill and experience to handle it.

The problem also goes beyond the ability of the system to carry out its functions correctly. The design of the system interface and the way in which information is presented to the controller will also play an important part in the future. If the controller may be misled by the manner of presentation then will he or she be at fault, or is it the system designers who made the specification? A design which works in some traffic configurations may not be effective in others. If the controller is required to continuously input data to the system in order for the system to properly detect conflicts and advise on solutions, will he or she be at fault if this is not done? What if the interface design did not permit the controller to keep up with the inputs, or the system response was too slow?

Verification Requirements

The human element is probably the most flexible component of a future system, but also the most difficult to predict and verify. A recent total failure of an ATC system was due to a human being short circuiting the back-up power supply during maintenance procedures and causing a power surge on the main power supply, which then failed. However, it was the human element inside the centre and at other ATC centres which coped with this problem situation and maintained adequate separation on all flights.

The verification of ATC systems must consider a considerable number of potential problems, and testing techniques must be elaborated to reduce these potential problems to a minimum. However, one of the principal problems is the human element within the system, which is difficult to verify in a formalized manner. It is therefore imperative, as new ATC systems are designed, that one of the vital aspects must be the inclusion of the users of the system, at all levels of design and verification; i.e., the controllers themselves. It is essential that the users are a formal part of the development of the techniques which they will be expected to use prior to being obliged to use them, and that they approve of the concepts and rules being defined. It may be necessary to seek approval from the professional guilds which represent the controllers.

The inclusion, then, of ATC personnel in the design and verification process is a prerequisite for safe design and implementation. This may be obvious, but there is often a trend to define a system and then give it to controllers who then have to make it work; their inclusion in the design and specification process is frequently too late to ensure provision of a system which they feel they can use efficiently.

However, there is also a problem that good design may require a change of technique compared to today's control methods. A change of technique also requires a re-training period, to be properly proficient in the new methods. Failure to train controllers adequately prior to the introduction of new automated features may be the biggest source of "system failure" which will occur. It will be, without doubt, a source of delays until controllers feel comfortable with the system. Training, in all its forms, documentation, films, lectures, computer based training, and simulation training is a vital aspect of good ATC system design and implementation. Controller training for new ATC systems is frequently the last aspect to be considered, and is rarely treated properly.

Conclusion

In earlier days, the controller provided all features of the ATC system: writing strips, passing estimates, and issuing clearances. There is already a shift away from this simple view of the responsibilities in present day ATC systems, and the future systems will increase this even more. It must be recognized that future ATC systems are a composite of the human and machine working harmoniously together, and that responsibility for the provision of safety is not for the controller alone. The machine must also bear responsibility, and that responsibility is shared by the system planners and designers, the software and hardware architecture and maintenance.

The system verification process must be conceived to ensure system reliability to the highest level of safety, but it must also consider the reliability of individual components which will contribute in the future to the complex interplay of the human and the machine.

A clear policy must be defined to fairly establish the responsibility of the controller in future ATC systems for any failures of part or all of the ATC system on which the controller will be dependent to carry out his functions. If conflict detection is an integral part of the future ATC system, the responsibilities of the conflict detection system (including the designers and implementors) must be the same as that currently assumed by the controller who, today, carries out this function manually. Such responsibility becomes even more acute when we consider resolution advisories.

Inclusion of ATC personnel in the design process is a pre-requisite for system reliability, as is properly planned and adequate training adjusted to individual controller needs. Professional controller associations and Guilds should be invited to approve the systems before and during implementation.

Although trials, simulation, and shadow operations may all contribute to the verification process prior to live application, there must also be a system of feedback and investigation to trace and analyze incidents which have occurred during normal operations; to this end a controller reporting procedure must be an obligatory part of any future ATC system. Such a procedure should encourage controllers to comment on their experiences without fear of retribution, even if they were at fault. The reason why they failed in a particular task may be due to inherent system design problems, which must be eradicated before more serious problems can occur.

Reliance on backup systems which provide alternative solutions in case of single or total failure will be unsuccessful if it is complex, or requires skills which may be current now, but will become lost in future automation.

While the above proposals are not new, there is every chance that the lessons learned in the past may be overlooked in the enthusiasm to bring in new ATC systems in the race for improved productivity and reduced delays. These lessons may become more vital in the future, and should be a standard part of any new system design, verification and implementation process of future ATC systems.

ATC is primarily concerned with safety, and the expedition of traffic cannot be allowed to compromise safety.

The views expressed in this paper are solely those of the author and do not necessarily represent the official opinion of the Eurocontrol Agency.

N95- 34778

Is There A Role For A "Test Controller" In The Development Of New ATC Equipment?

51-54
[illegible]
[illegible]

Ron Westrum

Eastern Michigan University

Introduction

In the aviation field, test pilots have long performed a valuable function in the evaluation and improvement of new aircraft. Through their special experience and training, test pilots are able to provide expert feedback for the development process. Although often glamorized by films and books, the role of the test pilot is basically that of a member of the engineering team. The test pilot checks out the plane in the air, explores the performance envelope of the various aircraft systems, notes "bugs" and other infelicities of the equipment, and makes suggestions for improvements. Test pilots have unusual piloting skills, but more importantly have training in systematic check-out and a high sensitivity to performance quirks that others might miss (Hallion, 1981).

Testing new ATC equipment necessarily involves similar skills to test-piloting. Check-out is expected to take place according to systematic protocols, and problems in operation are expected to be spotted and removed. Who is qualified to do this? And what impact will the involvement of controllers at various stages of the development process have on the effectiveness of equipment finally released? Patrick Dujardin (1993) has suggested that early involvement of controllers in the R & D process may discourage important advances, since controllers will feel comfortable only with equipment that seems familiar. There is broad agreement, in fact, that early involvement of working controllers is likely to lead to compromises or kludge designs. The regular controller is unlikely to want to "push the envelope." Many observers have remarked that equipment used by the FAA, both currently and in the near future, reflects this conservative attitude.

A test controller, however, would not share the same bias against new equipment. Note that this is a different role from the evaluation of finished systems. The "test controller" would be used to seeing equipment in raw form, just as a test pilot would be. Again following the analogy with airplane development, a test controller would have to be recognized as a top practitioner, with the respect of other controllers. Such a person's certification of the equipment to the working controller, then, would be one guarantee that the equipment, if not trouble-free, would be at least safe and efficient to use.

New hardware and software now face stiff resistance if they originate from someone other than facility automation specialists. For instance, FAA-produced software for airport ATC systems has many credibility problems. It is not perfect, and in any case needs to be customized

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PRECEDING PAGE BLANK NOT FILMED

PAGE 220 INTENTIONALLY BLANK

to handle site-specific problems. Currently, having capable controllers is the best guarantee against software's inadequacies. Most controllers use "so-so software developed by someone" as Jim Schmidt of Martin Marietta puts it. Controllers must carry on from where the software leaves off, bridging between it and the operating situation. Potentially, certification by a "test controller" that new equipment satisfies human factors requirements might give controllers the confidence they need to master complex new equipments and procedures.

Another important feature of the test controller is checking out the "far corners of the envelope." In the R & D process, early efforts are focused on getting the system to work. But test controllers need to try to make it fail, to exercise it, as it were, beyond ordinary limits, to eliminate the hidden bugs. Ideally, of course, a better process would be developed for getting error-free software. In real life, however, automated systems are likely to possess "glitches" difficult to eliminate. For instance, a recent *Wall Street Journal* article reported that the Honeywell autopilot installed in Boeing 747's behaved in mysterious ways. The FAA noted about 30 incidents, including a recent near-crash over Thunder Bay, Canada, involving malfunctioning autopilots. Experts have been unable to isolate the fault (Carley, 1993). Thus, test controllers need to check the system out using "impolite" actions. This is very similar to what Sir Karl Popper recommends in his book *The Logic of Scientific Discovery:* propound bold hypotheses, and then give them severe tests (Popper, 1961).

Before we go further, however, we need to consider the innovation process itself, since a test controller will have to fit into it. Innovation in the United States Federal Aviation Administration is a very problematic process. We need to examine it in a bit more detail before going further.

Innovation: A Long Haul

Charles Franklin "Boss" Kettering once said that "getting a new idea into a factory is the greatest durability contest in the world." He might have said this about Air Traffic Control. The current process by which new ATC equipment is introduced is slow and inefficient in many ways.

1) There are excessive delays, on the order of a decade or more, from the time new equipment is developed until it is actually used. Thus, by the time the equipment is installed, it has usually been obsolete for years. It may nonetheless represent a real advance over what was used before. Many airports function with equipment that controllers think properly belongs in museums. This is demoralizing both to innovators and operators.

2) An incremental approach is used. This approach forces new equipment to be compatible with current equipment, often leading in the end to an inelegant, "kludge" design, rather than an optimal re-design from the ground up. While each new piece of equipment may function well on its own, nothing guarantees either its compatibility or lack of redundancy with current equipment.

3) Use of political fiat is sometimes used to impose "quick fixes" that need better testing before implementation takes place. In many cases these programs fail to work as planned and thus increase barriers to further innovation. Problems include failure to introduce new equipment effectively, to take learning curve considerations into account, appropriate check-out by "test controllers," and well-conceived instructional methods.

4) Very high stakes are involved in securing government contracts, leading to intense struggles on the part of private firms to get their product accepted. Because competing parties often resort to legal action to block or reverse decisions already made, delay is common. As with defense contracting, the long haul involved and the "winner-take-all" outcomes often result in selection of contractors who are good at lasting through the many rounds necessary to win a contract; these are not necessarily the contractors with the best systems.

The current system is designed to include three parties: private firms that do the actual hardware development; FAA higher officials, who make decisions about which devices to install; and controllers, who will actually use the devices, once they are officially accepted. In principle, controllers develop needs, these needs get expressed as FAA requirements, and private industry responds to the requirements by hardware or software innovation. But there is a built-in paradox. The paradox is that controllers do not know what they can ask for until they know what can be developed. Vendors, on the other hand, often do not understand what controllers need. FAA higher authority, trying to bridge the gap between needs and products, is hemmed in on one side by political and legal constraints and on the others by vendors jockeying for contracts and FAA facilities fearful of clumsy automation.

Something of the complexity of bringing a new system on line is revealed by the failure of the IBM® Advanced Automation System to be implemented on schedule (Burgess 1993). The Advanced Automation System (AAS) will cost something over $4 billion to develop in a joint effort between IBM and the FAA. It will replace the current generation of mainframe-generated pictures for controller positions with personal computers, and will offer much more flexibility. But the FAA continually revised the specifications, and IBM seems to have created its own delays. The system will require something like 1.6 million lines of code, about the same order as "Star Wars." The specification documents themselves would form a stack about three and a half feet high. In basic terms, the contract involves the normal delays and overruns of the typical big-ticket military weapons system project. This is not a good sign.

Since the system has long delays, attempts to get around the normal channels are common. As Lee Paul (1979) has written, "The number of years required by an orderly development process results in irresistible pressures to bypass the system." This often leads to ill-considered moves, including "designs by fiat" that not only fail but also prejudice future attempts at innovation. The formal system also largely ignores the automation specialists (see below) and other members of the system who often have excellent ideas, but who are not considered partners in the innovation process. On the other hand, there has been long-term involvement of controllers on work teams.

These complex dynamics do not bode well for getting the right control equipment to the right people in the right time frame. A top priority for the FAA might well be to examine its own innovation process,

Patching It Up

Ironically, the controllers often seem to do better themselves through informal networking when it comes to customizing ATC software. While some software can be originated locally, hardware on the other hand necessarily is produced off-site and centrally tested and introduced. Still, because each site has slightly different requirements, software can be generic only up to a certain point. Beyond this point, software must be customized for the specific site. This is done through "patches": software oriented to site-specific problems, and written by local "automation specialists." For instance, at Detroit Metropolitan Airport (DTM) the FAA-produced A-305 ARTS-III software (introduced in 1993) was voluminous. Documentation for the software was four large volumes, each a folio volume the size of a desk encyclopedia. The cross-reference book alone weighed 20 pounds. Yet the software required 190 patches to adapt it to local conditions. The automation specialists on the staff estimated that development of this supplemental software required several months to complete, exclusive of already existing site-specific software, which also had to be changed. In principle, software received from the FAA is supposed to be implemented "as is." The reality is that this cannot be.

At Chicago O'Hare Airport, for instance, there was at one time a rule that strictly limited the number of local patches. This rule, however, did not make sense, and so was constantly bypassed. Each time local controllers asked for a specific change, the patch would be added to an existing patch, which clearly flouted the spirit of the rules, though superficially legal. This bypassing of the system was never formally acknowledged. Nonetheless, the local programmers thought FAA officials must have been aware of it.

While sites are often different, many sites share the same problems. The obvious thing, then, is to make sure that a site has available to it any patch in the system that will solve its problems. Although all patches used anywhere are included on a list sent to all "automation specialists," this list is seldom seen by controllers, and is hard to interpret in any case.

There are about 200 automation specialists in the United States. They are former controllers now responsible for software management at the control centers. They are expected to act as the local interface between the needs of controllers on one hand and the provision of new automation through borrowing patches or getting local technicians to do the programming. However, they usually have their hands full with programming responsive to demands by the local controllers for various kinds of minor fixes. One major airport had a list of 30 such patches waiting to be programmed; this is fairly typical. These demands are often either made by top management or presented through union channels, which makes them hard to ignore. Useful patches, then, may often get lost in the system's complexities because they are not available in a user-friendly way.

The automation specialists have more credibility because they are former controllers. Knowing the job that the controller must do in some detail makes their products far more user-friendly than it might otherwise be. But few have college degrees in computer science. Some do not have college degrees at all. They get considerable in-service training from the FAA, but both they and others believe that their programming would be superior if they had more computer training.

No one knows how much of the innovation in the system is actually due to the local automation specialists. For instance, a program called Cenrap allows the local facilities to get radar screen pictures even if their antenna goes out, by getting information on plane positions from more powerful Regional Center (ARTCC) radars. This program was reportedly suggested in about 1985 by either an automation specialist or a technician who realized that capabilities already in use, with a little extra work, could provide back-up radar pictures for facilities that lost their radar but not their system (ARTS-III) software.

Local Content

To compensate for the system's inadequacies, informal networking often provides the primary channel for patches to travel from one center to another. In the FAA southern region, for instance, an informal computer bulletin board provides information about patches. Other information comes about through individuals who move from one site to another, or who through union duties or curiosity circulate through the system. Actually watching a patch in operation may be more valuable than reading an abstract about it.

Yet informal networking clearly is a second-best to a user-friendly system for spreading information about patches. Why is there not a dedicated "patch specialist" who knows what is available who travels through the various sites?

Similarly, site development of patches is often done partly *sub rosa*. Legally, patches must be run through Washington D.C. for regulatory approval before they are used. The formal approval process takes about a year. After approval, the patch is sent back to the site for on-line testing. But local automation specialists do not want to send a patch through the system until they know it works. And how do they know it works? They try it out. To try it out, they need a computer. Often the only computer available for the purpose is the Center's main computer. It would be best to try the patch out on a mock-up computer off-line, but mock-up computers can cost as much as the main computer. So the patch is run on the main computer at a lull time, such as 2 AM. Controllers will almost never try to control aircraft with experimental software. Locking up the system would be both dangerous and would jeopardize their jobs. But without a realistic (i.e., live) test, they do not want to release the software. So while planes are controlled through some other method, the patch is tried out. Once it is known to work, it is sent through the formal process.

[I was unable to gain any information regarding site-generated patches for the regional (en route) centers. Ostensibly, all patches for regional centers must originate from the FAA Technical Center in Atlantic City. To proceed otherwise risks severe legal sanctions.]

Higher echelons of the FAA must know that this kind of covert experimental activity goes on, although they cannot publicly either acknowledge or condone it. However, while obviously better than a paper check-out of the software, this "skunk works" approach has some dangers. One of the problems is that fewer programs result from it than would if it were openly acknowledged. Controllers and automation specialists would both get in serious trouble if they were caught operating with an illegal patch. It would be better if the test were carried out openly. But the best would be creation of what Lee Paul calls a "more forgiving environment," where experiments with patches could be run off-line, a full-scale simulation facility that could be customized temporarily to run a Center's software.

Controllers' experience with innovation has largely been negative. Good ideas by those lower down in the system often seem to get stone-walled or put on the back burner. Ideas that come down from the top are often half-baked or flawed. But the strongest message about innovation is the equipment with which controllers in the U.S.A. are forced to use. State-of-the-art aircraft are controlled by ATC equipment which is often two or three generations out of date. Whether the explanation for this state of affairs is politics, bureaucracy, or sheer conservatism, the message it sends is one of stagnation and indifference. "Good enough for government work" seems to be the limit the controller can expect. Controllers have to be good; their equipment is not.

Test Controllers

The innovation process for new hardware and software occurs along a timeline that can largely be considered in three phases: research and development, preliminary testing, and full-scale deployment. There is a role for controllers in testing new hardware and software in each of these phases.

Research and Development. A role in R & D means that the controller would act in the same role as a test pilot. He or she would encounter new equipment in its formative stages and would be able to help suggest improvements which would move the system from the prototype to the operational stage. The FAA currently is using controllers on its work teams for the new consoles that IBM is developing in Rockville. Some of these controllers have been on the teams for ten years. However, unlike the system for test pilots, there is no way that the experience of these controllers as test controllers is recorded other than in the design of the equipment. They are simply sent back to their centers after they finish their tasks. Note also that at many ATC facilities, local software will be tested by individuals who serve the role of test controller for that facility, even though the term as such may not be used. Donald Pate at the Standards Development Branch of the FAA in Oklahoma City similarly uses journeymen controllers for his experimentation and standard-setting.

One anonymous observer pointed to a problem with the use of ordinary controllers in the R & D process. Early involvement of routine users tends to lower the team's sights, and thus may lead to incremental changes rather than re-design from the ground up. An example is the FAA's use of a "Sector Sweep Validation Team" involving ordinary controllers early on in the process. Ultimately, the console produced by the team was a kludge design, according to this individual. Thus early involvement can lead to dangerous compromises. A "test controller," in principle at least, would have enough experience with the innovation process to be less bothered by radical innovation.

A second problem to which union representative Larry Barbour called attention is the attrition of skill among controllers who are promoted to supervisor. Several individuals noted that supervisors could no longer be considered proficient in acting as controllers, once promoted. During the PATCO strike, many of the supervisors actually had to do some controlling. This was a frightening experience for some of them who had lost their skills. "I remember watching some of these guys with sweat pouring down their backs," said one observer. Yet several of the IBM-design work teams contained a majority of supervisors by the time the project was finished. One work team had one controller and eleven supervisors! This, however, would be a problem for test controllers, too. Some method for alternating actual control experience and innovation activities would be necessary.

Preliminary Testing. In this phase, the overall design of the software or equipment is fixed, and the purpose of testing is to eliminate any remaining "bugs". The role of controllers here is to act as intelligent customers rather than test pilots per se. It is often during this phase also that software can be customized for a particular site. Hugh Bergeron and Harold Heinrichs report their experiences in using controller "cadres" first to test software, and second to act as trainers, both at Denver and at Dallas/Fort Worth. These experiments seem to have been very successful, though the system was not completely ready for them (Bergeron and Heinrichs, 1993).

Bergeron and Heinrich's cadres might be seen as somewhat analogous to the New Equipment Introduction Details (N.E.I.D.) used by the U.S. Signal Corps in World War II. The Signal Corps, finding that newly developed equipment typically was not accepted in the field, developed a kind of special detachment under the leadership of a Lt. Col. Jensen. The detachment included no one without a uniform, and no one ranking above a major or below a sergeant. It always accompanied the equipment from the point of origin (the factory) to the field with no hand-offs. The Signal Corps discovered that this scheme was so successful that it could not get equipment into the field without an N.E.I.D. Unfortunately, the value of this device was not recognized after the war, and so no one studied it. Its only mention in the official history of the Signal Corps in World War II is a tiny footnote.

Full Scale Deployment. During this phase, controllers are still important as intelligent users. Cadres who have been used in phase 2 to work out bugs can act as brokers between laboratory and ATC facilities to transmit information backwards and forwards between designers and users. As equipment and software is given a full-scale test, limitations and bugs will become apparent. Often this may mean moving the novel hard- or software to a new location, with new demands. IBM, using Seattle Regional Center (ARTCC) as a test site, is rumored to have eliminated phone jacks from the controllers' positions which a "D man," used extensively in the busier centers to work computers and assemble flight strips, could use. Protests by Cleveland Center (and others) quickly got the jacks back. Developing effective channels for user feedback is thus very important.

Discussion And Conclusion

Earl Wiener points out that human factors problems fixed during the R & D stage are paid for once. When they are not fixed during R & D, they are then paid for every day. How users are involved in the R & D process to assist in developing equipment is a critical issue. Effective involvement can produce real improvements. Ineffective involvement can produce inefficient kludges or systems that are actually dangerous.

The underlying problem is the management of information and ideas. To develop a really generative system (see Westrum 1993) a great deal would have to change in the way that the FAA innovates. Use of test controllers would solve only some of the problems. For instance, we have cockpit resource management now for pilots; we may have it soon for controllers. But the management of ideas in the innovation process also needs intellectual resource management. Simply involving users is not enough. Brought in at the wrong point in the development process, users can block or compromise innovation. User involvement must be carefully considered. A test controller may be one solution to this problem. It might be necessary to have several kinds of test controllers (en route versus TRACON, for instance). No doubt further problems would surface in getting test controllers into operation.

I would recommend that the FAA engage in a series of case studies of controller involvement in the innovation process. A systematic comparison of effective and ineffective cases would do much to clarify what we ought to do in the future. Unfortunately, I have been unable to find any cases where test controllers have been used. Perhaps we need to create some, to see how they work!

Acknowledgement

Many individuals took considerable time and effort to assist the author in writing this paper. These persons include Larry Barbour, Alex Becker, John Dill, Dick Edsall, Ron Leoni, Don Pate, Lee Paul, Willis Richardson, Jim Schmidt, Steve Walter, Marty White, and above all, Dud Tenney. While these persons contributed, the opinions expressed in this paper are solely those of the author and do not necessarily reflect either the views of the meeting's sponsors or of any individual who assisted in the paper's preparation.

References

Bergeron, H., & Heinrichs, H. (1993). A training approach for highly automated ATC systems. *Seventh International Symposium on Aviation Psychology*, Columbus, Ohio.

Burgess, J. (1993). Out of control contract. *Washington Post*, Washington Business Section, March 8, pp. 1, 22, 23.

Carley, W. M. (1993). Jet's near-crash shows 747's may be at risk of autopilot failure. *Wall Street Journal*, April 26: A1, A6.

Dujardin, P. (1993) The inclusion of future users in the design and evaluation process. *Le Transpondeur*, April, pp. 36-39.

Hallion, R. P. (1981). *Test pilots: The frontiersmen of flight*. Garden City, New York: Doubleday.

Paul, L. (1979). How can we learn from our mistakes if we never admit we make any? *Proceedings, 29th Air Traffic Control Association Conference*, Fall, pp. 194-198.

Popper, K. (1961). *The logic of scientific discovery*. New York: Science Editions.

Westrum, R. (1993). Cultures with requisite imagination. In J. A. Wise, V. D. Hopkin, and P. Stager (Eds.), *Verification and Validation of Complex Systems: Human Factors Issues*. New York: Springer-Verlag.

N95- 34779

Towards a Framework of Human Factors Certification of Complex Human-Machine Systems

S5-54

Birgit Bukasa

Austrian Road Safety Board

Background

The recognition of the importance of human factors to system safety, especially in aviation, is constantly increasing. Foushee (1993), in his keynote address at the International Civil Aviation Organization (ICAO) Flight Safety and Human Factors Symposium, even talked about the "human factors revolution", emphasizing how quickly human factors thinking has infiltrated the world of aviation and high technology in some parts of the world. Consequently, human factors concepts have to become institutionalized into the aviation culture. In order for this to occur, the FAA (in its 1990 National Plan for Aviation Human Factors) amongst others is placing stronger emphasis on human factors as part of aircraft and avionics certification requirements (Foushee, 1993).

This claim for stronger consideration of human factors principles in the designs of complex and integrated man-machine systems, or at least of the human computer interfaces, is the reaction to changes and foreseen further future changes in the aeronautical world. These changes have and will have a great impact not only on the aviation community, especially the operators such as pilots or air traffic controllers, but also on the society as a whole.

Advanced automation in aviation, including its implications for the users as well as for safety concerns, is the main point at issue. It has evoked considerable controversy among the different groups involved, the users, the manufacturers, the scientists and the regulatory agencies. Scardigli (1991) identified antagonistic mental representations of designers, pilots and air traffic controllers concerning their present and future role, their vision of the ideal flight and desirable future changes as major roots for this controversy illustrating the struggle for power in the aeronautical world at the same time.

Besides, aircraft accident/incident analysis in aviation, as well as in other non-aviation environments (e.g. nuclear power accidents, ferry, tanker or train accidents), identify human performance problems, the so-called "human error", as major contributing factor. Taking the expected future air traffic growth into consideration, this would lead, in the next couple of decades, to a major aviation accident every week despite considerable improvements in technology (Foushee, 1993).

Undoubtedly, this very broad categorization called "human error" caused a lot of misunderstandings. Berninger (1991), for example, tried to clarify the role of human error in aircraft accidents by stating that the conclusion of human error only proves that the human

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

could have prevented the accident but not that the human (pilot) caused it. Instead, he argues that the system characteristics working against human performance cause the human to fail.

According to Foushee (1993), there is enough evidence that the automation philosophy – automation being an easy way to remove human error from the system – has to be critically examined and that new, more "human-centered" approaches to automation have to be considered. Berninger (1991) emphasized that systems which are compatible with humans seem to be a promising approach to further system safety improvements.

The idea of human factors certification has to be seen in this context. One approach to reaching the goal of generally more human-centered automation or technology is by putting pressure on system designers to incorporate human factors considerations into the design process.

Definition of Terms

According to the Austrian law of accreditation (Austrian Standards Institute, 1992), certification is defined as the formal certificate of conformity carried out by accredited independent representative impartial third persons or bodies. Certification is a forced legal act based on documented and accepted rules, procedures and processes. It is the end and result of a process of checking whether the subject of certification fulfills defined requirements.

Certification is, above all, a measure of quality assurance, often connected with safety goals. Only those products or systems which have proven their conformity to safety goals are allowed into operation. Thereby, the period of validity of certification might be limited or unlimited according to the subject of the certification.

Human factors certification is just a specification of the subject that has to be certified. It is meant in the sense of certification of human factors. Following this understanding, human factors certification means to certify human factors issues as part of general system certification. Therefore, the above mentioned general definition and principles of certification are valid too.

Problems of Human Factors Certification

At the Congress of the German Psychological Society held in Zurich in 1980 Bischof called for the "Galileo of psychology" meaning that psychology is still in the stage of astronomy at the end of the 15th century (cited after Barglik, 1993). While the natural sciences or technical disciplines progress in a more and more accelerating way, plunging into areas like artificial intelligence, virtual reality or non-linear self organizing systems, the behavioral sciences in general have no substantial impact on research programs determining man's presence or future.

This leads to the question of whether psychology or behavioral sciences in general already have a sufficient level of sophistication in order to carry out human factors certification. Are there ways to asses human factors other than by certification? It might be more fruitful to improve established procedures instead of introducing human factors certification. Especially in the aviation environment, there are established procedures concerning aircraft, accidents, operation and people which might be corrected or extended (e.g., Paries, 1994).

What makes human factors certification so difficult?

One of the particular problems of behavioral sciences are the so-called "soft data." Human behavior is not deterministic but rather probabilistic by nature. Human behavior depends on a lot of conditions from inside the individual and the outside world, from past experience and future plans and expectations, from man's interactions with others but also from social, political and economical conditions and developments. Humans are in a permanent process of adapting themselves to the external world as well as adapting the external world to their needs. Therefore, measuring methods, results and predictions are not that exact in behavioral sciences compared to natural sciences.

This leads to basic questions about human factors certification: which degree of certainty is certain enough, how to define human factors standards, how to consider cultural differences, where to set cut-offs, is it enough to identify components which are negative or is it necessary to distinguish between optimal and still acceptable solutions and what is the price for it, e.g., the loss of safety?

Undoubtedly, focusing on human factors certification is something new. Until now, the activities of the national and international standardization bodies are concentrated on testing, inspection and certification of products, processes, services and quality systems from a machine-centered engineering point of view. Human factors aspects primarily flew into ergonomical standards, e.g., standards for the optimal physical layout and anthropometry of operators' workstations. Concerning standards for quality management and quality system elements (e.g., ISO 9004-2), human factors are restricted to education and training requirements as well as to measures enhancing the motivation of service personnel.

Approach to Human Factors Certification

Starting with human factors certification, there is a great deal of work to do. The what, how, when, where and who questions have to be thoroughly discussed during the workshop, but the workshop was just the beginning. Therefore, this paper cannot present solutions to all the open questions. It can only point to some aspects which seem to be important.

Guidelines of Human Factors Certification

As a first step, guidelines of human factors certification have to be established by a multidisciplinary working group (including human factors experts, designer, user, representative of regulatory authority, expert from Standards Institute). These guidelines have to refer to the prerequisites of certification which are the goals, criteria, measures, methods, processes, standards, testers, test procedures, check and examination system and penalties. It is necessary to keep the specification of these prerequisites general and descriptive, not only in order to be valid for different man-machine applications, but also because of lack of experience with human factors certification considering the problem of soft data as well.

At the beginning, human factors certification should aim at minimum standards and at eliminating poor design. It should not aim at average or maximum standards which are far more difficult to check. This is according to the certification practice in general.

Model of Human-Centered System Issues

In the context of what should be certified, a general model of human-centered system issues which can be applied to different man-machine systems is an important aid. Harwood (1993), for example, introduced an approach distinguishing three broad categories of human-centered system issues which all have to be considered. These categories, which have to be specified according to the context of domain (e.g., ATC, flightdeck, power plant), are technical usability, domain suitability, and user acceptance.

Technical usability refers to perceptual and physical aspects of the human computer interface. Domain suitability refers to the content of information and display representation for domain tasks as well as functionality and decision-aiding algorithms. Finally, user acceptability refers to the ease of use and suitability of the system for supporting cognitive task requirements as well as to aspects of job satisfaction (Harwood, 1993).

Based on such a model, a comprehensive list of relevant human factors requirements can be elaborated for each new system.

Tools for Human Factors Certification

Common tools for certification purposes are handbooks which are available for human factors. Yet, they are not sufficient.

Even if they cover a very broad area of knowledge, human factors is evolving so rapidly that any test more than a decade old cannot do justice to its current state (Kantowitz, 1993). Moreover, considering the specificity of human behavior and the strong influence of the physical and social environment on man's behavior which was demonstrated by the work of Mischel (e.g., 1971, 1973), handbooks are of limited value in order to evaluate new system from a human-centered point of view. General human factors principles have to be adapted, modified and evaluated in regard to the specific application.

Due to the limited value of handbooks, it is necessary to strongly focus on empirical data, revealed by testing and evaluation, especially on validation studies.

Validation was thoroughly discussed at the NATO ASI in Portugal, 1992 (Wise, Hopkin, & Stager, 1993). Based on this, a lot of information is available on how to check and evaluate validation results.

Institutionalization of Human Factors Certification

There are several possibilities regarding how to institutionalize human factors certification. At least from an Austrian perspective, the most realistic way is to establish a collaboration with the existing national standardization bodies which are members of the ISO (International

Standardization Organization) in order to link human factors certification as close as possible to existing certification.

Persons or institutes who or which are qualified should be accredited. One might refer to guides like the EN 45011-13 or the ISO/IEC Guide 40 on general requirements for the acceptance of certification bodies.

Conclusions

As far as total automation is not realized, the combination of technical and social components in man-machine systems demands not only contributions from engineers but at least to an equal extent from behavioral scientists. This has been neglected far too long. The psychological, social and cultural aspects of technological innovations were almost totally overlooked.

Yet, along with expected safety improvements the institutionalization of human factors is on the way. The introduction of human factors certification of complex man-machine systems will be a milestone in this process.

References

Austrian Standards Institute (1992). *Law of Accreditation.* Vienna: Austrian Standards Institute.

Barglik, W. (1993). Can the theory of nonlinear systems give input to models of development of tests? In B. Bukasa & U. Wenninger (Eds.), *3.ART-90 Workshop.* Vienna: Austrian Road Safety Board.

Berninger, D. J. (1991). *Understanding the role of human error in aircraft accidents.* (Transportation Research Record 1298, 33-42).

EN45011-13 (1989). *General criteria for certification bodies operating product certification.* Brussels: European Community.

Foushee, C. (1993). The human factors revolution: Meaningful change or temporary infatuation? In ICAO (Ed.) *Human Factors Digest No.9. Proceedings of the Second ICAO Flight Safety and Human Factors Global Symposium* (ICAO Circular 243-AN/146, 11-30).

Harwood, K. (1993). Defining human-centered system issues for verifying and validating air traffic control systems. In J.A. Wise, V. D. Hopkin, & P. Stager (Eds.), *Verification and Validation of Complex Systems: Human Factors Issues, NATO ASI Series.* Berlin: Springer-Verlag.

ISO/IEC Guide 40 (1983). *General requirements of the acceptance of certification bodies.* Geneva: International Standards Organization.

ISO 9004-2 (1991). *Quality management and quality system elements--guidelines for services.* Brussels: European Community.

Kantowitz, B. H. (1993). Human factors knowledge requirements for flight crews. In ICAO (Ed.), *Human Factors Digest No.9, Proceedings of the Second ICAO Flight Safety and Human Factors Global Symposium* (ICAO Circular 243-AN/146, A-29-34).

Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, *80*(4), 252-283.

Mischel, W. (1971). *Introduction to personality*. New York: Holt, Rinehart & Winston.

Paries, J. (1993, July). *Some inadequacies of current human factors certification process of advanced aircraft technologies*. Paper presented at the International Workshop on Human Factors Certification of Advanced Aviation Systems, Bonas.

Scardigli, V. (1991, October). Automation in aeronautics: A French research program. In M. C. Dentan & P. Lardennois (Eds.), *Human Factors for Pilots. Report on the XIX Conference on the Western European Association for Aviation-Psychology.*

Wise, J. A., V. D. Hopkin, & P. Stager (Eds.), *Verification and Validation of Complex Systems: Human Factors Issues. NATO ASI Series.* Berlin: Springer-Verlag.

Reflections of Certification in Aviation

N95-34780

The Successful Management of Programs for Human Factors Certification of Advanced Aviation Technologies

51-54

Rod Baldwin

Baldwin International Services

Introduction

In recent years there have been immense pressures to enact changes on the air traffic control organisations of most states. In addition, many of these states are or have been subject to great political, sociological and economic changes. Consequently, any new schemes must be considered within the context of national or even international changes.

Europe has its own special problems, and many of these are particularly pertinent when considering human factors certification programs. Although these problems must also be considered in the wider context of change, it is usually very difficult to identify which forces are pressing in support of human factors aspects and which forces are resisting change.

There are a large number of aspects which must be taken into account if human factors certification programs are to be successfully implemented. Certification programs would be new ventures, and like many new ventures it will be essential to ensure that managers have the skills, commitment and experience to manage the programs effectively. However, they must always be aware of the content, and the degree of certainty to which the human factors principles can be applied – as Debons and Horne (1993) have carefully described.

It will be essential to avoid the well known pitfalls which occur in the implementation of performance appraisal schemes. While most appraisal schemes are usually extremely well thought out, they often do not produce good results because they are not implemented properly and staff therefore do not have faith in them. If the manager does not have the commitment and interest in his/her staff as human beings, then the schemes will not be effective.

Thus, one aspect of considering human factors certification schemes is within the context of a managed organisation. This paper outlines some of the management factors which need to be considered for the air traffic control services. Many of the points received attention during the plenary sessions while others were covered by the working groups when the question arose of how various aspects of human factors certification programs would be managed.

Management and organisational issues will certainly need to be included in any frame of reference by those who may be involved in developing certification programmes.

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PRECEDING PAGE BLANK NOT FILMED

PAGE 236 INTENTIONALLY BLANK

Definition of Human Factors

The concept of human factors issues is broad and still somewhat vague as the subject tends to include any aspect of human behaviour. However, experience in the design, operation and maintenance of large and advanced systems has shown that there is a human element which needs to be more carefully considered if the system is to perform as required. Unfortunately, there are still too many glaring examples of poor human factors aspects, and delegates to this workshop described numerous examples.

A formal definition of human factors is difficult, but many groups, such as The European Study Group for Human Factors in ATC (1991), have accepted the MANPRINT (1991) definitions of human factors as a useful guide:

- Manpower
- Personnel
- Training
- Human factors engineering
- System safety
- Health hazard assessment

Attitudes

There is no doubt that most management and certainly most employees in air traffic control have negative attitudes to the concepts of human factors and human-management techniques. Most instructors of management courses, especially those for technical personnel, know the problems which arise when it is suggested that there are theories for dealing with people. There is an initial suspicion that the instructor is suggesting that the theories will provide answers for dealing with Life, the Universe and Everything (Adams, 1982). It takes time to persuade them that the theories serve as a framework for putting the "human" problems into context and some theories can provide guidelines for dealing with the problems. When the term "psychological theory" is introduced, unease appears due to the trainees not appreciating the difference between psychologists and psychiatrists. The former implies that the other person is the problem, but that is OK, while the latter implies that I am the problem, and that is certainly not OK!

Such attitudes are varied between types of industry and the type of staff employed for certain types of work. Again, this varies from country to country and it is interesting to see the different responses to situations, such as those described above, by controllers and technicians from the various European countries in comparison to those from other countries.

It would be interesting to study the more highly educated leaders of the air traffic services, and see how receptive to human factors certification ideas they are, then compare them with those of more modest academic levels of achievement. Any such assessment would, of course, run into the problems of cultural issues. The latter aspects consistently arose during discussions, and there is no doubt that much more needs to be done in this area with respect to ATC if European integration is to be achieved successfully.

The matter will certainly need to be addressed if such schemes are to have support from the tops of the organisations. With the decline of the renaissance man there are now too many managers who confuse the issue of human factors with the old themes of the Humanists and their suggestion that achievers are not sufficiently interested in the human being.

A regular comment at the workshop was the difficulty that human factors experts had in trying to communicate with top management. In practice, much of the misunderstanding and confusion is a result of fear that manifests itself in outright antagonism to acceptance of human factors issues. But of what are they afraid? Is it that the human factors experts are seen in the "parent" role? Perhaps they are identified with those other "parent" figures, the teacher/instructor/ lecturer/professor, who knew them so well and perhaps knew that they were not up to the job?

As usual, the writers of fiction are ahead of the managers of reality, as novelists can express their views without having to substantiate their comments or deal with the practical issues. Nevertheless, the comments can be valid and pertinent as is shown in the following quotation from Thomas Mann's "The Magic Mountain" (1960) where Settembrini says:

> We humanists have all of us a pedagogical itch. Humanism and schoolmasters – there is a historical connection between them, and it rests upon psychological fact: the office of schoolmaster should not – cannot – be taken from the humanist, for the tradition of the beauty and dignity of man rests in his hands. The priest, who in troubled and inhuman times arrogated to himself the office of guide to youth, has been dismissed; since when, my dear sirs, no special type of teacher has arisen. The humanistic grammar-school – you may call me reactionary, Engineer, but in abstracto, generally speaking you understand, I remain an adherent –

If we are to convince antagonists, we will need to answer such questions as: have managers rejected the values and urgings of the priest, will human factors experts have to overcome consequent feelings of guilt and is antagonism based on a feeling that if the human factors expert was good enough he/she would be in a line managerial job?

If the above ideas are correct, then the human factors experts will have to:

- Mount a massive public relations exercise to convince the line management staff that human factors issues are important,
- Give concrete evidence to line management staff that human factors schemes will increase performance, quality, capacity, etc.,
- Be prepared to give firm advice and say if human factors aspects are not satisfactory.

Organisational Tasks

Strictly speaking, the whole organisation should be analysed as a total system of interacting parts. However this does require great effort and it is questionable whether such a large "picture" is indeed meaningful.

However, this question may be answered by the present *National Plan for Aviation Human Factors* of the FAA (1990). The program is certainly comprehensive and realistically aims to put implementation responsibility down into each line unit of the organisation.

It is often more appropriate to select a few key tasks of the organisation, starting with the end user of the system, and concentrate on the most significant interactions.

For instance, what do air traffic controllers actually do and how do they do it? When these aspects are established, we can provide better management for the activities in general and ensure that they receive all aspects of information support which they need – as opposed to what they think they need.

At present, there are a number of programs which are intended to address these matters. Unfortunately there was no mention at the workshop about the FAA's efforts with the SASHA programme, those undertaken for Eurocontrol by PA Management consultants as part of the EATCHIP programme and the studies, for the CAA of the United Kingdom being carried out at Roke Manor Research by Day (1993), and others.

However, the definition of the tasks must be related to the nature of the organisation, the type of management, and the leadership styles. The effect of the latter on the other aspects has been extensively studied (Handy, 1985) with particular attention given according to directive against consultative styles of management.

It is interesting that most staff seem to believe that a more consultative style of management will produce the best fits for achieving task requirements in conduction with the needs of the organisation, the team, and the individuals. That this is not necessarily so was indicated by Fiedler (1967), who has shown that both styles can be effective if used appropriately. He also states that an individual's style cannot be changed through training, as style is a stable characteristic of the individual. However, this view is refuted by Vroom (1973), who believed that a leadership development programme can enable a person to widen his or her range of management styles and so be more appropriate to a particular situation.

There is a need for a study of this contingency approach to leadership and management in the air traffic services in order to establish more understanding about the relationships between the different tasks. A very subjective attempt to place these factors onto a bar chart (Figure 1) shows up the wide misfit which can occur, and possibly explains why there are so many problems in the air traffic control services. These occur despite the evidence that controllers actually enjoy the detailed workings of their job, and that their monetary rewards are generally towards the higher end of the remuneration scale.

Organisational Structure

The type of organisation will play a significant role in how a certification programme is carried out. Unfortunately, many organisations have not fully clarified their objectives, how they are to be achieved, and how the various elements of the organisation relate to each other. Thus, misunderstandings and dissatisfaction will occur when an organisation attempts to introduce new schemes without having a clear understanding of the context into which the new scheme must fit.

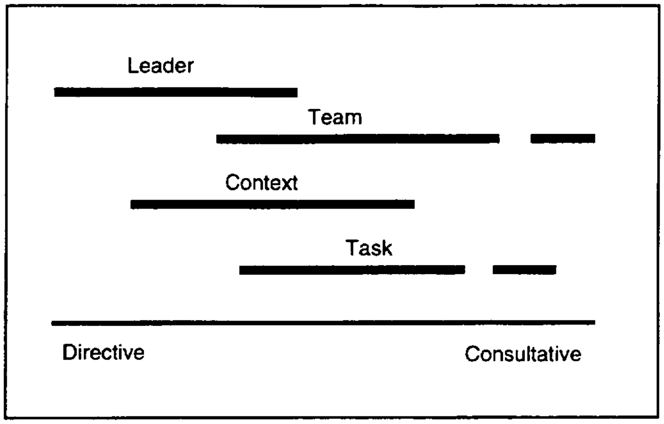


Figure 1. Management styles in ATC

Organisations can be classified in many different ways depending on performance or other criteria which need to be analysed. Burns (1968) distinguished three forms of organisation as mechanistic (i.e., bureaucratic), organistic (i.e., flexible), and pathological. Westrum (1993), however uses the more colloquial terms of normal, healthy, and sick!

Modern practice is to consider that individuals and teams can operate more effectively and with more motivation in organic style organisations as opposed to the more mechanistic forms. Thus, organisational development should start with a careful definition of the task or tasks to be achieved, and then consideration should be given to the individual and team requirements for maximum efficiency in achieving the task.

Most air traffic control organisations, however, have traditionally had a mechanistic structure, even, to some extent, with a slightly militaristic attitude. Now, many of them are (or have done so) attempting to change (with varying degrees of success) into organic forms in order to meet modern task and personnel requirements.

One of the dangers in attempting to change an organisation is the problem of becoming stuck in the pathological form. This can often come about by upper management trying to move the organisation from the mechanistic to the organic, but without wishing to lose the benefits of the former. Of course, "benefits" can imply material benefits or intellectual ones which provide a comfort zone based on previous experience.

This situation is particularly evident where government controls the organisation. This was recently seen in the case of Aer Lingus (1993) being required to become a realistic commercial operation but at the same time being directed by the government to operate its transatlantic flights via Shannon, which increased costs.

The present interest in privatisation – or more correctly "commercialisation" – of the air traffic control organisations has provided additional pressure for a move to more organistic organisational forms. But this in turn has necessitated a change in attitude among many staff who were formally civil servants. Unfortunately, in some cases a new problem has occurred through staff sectorisation because some of the changes have been more beneficial to certain personnel than others. Consequently, many of these organisations may be actually moving into the pathological state while offering a window dressing of go-getting private attitudes!

The Team

In any organisation, there are many teams operating at different levels. These are mostly interdependent but effort has to be expanded by the management to ensure that each understands its role and its relationship to the other teams.

Definition of the organisation's task should start with the end user team, but this should not necessarily imply any superior position. In fact, each team should see itself as an end user of another teams product and a supplier of a product (perhaps as in-house consultants) to another team or teams.

However, superior teams will arise if certification programs are introduced for some personnel but not for others. Initially there may be acceptance of the situation, but whereas equipment has an accepted life span and generally keeps to that if the original specifications have been adhered to, humans have a habit of changing according to circumstances. Thus, certified but poorly performing individuals introduce accentuated problems into the teams.

Hence, certification programs must give close attention to team composition and the shelf-life of the certified personnel! This therefore introduces the need for currency checking and the consequent costs of maintaining such programs. In addition, consideration must be given to the degree to which the certification process can slow down an operation and whether the process is counterproductive to ensuring that an organisation is adaptable to changing requirements as Hancock (1993) strongly urged.

The role of human factors auditing programmes needs much more consideration as a way to overcoming many of the disadvantages of certification as this process would also leave line managers in control of their operation.

Regulation

In Europe, there are a wide variety of certification standards and training schemes, and so mobility of staff is either difficult or impossible in many cases. There are two possible solutions if common acceptance is to be achieved:

- Each state could recognise the qualifications of personnel from other states; how the training was performed is the concern of each state. This, of course, requires a careful definition of what each qualification allows the holder to do and the context in which the task can be performed.
- The qualifications and training programmes are integrated throughout the European states. This requires very careful definition of the knowledge and skills which are required for the qualification, establishing other conditions which must be applied, and how standards are specified, set, and checked throughout the large number of states concerned.

At the moment, there appears to be pressure in Europe for the latter solution and a feeling that the problems can be overcome. The problem of maintaining common standards for training air traffic controllers was addressed by Baldwin (1988) some years ago, when he suggested the establishment of external examiners, where "external" means from another state. Initially, the external examiner would only act as an observer, but offer comments on how a similar situation would be addressed and dealt with in his/her own organisation. Naturally, there would be problems because each ATC system has grown up in its own way and comparisons are not very easy. However, as Europe moves towards an integrated system, then a common certification program becomes feasible.

A major requirement for these external examiners will be that they have a strong sense of tact and a serious desire to study the other country's system, appreciate how it is structured, how it operates, and the context of the operating system.

A similar scheme for the certification of aircraft is presently run by the Joint Aviation Authority (JAA), which is an associated body of the European Civil Aviation Conference (ECAC) through its MAST program. This consists of several expert groups being formed from staff of the member states but with a briefing from the Technical director of the JAA. These groups, in fact, are integration groups for the maintenance of standards and are therefore very welcome by each member state.

The JAA scheme is appropriate as aircraft are common to the various airlines of the different countries, and thus common requirements are relatively easy to apply. The same conditions will not apply to the air traffic control until there is an integrated system. However, a start will need to be made and there is no reason why some form of central bureau should not be established.

In fact, during the workshop an interesting exercise compared possible requirements for the certification of advanced air traffic systems with those presently applied to the certification processes of developing a new aircraft. This showed that much more thought needs to be given to which stages require the involvement of human factors specialists, to what extent they should be involved and who the specialists should be. In short, there appeared to be a need for a progressive change from human factors certification to checking, and then auditing as an overview.

Of course a major problem is that present organisations such as Eurocontrol, ECAC, the European Commission and the JAA need to redefine their terms of reference with respect to what they can best achieve, what they are good at, what authority they can wield and how they relate to each other.

The Leader and Technological Change

At one time, the leader could develop his or her technical skills, expand into leadership skills and then apply them until retirement. However, with the present rate of change, this method of working is no longer possible and, in fact, can cause obstruction to new schemes from the top.

Although this phenomena is not particularly new, it is so in terms of the rate of change together with the need to place more emphasis on the human factors elements. However, the workshop did bring out the fact that whilst many managers are expert in dealing with the component parts of systems, they might not be expert in dealing with the special aspects of the whole system – especially when it is advanced and large. As certification of the system therefore becomes particularly difficult, we will need to reconsider who should be carrying out the certification process, the methodology used and who endorses the certificates.

Toffler (1970) made an impact with his book *Future Shock* at the beginning of the seventies. In one section, entitled "Taming Technology" he noted:

> Given that a majority of men still figuratively live in the twelfth century, who are we even to contemplate throwing away the key to economic advance? Those who prate anti-technological nonsense in the name of some vague 'human values' need to be asked 'which humans'? To deliberately turn back the clock would be to condemn billions to enforced and permanent misery at precisely the moment in history when their liberation is becoming possible. We clearly need not less but more technology.

While many of these terms would need to be changed depending on the topic of discussion and the environment under consideration, there is no doubt that the views still echo much of the thinking in the air traffic services. Certainly, at the general discussion of the 1992 Advanced Study Institute (Wise, Hokin, & Stager, 1993) there were many views expressing resistance to the introduction of advanced automation. At times, there appeared to be a mind-set on the role of humans and whether they should adapt to automation. The main problem was the mind-set. Are we locked into set opinions, politically correct thoughts? Will these proceedings be able to break out of this straightjacket?

For this workshop the constraint for consideration was the degree to which human factors experts would commit themselves to certification because of the question concerning validity, or just not wanting to commit themselves in the way that operational staff have to. To some extent the matter was addressed, but there is still a suspicion that discussion and advising is easier that giving firm recommendations!

Westrum (1991) has taken the theme further (chapter 8 of section 3, Originators and Managers of Technology) by analysing the intellectual resistance to innovation in terms of failures of imagination and failures of nerve. The examples given from military operations make useful parallels for air traffic where safety is the priority factor.

However, at a recent ATCA Conference (General Discussion, 1993), the message from many speakers was that we now have enough technology to meet whatever technical task is required. What is needed now is the individual and team ability to use the technology. Baldwin (1991) pointed out that if the human element is ignored in the present massive European expenditure on new air traffic control systems and equipment, then the result might be no air traffic control capacity increase at all!

Conclusion

The corner has been turned, but the human factors experts must conduct their work and present their results in forms which are readily understandable to the leaders, managers, engineers and operators. That is, the Human Factors experts must now study themselves and how they relate to their clients!

This message has been made several times in this paper and it was made strongly on the last morning of the workshop.

References

Adams, D. (1982). *Life, the universe and everything.* New York: Simon and Schuster, Inc.

Article (June 28-July 11, 1993), Aer Lingus Moves to Sell Hotels, Other Businesses, *Commercial Aviation News*, U.S.A.

Article (June 28-July 11, 1993), Europe Without Borders, *Commercial Aviation News,*U.S.A.

Baldwin, R (1988), *ATC 2000,* Eurocontrol Institute, Luxembourg.

Baldwin, R (1991), *Humans are the limit!*, Flight International, 28 August-3 September.

Burns, T., & Stalker, G. H. (1968). *The Management of Innovation.* London: Tavistock.

Day, P. O., Hook, M. K., Warren, C. & Kelly, C. J. (1993). The modeling of air traffic controller workload. *Workload Assessment and Aviation Safety.* Royal Aeronautical Society Conference (27/28 April), London, U.K.

Debons, A., & Horne, E. H. (1994). Information system certification: Purview, perspective and projections. In J. A. Wise, V. D. Hopkin, & D. J. Garland (Eds.), *Human Factors Certification of Advanced Aviation Technologies.* Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Federal Aviation Administration. (Draft, 1990, December). *The National Plan for Aviation Human Factors.* Washington, DC: U. S. Department of Transportation.

Fiedler, F. E. (1967). *A Theory of Leadership Effectiveness.* McGraw - Hill.

General discussion. (1992). *Verification and validation of complex and integrated human-machine systems.* Vimeiro, Portugal: NATO Advanced Study Institute.

Hancock, P. A. (1994). Certifying life. In J. A. Wise, V. D. Hopkin, & D. J. Garland (Eds.), *Human Factors Certification of Advanced Aviation Technologies.* Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Handy, C. (1985). *Understanding Organisations.* London, U.K.: Penguin Books.

Helms, J. L. (1993). *Air Traffic Control Association (ATCA) Conference.* Geneva, Switzerland.

Mann, T. (1924). *The Magic Mountain* (Translation). Penguin Books. (Original work published 1960).

The European Study Group for Human Factors in Air Traffic Control (1991). *Terms of Reference.*

Toffler, A. (1970). *Future Shock.* London: Pan Books.

U.K. Ministry of Defense Army Department DTI. (1991). MANPRINT. London: HMSO.

Vroom, V. H. & Yetton, P. (1973). *Leadership and decision-making.* University of Pittsburgh.

Westrum, R. (1991). *Technologies and society.* Belmont, CA: Wadsworth Publishing Company.

Westrum, R. (1993). Cultures with requisite imagination. In J. A. Wise, V. D. Hopkin, & P. Stager (Eds.), *Verification and Validation of Complex Systems: Human Factors Issues.* Berlin: Springer-Verlag.

Wise, J.A., Hopkin, V.D., Stager, P. (1993). *Verification and Validation of Complex Systems.* Berlin: Springer-Verlag.

N95- 34781

Evaluation in Context: ATC Automation in the Field

Kelly Harwood & Beverly Sanford

Sterling Software

Introduction

Certification is defined as "attesting as certain" (Flexner & Hauck, 1983). "Certainty" however may be a rare commodity when the introduction of new technology into an existing system can "...destroy the blanket of established know-how" (Rasmussen & Goodstein, 1988; p. 179). It is impossible to foresee all emergent properties and interactions between system components and their implications. A complete set of requirements and criteria for safe and efficient system functioning is difficult, if not impossible, to define in advance of system implementation. Once the system is in an operational environment, requirements may need to be rejuvenated due to our imperfect foresight and lack of understanding. Christensen (1958) has referred to this dilemma as the "omnipresent criterion problem."

One way to tackle this dilemma is to incorporate field testing early in the system development cycle. This paper describes the field assessment process that has been applied to the development of an advanced ATC automation system, the Center/TRACON Automation System (CTAS). Field testing provides insight into the true characteristics of the system; that is, how it actually operates and any emergent properties as a function of being integrated into the operational environment. Such insight provides guidance for capturing and refining meaningful requirements for system verification and certification. By delaying field testing until late stages of development, solutions to design problems are likely to be technology driven with validation, verification, and certification relying on context-free guidelines for human-computer interaction.

Field testing conducted early during the development and demonstration phase of system development affords exploration of the user's experiences with the system in the context of their work domain. It provides the opportunity to understand the implications for system design of the interdependencies between the physical environment (lighting workplace layout), task domain (goals/functions of the domain) and work activities (social aspects of team coordination; sources of motivation and job satisfaction). The richness and complexity of these context-based factors and the relationships between them is not accessible through design guidelines or standards. Guidelines and standards cannot provide insight into effective design solutions when system performance is highly contingent on context (Meister, 1985; Gould, 1988). Early field testing promotes the development and validation of a tool as a problem-solving instrument (Woods, Roth, & Bennett, 1990), thereby increasing the likelihood of a match between the system's capabilities and its context of operation (Rasmussen & Goodstein, 1988; Bentley et al., 1992).

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PRECEDING PAGE BLANK NOT FILMED

PAGE 246 INTENTIONALLY BLANK

The FAA TATCA Program recognizes the importance of early field testing for the development and validation of advanced ATC automation. It is presently using rapid prototyping and early field exposure as part of the development of CTAS, using on-site system evaluations with active controllers and representative traffic flows and conditions. Iterative field testing is regarded as integral to the development process, with the objective of achieving a match between the system and context for its use. This approach deviates from traditional approaches to ATC system development and will expedite a possible national deployment of CTAS. Embracing the context of the ATC domain is particularly important because of our limited knowledge of the impact of advanced information technology on controller/team job performance and the stringent requirements for maintaining ATC system safety and continuity during system transition (Harwood, 1993).

The first section of the paper provides a brief description of CTAS, followed by an overview of the field development and assessment process in the second section. In the third section, particular attention is paid to the *structured* assessments of CTAS. These assessments take a principle-driven approach, drawing on principles, perspectives, and methods from human factors engineering, cognitive engineering, and usability engineering. Activities are described that include the identification of human-centered system issues to help guide the collection and interpretation of data, method selection and tailoring, data collection, data analysis, and interpretation. Examples are provided of the types of findings that are a consequence of this development and assessment process. The fourth section discusses requirements definition and rejuvenation. This paper is not a comprehensive review of all possible methods that could be used, but rather a description of those that have been applied in tailoring a process to bring CTAS functions to a level of stability and usefulness. Emphasis is on the mechanics of executing the process, with mention made of the nuances of conducting development and assessment at an operational field site.

CTAS

CTAS is an integrated set of automation tools, designed to provide decision-making assistance to both Terminal Radar Control (TRACON) and Center controllers via planning functions and clearance advisories. CTAS consists of three sets of tools: the Traffic Management Advisor (TMA), the Descent Advisor (DA) and the Final Approach Spacing Tool (FAST). CTAS development has involved thousands of hours of laboratory simulation with controllers to refine and extend algorithms and to enhance the user interface. In order to bring system functions to a level of operational stability and to provide information to air traffic and system development organizations on a possible national deployment decision, further development and validation is being conducted at FAA ATC field sites. TMA is the first CTAS component to undergo the field development and assessment process and will be the focus of discussion for this paper. (For further information on CTAS see Erzberger & Nedell, 1989; Tobias, Volcker, & Erzberger, 1989; Davis, Erzberger, & Green, 1991; ATC Field Systems Office, 1992; Erzberger, 1993.)

TMA has been developed for use by the traffic manager at traffic management units within Air Route Traffic Control Centers and TRACON facilities. Unlike controllers, traffic managers do not control traffic directly. Instead, they monitor the demand of arrival traffic into the center and terminal areas, coordinating with TRACON, center, and tower personnel, making decisions to balance the flow of traffic so that traffic demand does not exceed airport and airspace capacity. Traffic managers use information about the arrival flow to decide whether the traffic should be delayed or metered, to distribute the load from one area to another, and to assign departure times for aircraft departing airports within the center's airspace that will enter the arrival flow for the metered airport. Information about the traffic situation is accessed from multiple sources, such as flight strips, weather displays, operational personnel, and aircraft situation displays. Often, there is no steady state in the traffic flow; the location of a single *heavy* aircraft can disrupt the scheduled flow of traffic, as can poor weather, equipment outages, and emergencies. Given the extent of coordination required, the variety of sources of information accessed, and the dynamic and often variable state of the traffic flow, context through early field testing is crucial to ensure the robustness of TMA and its effective integration into the traffic management unit.

Representations of traffic flow are conveyed on the TMA by configurable moving timelines. Aircraft data tags move down the timelines and are color coded to portray landing schedule and sequence status information. The traffic manager can override TMA's automatically generated schedule at any time by resequencing aircraft, inserting slots for additional aircraft, or changing the airport acceptance rates. A traffic load display provides a graphical representation of various traffic load characteristics, and several configuration panels are available for modifying timeline displays and setting schedule parameters. The workstation consists of a SUN4® SPARC™ workstation with keyboard and mouse input devices. TMA presents the traffic management coordinator with new capabilities that are a significant departure from the current traffic management system. The next section describes the process that has been applied for developing and assessing TMA at an operational field site.

Field Development and Assessment Process – Overview

Development and assessment of CTAS is currently underway at two FAA ATC field sites. This paper focuses on the development and assessment of TMA at the Center and TRACON of one of the field sites. TMA is accessible in the traffic management units at the Center and TRACON. A one-way interface with the current HOST system is available so that TMA can reflect the current traffic situation. Traffic managers reference TMA during traffic rush periods and in off times to explore its capabilities and to understand how the tool can be used to solve traffic management problems.

The field development and assessment process is geared for system refinement and enhancing our understanding of the potential impact of TMA on traffic management problem-solving and inter-facility coordination. The process also provides insight for various program objectives, such as operational procedures and requirements definition. To do this expediently, two mechanisms are in place to allow the timely transfer of information from the field site back to the primary development site at NASA-Ames. These are "unstructured" and "structured" assessments. Both are described briefly here, and key aspects of the structured assessments are elaborated further in the third section.

Unstructured assessments are performed by traffic management personnel on a daily basis, during traffic rush periods and during off-times. Here traffic managers access TMA representations, comparing data between TMA and the existing system at the Center, or with decisions made from compiling many separate sources of information together at the TRACON.

Human factors engineers and development personnel may observe TMA-use for the purpose of understanding the system, but goal-directed data collection does not occur at this time. Unstructured assessments are for traffic managers to experience the system and provide feedback without an audience or intrusion. Having the system available on a continuous basis provides exposure to a variety of traffic flow and weather situations, allowing the users to "shape" TMA-use to fit the problem-solving demands of their environment. This process is instrumental for engendering trust in the system.

Structured assessments are conducted to systematically investigate tool use and to capture the user's experience with TMA. How the traffic manager uses the tool in response to various problem-solving demands is an important gauge of the match between TMA features and functions and the context for their use. It has been argued that a major cause of system failure is a mismatch between the system capabilities and the demands and constraints of the operational environment (Bentley et al., 1992). Calibrating the match is thus a key activity during development for ensuring system success. Structured assessments are conducted by human factors engineers and development personnel and provide feedback for further development and program milestones. Methods and approaches for structured assessments are described further in the third section.

Quick transfer of information from the users to the development site and back to the users again is critical for continuity at the field site. Timely feedback to the traffic managers on their questions and suggestions during unstructured assessment is essential for maintaining their interest and involvement as well as for streamlining the development and assessment process. An electronic-mail system connecting the traffic managers at the center and TRACON to NASA-Ames and field-site development personnel has facilitated timely information transfer. Questions are addressed immediately, and design issues from structured and unstructured assessments are entered into a data-base managed at NASA-Ames. Resolution and augmentation of TMA features and functions are decided by committee, with representation from program management, developers, human factors engineers, and testing personnel. Issues are categorized and prioritized by the committee according to their pragmatic and technical implications; for example, implications for system usability and operational suitability, availability of development resources, the need for further analysis, and objectives of an upcoming structured assessment. Refined software is shipped back out to the field sites on a near-monthly basis.

Structured Assessments – Methods and Approaches

Structured assessments of TMA take a principle-driven approach, drawing on principles, perspectives, and methods from human factors engineering, cognitive engineering, and usability engineering. These fields provide a knowledge base from which methods and approaches for validating system designs may be derived. Structured assessments focus on specific aspects of the users' experience with TMA and consist of several activities:

- Issue identification
- Method selection
- Data collection
- Data interpretation
- Analysis, inferences and implications.

These activities are described next. Each activity relies heavily upon the operational context of the traffic management unit, focusing on exploration and discovery as well as assessment.

Issue Identification

Operational requirements for TMA are currently being defined and thus are not yet available for verifying the system design. This lack of guidance is compounded by the generality of ultimate criteria for ATC – namely, safe, orderly, and expeditious flow of traffic, and the general lack of knowledge regarding performance of individual and controller teams in current and future ATC environments. In the absence of requirements and criteria, there is a risk of collecting data that may be expedient but inappropriate (Parsons, 1972; Hopkin, 1980). To compensate for this knowledge gap and to systematically guide the collection and interpretation of data to support the refinement and validation of TMA, we focused on three broad categories of human-centered system issues:

- Technical usability
- Domain suitability
- User acceptance.

These three categories are also of interest to the FAA for its specification of operational requirements and for formal operational test and evaluation. Others have distinguished previously between two or three of these categories (e.g., Hopkin, 1980; Gould, 1988; FAA, 1989; Rasmussen & Goodstein, 1988). Categories and approaches for defining issues are described briefly. Further details can be found in Harwood (1993).

Technical Usability. Technical usability refers to perceptual and physical aspects of the human-computer interface such as display formatting, graphics, and human-computer dialog as well as anthropometric characteristics of the workstation. Issues in this category address the general question: Can the users extract and access the data needed to do their job? A tremendous amount of research in human factors engineering and human-computer interaction has contributed to the development of principles and guidelines for designing and evaluating human-system interfaces (See Department of Defense, 1989; Shneiderman, 1987; Smith & Mosier, 1986; Van Cott & Kincade, 1974). These principles constitute the basis for defining technical usability issues.

Domain Suitability. Simply addressing issues of interface usability does not necessarily provide insight into the suitability of the automation tool for the domain. Here it becomes necessary to address domain suitability, which refers to the content of information and representations for domain tasks as well as functions and decision-aiding algorithms. Issues in this category address the general question: Does the representation support the problem-solving requirements of the domain? In contrast to technical usability, which is driven by issues of technology utilization, domain suitability requires an understanding of the "cognitive problems to be solved and challenges to be met" (Hollnagel & Woods, 1987, p.257; Rasmussen, 1986; Rasmussen & Goodstein, 1988).

The fundamental basis for understanding the types of cognitive demands that can arise is a description of the domain in terms of the domain goals to be achieved, the relationships between these goals, and means for achieving goals (Rasmussen, 1985, 1986; Woods &

Hollnagel, 1987; Rasmussen & Goodstein, 1988). This sort of system description, in terms of a goal-means decomposition, is particularly useful for system evaluation: it guides the description of the cognitive situation that the design must support and it guards against narrowly focusing on problem-solving demands in only one aspect of the work domain.

User Acceptance. User acceptance is obviously enhanced by the ease of use and suitability of the system for supporting cognitive task requirements. Yet user acceptance also depends upon job satisfaction, professionalism, esteem and opportunities for demonstrating individual merit. Hopkin (1980, 1992) has argued that such issues are usually overlooked in the context of technology upgrades, but may have serious consequences for ultimate system safety and efficiency. Attention must thus be devoted to disclosing issues associated with the impact of new technology on ATC job satisfaction.

Context is critical for understanding the impact of new system upgrades on sources of job satisfaction and professional merit. What is satisfying and motivating about a job is as much a factor of the individual as it is the nature of the tasks and work domain. Ethnographic techniques for understanding the work environment are thus instructive for capturing valid descriptions of sources of job satisfaction. Such techniques are geared to the study of complex social settings to understand what aspects of activities are important and relevant to individuals. In general, ethnographic techniques have been recognized as essential to understanding, designing, and evaluating complex systems (Bentley, et al. 1992; Hughes, Randall, & Shapiro, 1992; Hutchins, 1992; Suchman, 1987; Whiteside, Bennett, & Holtzblatt, 1988; Suchman & Triggs, 1991).

Issue identification for TMA has been based upon hundreds of hours of observing traffic management activities, reading operational documents on traffic management, and interacting with traffic managers. Approaches for identifying issues are contextually based, that is, based upon an understanding of the physical characteristics of the environment, causal relationships between goals and functions in the task domain, and characteristics of the work activities. Focusing on only one or two of these factors risks collecting data that will not provide insight into sources of design deficiencies or provide a basis for defining meaningful human factors system requirements.

Method Selection

The operational field-site is important for gaining insight into the match between an automation tool design and its context for use. The complexity of the operational environment, with its inherent task demands and the access to operational personnel, allows discovery of unexpected feature use and assessment of the extent to which the tool will support its users. However, while testing at an ATC field site offers a unique perspective on system effectiveness, it also presents a number of constraints that preclude typical laboratory practices and techniques (Johnson & Baker, 1974).

The availability of controllers and scheduling and resource constraints can severely restrict the extent to which different conditions or system configurations can be investigated. In addition, sample sizes may be small, with the number of replications limited to a single trial. The physical environment is natural and intrusive factors are uncontrolled. Variables are driven by the system, not the experimenter, and the units for measurement are macro-units in the order of minutes. Measures are more often qualitative rather than quantitative.

Given these constraints, our expectations of field assessment must be adjusted appropriately. Field assessments provide an opportunity for capturing the users' ongoing experience with the tool, discovering how new functions will be used and where mismatches occur between the capabilities of the technology and the user's needs. Field assessment provides insight into the integration of a new automation device into an existing environment, indicating issues for transition training and operational procedures. However, field testing is only one level of system evaluation, often augmenting simulation and laboratory testing. Field testing is not a panacea, but it provides an important and necessary perspective for achieving system success.

To accommodate the constraints associated with field assessment and to maximize the opportunity of accessing the operational site, methods must be tailored accordingly. Several criteria guided the selection of methods for assessing TMA:

- Methods must capture the user's ongoing response to the system
- Methods must be sensitive to design deficiencies
- Methods must provide opportunities for discovering new strategies and system functions
- Methods must not disrupt traffic management operations.

Context-sensitive data collection techniques, that is, techniques based on observation and interpretation in the context of the user's work environment, meet these criteria (Whiteside, Bennett, & Holtzblatt, 1988). Such methods include observation and contextual interviews with active involvement of the users in the interpretation of the observations.

Field assessments of TMA, to date, have focused on capturing the traffic managers' experience with the TMA. Whiteside and his colleagues have argued for the importance of the users' experience as valuable information for engineers about users needs. The appropriateness of features and functions "...exists in the experience of the user, and experience is driven by the context in which it occurs" (Whiteside, Bennett & Holtzblatt, 1988; p. 809). Capturing the users' experience with the tool is especially important for complex, ATC automation systems, where the implications of the interactions between system components and emergent properties are largely unknown prior to implementation. When validation of system designs rests on reconciling technological possibilities with work needs, the users' experience assumes an important role.

Data Collection

Assessments are conducted in the traffic management areas at the center and TRACON. This location serves both technical and pragmatic interests. Traffic management involves extensive coordination with other traffic managers and area supervisors, communications with other facilities, and accessing and integrating information from a variety of different sources, such as weather displays, aircraft situation displays, and flight strips. Accessing TMA-use in the context of these operational activities is essential for addressing domain suitability and user acceptance. In addition, access to operational lighting conditions is desirable for validating such technical usability issues as color discrimination and readability. Lighting in the operational area is complex, with overhead lighting located in high ceilings and local lighting on work surfaces.

From a more pragmatic perspective, the location of the test area accommodates resource constraints and works well with the culture of the unit. To date, it has not been possible to schedule participants prior to the assessments. Instead, the supervisors on duty release traffic managers when staffing and the traffic demand allows. Having the supervisors control access to the traffic managers minimizes the impact on the unit, thereby increasing acceptance of the assessment process. Supervisors release and summon traffic managers as the conditions permit. Modular organization of data-collection materials, and non-intrusive observation are also flexible to accommodate this scheduling constraint.

Several different methods are used to collect data for assessing TMA. Scenario-driven surveys using prerecorded traffic data are used to assess technical usability. Shadowing of traffic management operations is used to assess technical usability, domain suitability, and user acceptance. These methods are described next, with particular attention given to the mechanics of their execution.

Scenario-Driven Surveys. Scenarios systematically guide the traffic managers through the display and interactive features of TMA and instruct them to view or manipulate different features. Pre-recorded traffic data are used to ensure that everyone views the same traffic conditions during the exercise. Associated with each scenario are validation statements that focus on specific technical usability issues, such as color discriminability, symbol detectability, and ease of interacting with the input devices. Traffic managers indicate whether they agree or disagree with the validation statement, and space on the survey is provided for comments and suggestions. A human factors engineer sits with the traffic managers as they complete the survey, answering any questions, and observing TMA use. Scenarios generally progress from being easy and simple to more difficult and complex. This is arranged to gauge the level of understanding of basic TMA features in the implicit check of TMA proficiency. If a participant is deficient in any area, instruction is provided, and the session is treated as training instead of as TMA assessment.

Technical usability issues are assessed for all TMA modifications, new features and new functions. The initial survey of all TMA features and functions lasted 2.5-3 hours per session, and subsequent assessments have lasted 45 minutes to an hour. The modular organization of the survey allows traffic managers to resume operational traffic management duties when necessary.

Shadowing Live Operations and Contextual Interviews. Shadowing involves a traffic manager using TMA to make traffic management decisions, mirroring the operational traffic management position. The shadowing traffic manager has access to all other sources of information in the unit except for the operational traffic management system. One observer observes and queries the shadowing traffic manager, and the traffic manager's ongoing commentary is tape recorded for later analysis. Another observer watches the operational traffic manager. Here, traffic management activities and decisions are observed in a more passive mode to avoid disrupting operations. Understanding and interpreting TMA use, at both the Center and TRACON, depend upon an understanding of the operational context. The second observer is critical in this regard.

Shadow-mode operations are effective for discovering unexpected and serendipitous tool-uses and for assessing issues of technical usability, domain suitability, and user acceptance. Methods for data collection are similar at the TRACON and Center but tailored for their unique constraints. Efforts are focused on capturing the traffic managers' ongoing experience with the system using context-based interviews (cf. Whiteside, Bennett, & Holtzblatt, 1988). This technique involves observing and questioning the users about the tool as they are using it for various planning and problem-solving activities. A critical aspect of contextual interviews is involving the users in the interpretation of their experience with the system. This aspect is discussed further in the next section on data interpretation.

An important aspect of data collection in the field is the period of acclimatization that precedes actual data collection. Prior to conducting structured assessments, we spent several weeks in the traffic management units at the Center and TRACON, simply observing operations and answering questions on the purpose of our presence and the TMA assessment process. This acclimatization period allowed the traffic managers to become comfortable with us, making our observations less intrusive. It also allowed us to work out methodology issues, (e.g., optimum observation positions, and an effective observation checklist) and allowed us to gain a deeper understanding of traffic management operations.

Subjective Ratings. Subjective ratings of a system's usefulness provide another avenue for capturing the users' experience with an automation tool. Following a traffic rush period, traffic managers rate the usefulness of various TMA features, on a scale ranging from 1 to 5, for different traffic management tasks. Ratings capture the users' cumulative experience with the tool, in contrast to the momentary experience captured by a comment made during a specific activity. As a consequence, discrepancies between ratings and comments are possible and are a cue to dig deeper: while a feature may be useful in one situation it may be perceived to be insufficient in another. Such discrepancies underscore the importance of conducting observations at different times of the day, over several days, and preferably during different seasons to capture changes in traffic flow and weather disturbances. TMA has been at the field site for over a year and assessments to date have been conducted during the summer and winter months, each lasting about a month.

Data Interpretation

Data interpretation occurs on and off the field site. Observation alone is not sufficient for exploring and assessing tool use. The observer's interpretations of the observations must be shared with the user to verify their truthfulness (Whiteside, Bennett, & Holtzblatt, 1988). Mutual understanding of the traffic managers' experience with TMA is achieved during the traffic rush and immediately following the rush. The traffic managers are questioned in a debriefing interview on feature use for various problems, their experiences with TMA and their impressions of the traffic rush. In turn, the observers' interpretations of TMA use and the traffic managers responses to questions are also verified. Specific questions and observations, during and immediately following the traffic rush, are guided by a set of general questions:

- What is/was the traffic situation?
- What decisions and planning activities are occurring/occurred?
- What information is/was accessed from TMA and non-TMA sources?
- How is/was TMA used to support various traffic management decisions?
- What information is/was lacking or hindered decisions?
- What improvements are necessary?

These questions provide a framework for systematically exploring and understanding TMA use in the context of traffic management operations. They also provide a basis for deeper probing of technical usability, domain suitability, and user acceptance issues; for example: Is the number of steps for a particular feature excessive given operational time constraints? Is sufficient information provided for determining whether the airport acceptance rate should be changed? Is the *right* information provided to support equitable decisions? All phases of the interview are tape recorded and conducted at the TMA, in the operational area, to provide a reference for discussing and interpreting the system. The merits of video, for this purpose, have been broadly extolled. Unfortunately, we were precluded from videotaping activities in the control room.

Something that proved helpful for data interpretation was for the observers to spend time each day, off the field-site, reviewing and discussing the observations. Dovetailing the different observational perspectives was useful for identifying knowledge gaps and for recalibrating the focus to further explore unexpected discoveries of tool use and possible emerging strategies. Any outstanding questions or new interpretations were taken up with the traffic managers the next day.

Analysis, Inferences and Implications

Surveys, observations, context-based interviews, and subjective ratings provide multiple windows on the traffic managers' experience with TMA. These methods and data provide a qualitative assessment of the match between TMA features and functions and the operational context for their use. The challenge lies in elucidating a tractable set of inferences from this large amount of data. To date, the focus of the TMA development and assessment process has been on identifying design deficiencies, discovering unexpected feature uses, understanding how the tool is used for various problem-solving activities, and defining operational requirements. Analyses have been geared accordingly. Frequency counts of negative responses on surveys provide insight into deficiencies and discrepancies. Content analyses of observations and interviews, coupled with subjective ratings, also provide insight into design deficiencies and discrepancies and enhance the understanding of tool use. Analyses, inferences and implications are described next in some detail as guidance for requirements definition evolves from these insights.

Identifying Design Deficiencies

Surveys. Scenario-driven surveys directly assess technical usability issues. Analysis is straightforward, focusing on negative responses to survey validation statements, which indicate difficulties in extracting, reading, discriminating, and accessing data from the TMA. Comments made by the users during the survey suggest resolutions to these deficiencies.

Survey data can provide diagnostic insight into users' possible difficulties when using the tool to make traffic management decisions. For example, a survey finding helped account for what appeared to be less efficient decision-making during shadow-mode exercises. The survey had revealed that a particular configuration caused crowding of data. Later, during the shadow mode exercises, this finding helped pin-point why particular traffic management decisions were being altered to what appeared to be less efficient decisions: Data congestion was causing traffic managers to overlook data, leading them to alter their decisions as the traffic situation progressed. This interpretation was confirmed by the traffic managers, and the problem was remedied in re-design.

Survey data provide only a partial window on system usability. Display clutter, color coding, and data entry may be assessed differently when the users are actively engaged, using the tool to solve traffic management problems. For example, a particular feature that required manual setting received a positive response on the survey, but negative comments during actual use. During shadow-mode operations, the traffic managers were too busy with other traffic management activities to manually re-set a feature to reflect changes in the traffic flow. Too time constrained, they had to extrapolate the actual setting, making the feature cumbersome to use. While survey data provides useful information about system usability, this example illustrates the importance of accessing the users' experience with the system from different perspectives.

Observations and Interviews. Design deficiencies are also accessible from observations and interview data collected during shadow-mode exercises. Analysis of these data is time consuming, but the richness of the findings would not be available otherwise and outweigh the cost associated with the time spent. Observations and interview data from the two observers are merged into a single chronological description of each traffic rush. Such a description is useful for capturing the context of TMA use and provides a basis for various content analyses. Content analyses are performed in order to make qualitative inferences about TMA as a potential traffic management tool.

It is important to select categories for the content analysis that reflect the objectives of the assessment. To date, traffic managers' decisions and actions have been categorized according to design deficiencies, feature use for various traffic management activities, and unexpected discoveries. (For a concise description of content analysis, see Weber, 1990.) Data interpretation with the traffic managers during the interviews greatly facilitates the categorizing of observation and interview data. Some examples of the kinds of design deficiencies that can be inferred from content analysis of interviews and observations are presented next. Feature use and unexpected discoveries are described in the following section.

Technical usability deficiencies are defined as observed or reported difficulties in accessing, interacting with, or reading data. Examples of findings include ineffective coding of information for data search, the need for labeling to reflect operations, and too many steps required to implement various functions. In some instances, usability issues revealed here support findings from the surveys; in other instances new issues are raised.

Domain suitability problems are defined as occasions where the traffic manager needs certain information that is not available, and where extracting information interferes with or hinders problem-solving or decision-making. Examples of findings include the need for organization of information on panels to reflect operational constraints, the need for display parameter settings to reflect current airport configurations, and the need for representation of specific categories of information to reflect the characteristics of the traffic flow.

User acceptance problems are not as easily accessible or apparent as usability and suitability issues because they tend to be incidental consequences of the information technology (cf Hopkin, 1980; 1992). Understanding the operational context is thus essential for identifying user acceptance problems. System upgrades can affect job satisfaction and opportunities for recognizing professional merit by either affecting what was satisfying about the job in the current system or by causing new situations to emerge that disrupt job satisfaction. Findings from the TMA assessments have provided insight into both of these possibilities. Earlier observations, prior to data collection, had revealed that an important source of job satisfaction is in making decisions and plans that strike an equitable balance of restrictions across facilities and aircraft. Findings from the assessments to date have revealed that a key source of information for ensuring equitable decisions was quite difficult to extract from TMA. This difficulty reduced the use and preference for the representation, and pointed the way for system refinement. In contrast, TMA representations have also created new situations that appear to enhance job satisfaction. One situation is the elimination of a time-consuming counting task, and another is the pulling together of previously disparate sources of information into a single representation. Comments from the traffic managers indicate that such features provide them more time for important planning and allow them to keep up with the dynamic traffic flow situation.

In addition to helping disclose design discrepancies, content analysis of observation and interview data provide insight into feature use. This insight is important for understanding the users' needs and the extent to which they are supported by the automation's capabilities. Feature use and discovery are discussed next.

Discoveries and Description of Feature Use

The introduction of advanced technology and innovative display and interactive features, like that embodied in TMA, alters the way work is done and how problems are tackled. Understanding these changes and how the features are incorporated into the flow of work is as important for assessing the match between the user's needs and system capabilities as is the identification of design deficiencies.

Patterns of feature use elucidated from the analysis of observations and interviews led to the discovery of two different strategies for making a particular traffic management decision. One strategy involved feature use that solved the problem in a similar way to current practices; namely by accessing information that managed traffic demand at the level of individual aircraft. The other strategy solved the problem in a new and different way, by relying on representations that provided information about the aggregate traffic demand. This second strategy was an unexpected discovery. Decisions made with both strategies were equally efficient, relative to those made by the operational traffic manager using the current traffic management system. At one level, the finding of different strategies of feature use suggests that the TMA representations are flexible enough to support different traffic management styles and preferences. At another level, the finding has broader implications for operational requirements because the information needed to support different problem-solving strategies must be identified.

One of the biggest changes to traffic management as a consequence of TMA involves the level of coordination between the Center and TRACON. With TMA, both facilities now have access to the same information about the traffic demand. Observations and interviews with the traffic managers indicate that this has enabled the TRACON to coordinate proactively with the center on decisions regarding the distribution of the traffic flow. Such coordination between the two facilities is essential to avoid overloading the TRACON and to maximize airspace capacity. At a more subtle level, TMA elevates the role of the TRACON traffic managers, allowing them to be more active players in traffic management. This elevated status has obvious implications for job satisfaction.

Another change to current practices is the impact of TMA on the exchange of information between facilities. Analysis of the traffic management communications between the center and TRACON indicates that well over 50 percent of the transmissions between the facilities involves the transmission of information that is accessible from TMA. This finding suggests that certain verbal transmissions between facilities could be eliminated, augmented, or reduced by electronic sharing of information via TMA. Changes in the level of coordination and exchange of information between the Center and TRACON alter the way traffic management is performed and has implications for operational procedures and requirements.

Requirements Definition and Rejuvenation

Requirements are the services to be provided by the system and the constraints under which it must operate. Complex domains, like ATC, with their myriad interdependencies and interconnections, are difficult to understand and thus a complete set of requirements is not likely to be available prior to development. Instead, definition of requirements is likely to evolve with development, and modifications and refinements will be necessary as an understanding of the user's needs improves. Field testing conducted early in development catalyzes the requirement definition process. Identifying mismatches between the user's needs and tool features is important for refining the design and for exposing system constraints that must be captured in the requirements. This dual purpose of design deficiencies is illustrated well by the following example. A particular TMA feature had been designed to require several steps to access information. When exercised during shadow-mode operations, the feature was deemed unsuitable because immediate access to the information was needed. This not only identified a design deficiency but also exposed a system constraint that must be captured in the system requirements: immediate access to information for a particular traffic management decision.

Feature use in context is also instructive for defining and modifying system requirements. New technology can create new cognitive problems and information requirements for these problems must be defined. Similarly, capabilities may emerge when the system is used in the operational environment that were not anticipated in the conceptual design; for example, the level of coordination and information exchange between the TRACON and Center with TMA. System constraints for such capabilities must be included in the system requirements.

It has been argued that the implementation of a system and its specification should not be kept separate because, in practice, they are "inevitably intertwined." Models of system development that require their separation deviate from reality and restrict the development of effective systems (Swartout & Blazer, 1982). A similar argument can be applied to requirements definition. System implementation through field testing, early during system development, can facilitate the evolution of system requirements for complex domains like ATC. As described in this paper, field assessments help highlight system constraints, enhance our understanding of the user's needs, and provide insight into the impact of new technology on existing operational practices. While systematic analyses, feasibility studies, and system modeling are necessary precursors for requirements definition, field development and assessment in the field can help augment the process.

Summary

The process for incorporating advanced technologies into complex aviation systems is as important as the final product itself. This paper described a process that is currently being applied to the development and assessment of an advanced ATC automation system, CTAS. The key element of the process is field exposure early in the system development cycle. The process deviates from current established practices of system development – where field testing is an implementation endpoint – and has been deemed necessary by the FAA for streamlining development and bringing system functions to a level of stability and usefulness. Methods and approaches for field assessment are borrowed from human factors engineering, cognitive engineering, and usability engineering and are tailored for the constraints of an operational ATC environment. To date, the focus has been on the qualitative assessment of the match between TMA capabilities and the context for their use. Capturing the users' experience with the automation tool and understanding tool use in the context of the operational environment is important, not only for developing a tool that is an effective problem-solving instrument but also for defining meaningful operational requirements. Such requirements form the basis for certifying the safety and efficiency of the system. CTAS is the first U.S. advanced ATC automation system of its scope and complexity to undergo this field development and assessment process. With the rapid advances in aviation technologies and our limited understanding of their impact on system performance, it is time we opened our eyes to new possibilities for developing, validating, and ultimately certifying complex aviation systems.

Acknowledgments

The authors are grateful for the support of Barry Scott, Manager FAA Engineering Field Office at NASA-Ames Research Center and David Jones, Branch Chief ATC Field Systems Office at NASA-Ames Research Center. C. Halverson, E. Murphy, S. Nowlin, and P. Stager are acknowledged for their insightful comments on earlier versions of this paper.

References

ATC Systems Field Office. (1992). *Traffic management advisor (TMA) reference manual.* Moffett Field, CA: NASA-Ames Research Center.

Bentley, R., Hughes, J. A., Randall, D., Rodden, T., Sawyer, P., Shapiro, D., & Sommerville, I. (1992). Ethnographically-informed systems design for air traffic control. *CSCW Proceedings.* (pp. 123-129).

Christensen, J. M. (1958). Trends in human factors. *Human Factors, 1* (1), 2-7.

Davis, T. J., Erzberger, H., & Green, S. M. (1991). *Design and evaluation of air traffic control final approach spacing tool* (NASA Technical Memorandum 10287). Moffett Field, CA: NASA-Ames Research Center.

Department of Defense. (1989). *Human engineering design criteria for military systems, equipment, and facilities* (MIL-STD-1472D). Washington, DC: Author.

Erzberger, H. (1992). *CTAS: Computer intelligence for air traffic control in the terminal area* (NASA Technical Memorandum 103959). Moffett Field, CA: NASA Ames Research Center.

Erzberger, H., & Nedell, W. (1989). *Design of automated systems for management of arrival traffic* (NASA Technical Memorandum 102201). Moffett Field, CA: NASA-Ames Research Center.

Federal Aviation Administration. (1989). *FAA NAS test and evaluation program* (Order No. 1810.4A). Washington, DC: Author.

Flexner, S. B., & Hauck, L. C. (1983). *The Random House dictionary of the English language*. Random House: New York.

Gould, J. D. (1988). How to design usable systems. In M. Helander (Ed.), *Handbook of human-computer interaction* (pp. 757-789). New York: Elsevier Science Publishers BV (North Holland).

Harwood, K. (1993). Defining human-centered system issues for verifying and validating air traffic control systems. In J. A. Wise, V. D. Hopkin, & P. Stager (Eds.), *Verification and validation of complex systems: Human factors issues* (pp. 115-130). Berlin: Springer-Verlag.

Hollnagel, E., & Woods, D. D. (1983). Cognitive systems engineering: new wine in new bottles. *International Journal of Man-Machine Systems, 18*, 583-600.

Hopkin, V. D. (1980). The measurement of the air traffic controller. *Human Factors, 22* (5), 547-560.

Hopkin, V. D. (1992). Human factors issues in air traffic control. *Human Factors Society Bulletin, 35* (6). Santa Monica, CA: Human Factors Society.

Hughes, J. A., Randall, D., & Shapiro, D. (1992). Faltering from ethnography to design. *CSCW Proceedings* (pp. 115-122).

Hutchins, E. (1991). *How a cockpit remembers its speed.* (Technical Report). San Diego: University of California, Distributed Cognition Laboratory.

Johnson, & Baker. (1974). Field testing: The delicate compromise. *Human Factors, 16* (3), 203-214.

Meister, D. (1985). *Behavioral analysis and measurement methods.* New York: John Wiley & Sons

Parsons, M. (1972). *Man machine system experiments.* Baltimore, MD: The Johns Hopkins Press.

Rasmussen, J. (1985). The role of hierarchical knowledge representation in decision making and system management. *IEEE Transaction on Systems, Man, and Cybernetics, 15*, 234-243.

Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering*. Amsterdam: North-Holland.

Rasmussen, J., & Goodstein, L. P. (1988). Information technology and work. In M. Helander (Ed.), *Handbook of Human-Computer Interaction* (pp. 175-201). New York: Elsevier Science Publishers BV (North Holland).

Shneiderman, B. (1987). *Designing for the user interface.* Reading, MA: Addison-Wesley.

Smith, S. L., & Mosier, J. N. (1986). *Guidelines for designing user interface software.* Bedford, MA: Hanscom Air Force Base, USAF Electronic Systems Division. (NTIS No. A177 198)

Suchman, L. A. (1987). *Plans, and situated actions: The problem of human machine communication.* Cambridge, MA: Cambridge University Press.

Suchman, L. A., & Trigg, R. H. (1991). Understanding practice: Video as a medium in reflection and design. In J. Greenbaum & M. Kyng (Eds.), *Design at Work* (pp. 65-89). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tobias, L., Volckers, U., & Erzberger, H. (1989). Controller evaluations of the descent advisor automation aid. In *Proceedings of the AIAA Guidance, Navigation, and Control Conference* (pp. 1609-1618).Washington, DC: AIAA.

Van Cott, H. P., & Kincaid, R. G. (1972). *Human engineering guide to equipment design.* Washington, DC: U.S. Government Printing Office.

Whiteside, J., Bennett, J., & Holtzblatt, K. (1988). Usability engineering: Our experience and evolution. In M. Helander (Ed.), *Handbook of Human-Computer Interaction* (pp. 791-817). New York: Elsevier Science Publishers BV (North Holland).

Weber, R. P. (1990). *Basic content analysis.* Newbury Park, CA: Sage Publications.

Woods, D. D., Roth, E. M., & Bennett, K. B. (1990). Explorations in joint human-machine cognitive systems. In S. P. Robertson, W. Zachary, & J. B. Black (Eds.), *Cognition, computing, and cooperation* (pp. 123-158). Norwood, NJ: Ablex Publishing Company.

N95- 34782

Integrating Human Factors Knowledge Into Certification: The Point of View of the International Civil Aviation Organization (ICAO)

53-54

Daniel Maurino & Vincent Galotti

International Civil Aviation Organization

Introduction

Human factors has matured into a core technology. This contention is best reflected in the attention the aviation industry has dedicated to this technology over the last decade. The aviation industry is going through what has been dubbed "the golden era" of aviation human factors, both in research as well as in application, through human factors training for operational personnel, human factors consideration in accident prevention, and investigation and in workstation design. Although skeptics and critics still exist in all segments of the international community who would state otherwise, the contribution of human factors study and application to aviation system safety and effectiveness remains beyond question.

The most important step, however, has yet to be taken. No matter how potentially applicable the research, no matter how performance-enhancing the training, no matter how sensible the investigation of accidents and, finally, no matter how much the ergonomic know-how is incorporated into workstation design, all of these advances can be deemed to be remedial measures. From the perspective of advanced technology systems safety, the major contribution of human factors will be fully realized only after the huge amount of existing and available human factors knowledge is translated into the certification process of such systems; that is, before the system is in operation. The incorporation of human factors requirements into the certification of new, advanced technology remains to be the challenge of the decade.

It has repeatedly been suggested that, unless awareness about a problem is first gained, for practical purposes, that problem does not exist. Therefore, this paper advances a justification of why human factors requirements should be given the same consideration and weight as traditional "hard core" requirements that exist during the certification process of equipment, procedures and personnel that make up advanced aviation systems. Such justification is considered vital to secure the understanding and the necessary commitment of designers as well as regulators involved in certification processes. Secondly, an approach to the inclusion of human factors certification requirements is proposed, based on existing ICAO regulatory requirements and guidance material. An understanding of the role of ICAO in the processes of regulation is therefore necessary for the reader to identify how its mechanisms may be useful in attaining the required goals of human factors certification of advanced systems through such mechanisms.

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

Although general in nature, such a proposal is viewed as one feasible way to make inroads toward the stated goal of human factors certification of advanced aviation systems.

The consideration of human factors certification requirements during the design stage of advanced, new technology systems may be seen as resting over a three-legged stool. The first leg, the equipment that a system will utilize to achieve its goals, has traditionally attracted ergonomic considerations associated with equipment design, usually centred around "knobs and dials." As of lately, this view has expanded to include the so-called other important aspect of human factor's study which deals with the cognitive, behaviourial and social processes of the human operators. Study on this area must be furthered. The second leg of the stool, the procedures required to operate the equipment, however, has been largely unaddressed. Procedures are not inherent to equipment, but must be developed. The importance of proper human factor's consideration in the design of procedures can not be overstated. Lastly, the third leg of the stool, certification of personnel who will operate the equipment, is very much underway but far from being complete.

Background

Over time, the contribution of the human factors profession toward advancing the safety and effectiveness of socio-technical systems, including aviation, has been hindered because of its piecemeal approach. Designers, engineers, trainers and regulators have historically favoured solutions which were biased by their professional backgrounds. This often resulted in a state of affairs which, although the individual components had been designed to the maximum level of available know-how and expertise, the aviation system as a whole received only marginal benefits from the technological progress. Furthermore, an unlimited trust in technology-driven approaches toward system development often overlooked the fact that the human component of the system still remains as the old "Mark I" version with basically the same limitations as existed 5,000 years ago. Indeed, the shortcomings of overeliance on technology to overcome system safety deficiencies have clearly been identified by research, as well as by the investigation of major socio-technical systems' catastrophes. However, rather than revising the wisdom of this approach, some would suggest that industry has renewed its commitment to technology with a blind euphoria over technical systems claimed to be "absolutely safe." One can not help but reflect on the "Titanic syndrome": "such marvelous technology could not possibly sink." History, however, suggests otherwise.

When pursuing safety and effectiveness in aviation, we have tended to think in individual rather than in collective terms. At the operational level, this is reflected by the pervasiveness of remedial actions addressed to individuals rather than to the system as a whole. A good example of this assertion is the omnipresence of the *pilot error* clause in accident investigation literature, which would seem to be a relic from World War II. Even with the introduction of human factors into existing protocols for accident investigation, the focus still often remains on those at the tip of the arrow. It is a common mistake to narrow down an investigation of human factors in accidents and incidents to the behaviourial and cognitive aspects as they relate to the performance of operational personnel involved in the actual occurrence. Such investigations should rather broaden their perspective to include overall system performance, including the behaviourial and cognitive elements of all of its human components and not just those at the tip of the arrow.

From the perspective of design, the picture does not vary substantially. Technology has been introduced piecemeal, producing excellent examples of equipment which is quite remarkable. However, because of the lack of appropriate macro-analysis at the time that individual technical designs are introduced into a system, the interface with other system components including, first and foremost, the human, are sometimes rather clumsily accomplished. This less than optimum interface has demanded, often times, significant adaptation efforts to accommodate a new piece of equipment into a system already in operation. Such adaptation often involves trade-offs and compromises which, at the end of the day, diminish the benefits of the new design in terms of its contribution to safety and effectiveness.

The introduction of high-level automation into flight decks and air traffic control units stand as good examples of the shortcomings discussed in the previous paragraph. Some would consider it naive to accept that the design and introduction of automation routinely progresses beyond the level of micro-design and micro-analysis. The consequences are well documented and the equipment intended to reduce human error has in many cases merely displaced it. Likewise, designs intended to alleviate workload often increase it at the most inappropriate times and reduce it during times where it may actually foster boredom and complacency. The absence of macro-ergonomic considerations has led many to contend that the introduction of automation into many advanced aviation systems (e.g., new flight deck design, advanced automated ATC systems, intelligent reasoning tools, etc.) reflects a regrettable, although quite preventable, failure of the human factors profession. The good news is that lessons have been learned, and these lessons will hopefully be translated into the design of the future global ICAO Communications, Navigation, Surveillance/Air Traffic Management (CNS/ATM) System. This system is briefly described in a later chapter of this paper because of its importance to the future of aviation and the criticality of addressing human factors certification concerns at an early stage in its development.

ICAO's Role in Regulation and Implications for Certification

Regulation remains the vehicle to ensure that the hard learned lessons do not vanish into the labyrinths of sectorial interests or into the fragility of human memory. The literature being produced on how to proceed in securing safety and effectiveness in new technological systems accumulates rapidly. The justification behind these new approaches is well-documented and beyond challenge. Seldom a week goes by without one major human factors event taking place in the industry's agenda. It could safely be stated that industry has pursued the goal of awareness and education to an extent that can reasonably be expected. Although there will always be room for relevant fora to exchange ideas and foster feedback, it is time to progress beyond the stage of just talking and exchanging exciting stories. One way to accomplish this progression is to ensure that important developments emanating from workshops, seminars, etc., be made available, accepted and implemented at the world-wide level. To accomplish this goal, it is imperative to establish and introduce a requirement to include human factors knowledge into the certification process of equipment, procedures and personnel through legislation that is binding to the larger community. Should such a development take place, there would be a measure of assurance that, by virtue of the imperatives of regulation, a macro-approach to high-technology systems design will replace existing micro-approaches.

Having been developed by the International Civil Aviation Organization (ICAO), the crux of this paper is toward the understanding, acceptance and importance of standards and practices relevant to the certification processes which would be implementable at the world-wide level. It is therefore important to discuss the role and processes of ICAO toward the attainment of these goals for the purpose of clarifying the virtues of international standardization and cooperation which points the way toward a wider level of implementation and acceptance of the important work currently being developed concerning human factors certification of advanced systems. For certification in aviation to have significant relevance, it is imperative that it have world-wide application and standardization. Considering the global aspect of the CNS/ATM System, which will be based on satellite technology, the need to efficiently integrate all of the elements of the system on a wide level can be seen. Whether we are discussing air traffic control procedures, pilot licensing, aviation training or maintenance procedures, it is essential that there be international agreement between nations for regulations to be effective, worthwhile and able to achieve all potentialities. Consider the following example: in an afternoon's flight, an airliner can cross the territories of several nations; nations in which different languages are spoken and in which different legal codes are used. In all of these operations, safety must be paramount; there must be no possibility of unfamiliarity or misunderstanding. In other words, there must be international standardization and agreement between nations in all technical, economic and legal fields. To accomplish these difficult goals, the nations of the world established ICAO to serve as the medium through which this necessary international understanding and agreement can be reached. It acts as the mechanism whereby global coordination and harmonization is achieved.

The main accomplishments of ICAO associated with regulation, and thereby certification, have been the agreement of its Member States on the necessary standardization for the operation of safe, efficient and regular air services. This standardization has been achieved primarily through the adoption by the ICAO Council, as Annexes to the Chicago Convention, of specifications known as International Standards and Recommended Practices. The 18 Annexes so far adopted cover the whole spectrum of aviation. A Standard is any specification for physical characteristic, configuration, materiel, performance, personnel or procedure, the uniform application of which is recognized as necessary for the safety or regularity of international air navigation and to which contracting states will conform in accordance with the Convention. In the event of impossibility of compliance, notification to the ICAO Council is compulsory. A Recommended Practice is any specification for physical characteristic, configuration, materiel, performance, personnel or procedure, the uniform application of which is recognized as desirable in the interests of safety, regularity or efficiency of international air navigation, and to which contracting States will endeavour to conform in accordance with the Convention (ICAO, 1990).

An Overview of the ICAO Annexes

This overview is developed so that an understanding of the Annexes and their application toward the development of certification standards for human factors may be achieved. The Annexes to the Convention on International Civil Aviation have reduced many of the complexities of air transportation to everyday routine. They govern the standards of performance required of pilots and air traffic controllers. The ICAO licensing requirements must be met by all contracting States and they apply not only to pilots, flight crews and controllers, but also to ground maintenance crews. International specifications also exist for the design and performance of aircraft themselves and the equipment aboard. The rules of the air, by which pilots fly, were formulated by ICAO. These include both visual flight rules and instrument flight rules. The weather reports, so vital to the safety of international travel, are provided to pilots and airport staff by a world-wide network of meteorological stations. The aeronautical charts used for navigation throughout the world are also specified by ICAO, and all the symbols and terms on these charts are standardized for uniformity so that no confusion can ever arise. Even units of measurement used in aircraft communications are standardized so that no pilot can be confused when he/she is given a distance by a ground-based air traffic controller. Aircraft telecommunications systems, radio frequencies and procedures are also ICAO's responsibility. The way in which aircraft are operated is regulated by ICAO so that an international level of safety is maintained. Uniform rules of airworthiness of aircraft are ensured by internationally agreed certification processes. Without that certification, no aircraft can be accepted for flight. Even the requirements for aircraft registration and their identifying marks are based on international standards. ICAO also gives continued attention to ways and means of ensuring better utilization of airspace in high traffic density areas and providing for greater air traffic handling capacity. Measures to regulate the flow of air traffic along heavily traveled air routes are being applied and refined to reduce or eliminate excessive delays in flight. To facilitate free and unimpeded passage of aircraft and their loads, ICAO seeks to speed up customs, immigration, public health and other procedures for passengers, crews, baggage, cargo and mail across international boundaries and to ensure that essential facilities and services are provided at international airports. ICAO has made significant progress in reducing aircraft noise by adopting stringent noise limitations for aircraft engines. This continuing effort has already resulted in a whole new generation of quieter aircraft. Safeguarding international civil aviation against unlawful seizure of aircraft, sabotage and bomb threats has received special consideration during recent years. The standards, recommended practices and guidance material on airport security have resulted in a marked decrease of such incidents. Finally, ICAO has recently introduced a comprehensive set of recommendations for the safe transport of dangerous goods to be applied uniformly all over the world. This latest Annex shows how remarkably well the Chicago Convention meets the continuing standardization needs, which are so intrinsically tied to certification, of civil aviation (ICAO, 1982). A list of the currently existing Annexes is at Attachment B.

The ICAO Global CNS/ATM System

A discussion about the ICAO CNS/ATM global system is necessary as it can be identified as a major and complex set of new systems that will be the major influence on all new systems developed for civil aviation. An opportunity exists for an early human factors input which should include human factors certification possibilities. The human factors profession should seek to ensure that all involved with the development and implementation of the CNS/ATM system are made aware of the possibilities regarding human factors certification along with the traditional human factors involvement. Human factors awareness should be sought after by all involved, especially those involved in international civil aviation and the development of international regulations.

The ICAO Global CNS/ATM concept was developed by the ICAO Future Air Navigation Services (FANS) Special Committee which was established by the ICAO Council at the end of 1983 to study, identify and assess new concepts and new technology in the field of air navigation (including satellite technology) and to make recommendations for the development of air navigation for international civil aviation over the period of the next twenty-five years. The Tenth Air Navigation Conference, held in Montreal in September, 1991, endorsed the findings of the FANS Committee which identified the major elements of the planned future system. These elements are Communications, Navigation and Surveillance which will be increasingly accomplished through the use of satellite technology. A resulting Air Traffic Management System would have the capability of resolving the shortcomings of the present air navigation system and of alleviating some of the air traffic congestion problems experienced in many parts of the world. The implications for human factors certification of advanced systems is evident and offers a further direction of study and discussion for future gatherings (ICAO, 1991).

The Tenth Air Navigation Conference endorsed the view that the planning and implementation of improved Air Traffic Management capabilities should include considerations of human factors impacts and requirements. They further stated that the many goals listed for the future Air Traffic Management System should be qualified in relation to human factors. The Tenth Air Navigation Conference recommended that work by ICAO in the field of human factors include studies related to the use and transition to the future CNS/ATM System and that ICAO encourage member States to undertake such studies. It also developed a list of considerations which are recommended for use when determining human factors aspects in relation to the CNS/ATM system. These are listed in Attachment A for discussion purposes. It is certain that human factors will be considered in all phases of implementation of the new systems. It is important to consider ways in which human factors certification can be introduced into this important development.

Discussion of Human Factors and the Certification Process

Consider what has recently been achieved concerning Personnel Licensing. In the last revision to Annex I, a requirement has been included to the effect that each applicant for a license must demonstrate appropriate knowledge regarding: "human performance and limitations relevant to.(the license being issued)" (ICAO, 1988). Annex I also includes an augmented requirement for the demonstration of certain skills in a manner which dictates increased attention to particular aspects of human performance. Thus, for example, the holder of an Airplane Transport Pilot License "shall demonstrate the ability to: exercise good judgment and airmanship, understand and apply crew coordination and incapacitation procedures; and, communicate effectively with other flight crew members."

This requirement effectively guarantees that, during the certification process (i.e., licensing and training) of operational personnel, human factors considerations must be duly accounted in the form of demonstration of human factors knowledge and skills. Other isolated efforts exist in the certification of procedures (Wiener, 1990). These efforts are valuable but are, nevertheless, random efforts toward pursuing an avenue of action which deserves serious consideration and the necessary framework for institutionalization which only regulation – international regulation – can assure. The initiative here should belong with the international community, and it is further contended that it should be pursued as such and within the context of existing structures.

From the perspective of ICAO, the inclusion of human factors requirements, following the licensing process of operational personnel as outlined in Annex I (example above) is viewed as one possible way to proceed. The Annexes are well established and internationally accepted documents covering all aspects of aviation and as such, provide an ideal structure. ICAO contracting States must observe and abide by their provisions if these States are to be accepted as members of the international community. It is difficult to identify a more solid foundation upon which to build the much needed framework necessary for legislating human factors. The ICAO Annexes are subjected to periodic revisions to ensure the relevance and applicability of their contents. It is suggested that each of these Annexes might be analyzed by dedicated groups of experts who would endeavour to pinpoint if and where the inclusion of human factors requirements in such documents would be appropriate. Should it be deemed feasible, the drafted requirements would eventually be included in the Annexes as they progress through their routine revision cycles. At this level of analysis and drafting, the requirements would be broad in nature and scope. Annex I stands as a leading example to follow.

Once the basic requirements are legislated, even in the broadest terms, the responsibility will by force switch to the international research, design and training communities who would then have to find the means and tools to implement such requirements into practice. And while the ICAO Annexes would establish the policies, the international community would devise the appropriate procedures to achieve such policies. This combination has worked quite satisfactorily and has produced consistent and successful results throughout the implementation of Annex I (i.e., certification of personnel).

Questions for Discussion

In addressing the questions put forward by the organizers of this workshop, some early and tentative answers can be advanced. They are simple and general and need refinement. It is felt, however, that often times simple realities must be reasserted to further progress. Secondly, they are viewed as a bridge to foster sound discussion and an open exchange of information. Thirdly, they are submitted as a preliminary attempt to provide a foundation to help orient discussion of a subject that appears, at first, abstract and controversial and upon which professional and cultural biases and preferences may have influence. Short answers to the questions upon which discussion during the workshop revolve are proposed hereunder.

Who should have the authority to perform the human factors certification process when considering advanced aviation systems? Certification should remain a responsibility of the national authorities. It is suggested, however, that the certification process could be facilitated if placed under the general umbrella of "ICAO human factors Requirements," and internationally accepted. Such requirements would provide a basic framework and would be established with the support of the appropriate sectors of the international community. The approval process (even if only an endorsement) by ICAO would enhance the credibility of the requirements at the world-wide level.

Where/How should certification be accomplished? Within the general guidelines as discussed in the answers developed in this paper. The review of the ICAO Annexes and the subsequent inclusion of human factors certification requirements by groups of experts from different States, called upon by ICAO, would allow national authorities to include the human factors certification requirements during the normal certification processes dealing with the "hard" components of an advanced system.

What should be certified? This question can be appropriately addressed only after the ICAO Annexes are reviewed by groups of experts and the relevant human factors requirements are incorporated into them.

Why should it be certified? The arguments advanced in the background discussion of this paper and at this workshop as a whole are considered as the justification and answer to this question.

Conclusion

It is appropriate here to repeat the analogy described in the introduction to this paper which is: That the consideration of human factors requirements during the design stage of advanced, new technology systems may be seen as resting over a three-legged stool. The first leg, the equipment that a system will utilise to achieve its goals, has traditionally attracted ergonomic considerations associated with equipment design, usually centred around "knobs and dials." Lately, this view has expanded to include the so-called other important aspect of Human Factor's study which deals with the cognitive, behaviourial and social processes of the human operators. Study in this area must be furthered. The second leg of the stool, the procedures to operate the equipment, however, has been largely unaddressed. Procedures are not inherent to equipment, but must be developed. The importance of proper human factors consideration in the design of procedures can not be overstated. Lastly, the third leg of the stool, the certification of personnel who will operate the equipment, is very much underway, but far from being complete. The real quest now, however, is to integrate these three legs into an indivisible one.

Finally, and most importantly, this workshop and its topic are extremely timely in that we are at the dawn of the most ambitious development ever undertaken in international civil aviation. This would allow us the rather unique opportunity to put theory into practice in the near future by ensuring that the concepts developed and furthered by this workshop and the follow-up are implemented in the design and certification of the ICAO future CNS/ATM systems described earlier in this paper. Now is the time to incorporate human factors requirements during the certification processes of these systems. This might act as a test to the feasibility of these ideas. Such endeavors represent a challenge for the research, engineering, training, operational and regulatory communities. But there is certainly more to be gained by attempting to meet the challenge rather than refraining from progress by decrying the difficulties involved.

References

DeGani, A., & Wiener, E. L. (1993). Philosophy, policies, procedures and practices: The four P's of flight deck operations. *Proceedings of the Sixth International Symposium on Aviation Psychology.* Columbus, OH.

International Civil Aviation Organization. (1982). *The Convention on International Civil Aviation: The first thirty five years.* Montreal, Quebec, Canada: Author.

International Civil Aviation Organization. (1988). *International standards: Personnel licensing (Annex 1).* Montreal, Quebec, Canada: Author.

International Civil Aviation Organization. (1990). *International standards: Rules of the air (Annex 2).* Montreal, Quebec, Canada: Author.

International Civil Aviation Organization. (1991). *Report of the Tenth Air Navigation Conference.* Montreal, Quebec, Canada: Author.

Wiener, E. L. (1990). *Human factors checklists and the four P's.*

APPENDIX A

LIST OF CONSIDERATIONS WHICH ARE RECOMMENDED TO BE USED WHEN DETERMINING HUMAN FACTOR'S ASPECTS IN RELATION TO THE CNS/ATM SYSTEM

a) The level of safety targeted for the future system should be defined not only with reference to various system statistics, but also with reference to error-inducing mechanisms such as human capabilities and limitations as well as important individual cases.

b) Definition of system and resource capacity should include reference to the responsibilities, capabilities and limitations of ATS personnel and air crews who must retain situation awareness and understanding in order to carry out all of their responsibilities.

c) Dynamic accommodation of three and four-dimensional flight trajectories to provide user-preferred routings while an ultimate goal for users may initially be restricted by human capabilities and the need to organize the flow of air traffic in an orderly manner in order to provide separation. The transition period will need careful research and evaluation on human factors aspects.

d) Provision of large volumes of potentially relevant information to users and ATS personnel should be limited to what is absolutely necessary and mediated by methods that effectively package and manage such information to prevent information overload while providing information pertinent to particular operational needs.

e) Human computer dialogues serving flight "air and ground requests" should be consistent in form and style with the ways in which air crews and controllers plan and negotiate.

f) A single airspace continuum should be free of operational discontinuities and inconsistencies between kinds of airspace and kinds of facilities that affect responsibilities and activities of air crews or ATS personnel at functional boundaries.

g) Organization of airspace in accordance with ATM procedures should also be readily learnable, recallable, and, to the maximum practical extent, intuitively understandable by air crews and ATS personnel.

h) Responsibilities of pilots, air traffic controllers and system designers should be clearly designed prior to the implementation of new automated systems and tools (e.g., conflict resolution advisories, data link, ADS, etc.).

In addition to the analysis and assessment aimed at the specific concerns just outlined, evolution of the ATM should be accompanied by systematic pre- and post-implementation evaluations of its more general human factors impacts. These assessments should encompass its effects on aircrew and ATS personnel workload and performance, as well as implications for their selection, training, career progression and health.

APPENDIX B

DESCRIPTION OF ICAO ANNEXES

ANNEX

1) Personnel Licensing
2) Rules of the Air
3) Meteorological Service for International Air Navigation
4) Aeronautical Charts
5) Units of Measurement to be used in Air and Ground Operations
6) Operation of Aircraft
   - Part I – International Commercial Air Transport-Aeroplanes
   - Part II – International General Aviation-Aeroplanes
   - Part III – International Operations-Helicopters
7) Aircraft Nationality and Registration Marks
8) Airworthiness of Aircraft
9) Facilitation
10) Aeronautical Telecommunications
11) Air Traffic Services
12) Search and Rescue
13) Aircraft Accident Investigation
14) Aerodromes
15) Aeronautical Information Services
16) Environmental Protection
17) Security
18) The Safe Transport of Dangerous Goods by Air

N95- 34783

Improving Air Traffic Control: Proving New Tools or Approving the Joint Human-Machine System?

519-54
p 13

Irene Gaillard[1] **& Marcel Leroux**[2]

[1]Action Recherches Application Matra Irit en Interface Homme Systéme, France
[2]Centre d'Etudes de la Navigation Aérienne, France

Introduction

From the description of a field problem (i.e., designing decision aids for air traffic controllers), this paper points out how a cognitive engineering approach provides the milestones for the evaluation of future joint human-machine systems.

The European air traffic control (ATC) system has entered a deep crisis: this system is unable to face a tremendous increase of the demand. This is not only the consequence of its inertia since such inertia is normal for any complex system. Short term measures to optimize the present tools appear to be insufficient; these tools have already reached the limits of their evolution capability. A large discussion on how to enhance ATC methods and tools is open. Very ambitious goals are assigned to the future systems, such as the French CAUTRA V project, which plans to double the capacity of the system by the year 2005 and to significantly increase safety. Numerous ambitious projects exist, but none of them has already proved its efficiency or even its feasibility. ATC automation is really short of effective solutions.

Obviously, major technology improvements (FMS, Data Link, 4D-Navigation, computational power) must be extensively used, but in the meantime, full automation cannot be a solution (at least for the next two or three decades); human controllers must remain in the decision making loop. As automation cannot replace human operators, it must assist them. This is a paradox of automationexisting within the ATC system: as long as full automation feasibility and efficiency will not be proved (i.e., as long as we will need controllers to make decisions, even in an intermittent way) it is essential to preserve the controllers' skills. No matter the tools that will be designed, in is imperative that human controllers continue to exercise their skills.

Human operators are marked by of flexibility, of capability to deal with unexpected situations, of creativity, and of safety, thanks to their capability to compensate for machine's failures or inadequacies. To preserve these capabilities, we may have to automate "less" than possible from a purely technological point of view.

In the mean time, the human operators are an error factor. From this observation, and for years, system designers thought that the more human operators to be put into a system, the more the risk of error would decrease. In fact, they add another kind of difficulty to the

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PRECEDING PAGE BLANK NOT FILMED
PAGE 274 INTENTIONALLY BLANK
C-4
N

supervision of the initial system: the difficulty of understanding the behavior of the automatisms that partly monitor the system. Thus, automation makes the operators loose their skills as they know less about the initial system. It creates additional sources of errors and, as reported in numerous examples, the consequences of these errors are much more important than the previous ones. Instead of eliminating human operators with the consequences of depriving the joint system of major benefits and of increasing the risk of errors, it seems more sensible to design a system which is error-tolerant. Such a system cannot be designed only from the technical advances: we must automate in a different way than suggested by the use of technology alone.

The consequences are very important for the design of the future system as well as for its validation. Intuitively, we must not only validate the machine components of the joint system, but we must also verify that the human-machine cooperation works well. As we need to take advantage of the human as a factor of safety and flexibility, we must prove that this requirement is fulfilled. Some aspects of validation can be less critical than with fully automated systems.

Verification and Validation From a Cognitive Engineering Point of View

Cognitive Engineering

Cognitive Engineering as a Method of Designing New Tools. Cognitive engineering has arisen from the expression of a need by numerous system designers: the need to understand what really makes the task difficult for the operators, how this difficulty impairs human performance in order to define the most effective aids (Rasmussen, 1986; De Montmollin & DeKeyser, 1985; Hollnagel 1988), etc. Cognitive engineering is about the multidimensional open worlds which are the effective working context of the operators (Woods & Roth, 1988). Its aim is to understand and describe the present mental activity of operators, given their present tools, and how these mental mechanisms decay under time pressure, fatigue and stress. Cognitive engineering also enables the human factors professional to anticipate how new technologies will modify the activity of operators.

We must not only elicit the knowledge of operators, we must first understand how this knowledge is activated and utilized in the actual problem solving environment. The central question is not to identify the domain knowledge possessed by the practitioner, but rather to point out under which conditions this knowledge is (or is no longer) accessible. This is the problem of defining a validity domain for human performance. Cognitive engineering must also identify and predict the sources of error and the mechanisms of error (Hollnagel, 1991).

When several agents can act on the system, under which conditions can they cooperate efficiently under time pressure? What are the mental resources that are involved and what is the cognitive cost of cooperation (Woods & Roth, 1988)?

Then we have to point out how the present tools are inadequate and determine how the operators compensate for the deficiencies of their tools. Thus, we have to examine how tools provided for an operator are really used by the operators.

Cognitive Engineering as a Method of Validating These Tools. This whole analysis enables us to explain the cognitive model of the operator which is central to defining a global approach to design effective decision aids. In the meantime, this model provides a guideline for validating the joint human-machine system. Validation has no meaning per se; we have to validate according to some criteria which one has to specify. Cognitive engineering enables one to point out the weak aspects of the system as well as those of the human-machine interaction. Thus, it enables the transformation of high level validation requirements into relevant criteria to test the joint human-machine system. It determines which aspects of the machine or of the human-machine interaction must be verified closely so as to guarantee an effective performance of the whole system or to prevent error. Then one can determine or assess the gains along these dimensions.

We must verify that the new joint human-machine system preserves the sources of strong performance and really improves the weak points from a safety point of view as well as from capacity. This suggests that we must assess the performance of the new system with reference to the previous one in real conditions, i.e., whatever the variability of the real world is.

In the meantime, some crucial questions one should always ask are: does the new system preserve the sources of good performance of the operator, or does it preserve the capability of the operator to deal with unanticipated situations or machine deficiencies? These questions become more important as we consider cognitive tools.

Cognitive engineering provides a relevant framework within which one might answer basic questions such as: who has to determine validation criteria, when do we have to experiment so as to verify these criteria, and how and where to experiment? As cognitive engineering is an iterative process at any cycle after designing new tools, the experiments enable one to determine how the global human-machine system evolves, how the bottlenecks in the operator's activity evolve, disappear, decay or are created; what kinds of problems are solved and created by the new system, the new human-machine cooperation philosophy, and what are the consequences of this philosophy on operators' training.

A Field Study: The Case of Air Traffic Controllers

Method

The following is the description of ERATO, a project which is aimed at defining an Electronic Assistant for en-route air traffic controllers. This electronic assistant will include several decision aids. Figure 1 shows that this project includes seven different phases.

At first, we have to elicit the cognitive model of the controllers (1). This model explains the mental mechanisms which are common to all controllers and which enable them to process data and to make real time decision. These mental processes are analyzed for the executive controller, the planning controller and then for both controllers as a whole in order to assess the consequences of cooperation on mental load as well as on global performance. The main goal remains to describe the mental mechanisms involved in decision making process and how these mechanisms evolve and decay under time pressure.

The bottlenecks assessment (2) is a diagnosis phase: we have to point out the sources of poor and good performances of the air traffic controllers given their actual working context. As long as the situation is not too demanding, controllers can compensate for these bottlenecks; but in very demanding situations these bottlenecks may severely impair the controllers' performance. The assessment of bottlenecks then enables one to specify the basic functions of effective decision aids (3). The specification of the interface (4) also depends on the cognitive model. We need to know in which context these tools will be operated so as to optimize their use.

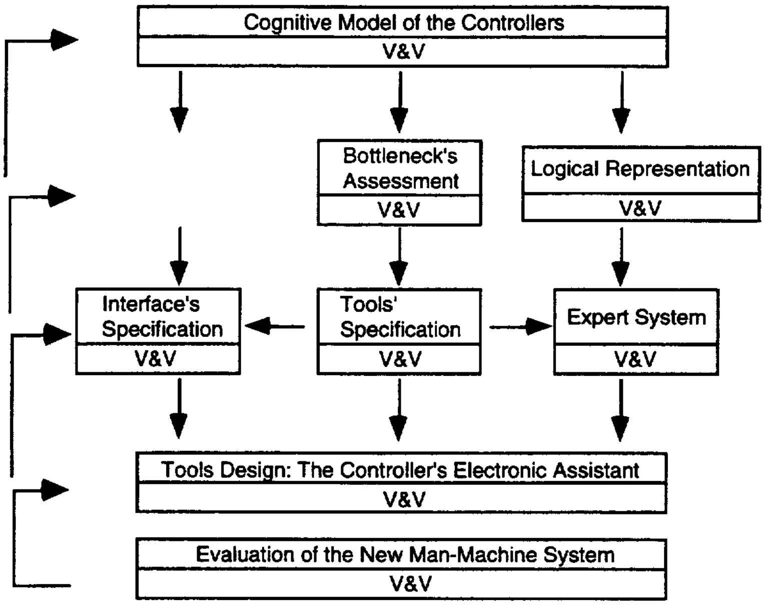

Cognitive Model of the Controllers
V&V
Bottleneck's Assessment
V&V
Logical Representation
V&V
Interface's Specification
V&V
Tools' Specification
V&V
Expert System
V&V
Tools Design: The Controller's Electronic Assistant
V&V
Evaluation of the New Man-Machine System
V&V

Figure 1. Seven phases for electronic assistant

To define a logical representation of the model (5), we need to combine different laboratory logic so as to build a logical tool adapted to formalize controller's knowledge. The design of the function defined during step (3) involves the use of an expert system which models large subsets of controller's knowledge (6). The expert system provides the two Electronic Assistants with relevant data, so we have to face the problem of the integration of the expert system to real time functions so as to design the two electronic assistants (7).

What Really Makes the Task Difficult

They have to process data which depend on the time factor for:

- Their value

- Their accuracy. When observing air traffic controllers, we found that they spent a lot of time, and a lot of cognitive resources, in eliminating ambiguity. A major reason why controllers are often unable to make a clear assessment of the situation is based on their representation of predicted time intervals: unless we sink into pure fatalism, we do not anticipate that an event will happen "at" a given time but "about" a given time. The difference is fraught with consequences for the operator.

- Their availability. All the data necessary to make a clear assessment of the situation are not available at a given time; some of them may be definitely unattainable. The operator may have to take decisions in a state of partial ignorance.

- Their flow. The controllers must deal with any cases comprising the dual aspects of both time-dependent information-processing and real-time decision making as: (i) dynamicity vs. inertia of the system or of any subsystem, (ii) lack of data vs. data overflow, and (iii) prolonged low workload vs. tasks overflow. We must also consider the question of how operators can adapt to sudden transitions from any of these aspects to the dual one, for example from a situation of lack of data to a data overflow.

- Data presentation is technology driven and uselessly bulky. Tasks and objectives are not well defined and may severely compete. Risk is so important that the controller has to guard against errors from all the actors (including the machines) in the system.

A Rapid Overview of the Controller's Cognitive Model

The controller anticipates, according to a "normal" routine, behavior of the aircraft in what is called the default world, with reference to the "default logic" that models this kind of reasoning. This default behavior is illustrated by controllers when they use sentences as "normally this aircraft going to Paris Orly will start descent about 30 NM before this fix." The controller does not know the top of descent, but from his experience, he knows that "this will normally happen about here," so he will ignore all potential conflicts that should happen if the given aircraft should start descending earlier to focus all his activity on the most probable conflicts. This is an efficient means of narrowing the range of contingencies to be examined and to increase efficiency: at first all the aircraft are processed as if their behavior always remain consonant with the "normal" behavior.

But to process exceptions, that is to guarantee safety, controllers monitor "sentry parameters." As long as these parameters remain in a "normal" range, all the previous diagnoses or decisions that are inferred from the default world remain valid. But if a sentry parameter drifts outsides the expected range, then all the previous plausible inferences have to be revised; some additional conflicts can be created, due to this abnormal behavior. In normal situations, this way of reasoning is an efficient and safe way of making decisions in a state of partial ignorance. But we can observe that, in very demanding situations, the monitoring task may no longer be performed by the controllers. Thus, when outside its validity domain (in too demanding situations), this mechanism may become a major source of errors.

Very often, diagnosis is the result of an ambiguity elimination mechanism. Even when remaining in the default world, the controller is often unable to assess a definite diagnosis. Controllers spend a large amount of time in ambiguity elimination processes. Allowing a doubt is a luxury for the controller; the mastery of doubt is an art. This is performed by associating one (or two) relevant parameter(s) to each undecided situation. To avoid a scattering of resources, these parameters will remain the only ones monitored.

Each conflicting situation, certain or potential, triggers several resolution frames. These frames are a part of the knowledge base common to all controllers. For example, let us consider the following situation: two aircraft converge over the same point, the second climbs through the level of the first. This canonical situation can trigger three resolution frames :

- Clear the climbing aircraft directly to the requested level, under radar monitoring
- Radar vectoring
- Clear the climbing aircraft to a safety level until both have flown over the fix.

A part of the controller's activity is devoted to choose the best frame. Each of these frames may be relatively demanding. In demanding situations, the cognitive cost becomes a basic criteria to choose a frame. Of course while resolving a problem, the controller may have to shift from an inoperative frame to a more relevant one.

According to the assessment of the workload, the controller can instanciate a resolution frame in a more or less efficient way. The controller can also abandon a more elegant frame to shift to a more efficient one: this is the consequence of his or her own resource management policy, according to the problems occurring at a particular time.

All of these mechanisms are a part of the real time data process. This process results in a problem driven organization of the raw data set.

To solve a given conflict, the controller may have to perform actions within far-off time intervals. These time intervals may be very short; if a controller misses the right time span to act on traffic, the very nature and complexity of the problem may change rapidly.

The only way to avoid this is to frequently monitor the position of the relevant aircraft. Obviously this mechanisms is very costly. The controller must shift frequently from one problem to another and, at each shift, must restore the resolution context. When conflicts are complex and under time pressure, this may become a critical task.

Most of the previous tasks can be performed by two controllers, successively or in parallel. Mental mechanisms involved in cooperation are an essential part of the model. Efficient cooperation between the two controllers relies on three factors.

They must have

- The same skills, knowledge and training
- The same representation of effective traffic requirements
- Simultaneously available cognitive resources to exchange information.

When demand increases, these two latter conditions may decay so much that cooperation may no longer be effective. Numerous near misses have been reported due to cooperation failure in too demanding situations. One controller did not even know that some tasks were urgent and important while the other controller thought that these tasks were normally performed. This points out the limits of cooperation based on implicit tasks delegation.

Conclusion: Validity Domain of all the Mental Processes

All these mental mechanisms have a validity domain. We can easily observe how their efficiency decays under stress, time pressure or fatigue.

For example, in demanding situations, the sentry parameters are less monitored and this may lead to errors when an abnormal behavior is not detected soon enough. We can also observe that the assessment of a given situation, including conflict detection and resolution assessment, needs a few tenth of seconds while it can lay on 7 to 8 minutes in very demanding situations; in this case, the controller is confronted with problems associated to numerous shifts from this conflict to concomitant ones, as described before and the risk of error (forgetting a relevant aircraft, choosing the wrong resolution frame, etc.) is high. The validity domain of the mental processes directly depend on the number of aircraft that have to be processed by the controller. This is the reason why we focused on a problem driven presentation of the information.

The efficiency of the mental processes also depends on the capability of the operator to focus his attention on the relevant problem at the right time.

Justification of the Tools

This analysis, which corresponds to phases 3 and 4 of the project, is the key point of the approach. It explains the reasons why each tool has been designed, and the improvements of the joint Human-Machine system that are expected. It also defines the criteria to experiment this system. At this level, there is a deep symbiosis between design and validation; however, this can no longer be true during the experiments.

Information Filtering

Justification. The aim of problem-driven information filtering is to reduce the number of aircraft to be considered at one time. By splitting a too demanding situation in several subsets of aircraft, we can expect that the controllers will have the capability to process these subsets of aircraft very efficiently. As we do not provide the controllers with the results of an automatic conflict detection and resolution, they will have to operate all their mental mechanisms to assess the situation. This should preserve their skills and their capability to deal with any unanticipated situation more properly then now. The expected gain is that, as they will be working on appropriate subsets of aircraft, these mental mechanisms will be much more efficient than now. Thus, we will have to verify that this human-machine cooperation philosophy enhances:

- The way they anticipate in a state of partial ignorance
- The associated sentry-parameters' monitoring processes
- The ambiguity elimination processes: the definite assessment of the situation should be made earlier than now.

- The choice of the relevant resolution frame should be made earlier than now and instanciated in a more "elegant" way.

- The cooperation between controllers should be improved. The information filtering is supposed to enhance the definition of the mental representation. Both controllers' mental representations of the situation should remain consistent over time, as they will be able to update it very easily.

Design Problems. The role allotted to the expert system is to provide the electronic assistants with adequate data in order to show how to organize the raw data in a problem-driven way. Information filtering techniques are under dispute (DeKeyser, 1987). The point is how to make sure that the operator will not need data that is hidden by the system. Data retention and access should not be a source of errors.

In order to answer these precise questions, the expert system must not only encode an exhaustive model of the controller, but we must also carefully define its role in the system.

Description of the Expert System. The first version of the expert system included about 3,000 Prolog first order rules. It processes the same set of data as the controllers have to process now (i.e., the information from the strips) and, when available, the radar information. It includes two main modules. The first computes the default representation of each aircraft. From this representation, the second module associates to each aircraft its relevant environment, called the interfering aircraft subset (IAS).

This environment is composed of :

- The subset of all conflicting aircraft. These conflicts may be certain or potential. This subset is not determined by means of a pure mathematical computation, but rather according to the current expertise of controllers.

- The subset of all the aircraft that may interfere with a "normal" radar resolution of the conflict; that is, all aircraft that may constrain conflict resolution. A normal resolution is a solution which is consistent with the current know-how of controllers.

The relevant environment of an aircraft is typically a problem-driven filtering of information. The IAS represents the relevant working context associated to an aircraft. Such an environment embodies traffic requirements and all information that may be useful to fulfill these requirements. The number of rules is explained by the need to represent current knowledge of the controllers so as to make sure that information filtering really meets the controller's needs.

The definition of the relevant environment is: "according to the traffic requirements, provided the EC works normally, he may need all, or a part of, the displayed data, but he will in no way need any other data."

Discussion. The discussion on the exhaustiveness and the relevance of data filtered by the expert system is central.

- The first answer consists in taking into account the default behavior of the aircraft in a more "prudent" way than the controller. This will result in the display of some aircraft that may be not relevant for the controller. In the most demanding situations (more than 25 aircraft in the sector), the most numerous IAS never include more than 12 aircraft. If one or two additional aircraft are displayed, this is not really a problem. In all cases, the number of aircraft displayed as a result of information filtering remains lower than the maximum efficient processing-capability (about 15 aircraft), while the initial number was largely above this figure.

- The system detects all potentially abnormal behavior of an aircraft in order to advise the controller as soon as possible and to update information filtering accordingly. In future versions, this mechanism should be performed using FMS/Data-Link capabilities.

- But these two first answers do not really solve the problem. The knowledge elicited in the expert system defines a set of "normal behaviors" of the controllers. But whatever the number of rules can be, it is impossible to represent the whole knowledge of all the controllers. Should we be able to do this, we should have to deal with controllers' errors or creativity. The solution defined in Erato consists of considering the expert system as a default representation of the controllers. To guard against the consequences of human error or creativity, (i.e., unexpected behavior) a monitoring process exists. This process is inspired by the natural sentry-parameters' monitoring process of the controllers. This monitoring process will detect any discrepancy between the actual position of all aircraft and any of the possible position as it could result from a "normal" behavior of the controller. When necessary, this process will trigger an alarm to advise the controller that the previous information filtering is no longer relevant and has been updated. We have to make sure that this mechanism is efficient in demanding situations, and that the controller is not overwhelmed by the warnings. In other terms, the expert system must be accurate enough.

This monitoring process associated with the expert system allows the electronic assistant to smoothly adapt to operator error and creativity. Such information filtering is error tolerant.

These results are used by several functions of the electronic assistant: simulation of problem's resolution, extrapolation, memorization aid, data transfer from one device to the other, cognitive-resources-management aid. The problem-driven information filtering allows the controller to concentrate on well-formulated problems in order to operate in a more efficient and creative way. This function substitutes a set of easily manageable problems for the initial complex situation. The basic information filtering will be used by all the functions of the Electronic Assistant.

The Extrapolation Function

Justification. This function will enable both controllers to improve problem formulation. As long as the aircraft are not in radar contact, the controller has to process data from the strips. In demanding situations, it is easy to observe that the strips are not read in an exhaustive way, which can cause severe errors. This function will display the situation at any future time, taking into account the default behavior of all the aircraft. It substitutes a graphical display for an alphanumeric representation of data, and this will improve the choice of the right resolution frames. It will also enable the controller to more easily assess where uncertainty lies, and consequently to more efficiently point out the parameters which are relevant to eliminate ambiguity.

Design Problems. It is commonly admitted that operators spend a significant part of their activity in compensating for tool deficiencies. An ill-adapted interface can significantly devalue the results of information filtering. We have suggested that very often the reference used by the controller is not a temporal one but a spatial one: the question "when will you act on this aircraft" is answered "There." Thus, the interface will enable the controller to drag an aircraft along its trajectory with the mouse; all the other aircraft will move accordingly. This interface responds in the way the controller anticipates. In some situations, the controller refers to a temporal reference. The interface will display simultaneously the time corresponding to the simulated position of the aircraft. However, should this interface have had only one of these references, the controller should have had to mentally convert distances into time intervals. In demanding situations this could represent a significant additional workload.

Simulation Functions. These functions will allow the controller to experiment with different resolution frames and to answer questions such as "what would happen if I climb this aircraft to..." or "Is it better to set course directly to..." The expert system will deliver the simulated information filtering. These answers will be updated until the controller has made a decision. For the time being, the controllers have no tool that assists them in performing this task.

Memorization Aids. The controller will have the ability to indicate the trajectory section where he intends to vector an aircraft. When the aircraft flies over this point (or abeam), a warning will be triggered; then, after having consulted the filtered information, the controller will initiate his decision. This should solve both problems of keeping onto memory the "right time to act" and context-resolution real-time updating.

Data Transfer Between Electronic Assistants

This function as aimed at improving cooperation between controllers. It enables them to transfer any filtered information or simulated result from one position to the other. Using this small set of data, the two controllers will have the same mental representation of the requirements of the situation. The capability to save incoming messages into a letter box until the controller has time to process them should help to solve the problem of simultaneous availability of both controllers to exchange information.

The Reminder

Justification. We observed how it may become difficult for the controller to focus complete attention on the right problem at the right time. The reminder consists of a specific window of the electronic assistant where each problem will be associated a label. A problem is defined as a conflicting situation involving two or more aircraft. The labels are positioned according to the urgency, and the display of the relative urgency of problems should enable the controller to avoid wasting cognitive resources on non-urgent and unimportant tasks while the short term situation decays. In normal operations, this should allow the controller to objectively manage all cognitive resources.

The aim of the reminder is to show what the traffic requirements are and their urgency to the two controllers. There are several ways to split a given situation into relevant problems. This variability can be observed for several different controllers as well as for any given controller, according to his cognitive resources management philosophy. The more the situation is felt to be demanding, the more the controller will split it in "little" problems and solve these problems in a very tactical way: with short term solutions. If the controller feels that the situation is mastered he or she will consider these elementary problems as a part of a whole and solve them in a more strategic way. Statistically, about a quarter of the problems may be more or less broken down, while the three other quarters are always described in the same way by controllers.

Design Problems. The reminder will propose labels by default. Most of these labels will correspond to the effective needs of the controller; some will not. Thus, the controller will have the capability of editing these labels.

When a situation can be described as a single problem or as several problems, the reminder will propose the simplest situation for three reasons:

- It corresponds to the need of the controller when the situation is the most demanding. If the controller wants to concatenate some sub-problems, the situation is probably not too demanding, and there is time to do this properly. This is a means of avoiding clumsy automation: the interface assists the operator more in the most demanding situations.
- It is easier to concatenate the labels then to split them.
- This enables the system to point out more information on each sub-problem.

Validation Techniques

Classically, we define dependability as that property of a computing system which allows reliance to be justifiably placed on the service it delivers (Laprie, 1987). We can point out four classical methods regarding dependability-procurement or dependability-validation: fault avoidance, fault tolerance, fault removal and fault forecasting. These definitions can be applied to a complex heterogeneous human-machine system as the ATC system as well as to any of its machine components. In the first case, the users are the airlines (or their passengers), while in the second case the user is defined as the controller or any subsystem.

The specification of decision aids relies on a philosophy of future human-machine cooperation, whether this philosophy is clearly defined or not. The central question is to make sure that this cooperation fulfills the initial requirements regarding capacity and safety. To answer this question, we have to choose the right parameters to be evaluated and then determine the minimum set of experiments to get a significant amount of data (Woods & Sarter, 1992). Some basic questions must be answered about the robustness of the joint human-machine system: is it error tolerant (Foster, 1992)? Does its organization allow a quick and easy correction of errors (Reason, 1992)?

The design of decision aids implies an analysis, either implicit or explicit, of operator deficiencies and of the most effective means to compensate for these deficiencies. The ultimate step of the verification and validation process should be the verification of these initial assumptions (Hopkin, 1992).

The validation of the Electronic Assistant is threefold (Leroux, 1992).

- The cognitive model has been verified through the analysis of the behavior of air traffic controllers during simulations of demanding traffic (Leroux, 1991). The observation of the controllers and their comments enabled us to point out the mental processes and the bottlenecks as described in the model.

- The philosophy of human-machine cooperation and the specification of the electronic assistants have already been presented to more than 180 fully qualified controllers.

- The expert system is used to provide the various functions of the electronic assistant with adequate data. To validate this knowledge-based module, we have to check that the outputs are acceptable by controllers on real conditions; i.e., that they always include at least the minimum set of information. Although the validation of the knowledge based module has not been carried out, it has been successfully confronted to the most usual problems. However, we will have to validate very carefully the monitoring process, as it guarantees the relevance of the outputs. We must prove that this mechanism detects all discrepancies between the observed state of the world and the expected one. The trust of the operator on the machine relies on the efficiency of this mechanism. Thus, the validation of the knowledge-based module will be twofold: from a pure dependability point of view and from a human factors point of view. This will be performed by two different teams with different techniques to experiment.

- The first level of validation of the Electronic Assistant will consist of:
  - Experimenting the interface of each function
  - Verifying that it really improves the target bottlenecks
  - And in making sure that it does not decay some sources of good performance. The criteria for experimenting with these functions are directly issued from the cognitive model, as previously described.

Then we will have to assess the validity domain of each function: are they really efficient for situations where the controller needs an effective aid?

- After that, we will have to validate the electronic assistant as a whole in order to analyze if the function-as-a-part-of-a-whole looses properties or acquire unexpected ones.

- Finally, we must assess the performance of the joint human-machine system and answer questions such as:
  - How is this cooperation philosophy is accepted by controllers?
  - How will it modify the activity of the controllers?
  - Does it enable them to work in a more efficient and creative way?
  - Does it provoke a loss of vigilance or of skill?
  - Does it improve the global performance, from capacity and safety point of views?
  - What are the consequences on training?
  - Does it enable a progressive and "soft" integration of technological advances in avionics?

References

Bainbridge, L. (1987). Ironies of automation. In J. Rasmussen, K. Duncan, & J. Leplat (Eds.), *New technology and human error.* John Wiley & Sons.

Celio, J. C. (1990, February 2). Controller perspective of AERA (MP-88W00015 rev1 / Project 1764c Dept W42). Mitre Corporation.

De Keyser, V. (1987). How can computer-based visual displays aid operators? *International Journal of Man-Machine Studies, 27,* 471-478.

De Montmollin, M., & De Keyser,V. (1985). Expert logic vs. operator logic. In G.Johannsen, C. Mancini, & L. Martensson (Eds.). *Analysis, design, and evaluation of man-machine systems.* CEC-JRC Ispra, Italy: IFAC.

Foster, H. D. (1992). Resilience theory and system valuation. In J. A. Wise, V. D. Hopkin, & P. Stager (Eds.), *Verification and Validation of Complex and Integrated Human-Machine Systems.* Vimeiro, Portugal: NATO Advanced Study Institute.

Hollnagel, E. (1988). Information and reasoning in intelligent decision support systems. In Hollnagel, Mancini, & Woods (Eds.), *Cognitive Engineering in Complex Dynamic Worlds.* London: Academic Press.

Hollnagel, E. (1991). The phenotype of erroneous actions: Implications for HCI design. In A. Weir (Ed.), *Human-Computer Interaction and Complex Systems.* London: Academic Press, Computers and people series.

Hopkin, V. D. (1992). Verification and validation: Concept issues and applications. In J. A. Wise, V. D. Hopkin, & P. Stager (Eds.), *Verification and Validation of Complex and Integrated Human-Machine Systems.* Vimeiro, Portugal: NATO Advanced Study Institute.

Laprie, J. C. (1987). Dependable computing and fault tolerance at LAAS: a summary. In A. Avizienis, H. Kopetz, & J. C. Laprie (Eds.). New York: Springer-Verlag Wien.

Leroux, M. (1991a). Premier bilan des experimentations du modele Erato, specification d'outils cooperatifs pour le controleur. Rapport CENA 91538.

Leroux, M. (1991b). Erato (en route air traffic organizer). *Proceedings of the Sixth International Symposium on Aviation Psychology,* Colombus (Ohio).

Leroux, M. (1992). The role of verification and validation in the design process of knowledge based components of air traffic control systems. In J. A. Wise, V. D. Hopkin, & P. Stager (Eds.), *Verification and Validation of Complex and Integrated Human-Machine Systems.* Vimeiro, Portugal: NATO Advanced Study Institute.

Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering.* New York: North Holland.

Reason, J. (1992). The identification of latent organizational failures in complex systems. In J. A. Wise, V. D. Hopkin, & P. Stager (Eds.), *Verification and Validation of Complex and Integrated Human-Machine Systems.* Vimeiro, Portugal: NATO Advanced Study Institute.

Roth, E. M., Bennett,K. B. & Woods,D. D. (1987). Human interaction with an "intelligent" machine. *International Journal Of Man-machine Studies 27,* 479-525

Villiers, J. (1968). Essai sur l'evolution du controle de la circulation arienne en route.

Villiers, J. (1992). L'homme face aux systemes techniques complexes. *Futuribles, 167,* 111-125.

Woods,D. D. (1986). Paradigms for intelligent decision support. In Hollnagel, Mancini, & Woods (Eds.), *Intelligent decision support in process environments.* New York: Springer-Verlag.

Woods, D. D., & Sarter, N. B. (1992). Field experiments for assessing the impact of new technology on human performance. In J. A. Wise, V. D. Hopkin, & P. Stager (Eds.), *Verification and Validation of Complex and Integrated Human-Machine Systems.* Vimeiro, Portugal: NATO Advanced Study Institute.

N95- 34784

Certification for Civil Flight Decks and the Human-Computer Interface

530-04
[illegible]

Andrew J. McClumpha[1] & Marianne Rudisill[2]

[1]RAF, Institute of Aviation Medicine
[2]NASA Johnson Space Center

Introduction

This paper will address the issue of human factor aspects of civil flight deck certification, with emphasis on the pilot's interface with automation. In particular, three questions will be asked that relate to this certification process: (1) are the methods, data, and guidelines available from human factors to adequately address the problems of certifying as safe and error tolerant the complex automated systems of modern civil transport aircraft; (2) do aircraft manufacturers effectively apply human factors information during the aircraft flight deck design process; and (3) do regulatory authorities effectively apply human factors information during the aircraft certification process?

Problems with Automation

Progressive automation is a feature of the modern civil cockpit and the trend toward more information, greater complexity, and more automated aircraft has the potential to significantly isolate the pilot from the aircraft resulting in less understanding and awareness. Billings (1991) reports that the dominant cause of aircraft accidents is human error and that the most important purpose automation can serve is to make the aviation system more error resistant and error tolerant. It is generally accepted that future developments in the civil flight deck environment will have substantial automation in order to provide cost-effective air transport. However, it is widely reported that the dominant factor in many civil aircraft accidents in the recent past has been automation and that automation has had a "debilitating" contributory role (Sheridan, 1991). For example, 65 to 70 percent of civil jet transport accidents are attributed to human error and of that amount the majority were controlled flight into terrain (Hughes, 1989). Accidents caused by controlled flight into terrain (CFIT) are of particular concern to human factors, since a primary contributor appears to be the pilot interface with automated aircraft (Lenorovitz, 1992a, 1992b).

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PRECEDING PAGE BLANK NOT FILMED

PAGE 288 INTENTIONALLY BLANK

Pilots criticize flight deck design. Poor human factors are cited (although often not explicitly) as a contributing factor in a wide range of aircraft accidents. Often these problems are caused through the pilot's interaction with "automation." The Air Accident Investigation Branch (AAIB) report on the Boeing 737-400 accident at Kegworth, U.K. (Department of Transport, 1989) noted that some time was lost as the copilot attempted unsuccessfully to program the flight management system (FMS) to produce the correct flight instrument display for landing at East Midlands Airport. From the cockpit voice recorder it is inferred that the first officer selected the route page and was entering the correct information (for the destination airfield) but as an enroute point. He did not notice the inadvertent selection nor understand the limitations of the available selections with respect to this information. An absence of appropriate feedback within the FMS allowed the error to remain. The first officer failed to select the arrival airfield page; this page is similar to the enroute data page in terms of data layout and data entry. In addition, there was no clear, unambiguous indication (e.g., a title) of the selected page within the FMS. The Airbus A320 FMS is also often criticized for not providing easy and efficient information access and selection. Pilots report having to select up to five different pages to obtain information necessary to complete a single action sequence. Both examples illustrate violations of well recognized, explicit guidelines that apply to the design of human-computer interfaces (HCI).

The results of recent surveys addressing attitudes towards glass cockpits have shown that pilots report an erosion of flying skills, increased workload at critical times of flight, and a sense of being "out-of-the-loop" as a result of automation (McClumpha, James, Green, & Belyavin, 1991; Weiner, 1989). An analysis of the comments made by respondents of the Royal Air Force (RAF) Institute of Aviation Medicine's (IAM) survey of pilots' attitudes toward automated aircraft highlighted two main types of comments about the human-computer interface as it relates to civil flight deck automation (Rudisill, in press). One set of comments related to specific systems and specific problems with the HCI on the aircraft. Although this information is particular to an aircraft type, it indicates human factors problems that could be used to help develop human factors assessments required for a certification program. The other set of comments, however, related to the broader nature of automation and the difficulty pilots experience interacting with automated equipment. For example, pilots reported difficulty in knowing what the automation was doing, what the boundaries or limits of performance were, and how to intervene effectively if problems arose. Pilots also reported that many of these types of problems emerge only after considerable line flying or with specific experience in unusual situations. Exposure to these types of situations forces pilots to appraise their ability to understand and feel competent with automation. These issues are the ones that present considerable difficulty for manufacturers in terms of how to design aircraft systems and their pilot interfaces. Automation also presents difficulties for regulatory authorities in terms of how to certify designs as safe and error tolerant.

Billings (1991) stated that only a subset of conditions and failures can ever be evaluated in systems as complex as modern civil flight decks. It is not, therefore, surprising that Weiner and Curry (1980) concluded that the rapid pace of automation is outstripping the pilot's ability to comprehend all the implications for aircraft performance. They reported that an uneasy balance now exists between accepting aircraft in which all the implications and potential for debilitating behavior are unknown and, on the other hand, training to support crew tasks when operating these highly automated aircraft.

Civil Aircraft Design and Regulatory Authorities

Regulatory authorities are responsible for providing certificates of airworthiness for all aircraft operating within their airspace. To that end, their role is to act as the arbiter for the aircraft acceptance process. They assure, among other things, that aircraft perform to engine power and thrust requirements and that airframe structures have appropriate mechanical integrity. They assess, to stringent requirements, the ability of aircraft to operate to a safe landing under circumstances unlikely encountered during normal flying. Therefore, the essence of regulatory authority certification is whether aircraft meet minimum acceptable standards identified as requirements for certification. The authority determines acceptance by defining a number of requirements and criteria and applying them during systematic aircraft evaluation.

Aircraft manufacturers take great care in the design process leading to certification. The Boeing® 757™ flew 3,942 hours of simulator missions before certification in 1982. The Boeing 767™ program involved 5,348 hours of simulator flying and the 747-400 accumulated 5,981 hours. A total of 369 pilots from 29 airlines flew the 747-400™ simulator and nearly 200 non flying personnel were involved in the design process (Hughes, 1989). A regulatory authority may require more than 1,000 hours of simulator flying and over 100 actual landings before certificates of airworthiness are issued. There is no doubt that certification of civil aircraft by regulatory authorities is approached thoroughly, diligently, and carefully and that aircraft manufacturers are responsive to certification requirements and design for safe, error tolerant systems.

Aircraft design, particularly flight deck design, is evolutionary. This evolutionary approach carries with it two strong implications from a human factors perspective. One is that a flight deck will not vary substantially from previous versions. This helps ensure that pilot training for that aircraft will be kept to a minimum (i.e., cost effective). The second and more disquieting implication is that flight deck design appears to have evolved without systematic adherence to any rigorous human factors scientific base and that this situation has been perpetuated through the evolutionary nature of aircraft design. Norman (1991) takes a strong position with regard to this issue. He argues that cockpit procedures, flight instruments, and regulations are "guided more by instincts, anecdotes, and reactions to individual incidents rather than by systematic, scientific analysis of the issues." There appears to be a failure to comply with a basic tenet of human factors which dictates that the user interface (which, broadly defined in this context, includes procedures, instruments, and regulations) should be derived from an understanding of the tasks, procedures, and information requirements of pilots. This understanding and information should then be mapped to the interface design in a way that exploits the pilot's sensory, perceptual, physical, and cognitive processes. However, certain flight situations (e.g., engine fire) require specific actions by crew members in clearly defined ways and the crew is legally responsible for following set operating procedures. Unfortunately, even in this situation, nowhere is it explicit that procedures, instruments, and regulations were derived from understanding the pilot's task, procedures, and information requirements and subsequently mapped to the interface design so as to support the pilot's cognitive activities.

Existing certification requirements do, however, establish the need to provide specific information to the crew. Unfortunately, neither the form or content of information nor the pilot's ability to access it easily and efficiently for required tasks appear as established certification requirements. A recent research program at the NASA Langley Research Center investigated information that pilots use and how long and when they use it. It is interesting that fundamental research of this nature was only underway as late as 1991. A case could be made that there has been a fundamental shift on flight decks in the pilot interface from one that is essentially electromechanical to one that is largely software based and automated, consequently driving a need for research in this area. Nevertheless, the pilot's task has not changed appreciably as a result of technology changes. Therefore, pilot information requirements have remained essentially unchanged and should have served as the basis for design, independent of final implementation technology.

Neither is there apparently a requirement levied on the aircraft designer by the regulatory authority for an aircraft design to conform to detailed and explicit requirements for good quality human engineering. The specific number of hours flown and/or tested in a development and certification program is not necessarily an indication of good design. On the other hand, what is examined, found, and remedied during the design process is of greater interest. Information, though, on the form and content of specific human factors examination does not appear readily available. More specifically, the pilot interface with aircraft systems does not appear to have been subjected to rigorous and critical examination based on existing human factors and human-computer interaction methods, standards, or guidelines. There is evidently no requirement to systematically evaluate the aircraft for adherence to good human factors criteria.

The certification authority, as the arbiter of civil aircraft acceptability, maintains responsibility for ensuring that the pilot interface provides the necessary facilities for safe flight. Besides ensuring aircraft conformance to the traditional requirements such as appropriate propulsion and structural integrity, they should also establish pilot interface conformance requirements that rely on good quality human engineering based upon established human factors design principles. Mechanisms and processes to evaluate human factors systematically do not appear to be in place with the authorities; it is suggested that they be established to help foster sound human factors flight deck design.

The certification process is flexible enough to allow a manufacturer to request a derogation from the certification requirements. Derogation or relaxation of requirements can be permitted at the discretion of the regulatory authority, but only when it is considered safe to do so. The question asked by a regulatory authority is whether an aircraft is likely to be as safe as a previous version. Unfortunately, human factors requirements are strong candidates for derogation because of the often subjective and generally implicit nature of human factors evaluation. There is also a concern with the possible cumulative effect on pilot performance of numerous separate derogations. The possible effect of any derogation on the pilot's ability to operate the aircraft safely should to be considered as part of the derogation analysis. However, as a result of the generally subjective nature of the human factors assessment, particularly with regard to HCI, it is perhaps not surprising that human error is identified as the major contributory factor in aircraft accidents.

We have described the exacerbating effects of progressive automation and the apparent absence of systematic human factors practice in the design of civil aircraft flight decks. The apparent absence of systematic evaluation of human factors by regulatory authorities for aircraft certification has also been commented upon. Two questions now remain: 1) is information available within the human factors community to support the processes of flight deck design and aircraft certification, and 2) what is required to achieve the goal of incorporating human factors into design and certification processes?

Role of Human Factors

The domain of classic human factors contains concepts, principles, and methods that are sufficiently objective and advanced to meet the need for certifying aircraft to specific standards, requirements, and criteria. This information is derived from the scientific base of experimental psychology, human factors, and physiological research. It describes human sensory, perceptual, and motor performance, and, to a lesser degree, cognitive performance. This empirically-based information is potentially available for the analysis and certification of sensory and perceptual characteristics, manual operation, and physical effects (e.g., noise, vibration, anthropometry, biomechanics) operating within the pilot interface environment. For example, anthropometric design standards can precisely define acceptable reach requirements for appropriate percentiles of the population. It would be highly unusual to design an aircraft flight deck that prevented the pilot from reaching required controls and displays. Neither would regulatory authorities certify such a design. A detailed knowledge of vision and character readability allows the equally precise definition of alphanumeric font size based on the seated eye position of the pilot and copilot. Reach envelopes, workspace layout, display readability, and the acceptable limits of force functions can all be predicted, established, evaluated, and certified, and specific consequences of derogation can be predicted. Many references are available for the design of these system characteristics and these could be adapted for use in certification programs (e.g., Boff & Lincoln, 1988; Department of Defense, 1989).

The situation is rather different for design and certification of the aircraft human-computer interfaces on today's automated civil aircraft. Automation per se and software-based interfaces in general have changed the locus of control within the cockpit from the physical domain to the cognitive domain. Therefore, much of the HCI today on the flight deck is related to human cognitive performance and it is this particular domain of performance that presents the true challenge for certification programs effectively dealing with human factors.

To date, regulatory authorities have generally based examination on theory and models followed by testing to criterion performance. Hence, mathematical modeling is used to predict the likelihood of system failure. There are two problems with this approach when applied to human factors certification, especially certification of the HCI. First, mathematical modeling will be used to predict the failure of individual elements of the system to a 10^{-8} error rate. Assessment of human error as a result of system design is the domain of certification test pilots who will assess how critical a potential error is likely to be. An assessment may suggest that a system has a low error rate and is therefore acceptable. What does not appear to be assessed, however, is the contribution of the error potential of an individual system component with all other aircraft systems nor the effects of combinations of errors created through poor interface design. It is unclear if mathematical modeling is an appropriate technique for modeling or for predicting failures of the pilot-vehicle interface. Indeed, the approach generally taken during HCI design involves designing the interface according to established good design practices rather than modeling after design and then assessing error potential.

It is our contention that regulatory authorities do not yet possess the understanding to effectively review or assess the demands that cognitive systems require. A typical "fallback" position of a regulatory authority is to adopt the "blame and train" philosophy. Human error is often treated as a training issue rather than as a reflection of underlying design deficiencies (cf. recommendation from AAIB report on Kegworth accident). However, the difficulty with this reasoning is that current methods and practices used by regulatory authorities do not appear to address the human-computer interface aspects of system design. Furthermore, they do not appear to support the process in a manner that can lead to improved understanding of potential problems and subsequent establishment of standards and requirements for design for reference by manufacturers. The cognitive aspects of system performance require very different methods and techniques from those applied to the classic domains of human engineering and impose unique methods for certification. The current form and status of HCI science requires a substantial change by regulatory authorities in terms of flight deck HCI design philosophy. These issues are dealt with in the following sections.

Contributions of Cognitive Psychology

In a paper addressing the role of cognitive psychology in designing useful and usable systems, Landauer (1991) addresses current limitations of theory. He suggests that a theory of human computer interaction good enough to allow prediction of system characteristics for design is likely to be impossible because of the complex behavior of systems and the cognition supporting user interaction with the systems. Landauer argues that laws, such as Hick's law, Fitts' law, and laws of visual and auditory perception, have indeed made an impact, albeit minor. The emphasis of this reasoning is that a theory that allows predictions and design decisions from domains other than classic human factors is unlikely to be available for some time, if ever.

The data, methods, tools, guidelines, and standards that would serve as the bases for analysis of more cognitive aspects of pilot operations and design of HCIs to support them do not yet exist in a form readily usable for certification. In a sense this is understandable given the complex nature of the underlying psychological domain. As has already been mentioned, sensory and perceptual processing and manual control are complex behaviors, but they are also more amenable to direct investigation. Systematic investigation of cognitive processing and development of coherent theories has occurred only recently. A base of objective data, sufficiently complex unifying theories of cognition, and an appreciation for how this information should be reflected in an HCI are now being formalized.

However, what is presently available for use in HCI design and certification is an explicit representation of what constitutes good HCI design practice, with supporting methods and validated guidelines and standards based on substantial supporting research. This work occurs in several forms: (1) a software design model that explicitly identifies HCI design tasks and requires iterative prototyping and review, (2) HCI guidelines and standards that embody principles of good HCI design derived from empirical research, and (3) instruments and methods, such as HCI checklists and usability testing tools, that may be used for HCI evaluation and certification. This approach can provide the foundation for effectively integrating HCI knowledge into civil flight deck design and certification processes.

HCI and the Design Model

A feature often incorporated within the military weapon systems procurement cycle is an explicit model of a design that includes consideration of HCI within the design cycle. The models require iterative HCI design, with prototyping and formal assessments integrated throughout the design cycle. Requirements for acceptance test procedures are also identified early. The basic, general procurement model is: (1) concept studies, (2) feasibility analysis, (3) project definition, (4) development, (5) build, and (6) in-service support. Human factors can, in principle, be integrated into this cycle at each stage with differing specificity and requirements at each stage. For example, NATO Standardization Agreement 3994 AI (NATO Headquarters, 1993) is an example of a procedural guide for integrating human engineering into system design and evaluation.

At present, the civil transport aircraft model of procurement and production is different and includes various roles and responsibilities for system designers, airlines, and regulatory authorities. There is apparently no explicit requirement for HCI prototyping and review by the user population. Certification typically occurs late in the design and development process and certification approval will be sought during the final phase of the development program. Furthermore, certification is formally planned late in the development schedule. If a problem is identified at this stage, it is usually too late to either address the problem adequately or implement a redesign. However, it is noted that certifying authorities can participate in reviews during the aircraft design process. In this way, design features that would prevent certification of the aircraft could potentially be noted and corrected.

The international nature of aircraft manufacturing, purchasing, and regulation exacerbates the problem. Typically, a regulatory authority from a country other than that of the airframe manufacturer will not formally interact with the manufacturer until the point at which certification is required. Experience from work with military systems procurement strongly suggests that this approach is likely to be problematic and issues identified at this stage of certification are usually too costly to remedy. The development model for civil flight deck design needs to be updated to include explicit consideration of the HCI. Aircraft regulatory authorities should more actively participate in the design and development process.

HCI Guidelines and Standards

A trend within the human factors industry has been the development of standards for human computer interfaces. In recognition of the importance of a user's interface with a system, Apple Computers requires that all applications for their products adhere to their HCI guidelines and standards. Recently other HCI guidelines and standards have been published and are becoming widely used in commercial applications (e.g., Open Software Foundation [OSF] Motif™; Open Look™; Presentation Manager™).

A design practice that has emerged on projects involving software-based systems is the provision of both project-specific and general HCI guidelines and standards. The purpose of the guidelines and standards is to provide a human factors-based, well designed, common HCI across all user-system interfaces. The guidelines and standards can be levied on all system components even if produced by several suppliers. However, this explicit HCI design practice does not appear to be well integrated into the airframe manufacturers' flight deck design approach or the regulatory authorities' acceptance testing and certification procedures. Evidence indicates that numerous designers are unaware of many relevant HCI design guidelines and standards. Evidence also suggests that many designers view guidelines and standards as a hindrance constraining design rather than an aid providing structure and boundaries to good design practice.

Airframe manufacturers' problems are further compounded because flight decks consist of a mix of avionics from a range of suppliers. Each separate avionics subsystem supplier is unlikely to have either in-house guidance for HCI design or its own proprietary standard for the "look and feel" of the system. Major airframe manufacturers integrate many subsystem components into a "coherent" whole, yet little or no effort is made to create a common look and feel across equipment groups. Often a flight deck will consist of a multitude of HCIs, a unique one for each piece of avionics hardware (e.g., there are cases where two different keyboard layouts exist on the same flight deck). Many general HCI design guides and standards are available that could be used for flight deck HCI design and certification (e.g., Boff & Lincoln, 1988; Department of Defense, 1985; Department of Defense, 1989, sec. 5.15; National Aeronautics and Space Administration, Space Station Freedom Program Office, 1991; Smith & Mosier, 1984).

HCI Evaluation Methods: Checklists and Usability Testing

Many HCI guidelines and standards can be represented as high level HCI evaluation criteria embodied within checklists. HCI checklists help formalize design evaluation by domain experts (i.e., human factors HCI design practitioners). For example, Ravenden and Johnson (1990) identify nine checklist criteria: visual clarity, consistency, informative feedback, explicitness, appropriate functionality, flexibility, control, error prevention and correction, and user guidance and support. There are over 100 checklist questions and the answers help provide a standardized and systematic method enabling those evaluating an interface to identify problem areas. Shneiderman (1986) identifies 20 major items in a short generic user-evaluation questionnaire for interactive systems. The Shneiderman questionnaire identifies a range of questions with bipolar semantically-anchored items that require users to evaluate aspects of system performance. The preceding examples are representative of generic audits and can be used in a variety of ways.

Usability testing has become one of the most well established methods of product evaluation used by human factors engineers. A recent series of studies funded under the European Strategic Program for Research in Information Technology (ESPRIT) Project 5429 (1992), referred to as MUSIC (Metrics for Usability Standards in Computing), have examined and identified methods and tools to measure the usability of products that require human computer interaction. It is not the intention of this paper to review HCI procedures but to simply point out that considerable work has been underway for some time in this area, that information is available, and that it is sufficiently refined for direct application to civil aviation.

Organizational Implications

The point was made earlier that regulatory authorities will need to adjust to different requirements imposed by cognitive elements of a system for effective design assessment and, hence, certification. At an organizational level, though, the form and function of human factors also need to be addressed. Rouse (1991) argues for the concept of human-centered design and makes two important points. He suggests that not only are the concepts, principles, and methods in human factors not sufficiently advanced, but, more importantly, that most current principles have little impact on design. This, he argues, is a result of the massive influence of the organization in general and the structure within which human factors is allowed to function. There are obvious parallels with Billings' (1991) human-centered approach to automation. Whereas Billings refers to users primarily as pilots, Rouse takes a system-oriented view and refers to users as any stakeholders within the ultimate system (e.g., regulatory authorities, airlines, maintainers, sales staff, pilots, and others). Rouse argues that human factors will have greater impact when these wider implications are considered. To this end we believe that regulatory authorities require a substantive change in philosophy concerning human factors. Four organizational changes are required.

The first organizational change is to recognize and integrate general human factors developments, particularly HCI design, into aircraft design and certification processes. In fact, this is a key message of this report. There is ample evidence that pilots, engineers, and systems designers typically do not understand the benefits that can accrue with a formal human factors evaluation. Regulatory authorities also need to recognize the unique contribution of human factors and formally establish a mechanism for its incorporation into the certification process. It should also be recognized that the HCI scientific base can be of a qualitatively different form than other more entrenched certification areas. In addition, regulatory authorities can set requirements by specifying good quality human factors and HCI design, even if these are established at a high level. Designers will have to attempt to meet those requirements or provide substantial, appropriate data for consideration of relaxing them. If this approach is adopted, examples of poor HCI design in civil cockpits may, in time, begin to disappear. A better representation of aviation and, particularly, non-aviation, related but immensely applicable, standards and guidelines needs to be achieved.

A second organizational modification concerns the model used for flight deck HCI design and the stage at which certification occurs. The detailed form and content of human factors certification support should be defined and established. Regulatory authorities should start early in the process and review the technicalities and implications of an aircraft's design. National regulatory authorities outside the manufacturer's home country should also be given consideration for reviewing proposed aircraft designs during the design process.

The third organizational change involves end users. Line pilots who daily operate the systems and who must interact with their aircraft through interfaces provided by the manufacturer should be formally involved in some way in the HCI review process. Neither aircraft manufacturer test pilots nor airline managers, who may be pilots themselves, can adequately represent typical line pilots' operational requirements.

Finally, the fourth and perhaps most ambitious requirement may be incorporation of expert human factors practitioners as regulatory authority "sign-off" administrators, as well as the manufacturer's design team to ensure that human factors is given appropriate consideration.

Conclusions

Pragmatically, it is assumed that aircraft design methods will not change unless regulatory authorities establish requirements and criteria for design verification that induce an aircraft manufacturer to change its organization; in fact, the regulatory authority itself must also change. To foster these changes, integration of knowledge from other human factors domains, other than those specifically related to aviation, must occur. The knowledge is in the form of a design approach that explicitly incorporates human factors and HCI considerations and uses empirically-based design guidelines and standards. In addition, the use of formal assessment methods and tools for usability testing and certification assessment must be applied. This paper mentioned the potentially valuable contributions from HCI studies and usability assessments; the form and content of that intervention was also described.

Additionally, a number of organizational alterations are required to allow application of data and methods for aircraft design and certification. Norman (1989) would argue for a complete revision of the aircraft design process. Such a revision is probably untenable. The effective management of HF in the civil aircraft certification process is the more tenable option and probably implies "a little and often." We recommend that regulatory authorities establish a better appreciation of the nature and form of HCI assessment. In particular, an appreciation of the type of assessment demanded by cognitive systems must be developed. There is also a strong case for encouraging regulatory authorities to examine systems and subsystems early in the design process. Authorities must also establish effective HCI requirements. Finally, formal recognition of the unique human factors contribution in the form of sign-off authority should be established. The preceding approaches are needed in order to ensure appropriate integration of human factors in the certification process for civil flight deck design.

The position of this paper is that a number of changes are required to the aircraft design process and the aircraft certification process in order to enhance flight deck HCI design. These changes include the use of a software design model which explicitly considers the HCI, empirically-based guidelines and standards during HCI design, and instruments and methods like checklists and usability testing tools for HCI evaluation and certification. This approach can provide the foundation for effectively integrating HCI knowledge into the civil flight deck design and certification process.

(The views expressed in this paper are those of the authors).

References

Billings, C. (1991). *Human centered aircraft automation: A concept and guidelines* (Report No. NASA TM 103885). Washington, DC: National Aeronautics and Space Administration.

Boff, K. R., & Lincoln, J. E. (Eds.). (1988). *Engineering data compendium* (Vols. I-III). New York: J. Wiley and Sons.

Department of Defense. (1985). *Human engineering guidelines for management information systems* (DOD-HDBK-761). Washington, DC: Author.

Department of Defense. (1989). *Human engineering design criteria for military systems, equipment, and facilities* (MIL-STD 1472D). Washington, DC: Author.

Department of Transport. (1989). *Report on the accident to Boeing 737-400 G-OMBE near Kegworth, Leicestershire on 8 Jan 1989* (Air Accident Investigation Report No. 4/90). London: Her Majesty's Stationery Office.

European Strategic Program for Research in Information Technology (ESPRIT) Project 5429. (1992). *Metrics for usability standards in computing (MUSIC).* Teddington, Middlesex, U.K.: National Physical Laboratory.

Hughes, D. (1989, August 7). Human factors research aids and glass cockpit design effort. *Aviation Week & Space Technology*, pp. 34-36.

Landauer, T. (1991). Let's get real: A position paper on the role of cognitive psychology in the design of humanly useful and usable systems. In: J. Carroll (Ed.), *Designing Interaction* (pp. 60-73). Cambridge, England: Cambridge University Press.

Lenorovitz, J. M. (1992a, February 3). Confusion over flight mode may have role in A320 crash. *Aviation Week & Space Technology*, pp. 29-30.

Lenorovitz, J. M. (1992b, March 2). French government seeks A320 changes following Air Inter crash report. *Aviation Week & Space Technology*, pp. 30-31.

McClumpha, A., James, M., Green, R., & Belyavin, A. (1991). Pilots' attitudes to cockpit automation. *Proceedings of the Human Factors Society 35th Annual Meeting, 1*, 107-111.

National Aeronautics and Space Administration, Space Station Freedom Program Office. (1991). *Human computer interface guide* (SSP 30540). Reston, VA: Author.

NATO Headquarters. (1993). *Application of human engineering to advanced aircrew systems design* (NATO Standardization Agreement 3994 AI, Edition 1). Brussels: Author.

Norman, D. (1989, July). *The "problem" of automation: Inappropriate feedback and interaction, not "overautomation"* (ECS Report 8904). San Diego: University of California, Institute of Cognitive Science.

Norman, D. (1991). *Cognitive science in the cockpit* (Vol. II, No. 2). Dayton, OH: Wright-Patterson Air Force Base, Crew Systems Ergonomics Information Analysis Center (CSERIAC).

Ravenden, S., & Johnson, G. (1990). *Evaluating usability of human-computer interfaces: A practical method.* New York: J. Wiley and Sons.

Rouse, W. (1991). *Design for success: A human-centered approach to designing successful products and systems.* New York: Wiley.

Rudisill, M. (in press). Pilot comments concerning the interface to automation: results from the RAF IAM survey (*NASA Technical Report*).

Sheridan, T. B. (1991). Automation, authority, and angst--revisited. *Proceedings of the Human Factors Society 35th Annual Meeting, 1*, 2-6.

Shneiderman, B. (1986). *Designing the user interface: Strategies for effective human-computer interaction.* Reading, MA: Addison Wesley.

Smith, S., & Mosier, J. N. (1984). *Design guidelines for user-system interface software* (MTR No. 9420). Hanscom Air Force Base, MA: Mitre Corporation.

Weiner, E. (1989). *Human factors of advanced technology (glass cockpits) transport aircraft* (NASA Report CR 177528). Washington, DC: National Aeronautics and Space Administration.

Weiner, E., & Curry, R. (1980). *Flight deck automation promises and problems* (NASA TM 81206). Washington, DC: National Aeronautics and Space Administration.

N95- 34785

Some Inadequacies Of The Current Human Factors Certification Process Of Advanced Aircraft Technologies

Jean Paries

Bureau Enquêtes Accidents

Introduction

Automation related accidents or serious incidents are not limited to advanced technology aircraft. There is a full history of such accidents with conventional technology aircraft. However, this type of occurrence is far from sparing the newest "glass cockpit" generation, and it even seems to be a growing contributor to its accident rate. Nevertheless, all these aircraft have been properly certificated according to the relevant airworthiness regulations. Therefore, there is a growing concern that with the technological advancement of air transport aircraft cockpits, the current airworthiness regulations addressing cockpit design and human factors may have reached some level of inadequacy. This paper reviews some aspects of the current airworthiness regulations and certification process related to human factors of cockpit design and focuses on questioning their ability to guarantee the intended safety objectives.

Current Certification Principles

Certification Purposes and References

According to Article 31 of the Convention on International Civil Aviation (Chicago, 1944), any aircraft involved in international operation shall hold a certificate of airworthiness delivered or validated by its state of registry. This is intended to achieve protection of other aircraft, third persons and ground property while an aircraft registered in state B flies into or over the territory of state A. Following article 33 of the Convention, Annex 8 to the Convention includes broad standards which define the minimum international basis for the recognition by states of certificates of airworthiness delivered or validated by other states. Furthermore, Annex 8 sets the minimum international degree of standardization called upon by article 37 of the Convention.

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PRECEDING PAGE BLANK NOT FILMED

PAGE 300 INTENTIONALLY BLANK

However, as a matter of fact, the main purpose of airworthiness certification has been the protection of the passengers and crewmembers for a long time. This is a much more demanding objective, which has been achieved trough the implementation of national codes of airworthiness containing the full scope of requirements considered necessary by the states to reach the target safety level.

Today, only two airworthiness codes form the potential reference for any transport category aircraft certification: the U.S. FAR 25, and the European JAR 25. Due to the pressure of a highly internationalized business, these two codes are only differentiated by minor differences, and can therefore be referred to as a single reference for the present discussion.

Basic Principles

The airworthiness requirements concerning the cockpit, as for any other subsystem of the aircraft, are *not* aiming at any "best possible" design, but they intend to specify the minimum objectives to be matched by an applicant design. This is a very basic principle of any certification.

As far as human factors in cockpit design and equipment are concerned, the minimum objectives currently set by the airworthiness code are more or less limited to the following:

- To guarantee that the minimum crew (i.e., after one crew member incapacitation) is still able to do the job without excessive workload or fatigue (FAR/JAR 25-1523)

- To provide the crewmembers with acceptable comfort and protection against outside conditions, so that they can do their job without excessive effort, concentration or fatigue (FAR/JAR 25-771)

- To provide the crew with a sufficient visibility to the outside (FAR/JAR 25-7)

- To minimize the risks of mistake in the controls use, particularly through a standardization of the shape and movements of the primary flight controls (FAR/JAR 25-777; 781)

- To minimize ambiguities in the information displayed by the instruments (FAR/JAR 25-1303; 1321; 1322)

- To provide the crew with relevant alerting information about unsafe functioning states of any equipment or system, and to allow appropriate crew action (FAR/JAR 25-1309).

However, these generic requirements are completed by a set of "special conditions," adapted to the specifics of each particular aircraft. These special conditions may well include extensive and detailed requirements for systems like CRT display flight instruments.

Demonstration of Compliance

The methodology used to check the compliance of a proposed design with a relevant airworthiness requirement heavily depends on the explicit versus implicit nature of the requirement.

Explicit requirements are directly expressed in terms of design characteristics. For example, FAR/JAR 25-781 quotes: "cockpit control knobs must conform to the general shape...in the following figures." In this case, the compliance of a proposed design is rather easy to check, and direct examination of descriptive material (drawings, scale models, mock-ups) can be used.

However, most of the human factors related issues are covered by implicit requirements, expressed in terms of general outcomes to be achieved. For example, FAR/JAR 25- 777 (a) quotes: "each cockpit control must be located to provide convenient operation and to prevent confusion and inadvertent operation".

In the later case, the *methodology* used to evaluate the ability of a proposed design to reach the objective obviously *is the critical part* of the certification process. A first possible source of difficulty is the interpretation of the regulatory objective itself. A second possible source of difficulty is the acceptable means of compliance with the (interpreted) objective. Consequently some regulatory requirements are complemented with advisory material, including *interpretation guidelines* and/or indications on *acceptable means of compliance*. Acceptable means of compliance more often than not are proven solutions, or sets of solutions, that have been shown by service history to be satisfactory.

The Test Pilot Judgment Methodology

As a matter of fact, one of the only tools currently in use to evaluate a new design human factors acceptability in a certification process is *test pilot judgment*. This judgment is based on regulatory and company test pilots comparative experience, following from actual or simulated flight exposure (several thousands of hours for major test programs) in the subject cockpit on the one hand, and previous experience with existing designs on the other hand. In other words, this judgment is based on extrapolations to the new design of expertise gained on the previous ones.

Furthermore, the certification process *cannot wait* until the first aircraft prototype is built to start up. No manufacturer would take the risk of becoming involved in such highly expensive development programs without reasonable guarantee that the projected designs are certifiable. Consequently new designs are submitted to regulatory authorities to get some "certifiability" agreement well before a prototype aircraft is built. In this situation, pilot judgment cannot be exercised in a real cockpit (real flight context) but has to be exercised in a mock-up or some other form of simulated environment.

At the end of major certification programs, certification authorities are nowadays calling for operational route proving programs. These programs are the occasion for evaluating the aircraft in an airline-type environment, including "natural" and artificially induced failures, with mixed crews being composed of airline pilots and manufacturer test pilots. They have included up to one hundred flights in one occasion.

Discussion

The history of automation related accidents or serious incidents (see accident/incident reports list in annex) includes conventional generation aircraft such as the Lockeed L-1011 Tristar, the

McDonnell Douglas DC-10, and the Boeing B-747. This type of occurrence even seems to be a growing contributor to the newest "glass cockpit" generation accident rate. This may be an incentive to question the current human factors certification process of advanced technology aircraft.

The protection ability of a certification process depends on several factors, including, but not limited to:

- The adequacy of the requirements in expressing relevant safety objectives
- The adequacy of the compliance checking methodology in use.

As far as the human factors certification process of advanced technology aircraft is concerned, it seems that critics may be surprised by both of these aspects.

Some Potential Biases in the Current Human Factors Certification Objectives

Airworthiness codes and their objectives are not floating by themselves in the vacuum. They are embedded in a global aviation safety system, including components such as personnel certification (selection, training, proficiency checking), or operational procedures certification. The design and functional characteristics of this safety system reflect *specific theories* about risk and safety in the aviation transportation system. These theories are far from being mere rational constructions, consistent with all the available scientific evidences. They also are ***historical and social outcomes***, conveying the current fears and faiths of the aviation community.

The cutting up of the different safety codes and their role distribution is a first indicator of the background safety approach. This is particularly perceptible with the human error question. On the one hand, it is widely claimed that pilot error is contributing to about 70 percent of air transport accidents. On the other hand, merely four paragraphs (§ 777, 781, 1303, 1309) of the airworthiness code, out of 330, explicitly or implicitly address pilot error. This one percent order score suggests an implicit assumption that pilot error is neither really associated with airworthiness nor a design concern, but is much more related to pilot certification and procedures and operational regulations. As a matter of fact, this is the prevailing theory in the airworthiness world, and all the publications of the human factors researchers have had little influence on it up to now, except perhaps for this ironical one: a shift as occurred from culpability-based theories (pilot should not make errors) to fatalism (errors are normal and will always occur, whatever the design – so let's reinforce cross-check procedures...and substitute automation for pilot action as far as possible).

A second illustration is offered by the workload assessment focus in the certification of the last ten years cockpit generation. Even with the same reference airworthiness code, large variations generally occur in the selection of items subject to specific attention in the certification process of different aircraft. Great departures from previous designs will normally be given a closer look. This is true for technical designs and for human factors aspects as well. The glass cockpit generation has progressively brought a drastic change to the previous pilot environment, including two-crew design, computer generated displays, sophisticated automated flight controls, and flight management computers. These changes presented a lot of challenges to pilots, such as autopilot active mode awareness, total energy awareness, crew communication, automation over-reliance, and computer interface problems. But in practice, one single question stands above everything else: the workload question. An ad hoc

Presidential Commission was set up in the United States in 1980 to endorse the concept. Since then, much effort, research and flight test time have been spent for evaluation, rating and judgment of workload levels during certification programs for the purpose of minimum crew complement assessment.

At this point, it is rather difficult to elude some paradoxical feeling about the situation. On the one hand, great efforts are made about work load certification, but there is no real history of overload related accident on glass cockpit aircraft. This could be interpreted as the nice benefit of a particularly effective certification process on cockpit designs. Unfortunately, workload certification programs take place so late in the certification process that it is hard to imagine any significant change in the cockpit at this time, except for associated procedures. (It is not the intention here to minimize the importance of minimum crew evaluation, but only to suggest that its perceived importance is also subjective, crystallizing the socially highly sensitive skip to the two-crew cockpit.)

On the other hand, there is a history of automation related accidents, but far less efforts are devoted to the certification of the related design aspects since they are not really felt to be airworthiness matters.

The Adequacy of the Compliance Checking Methodology

As already stated previously, nearly the only tool currently in use in a certification process to evaluate the acceptability of a new design in terms of human factors is test pilot judgment. This judgment is based on extrapolations to the proposed new design of expertise gained on the previous ones. Furthermore, as new designs are submitted to regulatory authorities to get some "certifiability" agreement well before a prototype aircraft is built, it often has to be exercised in simulated environment.

This test pilot assessment methodology is by nature affected by some biases.

First, it is based on the assumption that the experience of test pilots on previous aircraft is transferable on the new one. This may not be true for great developments in human-machine interface design. Secondly test pilots are not a representative sample of the airline pilots population. They have a very specific knowledge of the aircraft, which leads to different mental models of the aircraft. They do not share the daily routine operations and the associated constraints, and therefore the cognitive processes involved at the crew/aircraft interface are very doubtfully the same for test pilots and for airline pilots. Furthermore, typical figures for the number of individuals involved is about ten or a few tens and the total exposure time about two thousand hours. This is to be compared to the frequency of the hunted critical events or combinations of events, which is more likely to be of the order of one per hundred thousand hours. And to make things worse, recent studies (Amalberti & Wilbaux,1994) indicate that cognitive behaviors evolve a lot during the training process on glass cockpit aircraft, and that the average experience needed for the training process to reach a maturity stage and stabilize the cognitive behaviors is about 800 hours, or one and a half years. This is by far a figure that no test pilot will reach during a typical test period with a new type of aircraft. As a consequence, test pilot judgment will be exercised within a cognitive frame and the typical errors encountered in testing, which is significantly different from the average airline pilot situation.

Finally, an individual judgment, even from a test pilot, would only provide rather soft grounds to refuse a proposed design.

Conclusion

The current aircraft airworthiness certification requirements addressing human factors issues are expressed in rather general terms, and they are subject to interpretation uncertainties. This leads to an increasing inconsistency between the certification requirements and the new types of human-machine interface problems brought in by the glass cockpit highly automated aircraft generation. Consequently, it seems that there is a need for a redefinition of the objectives of the human factors certification process.

The current evaluation process for the certification of human factors related aspects of cockpit design rely almost entirely on test pilot judgment. This method has proven satisfactory for the past years, although it is marred by several biases. However, the changes induced by the new design of the pilot-aircraft interface has recently increased the effects of these biases to a significant degree. Consequently, it seems that there is a need for the development of human factors assessment protocols to complement test pilot assessment methodology.

References

Amalberti, R. & Wilbaux, F. (1994). Advanced automated glass cockpit certification:Wariness about human factors. In: J. A. Wise, V. David Hopkin, and D. J. Garland (Eds.), *Human Factors Certification of Advanced Aviation Systems.* Daytona Beach: Embry-Riddle Aeronautical University Press.

Aircraft Accident/Incident Reports

(1) Eastern Airlines flight 401, Lockeed L-1011, Miami, Florida, December 29, 1972. NTSB accident report no AAR/ 73-14.

(2) Aeromexico flight 945, Mc Donnel Douglas DC10-30, over Luxembourg, Europe, November 11, 1979. NTSB incident report no AAR/ 80-10.

(3) SAS flight 901, Mc Donnel Douglas DC10-30, New-York JFK, February 28, 1984. NTSB accident report no AAR/ 84-15.

(4) China Airlines flight 006, Boeing 747-SP, 300NM northwest of San Francisco, California, February 19, 1985. NTSB accident report no AAR/ 86-03.

(5) Indian Airlines, Airbus A320, Bangalore, India, February 14, 1990.

(6) British Midlands Airways, Boeing 737-400, East Midland, UK, January 8, 1989. AAIB accident report.

(7) Lauda Air , Boeing 767-300, near Bangkok, Thailand, May 26, 1991.

(8) Air Inter flight 148, Airbus A320, Strasbourg, France, January 20, 1992. Rapport préliminaire de la Commission d'Enquête, February 1992.

(9) Thai International Airways, Airbus A310-300, Katmandu, Nepal, July 31, 1992. Accident report by the nepalese investigation commission.

Issues in Future Design and Certification of Complex Systems

307

N95- 34786

Advanced Automated Glass Cockpit Certification: Being Wary of Human Factors

52-54

Rene Amalberti & Florence Wilbaux

Direction Générale de l'Aviation Civile

Summary

This paper presents some facets of the French experience with human factors in the process of certification of advanced automated cockpits. Three types of difficulties are described: first, the difficulties concerning the hotly debated concept of human error and its non-linear relationship to risk of accident; a typology of errors to be taken into account in the certification process is put forward to respond to this issue. Next, the difficulties connected to the basically gradual and evolving nature of pilot expertise on a given type of aircraft, which contrasts with the immediate and definitive style of certifying systems. The last difficulties to be considered are those related to the goals of certification itself on these new aircraft and the status of findings from human factor analyses (in particular, what should be done with disappointing results, how much can the changes induced by human factors investigation economically affect aircraft design, how many errors do we need to accumulate before we revise the system, what should be remedied when human factor problems are discovered at the certification stage: the machine? pilot training? the rules? or everything?).

The growth of advanced-automated glass cockpits has forced the international aeronautical community to pay more attention to human factors during the design phase, the certification phase and pilot training. The recent creation of a human factor desk at the DGAC-SFACT (Official French services) is a direct consequence of this.

The paper is divided into three parts. Part one debates human error and its relationship with system design and accident risk. Part two describes difficulties connected to the basically gradual and evolving nature of pilot expertise on a given type of aircraft, which contrasts with the immediate and definitive style of certifying systems. Part three focuses on concrete outcomes of human factors for certification purposes.

What Model for Human Error and What Links Between System Design and Accident-Risk?

The Goal of Aircraft Certification

The goal of aircraft certification is simple: guarantee that an aircraft fits the legal flight safety requirements when flown by qualified standard pilots who are as representative as possible of

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PRECEDING PAGE BLANK NOT FILMED

PAGE 368 INTENTIONALLY BLANK

end-users. The crews taking part in the certification campaign are used to bring systems into play and are at times as co-appraisers, but are never examined in their own right in the certification campaign. Thus, human errors observed during the certification campaign used to be classified in two categories. First, there are errors related to system-design: e.g., input errors, such as stick and throttle errors and inappropriate settings. These errors were taken into consideration in the certification process. Second, there are errors resulting from pilot attitudes, air-traffic control dialogues and clearances, or individual weaknesses related to general aviation know-how. These second type errors were not considered to be relevant for aircraft certification purposes.

Why Change the Current Procedures of Certification?

The present level of flight safety is very good. The risk of accident is about one per one million departures in industrial nations. However, this value has been virtually stable for fifteen years. It was attained before the growth of automated aircraft and before human factors became a target objective of FAA and the international aeronautical community.

In this context, why should the certification process bother with human factors?

The answer is twofold. First, advanced-automated aircraft have fewer dramatic failures, but the rate of accidents is not decreasing. Accident causes are increasingly due to the cognitive failure of crews. This introduces new problems of interference between system reliability and human reliability. In other words, the technical improvement of a system safety based on automation could result in a negative outcome for human reliability. Moreover, since these technical changes are barely related to the previous experience acquired on standard aircraft, certifiers, whether they are expert pilots or engineers, must ask human-factors specialists for scientifically-based assistance to evaluate systems better. Second, with a constant rate of accidents, any increase in traffic volume will result in an increase in the total number of accidents. Moreover, the negative impact of each (rare) accident on customers (passengers and also companies) is multiplied by modern information networks. Naturally this is undesirable, and the improvement of this situation is the stated objective of the aeronautical community. The only solution to remedy this situation is to continue to increase flight safety up to the present level. However, it is clear (given the current level) that gains will not be easy in previously examined domains and also that new domains will have to be taken into consideration. Human factors is one the most important of these new domains aside from the future of ATC communications.

Let us examine how these objectives impact the relationship between human error, system design and accident risk, and hence impact certification procedures. But of course the first problem to tackle is this of defining human error without ambiguity.

Wariness of the Definition of Human Error

The ergonomics literature has proposed numerous definitions and classifications of human error from process control. The dominant definition considers human error as a deviation from the norms, whether these norms are written or assumed from practice. This definition is both the easiest to use and the most debatable. Most recent cognitive ergonomics field-studies (Amalberti, 1992; deKeyser, 1986; Rasmussen, 1986) show that novices continuously interpret norms to make the job feasible with their limited resources and knowledge and that experts interpret norms still further with routines, shortcuts and violations. To sum up, norms are never respected, although they serve operators as references both to give and to limit degrees of freedom adapting to the on-going situation. Human operators are experts in piloting this derivative of norms to fit the situation requirements with minimum workload.

In contrast, another way to define error would be to consider error as a deviation from operator intention. This approach is probably less biased than considering error as a deviation to norms. However, operator intentions are hard to evidence, especially after a lengthy delay between action and analysis, and in the end human error still continues to be considered as a deviation from norms during the process of certification.

Experience acquired in participating as a neutral observer in the minimum crew campaign of advanced automated glass cockpits shows that ambiguities in human error definition lead the certification team, poorly trained in human factors, to make numerous mistakes in classifying and interpreting the crew errors. These mistakes are threefold. First, errors immediately corrected by the crew tend to be ignored. However, many of these errors point to incorrect system design, especially when they are repeated almost systematically. Second, deviations from norms are too often considered as errors although they are not. In many cases, they represent pilots' attempts to conserve resources or to manage the system and the task more conservatively. Third, some deviations where there are no specific procedural norms are ignored, although they potentially endanger the flight. This is typically the case of poor synergy and poor crew coordination which can result from the system design as well as procedure or input errors.

Modelling Relationships Between Human-Error and System Design

Ergonomics has always argued that systems should be designed primarily for end-users (human-centred design). This is a very basic and central value of ergonomics. Nevertheless, the concept of the human-centred system, and in a certain sense of a "good system," has significantly shifted over the last ten years with changes in technologies and ergonomics theories. Let us examine these changes.

A good ergonomics design has long been considered to be a design that prevents errors and facilitates good performance with as low a workload as possible. In the 1940's, the main interest was in unambiguous commands designed to minimized errors; e.g., confusion between gear and speed brakes (Fitts, 1947). This type of ergonomics, which is dominant in the USA, is termed *Classical Human Factors*. The basic philosophy draws on the central idea that human error is avoidable if the design respects human limitations and capacities, and this leads to the concept of "fault-preventing system design." It has been extensively and successfully applied to cockpit design and is currently used.

However, several factors have contributed to a recent decline of this approach: new technologies, new needs, and new ergonomics theories.

New Technologies and Classic Human Factors

Cockpit automation has enhanced flight performance in many domains such as precision approach, flight accuracy, engine performance, and pilots' situation awareness (with the introduction of map display). Automation has also mechanically reduced a great number of human errors resulting from improper power or stick settings and system handling simply because these tasks do not longer pilot-dependent.

However, the drawbacks of automation for human behaviour are as numerous as the advantages described above.

Cockpit automation and its consequences for cockpit layout have considerably reduced the benefits expected from a "simple human factors-based system-design."

The Flight Management System (FMS), with its undifferentiated and multiplexed keyboard, is a blatant example of this (Pelegrin, 1993; Sarter, 1992). This design is the source of many input errors in programming systems.

Criticisms have also been directed to information displays. The ability to display, aside from classical dials, much more information in various new forms (such as drawings and texts) has led designers to a series of poor ergonomics solutions regarding the capacities and limitations of human perception. Perception time tends to be increased with the use of textual information, perceptual feedback is reduced in peripheral vision (due to loss of motion of sticks and throttles and also due to cockpit architecture which requires the other crew member to move less), and auditory and kinesthetic sensations are also reduced (due to computer program smoothing system reaction to improve passenger comfort).

But of course the expected benefits of automation for ergonomics are elsewhere. Situation awareness has been improved with the use of map displays (MD), Primary Flight Display (PD) and ECAMs or EICAS.

To sum up, the new cockpit layout (glass cockpit) is assumed to enhance the pilot's situational awareness, but the solutions chosen to reach this goal are clearly to the detriment of classic sensory-motor human factors.

Evolution of Theories: Cognitive Ergonomics, Another Way to Consider Ergonomics

The change in goals for cockpit design have prompted new developements in ergonomics theories in the eighties. This is the domain of cognitive ergonomics, which draws heavily on the European ergonomics tradition. The value of such ergonomy is pilots' cognitive modelling focusing on their strategies and know-how. Numerous field studies have shown the advantages of this type of operator's cognitive model (Amalberti, 1992; Bainbridge, 1989; de Keyser, 1986; Hollnagel, 1993; Reason, 1990). The following section summarizes the main characteristics of one pilot's cognitive model.

Professional pilots generally have satisfactory procedural knowledge of their work domain and remarkable reasoning capacities, but they are resource-limited and cannot use all the knowledge and the reasoning capacities they would like to in time-related situations. The true task of pilots is to develop strategies to get the job done with respect to this resource-limitation bottle-neck.

Solutions Call for Planning, Anticipation and Risk Taking

Because their resources are limited, pilots need to strike a balance between several conflicting risks: an objective risk resulting from flight context (risk of accident) and a cognitive risk resulting from personal resource management (risk of overload and deterioration of mental performance).

To keep resource management feasible, the solution consists of decreasing outside risks, simplifying situations, only dealing with a few hypotheses, and schematizing reality. To keep outside risk within acceptable limits, the solution is to take as many preflight actions as possible in order to simplify the flight.

Any breakdown in this fragile and active equilibrium can result in unexpected situations, in which pilot performance may be decreased. Evidence shows that human errors result from internal characteristics of cognitive models and are not suppressible (Reason, 1990; Senders, 1991). Because resources-limitations force the pilot to make a series of compromises between what the situation should ideally require and what he is capable of doing, errors are the logical consequence. Moreover, expertise results from experiencing errors, (Anderson, 1985) and errors are generally profitable when the pilot receives immediate feedback from his errors.

New technologies confirm this general picture. As Wiener and Bainbridge point out (Bainbridge, 1989; Wiener, 1980), automation does not reduce the number of global errors, but merely changes error types. There are more routine errors and representation errors.

The consequences for ergonomics and certification purposes are twofold. First, the concept of "fault-tolerant system design" replaces the one of "fault-preventing system design." Fault-tolerant system design does not aim at limiting local errors but merely at improving pilot's awareness, giving as clear feedback as possible of error, and possibly correcting the immediate consequence of error when the flight is endangered; i.e., logical testing on FMS inputs or safety envelope of flight-laws (alpha-floor). Second, it is clear that in this theoretical framework it is no longer satisfactory to measure human performance as a simple error rate. More complex approaches are required to efficiently serve the certification process.

Should Error Analysis be Restricted to Human-Machine Interaction?

Standard certification procedures do not deal with crew errors which are not directly related to system design. However, new technologies could lead to a change in this position. The interdependency of any component of the aeronautical system, aircraft, crews, ATC, or maintenance, makes the analysis of causality between design and consequence of design much more complex than on a simple system. Any change in system philosophy influences the way operators carry out the task, even for actions not directly related to system interface. This is typically the case for crew coordination in glass cockpits.

Glass cockpit layout is generally assumed to make crew coordination more difficult. The reasons are threefold. First, as described above, communications require more and more central vision and active vocal dialogue to read the written information and to remedy the relative deprivation of sensory inputs. Second, new cockpit architectures, with independent access to information and commands, facilitate desynchronization. Pilots can display modes, change modes, or change parameters on their channel which the other crew member is totally unaware of. Third, problems of language emerge because more and more information is written. The standard language of aviation is English, although most pilots in the world have a different native language and do not speak perfect English (Pelegrin, 1993).

What emerges from these various difficulties is that many situations of poor coordination in a glass cockpit can be related to system design although they are not directly related to a specific action on the interface. This level of causality challenges the philosophy of the system and calls for complex corrections. It is easy to understand that designers are very reluctant to consider that these errors are related to system design and prefer to pass on the problem to trainers.

Relationship Between Human Error and Accident-Risk

Safety is a central concern of aircraft certification. System-failure classically serves to measure system reliability, and human error serves as an equivalent measure of human reliability. Measures could be quantitative or qualitative (type of error), but it is explicitly admitted that a good design and a safe system would provoke less errors than an unsatisfactory design, and would therefore result in fewer accidents.

This is only partially true. We have seen that human error is not totally avoidable. Moreover, it is important to remember that the accident-rate is 1 accident per million departures, and that there are over 5 human errors per flight which are not detected and corrected immediately by crews (Amalberti, unpublished report; this value comes from numerous flights made in 1992 on glass cockpits in the observer-position). Thus, even though human responsibility appears to have risen in glass cockpit accidents, the relationship between human error and accident is far from being trivial.

The key-point is that the relationship between system design, human-error and the risk of accident is not linear from great risks to no risk. Obvious bad system-design or/and unadapted regulations or training will cause numerous human-errors and will increase the risk of accident. However, even if the design, training and regulation are perfected, numerous human-errors and a non-decreasing risk of accident will remain. Without strong relations between the two arguments, remaining accidents are poorly linked to human-errors. They are better linked to a matter of circumstances, a dramatic combination of unexpected events in which human-error can occur but not seen as decisive factor. This picture specifically applies to rare accidents arising in the context of high reliability.

For certification concerns, this non-linearity between human-error and risk of accident can become a problem. The central concern is to have enough references on human-error theories to clearly separate what should be a "normal-rate and type of human error" from an "abnormal rate and type of human error" due to poor system-design, training or regulation. This is typically a domain in which human factors could improve the current process of certification.

The Evolving Nature of Pilot Expertise and the Immediate and Definitive Style of Certifying Systems

Although computer technology clearly enables software modifications during and after the end of certification, aircraft philosophy and most sub-system designs are considered to be stable and definitive at the beginning of the certification process. This is not the case for pilots' expertise. Most official pilots in charge of flying the aircraft during the certification process have less than 200 hours experience in the aircraft. This is far from having stabilized expertise in glass cockpit. Results from field experiments (Sarter & Woods, 1992; Pelegrin & Amalberti, 1993) all indicate that pilot expertise for flying an aircraft with a glass cockpit shifts significantly up to 800 flight hours and perhaps more.

These values are almost double the values observed for those required for expertise in a standard cockpit. The problem is that behaviours change with experience. Errors also change in nature.

The lengthy period required to stabilize expertise in glass cockpits creates difficulties for pilots becoming used to the system. Pilots cannot easily grasp the enormous possibilities of the system. With pilots of up to 400 flight hours, the main risks are overconfidence or excessive doubt concerning their own capacities which respectively result in engaging the system in unknown domains or hesitating to make the right choice of action.

Once pilots gain confidence with the system, routine errors and violations are multiplied. The risk of accident still exists but changes in nature. It is clear that some relationships between human error and system design will only emerge in experienced pilots.

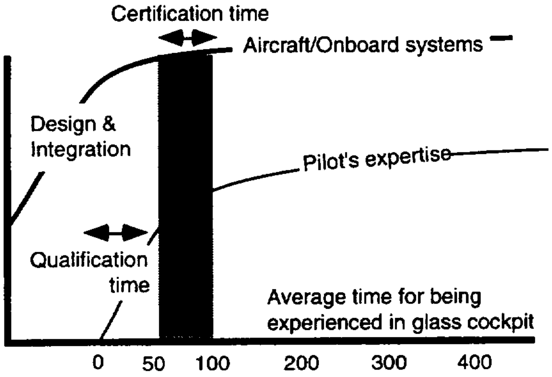


Figure 1

This picture raises a key-question on certification: do we test all aspects of system design and system safety with novices? If not, do we have to consider various levels of pilot expertise? A positive answer would result in envisaging a "double track certification," one initial test comparable to what is done currently and one operational complement administered after a few months of experience on the system.

Another solution would be to use official pilots who are already experts on similar machines (machine of the same family) to gain time and experience. In this case, novice official pilots in glass cockpits will also be required to represent this class of pilots and their specific problems.

Note that, in any case, the choice of official pilots who represent the future range of company pilots and the composition of crews is a key for efficient certification; e.g., take into account the pilots' level of experience, form a representative panel of pilots, avoid crews made of two captains, etc. This area could improve greatly in the future.

Practical Outcomes: How to Improve Certification With Human Factors Considerations

When to Bother With Human Factors in the Certification Process?

A good answer would be: "anytime the human factors perceptive gives a plus to the standard approach which is already being used." The objective of integrating human factors should not be to make a revolution in certification, but merely to support and improve the current way of doing certification.

The expected benefit is threefold. First, it must be in terms of the identification of undetected system weaknesses. Second, it is in terms of giving a rationale for pilots' problems and relationships to system design in the risk of accident. Last but not least, an aircraft can no longer be viewed as an isolated system. Certification typically concerns the integration of aircraft into operational conditions. Therefore, outcomes of certification must concern system-design as well as pilot training and regulations in order that sub-components of the macro-social system might be included in the evaluation. For this specific concern, psycho-sociology can effectively support certifiers to make the relevant decisions.

What to Certify?

The complexity and the novelty of systems lead one to consider that the integration of human factors in certification should overpass the simple evaluation of system-performance and the reliability of an end-product. It should extend to design-procedures, based on general principles which have proven to be efficient, and should begin with prototype evaluation at a period where changes are effectively possible.

What to Measure?

The starting point should be to analyse pilots' activities and detect human errors. However, the analysis of these human errors must vary according to the goal: assessing the risk of accident or testing pilot's ease with the system design. This is the reason why no unique classification of human error can figure out all questions raised by certification. Multiple classifications are required. Further analysis would concern the assessment and possible measurement of mental reasoning, mental workload, communications, crew coordination, to sum up all cognitive activities which serve pilots and set up a relevant representation of the on-going situation.

Who Makes the Evaluation?

We saw in previous sections of this paper that the selection of the panel of official pilots and engineers participating into the certification campaign is a key-factor in obtaining relevant results. One could suggest that this panel should be as large as possible to grasp a great variety of opinions, and also to avoid making certifiers co-designers due to a (too) long relation the certifier develops with designers. Whatever the panel, it seems useful to require a minimum human-factors background for people in charge of certifying.

What Limitations for Human Factors?

Many human factor aspects of cockpit automation are beyond the traditional certification process. We have seen that some of them could be easily better taken into consideration. Yet, numerous others which relate to psycho-sociology, work-organization, careers, trades, or companies will also be beyond human factors investigation during the certification campaign.

However, they should be crucial to system acceptability and risk acceptability, but this is another story.

Assuming that there is the integration of a human factors specialist in the certification process, another clear limitation should be the legal responsibility of this specialist in case of accident. Human factors cannot answer all the cockpit-problems, either because of the lack of knowledge or just because of the lack of time to apply relevant methodology. Therefore it shall probably be required to specify in writing what domains are relevant for human factors actions and what mix of responsibilities will draw on human factors specialists and on other certifiers.

What Should Change?

In most cases, the modifications in system-design required by the certification are small. The reason is obviously the financial cost. Therefore, when problems are observed, they tend to be solved by putting effort into pilot training and regulations.

Software technologies have changed this picture a little by introducing a greater level of flexibility, but it is clear that the underlying system philosophy remains unchanged.

Even though this is an acceptable outcome and even though experienced pilots rate the system as very good (this is the case of modern glass cockpit), human factors specialists worry about this increasing reliance on training solutions. What will occur for flight safety if we continue to produce opaque systems which require over 1,000 flight hours before pilots are experienced? Is this realistic?

Similarly, the presence of various generations of aircraft poses unsolved problems at this time: what will occur with pilots flying successively old and new aircraft? What will occur with multiqualified pilots flying almost identical aircraft with just a few differences, in particular as regards Cross Crew Qualifications (CCQ)?

In both cases, the questions overlap systems certification, introducing new types of questions to investigate and new constraints in forming the panel of pilots called for testing the system.

Systematic flight analysis of current modern aircraft is a fantastic tool to anticipate most troubles pilots will have with next technology. This is a central direction for improvements for all the aeronautical community, and it can be useful (for certification purposes) to ask the designer to take into account lessons from the previous design.

But again, any envisaged modification in a new machine will have to be investigated with a human factors perspective not really for itself, but for the possible negative consequences it will introduce when flying similar systems with bi-qualification.

Finally, one should remember that certifiers do not have to overpass the mandatory mission they are paid for (assessing that the system fits safety and minimum requirements). Once these minimum requirements are established and respected, designers will be free to create and certifiers will not have to officially judge a design in terms of being good or bad. In the context of a free market there is competition between manufacturers and the success or failure of a product remains a decision of customers.

Closing Notes

The international aeronautical community aims at introducing human factors more efficiently in the certification loop because of the desire to reduce the risk of accident and because of new technologies which have negatively impacted on human performance in a few domains. Analysis shows that this improvement cannot be made without a reconsideration of the concept of human error before (flight analysis), during, and after the certification phase (feedback and accident analysis).

One last and chronic source of misunderstanding in bridging knowledge between human factors specialists and engineers is that engineers superimpose human error and system failure upon one another. This makes no sense. Humans are intelligent and flexible. They can be perfectly adapted, whatever the complexity of situation and can ensure a very high level of safety. They learn from errors, cover billions of domains and can adapt to unknown domains. However, errors are always possible and always occur. These errors are poorly predictable and tend to occur at times, in areas and with people that nobody would have predicted. On the other hand, machines are rigid, unintelligent, and repetitive, and failures are predictable and curable. Because of their stability, machine reliability appears to be easily modeled and also more reliable than human reliability. The result is that engineers give a systematic priority to machines to the detriment of pilots because they feel this is the only way to improve and control safety.

All human factors findings show that they are wrong. Human and machine reliabilities are simply different and must work in synergy to reach a better level of reliability. Unfortunately, the solutions chosen at this time to increase system reliability interfere with human reliability and lower this human reliability.

Thus, it would not be realistic to discuss a very detailed point in the interface, although fundamental points are being ignored. A French maxim perfectly summarizes this point: "It is not good that the tree hides the forest."

The ideas expressed in this paper are aiming at launching debates. They are not firm directions decided upon SFACT, but only preliminary thoughts.

Future decisions of French official services will take into consideration, in addition to some ideas expressed in this paper and other technical ideas, all legal, international and sociotechnical aspects of the problems which have not been mentioned in the paper.

References

Amalberti, R. (in press). Safety in flight operations. In B. Wilpert, & Qvale (Eds.), *New technology, safety and systems reliability.* Hillsdale, NJ: L. Erlbaum.

Amalberti, R. (1992). Safety in risky process-control: an operator centred point of view. *Reliability Engineering & System Safety, 38*, 99-108.

Amalberti, R., & Deblon, F. (1992). Cognitive modelling of fighter aircraft's control process: a step towards intelligent onboard assistance system. *International Journal of Man-Machine Studies, 36*, 639-671

Anderson, J. (1985). Development of expertise. In Freeman (Ed.), *Cognitive psychology and its implications* (pp. 235-259). New York.

Bainbridge, L. (1989). Development of skill, reduction of workload. In Bainbridge & Quintinilla (Eds.), *Developing skills with new technology.* London: Taylor & Francis.

Bainbridge, L. (1987). Ironies of automation. In Rasmussen, Duncan, & Leplat (Eds.), *New Technology and Human Errors* (pp. 271-278). New York: Wiley.

de Keyser, V. (1986). Technical assistance to the operator in case of accident: some lines of thought. In Hollnagell, Mancini, & Woods (Eds.), *NATO Series* (pp. 229-254).

Fitts, P., & Jones, R. (1947). *Analysis of factors contributing to 460 "pilot error" experiences in operating aircraft controls* (Report TSE AA-694-12). Wright-Patterson Air Force Base, MA.

Hollnagel, E. (1993). *Reliability of cognition: Foundations of human reliability analysis.* Amsterdam: Elsevier.

Pelegrin, C., & Amalberti, R. (1993). *Pilot's strategies of crew coordination in advanced glass-cockpits; a matter of expertise and culture.* Second Flight Safety International Congress, Washington, DC.

Rasmussen, J. (1986). *Information processing and human-machine interaction.* Amsterdam: North Holland.

Reason, J. (1990). *Human error.* Cambridge University Press.

Sarter, N., & Woods, D. (1992). Pilot interaction with cockpit automation: operational experiences with the flight management system. *International Journal of Aviation Psychology*, *2*(4), 303-321.

Senders, J., & Moray, N. (1991). *Human error.* New York: L. Erlbaum.

Wiener, E., & Curry P. (1980). Flight desk automation: promises and problems. *Ergonomics*, *23*, 988-1011.

Wiener, E. (1985). Beyond the sterile cockpit. *Human factors*, *27*(1), 75-80.

N95- 34787

Beware of Agents when Flying Aircraft: Basic Principles Behind a Generic Methodology for the Evaluation and Certification of Advanced Aviation Systems

Denis Javaux, Michel Masson, & Vèronique De Keyser

University of Liège

3-

Introduction

There is currently a growing interest in the aeronautical community to assess the effects of the increasing levels of automation on pilots' performance and overall safety.

The first effect of automation is the change in the nature of the pilot's role on the flight deck. Pilots have become supervisors who monitor aircraft systems in usual situations and intervene only when unanticipated events occur. Instead of "hand flying" the airplane, pilots contribute to the control of aircraft by acting as mediators, instructions given to the automation.

By eliminating the need for manually controlling normal situations, such a role division has reduced the opportunities for the pilot to acquire experience and skills necessary to safely cope with abnormal events (Bainbridge, 1987).

Difficulties in assessing the state and behaviour of automation arise *mainly* from four factors:

- the complexity of current systems (e.g., Billings, 1991) and consequent mode-related problems (Sarter & Woods, 1993)
- the intrinsic autonomy of automation which is able to fire mode transitions without explicit commands from the pilots (e.g., Sarter & Woods, 1992)
- the bad quality of feed-back from the control systems displays and interfaces to the pilots (e.g., Norman, 1990 ; Sarter & Woods, 1992), and
- the fact that the automation currently has no explicit representation of the current pilots' intentions and strategy (Onken, 1992 a; 1992 b).

The conjunction of those factors induces a large set of crew-automation interaction problems that pose questions to the current research: difficulties in anticipating computer generated mode changes, difficulties assessing the implications of changes to previously given instructions,

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PRECEDING PAGE BLANK NOT FILMED

PAGE 326 INTENTIONALLY BLANK

difficulties in reacting to unanticipated events and to command changes, difficulties in finding, integrating and interpreting relevant data for situation assessment and difficulties in building extended and refined mental models of how automation is working and how instructions have to be input (Sarter & Woods, 1992).

For pilots, the consequences of those difficulties are an increase in cognitive workload and the development of "unofficial" strategies to override or "hijack" the automation, in an attempt to satisfy "official" goals (Amalberti, 1992).

As a result, *certification is facing a range of new and complex problems* that challenge the aeronautical community to predict and account for all kinds of pilot-automation interaction patterns arising from the introduction of new and sophisticated technologies in cockpits.

> The rapid pace of automation is outstripping one's ability to comprehend all the implications for crew performance. It is unrealistic to call for a halt to cockpit automation until the manifestations are completely understood. We do, however, call for those designing, analysing, and installing automatic systems in the cockpit to do so carefully; to recognize the behavioural effects of automation; to avail themselves of present and future guidelines, and to be watchful for symptoms that might appear in training and operational settings. (Wiener & Curry, 1980) (Mentioned by Billings, 1991, p. 67)

In particular, this paper tries to characterize the added complexity and problems created by the introduction of autonomous agents as (intended as automated resources) in new generations of aircraft.

As an example of the potential for catastrophic consequences of these problems, we would like to refer to the China Airlines B747-SP accident, 300 miles Northwest of San Francisco, of February 19, 1985, using the accident report proposed by Billings:

> The airplane, flying at 41,000 ft. enroute to Los Angeles from Taipei, suffered an inflight upset after an uneventful flight. The airplane was on autopilot when the n. 4 engine lost power. During attempts to relight the engine, the airplane rolled to the right, nosed over and begun an uncontrollable descent. The Captain was unable to restore the airplane to stable flight until it had descended to 9500 ft.
>
> The autopilot was operating in the performance management system (PMS) mode for pitch guidance and altitude hold. Roll commands were provided by the INS, which uses only the ailerons and spoilers for lateral control; rudder and rudder trim are not used. In light turbulence, airspeed increased. As the airplane slowed, the PMS moved the throttles forward but without effect. The flight engineer moved the n. 4 throttle forward but without effect. The INS caused the autopilot to hold the left wing down since it could not correct with rudder. The airplane decelerated due to the lack of power. After attempting to correct the situation with autopilot, the Captain disengaged the autopilot at which time the airplane rolled to the right, yawed, then entered a steep descent in cloud, during which it exceeded maximum operating speed. It was extensively damaged during the descent and recovery (1991, p. 98).

As noted by the author, the NTSB concluded that:

> ... the probable cause was the captain's preoccupation with an inflight malfunction and his failure to monitor properly the airplane's flight instruments which resulted in his losing control of the airplane. Contributing to the accident was the captain's over reliance on the autopilot after a loss on n. 4 engine. The Board noted that *the autopilot effectively masked the approaching onset of loss of control of the airplane.* (ibid., p. 98.).

Without stating too much about the concepts that will be developed in the following sections, yet in contrast to the first elements of analysis retained by the NTSB, this paper claims that this accident's main contributing factors are *flaws in the design* of the information and control systems combined with the presence of *agents* that *operate independently* of any pilot's control action but *without adequate feedback.*

More precisely, as revealed by this incident, the breakdown of the pilot-automation system onboard this aircraft – which is typical of a design currently known as "technology centered automation" – is mainly due to a lack of *controllability* of the automatic systems involved, coupled with a lack of *visibility* and *predictability* of those systems' status, effects and interactions over the considered flight phase, and an engine failure.

Assuming certification has among its major goals to guarantee the passengers' and pilots' safety and the airplane integrity under normal and abnormal operational conditions, the authors suggest it would be particularly fruitful to come up with a *conceptual reference system* providing the certification authorities both with a theoretical framework and a list of principles usable for assessing the quality of the equipment and designs under examination.

This is precisely the scope of this paper. However, if the authors recognize that the conceptual system presented is still under development and would thus be best considered as a *source of reflection* for the design, evaluation and certification processes of advanced aviation technologies.

The Multiple Resources of Automation

> We consider automation to be a tool or resource – a device, system or method by which the human can accomplish some task that might be otherwise difficult or impossible, or which the human can direct to carry out more or less independently a task that would otherwise require increased human attention or effort. (Bilings, 1991, p. 7)

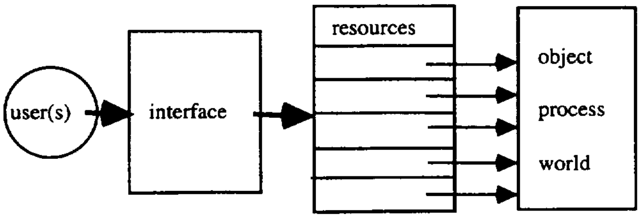


Figure 1. A simplified diagram of automated control of automation

Four components define the classical automated control situation (Figure 1) (e.g., Sheridan, 1988):

- A set of users, operators or pilots with some goals or tasks to achieve

- An object, a process, or a world, characterized by a set of stated variables, upon which the users want to act

- A set of automated resources which possess the capability to change the state of the world on the behalf of the user

- An interface which provides the user with the means to activate and control these resources.

It is clear from everyday life experiences (Norman, 1986) that resources can display very different behavioral characteristics, and that this influences the way we use them as well as the type and amount of knowledge we need to do this efficiently.

The following three essential categories of resources can be identified according to their different behavioral characteristics.

Functions constitute the simpler type of resources and affect the state of the world in a straightforward way. Their effect only depends on the state of the world prior to their activation. Moreover, this effect can be described by a simple state transition: the state of the world before and after the activation of the function. Functions are thus usually extremely predictable resources (e.g., manual seat-belt and non-smoking signs activation, manual throttle control, etc.).

Functional patterns constitute the second type of resource. The behaviour of functional patterns is also only dependant on the state of the world prior to their activation. Nevertheless, contrarily to functions, their effects are not described as simple state transition but as sequences of successive states. Predictability of these patterns is still high, but requires more information than with simple functions (e.g., landing gear extraction and retraction, flaps retraction).

Agents finally are described by sequences of successive states. Nevertheless, with agents sequences are not only influenced by initial conditions but also by conditions varying during execution of the sequences themselves (e.g., agents range from low-level automatisms [attitude stabilizers...] to high-level pilot aiding devices [the Flight Management Systems or the Performance Management Systems]):

> In more automated systems, the level of animacy of *machine agents* has dramatically increased. Once activated, systems are capable of carrying out long sequences of tasks *autonomously.* For example, advanced Flight Management Systems can be programmed to automatically control the aircraft from takeoff through landing. (Sarter & Woods, 1993, p. 6)

As suggested by the previous examples, automated functions, functional patterns and agents are present in most technological contexts. Processes and agents are especially useful in task-intensive situations and have come to progressively replace functions in modern airplanes. Several reasons account for that evolution.

The first is that, as any human operators, pilots are limited in their perceptual abilities (e.g. they cannot fine control the airplane's attitude without assistance or manually react to events in milliseconds) and in their capacities to process information (cf. the classical concepts of bounded rationality (Simon, 1957)* and short term (Miller, 1956) or working memory (Baddeley & Hitch, 1974; Baddeley, 1986; Reason, 1987) limitations in cognitive psychology).

Some external resources can be introduced to cope with these limitations, but it should be clear that purely functional resources cannot suffice in highly dynamic situations such as piloting an airplane. Because of humans' limited bandwidth I/O channels and because of their limited and rather slow processing capabilities, it is not possible to ensure correct coordination and activation of multiple functional resources. Agents, on the other hand, because they can be considered as functions with autonomy, display that ability to coordinate, at least locally, several specialized functions (Maes, 1989). Agents integrate the logic behind functional integration and activation (acting on the process through functions – see the notion of competence – or recursively through simpler agents).

Producers (airplane designers) and consumers (commercial airlines) have extended the scope of the tasks expected from the global system crew/airplane/ATC. The necessity to enhance safety and performance while flying in highly dense and crowded airspace are among the main motivations for the introduction of agents in airplanes. As a result, the complexity of the flying task has grown to such levels that it has become necessary to extend the perceptive, motor and processing capabilities of the crew. The task itself has been broken down into simpler primitive subtasks that have been allocated to specialized agents.

Thus, there has been a continuous trend in aeronautics to introduce more and more automation into cockpits. However, some problems related to this approach have been described by several human factors experts, like Sarter and Woods:

> New automation is developed because of some payback (precision, more data, reduced staffing, etc.) for some beneficiary (the individual practitioner, the organization, the industry, society). But often overlooked is the fact that new automated devices also create new demands for the individual and groups of practitioners responsible for operating and managing these systems. The new demands can include new or changed tasks (setup, operating sequences, etc.), and new cognitive demands are created as well. There are new knowledge requirements (e.g., how the automation functions), new communication tasks (e.g., instructing the automation in a particular case), new data management tasks (e.g., finding the relevant page within the CDU page architecture), new attentional demands (tracking the state of automation), and new forms of error or failure (e.g., mode error). (1992, p. 17)

This kind of role sharing and interaction pattern has the long term effect of removing the pilot from the loop and decreasing system awareness especially as feedback on automation status and behaviour is poor and difficult to obtain.

While our goals are not to ignore these very important problems, we would like to draw the attention to the problems specifically related to the interfacing of functions and agents in modern aircrafts. We especially believe that some of the problems encountered in modern glass-cockpits

* " The capacity of the human mind for formulating and solving complex problems is very small compared with the size of the problems whose solutions is required for objectively rational behaviour in the real world - or even for a reasonable approximation of such objective rationality. "
(Simon, 1957, quoted by Reason, 1987, p. 76).

appear because the "agent" character of some resources has not been sufficiently recognized, and that in some case agents have been interfaced as if they were mere functions.

As will be shown later, the main question behind usable interface design is:

> *How do we provide the user with the necessary knowledge and the means to interact with the resources in order to act and react in the world according to goals or tasks?*

We will first show how the approach adopted by the classical HCI community regarding the interfacing of functions on a static setting has succeeded in its attempts to answer to this question, how it has provided designers with principles to support evaluation, certification and designs methodologies and, in the end, end-users with highly usable forms of interfaces. We will then show how such a strategy could be applied to interface agents in dynamic worlds.

In the end, we will have provided the reader with two sets of principles, respectively for functions and agents interfacing, that could influence the way evaluation and certification of interfaces incorporating these two types of resources are performed.

Interfacing Functions in Static Problem-Spaces: Classical HCI

The now classical domain of Human-Computer Interaction has proven its ability to solve interfacing problems with powerful computerized tools.

Such successes must be related to three factors:

a) cognitive theories of human-computer interaction have been produced

b) some general principles that interfaces have to verify have been defined, either as a subproduct of the cognitive theories of the interaction, or of empirical data (controlled experiments, analysis of errors, etc.)

c) some generic forms of interfaces conforming to these principles have been designed and have received a wide acceptance.

Cognitive Theories of Interaction

Cognitive theories of interaction between users and computers have existed for several years now. Strongly influenced by the early attempts of Artificial Intelligence to produce models of problem-solving and planning (such as GPS, Newell & Simon, 1972), nearly all rely on the same approach and assume that the user achieves goals by solving sub-goals in a divide-and-conquer fashion (Dix, Finlay, Abowd, & Beale, 1993): GOMS (Card, Moran,.& Newell, 1983), CCT (Kieras & Polson, 1985), TAG (Payne & Green, 1986).

The GOMS model, which has served as the basis for major research in cognitive modelling applied to HCI (Dix et al., 1993), considers for example that an interaction situation can be described in terms of Goals, Operators, Methods and Selection.

Goals are the user goals; they are "what has to be achieved".
Operators are the basic operations the user can perform to affect the system state (represented as state transitions).
Methods describe how alternative sub-goals decomposition can help the user to reach the same goal.
Selection rules attempt to predict which methods the user will use to reach goals depending on the user itself and the state of the system.

In such models, the computer is clearly considered as a static setting; that is, one whose state only changes as an effect of the actions of the user considered as the application of operators.

To illustrate how the distinction between a static problem-space and its related operators or functions encounter personal experience, we will analyse how the file management problem is treated on most personal computers.

Interface designers confronted with the file management problem have to define ways to represent files as they appear on some physical support (a hard disk for example) and provide users with the means to manipulate them. Files are usually organised on this support according to a hierarchical structure (a tree). This structure is static; it remains as it is unless the user attempts a modification. Files and directories can be moved, copied or deleted. Individual files can be transformed thanks to applications (word processors, spreadsheets,...) that change their internal structure. All these operations are under control of the user.

The desktop metaphor (Booth, 1989) elegantly solves this problem:

a) *The static problem-space:* the desktop metaphor is a possible alternative to the problem of representing static problem-space. Files and directories are represented by icons or names in lists. Files which are in use are represented by open windows.

b) *Functions or operators:* most of the functions are accessible directly on the desktop (Direct Manipulation Interface; Hutchins, Hollan, & Norman, 1986). File displacement and deletion are operated through moves of the mouse or function activations through menus. Activation of an application on a specific file is possible through double-clicking commands or menus.

General Principles

Thanks to the coherent framework provided by the analogy with problem-solving or planning on static problem-space, it is possible to produce a structured and theoretically sound (contrarily to most of the guidelines) set of principles about properties of usable interfaces.

These principles rely on four underlying ideas.

a) In order to act efficiently on a static problem-space, the user must have access to some knowledge about the problem-space itself and the functions that can be applied.

- The user must be able to assess the *current state* of the problem-space;
- He/she must know which *operators* or *functions* can be applied to this state; and
- What *transition* will occur if an operator or function is applied;

Without this information, the goals cannot be reached (one would say, in terms of problem-solving or planning theory, that the problem cannot be solved).

b) Part of this knowledge is related to the static problem-space and the other part concerns the functions themselves.

c) The knowledge required to interact with static problem-spaces can be distributed within the interface/user system. Well designed interfaces provide the user with a lot of knowledge about the current state of the problem-space (visibility), the functions that can be applied (availability) and the related transitions (predictability). To paraphrase Norman (Norman, 1988) in such interfaces, "information is in the world." In badly designed interfaces, the current state of the problem-space is not visible and a lot has to be remembered (in short-term memory). It is hard to tell which functions can be applied or what will be their effects. In such interfaces, "information is in the head."

d) Principles (necessary principles) that warrant the presence and availability of the necessary knowledge can be stated. Secondary principles, considered less important, can be proposed to indicate how to make the interface more usable or how to support the user in its tasks.

Principles for Static Problem-Spaces

Visibility. Can I see what I need to see? The goal of this principle is to ensure that the user might have a full and accurate representation of the current state of the problem-space.

Interpretability – Do I see what I'm supposed to see? It is not sufficient for the user to have access to a representation of the problem-space. A representation conveys some meaning about some real situation which is abstracted into symbols, and have to be interpreted by the user. This principle ensures that the user correctly interprets the representation. Some simpler but nevertheless essential principles usually support interpretability: consistency or coherence of the symbols and of their interpretation, familiarity and generalizability of the symbols.

Flexibility – May I change the way I see? The possibility of tuning the representations, in particular to modulate the informational flow according to the bandwidth of the human cognitive processing limitations and the particular needs of the current situation, is a specially desirable property of usable interfaces.

Reliability – Is this thing the real picture? This one deals with a critical feature of any interface. It must be reliable, and the user must be confident with the information it provides or the resource it helps to use. When applied to problem-space representation, the reliability principle wonders whether the representation presented to the user constitutes an accurate and reliable representation of the problem-space and how this can be assessed by the user itself.

Learnability – Can I avoid headaches? This principle is important because of the way users accept new products or interfaces is influenced by their learnability. In the case of a static problem-space representation, how easily can the user learn the rules that help to interpret correctly the representation. Once again, simpler principles such as consistency, familiarity and generalizibility strongly contribute to facilitate learnability.

Principles for Functions

Availability – What can I do? In order to apply functions on the static problem-space as if they were operators, the user must be in a position to decide which functions can be applied on the problem-space. General availability refers to the complete list of functions provided by the interface. Local availability concerns the limited list of functions applicable to specific states of the problem-space. Knowledge concerning both types of availability should be accessible to the user.

Accessibility – How can I do it? Once the user has gained some knowledge about which functions can be applied on the problem-space and has chosen one or a sequence of them to apply, he/she has to specify it for the interface. Knowledge about how to access functions and how to activate them on the correct objects should available to the user. Consistency, familiarity and generazibility are once again among the simpler principles that help the user to access functions.

Predictability – What will happen? Predictability is without any doubts *the* essential principle to conform to. In problem-solving and planning models, the ability to predict how the state of the world will change when an operator iterface with a machine is crucial for resolution or task satisfaction. The user must possess the necessary knowledge to be able to generate plans or sequences of actions on the interface that in the end will meet its goals. Modes, if any, have to be made visible to the user because they influence, by definition, the way functions behave and thus constitute a threat to predictability.

Feedback – How is it working and what are the effects? Feedback is essential because it permits the user to assess that the intended state has been reached, and hence that the activated function has been applied correctly. Feedback is thus associated to error detection, but also to the ability to learn (see learnability principle) the necessary knowledge to predict functional effects (see predictability principle). Two forms of feedback are usually encountered. The first type concerns the visibility of the function status (progression bars) and help to confirm that the access to the function has been successful (see accessibility principle). The second type of feedback ensures that the effects of the activated functions are visible. In the strictest sense, this second form of feedback is more concerned with visibility of the representations, and is thus not a pure functional principle.

Controllability – How the hell do I stop this thing from erasing my hard disk? As dramatically stated by the previous sentence, controllability is a particularly desirable feature. Nevertheless, in general, interfaces provide a very limited set of interactions between a running function and the user (otherwise, it would be an agent). Control is usually limited to interruption (either temporary or definitive) of the function execution.

Flexibility – Can I do it the other way? Users are not machines, and they are faced with very variable tasks. Moreover, users all differ. They have different backgrounds, different cognitive styles, and usually different goals. For such reasons, while not resorting to the major, necessary principles, flexibility is generally appreciated by users.

Automatibility – Can I automate this sequence of operations? There are two facets to automatibility whose advantages are obvious. Machine-supported automatibility refers to the possibility for the user to define "macros," automated sequences of functional activations, with or without parameters. Cognitive automatibility concerns the ability of the user to automate the motor and cognitive processes that support its access to functions. This form of automatibility is strongly conditioned by good visibility of the problem-space and easy and consistent access to functions.

Task Conformance – Does it really cover all my needs? This principle concerns the scope of the available functions regarding the nature of the task they are to perform. It can be considered from a general point of view (the global availability) or more locally according to specific situations (the local availability compared to the local task): i.e., is the function available when needed?

Error Management – What if I err? Users are fallible (Reason, 1990). Good interfaces have to take this into account and exhibit error resistance (prevent users to make errors, e.g. Masson & De Keyser, in press) and error tolerance (help users to correct effects of errors through reversibility, escapability, recoverability of function applications).

Reliability – Is this stuff really working as it is supposed to? While being extremely reliable systems, modern computers are nevertheless mere material human artifacts and consequently suffer from design errors as well as from the usually hidden effects of the second-law of thermodynamics. At the software level, bugs are present in any complex application. At the hardware level, problems and troubles sometimes occur due to heat, dust, fatigue or even failure of a component. Interfaces should furnish the user with means to ensure that the functional resources effectively affect the state of the problem-space as reflected by its representation and the different feedback mechanisms.

Learnability – Can I avoid headaches? Learnability of functions is essential. As already stated for problem-space related principles, it is generally a necessary condition for the acceptance of an interface. Several aspects can be learned and thus lead the user to eliminate exploratory or information seeking behaviors. Every piece of the necessary knowledge related to the primary principles (availability, accessibility, predictability) can be learned. Some rules that help the user to deduce such essential pieces of information from the representation of the problem-space can also be abstracted and then greatly contribute to simplify the activation of the functional resources; hence the pervasive character of the consistency principle.

Generic Forms of Interfaces

Classical HCI has also succeeded in its attempts to apply these principles to interface design. Graphical user interfaces, and especially WIMP (windows, icons, menus and pointers) interfaces (Dix et al., 1993), which constitute the standard interface for interactive computer systems (Macintosh, Windows-based IBMs and compatibles, desktop workstations) have proven their usability to millions of end-users.

Such kinds of interfaces indeed provide users with an excellent visibility over the current state of the problem-space (e.g., the desktop of the Macintosh) as well with consistent and familiar rules to interpret the representation (the desktop metaphor). Users habitually have the opportunity to tune these representations (e.g., different ways to display files in a directory) and this contributes to the interface flexibility. Moreover, such interfaces are highly learnable, especially because of their coherent and metaphoric nature.

GUIs and WIMPs equally perform at their best regarding functions. Availability is usually very well documented by the interface. This is a least true of the most used functions. Less common functions are not well known to users, especially in case of very powerful tools such as word-processors that provide users with hundreds of functions. Accessibility is extremely good, thanks to the mouse and its clicking commands and to menus (that also contribute to availability). Predictability is good (at least for simple operations on the desktop) because of the coherence of the access rules and the already quoted metaphoric nature of the interface. Feedback is immediate, but restricted to objects visible on the desktop. Controllability is limited, but it is enhanced for functions that have destructive effects on the desktop. Flexibility is usually good, thanks to the several different ways to perform operations (directly on the desktop or through menus). Macros are provided as default features or can be added thanks to dedicated applications. Task conformance is the principle where these graphical interfaces are at their worst: the possible scope of what can be done is somehow limited, especially if compared to very powerful command-languages (e.g. Unix) dedicated to files management. Errors are handled differently by the manufacturers of common GUIs. Operations that imply a displacement of files between two places (a move operation) can generally be undone, but file deletion is sometimes an operation that cannot be reversed without specifically dedicated tools. Learnability, finally, is usually extremely good (perhaps it is in the end the main reason for the success of these interfaces within a computer-illiterate population) especially because of the so-praised consistency of the interface (even between applications) and the metaphor of the desktop.

Interfacing Agents in Dynamic Problem-Spaces: HCI Goes to the Real World

We have carefully analysed the approach followed by classical HCI to solve the problems related to the interfacing of functions in static problem-spaces. Now we would like to see how such a strategy can be applied to interface agents in dynamic problem-spaces. According to other authors (Kay, 1990; Laurel, 1990), the interfacing of agents is *the* challenge of tomorrow's HCI. We already have shown how agents constitute invaluable resources for users, operators or pilots in their respective tasks. That is why manufacturers and designers have introduced them at several different levels of automation used in process control.

Cognitive Theories of Interaction

There has been for a few years an emergent interest about ideas related to the integration of a distributed work or processing force into a coherent and goal-oriented whole. Computer Science, for example, has already produced numerous formal studies about parallel processing and synchronization problems. Distributed Artificial Intelligence (DAI) aims designing systems or societies of agents (multi-agent systems) that collectively exhibit the ability to solve complex problems with more robustness than classical approaches (Maes, 1990). On the linguistic side, Winograd and Flores (Flores, Graves, Hartfield, & Winograd, 1988) have developed linguistic-based theoretical perspectives for analysing group actions. Coordination Theory (Malone & Crowston, 1990) as a general and abstract theory tries to establish the connections between several different disciplines that are concerned with similar coordination phenomena. On the applied side, Computer Support to Cooperative Work (CSCW) is aiming at providing organizations or groups of users with better ways and tools to work together (Dix et al., 1993). As demonstrated by the next excerpts, concerns about agents and modelling human/agent interaction have even been expressed in aeronautics by human factors authors.

> Pilots were surprised when the aircraft did not respond as expected; they did not realize or understand why their instructions to the automation had not resulted in the desired change. In some sense, this is a good example to show how pilots try to communicate with the system in a way analogous to communication with another human agent. They assume that entering the desired target value is sufficient for the system (as it would be for a human crew member) to understand that it is supposed to achieve this new target and how it is supposed to do so in detail. (Sarter & Woods, 1993, p. 12)
>
> (This) direction is to consider supervisory control of automated resources as a kind of cooperative or distributed multi-agent architecture. (Sarter & Woods, 1993, p. 12)

Despite these efforts and remarks, there is nothing today like a single and coherent body of theory about coordination between agents (Malone & Crowston, 1990), and it hard to think of any integrated cognitive theory of interaction between humans considered as agents, or between humans and automated agents. Nevertheless, there is more and more awareness of the similarities between the problems encountered by researchers involved in these approaches to cooperative systems as is witnessed by the increasing number of workshops or conferences on the topic. On the cognitive side, expectations about future progress will rely on domains such as social or developmental cognitive psychology as well as psycholinguistics to produce a coherent and integrated theory of human interaction with agents.

General Principles

Designers faced with the problem of interfacing agents are still left without the sort of powerful framework they used to rely on when designing functional interfaces. Nevertheless, some important principles that interfaces with agents should verify can already be stated, thanks to extensions of the basic principles for functional interfaces, reflections about the necessary knowledge required for usable interaction, and to recommendations formulated by analysts when incidents with such interfaces were reported. We will thus try to rely on these excellent studies of problems and incidents encountered with automation in modern glass-cockpits as major sources for defining general principles.

On the epistemic side, it is at least clear from a formal point of view that more knowledge (distributed between the user and the interface) is needed to control a dynamic problem-space than a static one. Anticipatory behaviors, of which some researches have shown the heuristic value (Van Daele, 1992), are only possible if the user, operator or pilot has some knowledge or ability to predict how the controlled system will naturally evolve if no action is taken. Interfaces to dynamic worlds or problem-spaces should provide the user with such knowledge or resource (see the predictability principle for dynamic problem-spaces).

More knowledge is also needed to interact with agents than with functions. Agents can be of numerous different types and differ in terms of complexity (ranging from reactive to cognitive agents; Erceau & Ferber, 1991). Whatever the importance of such factors, the main difficulty with agents certainly comes from their flexibility (complex agents can exhibit different behaviours in similar situations) and from their autonomy (agents incorporate their own logic behind functional activation and act autonomously on the world). As a consequence, agents must be considered as generally less predictable resources than functions.

Respective Scopes or Competences and Cooperative Modes. A supplementary and rather essential distinction must be introduced before devoting some attention to the principles. It concerns the distribution of competence between the user, the operator or pilot and the agent. To use a multi-agent terminology, one would say that only two cooperation modes are possible: either the job is done by the function or it is done by the user. The situation is quite different with agents. Because such resources display possibilities for an extended amount of controllability these resources provide the capability for more complex cooperation modes.

Three classes of cooperation modes have to be considered:

a) The job is done by the user. The agent is not active or works on another task.

This corresponds to the concept of "direct manual control."

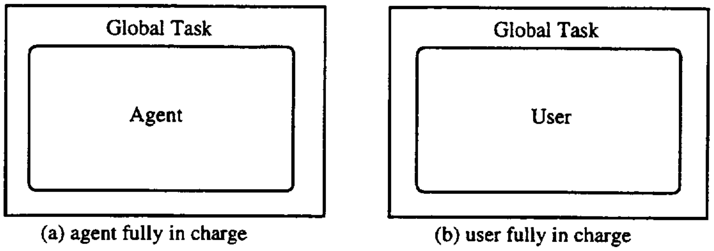
According to Billings (1991, p. 27) *direct manual control* is characterized by the pilot's direct authority over systems, manual control using raw data, unaided decision making and manual communications.

However, as pointed out by the author, no modern aircraft can be operated entirely on that mode. "Indeed, an aircraft operated even by direct manual control may incorporate many kinds of control automation, such as yaw dampers, a pitch trim compensator, automated configuration warning devices, etc." (Billings, 1991, p. 26). For example, landing gear retraction and extension are still manually operated in all transport aircrafts.

b) The job is done by the agent. The user is not active or works on another task.

This is precisely the meaning of the "autonomous operation" concept. As summarized from Billings, autonomous operation is characterized by the fact that the pilot has no role to play in operation, that the pilot has normally no reason to intervene and that the monitoring is limited to fault detection (Billings, 1991, p. 26).

Until the introduction of the A32O and MD11, very complex systems were operated in a full autonomous fashion. In those new aircraft, however, major systems operate this way. For example, in the MD11, failure detection and subsystem reconfiguration are performed autonomously. Longitudinal stability and control wheel steering are also autonomous operations.



(a) agent fully in charge (b) user fully in charge

Figure 2: Agent and User interaction.

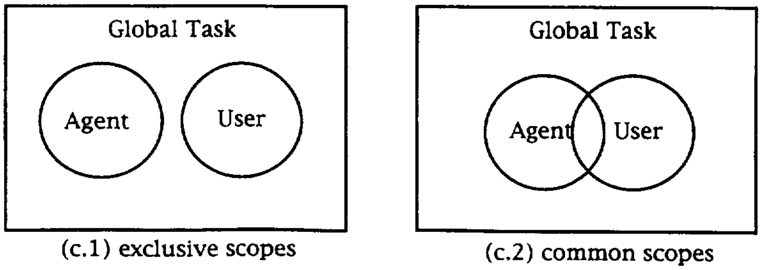
c) The job is done by both the user and the agent. Each of them has its own part of the task. Two situations have to be distinguished:

1) The two tasks are exclusive. This occurs for example when the agent and the user work on two different sub-systems. If the two tasks are interdependent (the two sub-systems interact) then the agent and the user have to synchronize their actions.

An example of such a sharing pattern is given by Billings: "the pilot may elect to have the autopilot perform only the most basic functions: pitch, roll and yaw control...he or she may direct the automation to maintain or alter heading, altitude or speed, or may direct the autopilot to capture and follow navigation paths, either horizontal or vertical...In all cases however, the aircraft is carrying out a set of tactical directions supplied by the pilot. It will not deviate from these directions unless it is capable of executing them." (1991)

2) The two tasks share a common part. This could occur when the agent and the user do work on the same sub-systems. In such cases, conflicts are likely to arise, and resolution techniques have to be provided.

For example, in the 320, the flight control system incorporates an envelope limitation system that operates at all times and interacts with pilot's commands, in order to guarantee that safety barriers are not overcome. For example, bank angle, pitch and angle of attack cannot be exceeded by the pilot unless the flight control computer is turned off.



(c.1) exclusive scopes (c.2) common scopes

Figure 3: Agent and User scopes.

Moreover; cooperation modes with agents cannot solely be considered in a static perspective (they are fixed and cannot be changed over a task session or a flight). Dynamic mode changes are also observed in modern cockpits (modes change over the course of a task session, either through agent or user instruction).

> An important characteristic of *automatic flight-path control* is the high degree of dynamism. Transitions between modes of control occur in response to pilot input and changes in flight status. Automatic mode changes can occur when a target value is reached (e.g., when levelling off at a target altitude), or they can occur based on protections limits (i.e., to prevent or correct pilot input that puts the aircraft in an unsafe condition). (Sarter & Woods, 1992, p. 306)

For such reasons, as stated by Billings, feedback (see the feedback principle) should be given to the user or pilot whenever an important mode change occurs.

> Automation should never permit a situation in which "no one is in charge"; pilots must always "aviate" even if they have delegated control to the autopilot. It is for this reason that autopilot disconnects are usually *announced* by both visual and aural alerting signals (Billings, 1991, p. 85).

To confirm the importance of issues related to cooperation modes, Sarter and Woods also describe how the pilot's inability to dynamically change modes can lead to some drastic measures.

> During the final descent, the pilots were unable to deselect the APPR mode after localizer and glideslope capture when ATC suddenly requested that the aircraft maintain the current altitude and initiate a 90° left turn for spacing. They tried to select the ALT HOLD and HDG SEL modes on the MCP to disengage the APPR mode and comply with the clearance, but neither mode would engage and replace the APPR mode. They finally turned off all autoglide systems (Sarter & Woods, 1992, p. 311).

This leads us to some critical remarks about the way evaluation or certification of agent-based interface relying on principles should be performed.

- The analysis should begin with a very careful study of the possible cooperation modes between the user and the agent.

- It should detail *who is in control* of
  - The cooperation mode changes
  - The relative scopes of the user and the agent within a given cooperation mode (task migrability).

- For each possible cooperation mode, consider how the duality user/agent is positioned according to the principles. However, due to the very different ways the task is conducted in different cooperation modes, principles have to be applied with some nuances in mind and be related to the current specificities of the current mode.

In a short-response time agent (a regulator), whose capabilities are far beyond those of the pilot, the cooperation mode is such that the task is exclusively under the agent control. The main principles in this situation are a) reliability, and b) the capability for the pilot to assess that the agent is working in its competence domain. Principles such as predictability, that used to be essential for functional resources, are hereby not necessary (e.g. the gyroscopic stabilizer of Maxim, 1891, the stability augmentation system of Wright, 1907, and their successors in modern autopilots).

Principles for Dynamic Problem-Spaces

Visibility – Can I see what I need to see? Visibility of the problem-space acquires hereby a special status due to its dynamicity. In static problem-spaces where the world does not change spontaneously, the user or pilot can rely on short-term memory to maintain awareness and orientation. In dynamic problem-space, updating is necessary and this is only possible through predictions or observations of the future states. In a particularly complex dynamic context with heavy task constraints, the concept must even be extended to meet the notion of "situation awareness" (Sarter & Woods, 1991). In such situations, it is not enough to provide the user with the means to perceive the state of the problem-state, but also to ensure that it will be *attended.*

> Situation awareness has recently gained considerable attention as a performance-related psychological concept. This is especially true in the aviation domain where it is considered an essential prerequisite for the safe operation of the complex dynamic system "aircraft." There are concerns, however, that inappropriately designed automatic systems introduced to advanced flight desks may reduce situation awareness and thereby put aviation safety at risk. (Sarter & Woods, 1991, p. 45)

The problem of the amount of information that must be visible is also addressed by Billings:

> "How much information is enough? How much is too much?" Though pilots always want more information, they are not always able to assimilate it. (Billings, 1991, p. 46)

As pointed out by the author, such a question should be answered (as suggested above) according to a clear consideration of the cooperative mode between the pilot and the agent and their respective involvement in the control task.

> Pilots need much less information when subsystems are working properly than when they are malfunctioning (Billings, 1991, p. 46)

Interpretability – Do I see what I'm supposed to see? No significant difference with its functional counterpart.

Flexibility – May I change the way I see? Flexibility applied to the representation of the dynamic problem-space means that the user or pilot is capable of adapting this representation to its current or future needs. Such a possibility is present in several subsystems displays (zooming features) of glass-cockpits.

Predictability – What will happen if I stop acting? The new principle is based as already stated on the heuristic value of predictory or anticipatory behaviours in most dynamic control situations (Van Daele, 1992). Good interfaces for dynamic problem-spaces should provide the users with means to anticipate future states. Several examples of such an approach already exist in aeronautic contexts. On ATC radar control screens, airplanes can be represented with small tails indicating their speed and direction. This helps operators to anticipate their future trajectory. On TCAS screens, vertical speeds of surrounding airplanes are represented by small arrows. In general any graphical representation of a trend indicator can contribute to the predictability of the dynamic problem-space.

Reliability – Is this thing the real picture? The problem of reliability related to dynamic problem-space is perfectly stated by Billings:

> It must be kept in mind that sensors, processing equipment or display generators can fail, and that when incorrect information is presented, or correct information is not presented, there is the potential for confusion in the minds of pilots. (1991, p.40)

An interface on a dynamic problem-space should help the user to ensure that it functions correctly, both in its ability to display correct and accurate information about the real state of the monitored systems and in its ability to inform the user about future states (support to predictability principle). Redundancy of display equipments or availability of displays related to interdependent subsystems can help the user or pilot to ensure that the informational interfaces are functioning correctly thanks to comparison or inter-display coherence checking.

Learnability – Can I avoid headaches? Here, again, there is no significant difference with functional counterpart.

Critical State Assessment – Is this the right place to stand still? This new principle concerns the peculiar problems associated with dynamic problem-spaces. In such spaces, states are rarely equivalent. Some of them require special care or attention, either because the monitored system ventures within space regions where irreversible damages could be observed, or because its intrinsic dynamic could lead the user or pilot to lose control. Interfaces provide critical state assessment support users and help them to enhance their performance.

Principles for Agents

Availability – What can I do? Availability as such refers to the capability the user has to decide whether a resource exists and is available. Users or pilots should be informed of the different agents they might use as resources (global availability) as well as when these can effectively be used (local availability), and in which cooperative modes.

Accessibility – How can I do it? Users or pilots willing to use an agent as a resource should have access to some knowledge about how to activate and configure it in the cooperative mode of their choice (if they are in control of this variable).

Predictability – What will happen? Predictability is, without a doubt, one of the principles that must be considered with extended caution when trying to interface agents. As previously stated, agents are autonomous systems. They consequently present less predictable behaviours than functions. Numerous examples of incidents related to a lack of predictability of some agents in glass-cockpits have already been reported. Sarter and Woods (1992) have realized a study through a questionnaire asking pilots to describe instances in which FMS behaviour was surprising, and to report modes and features of FMS operations that were poorly or not understood. 135 B-737-300 line pilots from an airline company participated to that survey (Sarter & Woods, 1992).

> Pilots indicated that the algorithms underlying the calculation of a VNAV are not transparent to them. They cannot vizualize the intended path; therefore, they are sometimes unable to *anticipate* or *understand* VNAV activities initiated to maintain target parameters...Several pilots reported that they have been *surprised* by VNAV when it failed to start the descent on reaching the top-of-descent (TOD) point...

The problem the user or pilot is faced with is one of agent modelling. Designers must ensure that they provide the user with a correct model of the agent. Two principal classes of models govern the theories about agents: mechanistic models and intentional models. In mechanistic models, the user relies on a kind of finite-state automaton approximation of the agent whose behaviour can be predicted thanks to the awareness of relations between some internal parameters or variables of the agent, the input it is actually processing and the resulting behaviour. In intentional models of agents, the user predicts future behaviors of the agent on the base of its goals or intentions. Consequently, and whatever the type of model hold by the user (depending on the type and complexity of the agent), it seems essential that any important autonomous change that might modify the way the agent will behave in the near future (a change of mode in mechanistic models or a change of goals or intentions in intentional models) is reported to the user.

Scope or Competence Awareness – What can this thing do and when? This new and essential principle concerns the competence of the agent: what it (can) do and in which circumstances. With purely functional resources, competence awareness is close to predictability. Functions induce simple state-transitions (what is does) from the states upon which they apply (in which circumstances). Due to the extended flexibility and autonomy of agents, this similarity does not appear and a new principle has to be introduced. Scope awareness is extremely important, at least when the user or pilot is in control of the cooperation modes and of task migrability: the pilot must be able to assess that the agent is performing reliably (reliability principle) and correctly (adapted to the task) in its domain (scope awareness) of competence. Designers must consequently provide the user or pilot with this knowledge through the interface, documentation, and/or training courses.

Feedback – How is it working and what are the effects? This is another essential principle for agents. Due to the new problems introduced with predictability of the agent and the correlated needs to model its behavior, the visibility of agent status increases in importance. As already reported, mode awareness (in mechanical models) is a condition for real cooperative work between the user or pilot and the agent.

> Pilots reported that they are surprised by uncommanded mode transitions that occur on reaching a target state or for protection purposes. Most often, the reports referred to the automatic reversion from vertical speed mode to LVL CHG mode, which occurs if the airspeed deviates from the target range due to an excessive rate of climb or descent.

> Pilots' reports seem to indicate that such uncommanded changes are difficult to track given current cockpit displays and indications. (Sarter & Woods, 1992, p. 311)

Visibility of the agent's effects are of equal importance. Because agents display autonomy, any change introduced on the dynamic problem-space by the agent should be reported or be at least visible to the user, especially in cooperative modes where both the agent and the user are in charge of the same task. It is also due to this visibility that the adequacy of the decision to activate the agent as well as its reliability can be assessed. See also Billings (1991, p. 85) and the concept of "fail-passive" control automation situations that describe hazardous conditions where visibility of the agent effects is lowered.

Controllability – How the hell do I stop this thing grounding my airplane? Because of the autonomy of agents, and their ability to induce disastrous effects on the controlled problem-space, controllability should remain high in every circumstances. Woods describes how "clumsy automation" can contribute to lower controllability in circumstance where it is especially needed.

> Clumsy automation is a form of poor coordination between the human and machine in the control of dynamic processes where the benefits of the new technology (i.e. additional tasks, forcing the user to adopt new cognitive strategies, new communication burdens, new attentional demands) occur during periods of peak workload, high criticality or high tempo operations. (Cook et al., 1990 ; Sarter & Woods, in press)
>
> Significantly, deficits like this can create opportunities for new kinds of human error and new paths to system breakdown that did not exist in simpler systems. (Woods, Cook, & Sarter, 1993) (Woods, 1993, p. 2)

It is a clear evidence that users or pilots should have the capability to disengage automation (agents) or at least change the current cooperative mode to some mode where they are more involved whenever they think it is needed. Billings states this very precisely:

> **Premise**
>
> The pilot bears the ultimate responsibility for the safety of any flight operation
>
> **Axiom**
>
> The human operator must be in command

Principles of Human-Centered Automation (extract). (Billings, 1991, p. 12)

The same author expresses serious concerns about recent examples of violation of such principles. The flight control system of the A320 and its envelope limitation operate at all times: they cannot be disengaged by the pilot. In the MD-11, major aircraft systems operate autonomously to a large extent:

> ...(civil aircraft) do on occasion have to take violent evasive action, and they may on extremely rare occasions need control or power authority up to (or even beyond) structural and engine limits to cope with very serious failures. The issue is whether the pilot, who is ultimately responsible for safe mission completion, should be permitted to operate to or even beyond airplane limits... (Billings 1991, p. 29)

Error Management – What if I err? As pointed by Billings, system operation errors are responsible for roughly two-thirds of air carrier accidents (1991, p. 24). It thus mandatory, as for functions, to design error-resistant and error-tolerant agent interfaces that attempt to minimize the effects of human error. Monitoring capabilities into the automation, system envelope limitations and procedural control are among the currently investigated techniques to enhance safety.

Task Conformance – Does it really cover all my needs? Here again, there is no significant difference with functional counterpart.

Flexibility – Can I do it the other way? The multiplicity of ways a given resource can be used is usually a rather desirable feature, especially because it provides the user or pilot with a choice within several different strategies to achieve the same goal.

> For example, an automated cockpit system such as the Flight Management System (FMS) is flexible in the sense that it provides pilots with a large number of functions and options for carrying out a given flight task under different circumstances. There are at least five different methods that the pilot could invoke to change altitude (Sarter & Woods, 1993, p. 2).

However, with complex cooperative agents, flexibility can strongly contribute to the "clumsiness" of automation and lead to very serious problems, as is witnessed by the same authors:

> This flexibility is usually portrayed as a benefit that allows the pilot to select the mode best suited to a particular flight situation. But this *flexibility has a price*: the pilots must know about the functions of the different modes, how to coordinate which mode to use when, how to "blumplessly" switch from one mode to another, how each mode is set up to fly the aircraft, and he has to keep track of which mode is active. These new cognitive demands can easily congruate at high tempo and high criticality periods of device use thereby adding new workload at precisely those time periods where practitioners are most in need of effective support systems.
>
> *Clumsy use of technological possibilities,* such as the proliferation of modes, creates the potential for new forms of human-machine system failure and new paths towards critical incidents, e.g. the air crashes at Bangalore (e.g., Lenorovitz, 1990) and Strasbourg (Monnier, 1992) (Sarter & Woods, 1993, p. 2).

Reliability – Is this stuff really working as it pretends? The reliability principle is extremely important with agents, especially because of their capability for autonomy and of the corresponding tendency of users to rely blindly on them.

As an example of overconfidence in automation, we would like to mention the accident of Scandinavian Airlines DC-10-30 occurred at Kennedy Airport on February 2, 1984. In this accident, the airplane touched down 4700 ft beyond the limit of an 8400 ft runway, was then steered to the right and landed in water 600 ft beyond the runway. The accident was due mainly due to a failure of the throttles to respond to the autothrottle speed control system commands and to the excessive confidence of the Captain in the reliability of that autothrottle system, in spite of a one month history of malfunctions. As noted by the NTSB among other causes, the "performance was either aberrant or represents a tendency for the crew to be complacent and over-rely on automated systems" (quoted by Billings, 1991, p. 99).

As pointed out by Billings, the ability to assess reliability is related to visibility of the problem-space, and to predictability of the agent behaviour (either based on mechanical or intentional models).

> It is thus necessary that the pilot be aware both of the function (or dysfunction) of the automated system, and of the results of its labors, on an ongoing basis, if the pilot is to understand why complex automated systems are doing what they are doing (Billings, 1991, p. 83).

However such a strategy might fail simply because of the stable condition the controlled process is in. Automatic monitoring of the agent reliability and visibility on its status is badly needed in such situations.

> "Fail-passive" control automation represents a particular potential hazard, in that its failure may not change aircraft performance at the time if the airplane is in stable condition. Such failures must be announced unambiguously to insure that the pilots immediately resume active control of the machine (Billings, 1991, p. 85).

Learnability – Can I avoid headaches? As with functional interfaces, comments must be made about the strong relation between the learnability of agent interfaces and their success measured in term of acceptance by users as means to access the full capabilities of the resources, in a safe and error-free fashion, and without the side-effects (clumsy automation, shift or loss of expertise, etc.) usually observed. Given the amount of knowledge that must be learned to interact cooperatively with an agent (this point will be developed later), learnability of agent interface *must* be (very) high. A few possible solutions will be described in the section about generic interfaces design.

Generic Forms of Interfaces

To begin with interfaces, some special points must be made about the amazing amount of knowledge required to interact fruitfully with agents. Users or pilots must be educated about the availability of agents (when they can be used), their accessibility (how they can be used), their scope or competence (what they can and can not do), their predictability (how they will behave or act on the problem-space under control), and the related mental models of their functioning, and finally about their controllability (how can they be controlled). Moreover, they must develop skills or mental processes dealing with how to communicate with them, how to evaluate their reliability through predictability and visibility of the problem-space, how to use or require feedbacks to enhance predictability itself, how to manage errors when they occur, etc.

To gain this knowledge or develop the means to access it is an extremely important task user/pilot face (hence, the "clumsy automation" problems and shift of workload toward more cognitive tasks reported by many authors). Moreover, to add to the task, the knowledge is required for each agent the user or pilot is interfaced with!

Our claim is that many problems described in modern glass-cockpits could be avoided if these simple – but overwhelming – considerations were taken into account.

A possible and promising solution, as already demonstrated with function interfacing through Direct Manipulation and metaphoric interfaces, is *to provide the user with a lot of the necessary knowledge* embedded *in the interface itself and with the means to* extract *it whenever needed.* Whether such interfaces should rely on graphical DMI-type interfaces or even more futuristic solutions (e.g., virtual realities) remains an open question.

A second and complementary approach is to reduce the *amount of knowledge* required to interact with agents. This is especially true at the level of the cockpit considered as a global work environment (or *macro-interface* with functions, functional patterns or agents provided as resources to interact with the airplane, the airspace and the ATC). Introducing intra- and inter-agent coherence into cockpits seriously contributes to limit the necessary knowledge to use them: agents can be classed according to the kind of cooperative modes they entertain with the crew and coherent communication protocols, feedback techniques and support to mental modelling can be established. The current situation with cockpits might be similar to the situation of interactive computers prior to the introduction of coherent GUIs, when every application had its own way to interact with the user.

Another important issue already considered by designers as decreasing the amount of knowledge that is not intuitive is familiarity. Thanks to the introduction into cockpits of more "natural" cooperative and communication modes (e.g., multi-modal and multi-media), the everyday life experience of interaction situations could be made more usable.

Conclusion

The influences of the introduction of new and sophisticated automation technologies in the last generations of commercial aircraft regarding the pilots-systems interactions has been extensively described by numerous experts in aeronautics and human factors engineering.

Technology allows a proliferation of interaction possibilities with increasing level of automation autonomy and poor feedback capabilities. These changes create new cognitive demands for the pilots, demands that turn to be he highest precisely during the most critical flight phases, where one would have expected the automation to be of the highest utility (Sarter & Woods, 1993. See also Moll van Charente et al., 1992, for similar results in the medical domain).

In summary, the complexity and lack of transparency of current automation challenge the pilot's ability to *cooperate* with the sophisticated systems he is provided with. At least three sets of measures can be explored to tackle the difficulties showed between current technologies and designs. The first set of measures would aim at improving the crew-automation interface as suggested above. A second approach to improve the quality of the cooperation is to decrease the cognitive demand on the pilot. More "natural" cooperative and communication modes are considered by cognitive psychologists as rather "effortless" processes, thanks to the many years spent to learning and automate them to interact with other humans. Improving mutual models of each other (it reduces the need to communicate), increasing reliability and the means to assess it, giving agents the possibility awareness of their own scope or competence, or providing dynamic feedback for important modes or intentional changes (e.g., Billings, 1991; Onken, 1992; Sarter & Woods, 1993) are among the several paths designers follow. The third set of measures is to conceive of the interactions between the pilots, the various automated resources, and even the ATC and other airplanes as a distributed cooperative multi-agent architecture in which each partner is engaged, in collaboration with all other agents, in the pursuit of a common system goal.

To sketch the current problems encountered with the "technology-centered automation," Wiener (1989) reports that the most common questions asked by pilots in glass cockpits are: "what is it doing?", "why did it do that?" and "what will it do next?"; to which Sarter and Woods (1993) add: "how in the world did I ever get into that mode ?"

We believe that all those interrogations could be reinterpreted in the light of the concepts and methodology developed in this paper.

According to the analysis made on the effects of current automation in cockpits, we suggest to extend that list by adding: *"how can I coax agents into performing what I want them to?"*

But as we have tried to highlight, this might not the right way to envisage operator-automation interactions. We here suggest that a shift in view could be fruitful, which would envisage both human and artificial agents as collaborative partners. And new technologies should facilitate that shift.

The question to be asked should rather be: "how can *we together* perform the missions I am in charge of ?"

Facing that new complexity, we suggest that the certification of future equipment and designs could benefit from a systematic methodology aimed at identifying the most critical problems in pilot-automation interactions. This paper constitutes one attempt to come up with such a methodology.

References

Amalberti, R. (1992). Safety in process-control: An operator-centred point of view. *Reliability Engineering and System Safety, 38*(313), 99-108.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.) *The Psychology of Learning and Motivation, 8.* London : Academic Press.

Baddeley, A. D. (1986). *Working memory.* Oxford : Oxford University Press.

Bainbridge, L. (1987). Ironies of automation. In J. Rasmussen, K. Duncan, & J. Leplat (Eds.), *New technology and human error.* United Kingdom: John Wiley and Sons Ltd.

Billings, C. E. (1991). *Human-Centered Aircraft Automation.* NASA Tech. memo N° 103885. Moffett Field, CA : NASA-Ames Research Center.

Booth, P. (1989). *An Introduction to Human-Computer Interaction.* Lawrence Erlbaum Ltd.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction.* Lawrence Erlbaum Associates, New Jersey.

Cook, R. I., Woods, D. D., & Howie, M. B. (1990). The natural history of introducing new information technology into a dynamic high-risk environment. In *Proceedings of the Human Factors Society, 34th Annual Meeting.*

Dix, A., Finlay, J., Abowd, G., & Beale, R. (1993). *Human-Computer Interaction.* Prentice-Hall International. Cambridge.

Erceau, J., & Ferber, J. (1991) *L' Intelligence Articielle Distribuée.* La Recherche. Volume 22. Juin.

Flores, F., Graves, M., Hartfield, B., & Winograd, T. (1988) *Computer Systems and the design of organizational interaction. ACM Transactions on Office Information Systems,* 6(2), 153-172.

Hutchins, E. L., Hollan, J. D., & Norman, D. A. (1986). Direct Manipulation Interfaces. In D. A. Norman and S. W. Draper (Eds.), *User Centered System Design,* pages 87-124. Lawrence Erlbaum Associates, New Jersey.

Kay, A. (1990). User Interface: A personal view. In B. Laurel (Ed.) *The Art of Human-Computer Interface Design.* Addison-Wesley Publishing Company.

Kieras, D. E., & Polson, P. G. (1985) An approach to formal analysis of user complexity. *International Journal of Man-Machine Studies,* 22:365-394

Laurel, B. (1990). Interface Agents: Metaphors with Character. In B. Laurel (Ed.) *The Art of Human-Computer Interface Design.* Addison-Wesley Publishing Company.

Lenorovitz, J. M. (1990). Indian A320 crash probe data show crew improperly configured the aircraft. *Aviation Week & Space Technology, 132 (6/25/90),* 84-85.

Maes, P. (1989). *How to de the right thing.* A.I. Memo N° 1180. Massachussets Institute of Technology, Artificial Intelligence Laboratory. December 1989.

Malone, T. W., & Crowston, K. (1990). What is Coordination Theory and How Can It Help Design Cooperative Work Systems ? *Proceedings of CSCW 90 conference.*

Masson, M., & De Keyser, V. (in press). Preventing Human Error in Skilled Activities trough a Computerized Support System. *Proceedings of HCI International '93, 5th International Conference on Human Computer Interaction,* August 8-13, Orlando, FLO.

Miller, G. A. (1956).The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81-93.

Moll van Charente, E., Cook, R. I., Woods, D. D., Yue, L., & Howie, M. B. (1992). Human-computer interaction in context: Physician interaction with automated intravenous controllers in the heart room. *Proceedings of the Fifth IFAC/IFIP/IFOEA/IEA Symposium on Analysis, Design and Evaluation of Man-Machines Systems.* De Hague, the Netherlands.

Monnier, A. (1992). *Rapport préliminaire de la commission d'enquête administrative sur l'accident du Mont Sainte* Odile du 20 janvier 1992.

Newell, A. & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, N.J. : Prentice Hall.

Norman, D. A. (1986). Cognitive engineering. In D.A. Norman and Draper S.W. (Eds) User Centred System Design: *New Perspectives on Human-Computer Interactions.* Lawrence Erlbaum Associates, Hillsdale, NJ.

Norman, D. A. (1988). *The Design of Everyday Things.* Double Day Currency, New York.

Norman, D. A. (1990). The "problem" with automation: Inappropriate feedback and interaction, not "over-automation." *Philosophical Transaction of the Royal Society of London,* B327.

Payne, S. J., & Green, T. R. G. (1986) Task-Action Grammars: a model representation of task languages. *Human-Computer Interaction, 2.* (2):93-133

Onken, R. (1992a). New Developments in Aerospace Guidance and Control: Knowledge-based pilot assistance. *IFAC Symposium on Automatic Control in Aerospace.* 8-11 September 1992, München.

Onken, R. (1992 b). Pilot intent and error recognition as part of a knowledge based cockpit assistant. *AGARD GCP / FMP Symposiom*, Edinburgh, 1992.

Reason, J. T. (1987). Generic Erro-Modelling System (GEMS): A Cognitive Framework for Locating Common Human Error Forms. In J. Rasmussen, K. Duncan and J. Leplat *New Technology and Human Error.* John Wiley and Sons Ltd, UK.

Reason, J. T. (1990). *Human Error.* Cambridge : Cambridge University Press.

Sarter, N. B., & Woods, D. D. (1991). Situation awareness: A critical but ill-defined problem. *The International Journal of Aviation Psychology, 1*, 45-57.

Sarter, N. B., & Woods, D. D. (1992 a). Pilot interaction with cockpit automation : Operational Experiences with the Flight Management System. *The International Journal of Aviation Psychology, 2* (4), 303-321. Lawrence Erlbaum Associates, Inc.

Sarter, N. B., &Woods D. D. (1992b). Pilot interaction with cockpit automation: an experimental study of Pilots' Model and Awareness of the Flight Management System (FMS). CSEL Technical Report N° 92-J-11. Submitted for publication.

Sarter, N. B., & Woods, D. D. (1993). "How did I ever get into that mode?" Mode Error and Awareness in Supervisory Control. CSEL Technical Report . Submitted for publication.

Schneiderman, B., et al. (1991). User Interface Strategies '92. Video tape courses. Instructional Television System. University of Maryland.

Sheridan, T. B. (1988). Task allocation and supervisory control. In *Handbook of Human Computer Intaercation.* M. Helander (Ed.). North-Holland, Amsterdam, NL.

Simon, H. A. (1957). *Models of Man.* New-York : Wiley, USA.

Sullivan, J. W., & Tyler, S. W. (1991). *Intelligent User Interfaces.* Addison-Wesley Publishing Co (ACM Press), Reading, MA, 560 p.

Van Daele. (1992). *La réduction de la complexité par les opérateurs dans le contrôle des processus continus. Contribution à l'étude du contrôle par anticipation et de ses conditions de mise en oeuvre.* Thèse de doctorat. Université de Liège, Belgique.

Wiener, E. & Curry. (1980). *Flight Deck Automation: Promises and Problems.* Moffett Field, CA : NASA TM 81206, June.

Wiener, E. (1989). *Human factors of advanced technology. ("glass cockpit")transport aircraft* (NASA Contractor Rep. N° 177528). Moffett Field, CA : NASA-Ames Research Center.

Woods, D. D. (1993). The Price of Flexibility. *Plenary paper in Proceedings of International Workshop on Intelligent User Interfaces.* W. Hefley and D. Murray (Eds). ACM, January.

Woods, D. D., Cook, R. I., & Sarter, N. (1993). *Clumsy Automation, Practitioner Tailoring and System Failures. Cognitive Systems Engineering Laboratory Report*, The Ohio State University, Columbus OH, prepared for NASA Ames Research Center.

Conclusion

PRECEDING PAGE BLANK NOT FILMED

PAGE 346 INTENTIONALLY BLANK

Human Factors Certification of Advanced Aviation Technologies: Overview

[illegible]

V. David Hopkin

Independent Human Factors Consultant

Introduction

This Workshop on certification had several origins. Among the most influential were the previous meeting on verification and validation which raised some certification issues directly, the inclusion in the United States National Plan for Aviation Human Factors of certification issues which have apparently not subsequently been pursued, and the recognition that certification is a job and therefore the application of human factors principles to it should offer prospects of benefits. Many of those at this Workshop brought broad and relevant knowledge to it but only a few had direct previous experience of applying human factors to certification because of the paucity of such work.

Human factors as a discipline can relate to certification in two distinct ways. In the first, existing certification processes provide the starting point and the objective is to apply human factors data and principles to improve the products of certification processes without significantly changing the certification processes themselves. In this application, the human factors specialist would work mainly as a member of an interdisciplinary team. In the second relationship, certification is considered primarily as human work, and human factors considers its methods, measures and processes in terms of their effectiveness as judged by their matching with known human capabilities and limitations. In this application, human factors is applied directly to certification, and the human factors specialist therefore would tend to be working independently of an interdisciplinary team. Both of these approaches were considered during this meeting, but most emphasis was on the first.

This is an overview rather than a summary. Its main focus is on broad human factors implications of certification that appear to be identifiable at the present time. In some instances, significant progress could apparently be made by examining the possible relevance to certification of human factors recommendations applied in other contexts.

This overview has four themes. The first considers some of the characteristics of existing certification which appear to have human factors implications. The second specifies some human factors issues that would be associated with certification processes. The third discusses briefly some characteristics of human factors as a discipline that would affect attempts to apply it to certification. The fourth mentions a few topics which appear to be relevant to any application of human factors to certification but were not discussed in detail during the Workshop.

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PRECEDING PAGE BLANK NOT FILMED

PAGE 348 INTENTIONALLY BLANK

Some Characteristics of Certification with Human Factors Implications

The existence of certification processes suggests a perceived need shared by those concerned with what is certified and by many other interested parties who desire to have some checking of advanced aviation technologies prior to their operational use. Human factors is one of many disciplines contributing to such checking processes, but it differs from all other disciplines insofar as part of its expertise covers the reasons why people may need the reassurance of certification and also the forms of certification that are likely to be most efficacious in providing reassurance.

One aspect of reassurance is that the certification processes need to be seen to be independent. Human factors as a discipline can contribute by suggesting conditions that would have to be met for certification to be generally conceded to be independent, and by explaining how judgments of independence are made, including both rational and irrational influences.

Whether certification is best treated as an iterative activity during system design and procurement or as an end product has been extensively discussed. If certification occurs during system evolution then this raises issues about how it should relate to other concurrent activities of testing, assessment, checking, auditing, verification, and validation also taking place during system procurement. If certification is an end product, it may be viewed by those concerned with system development and procurement as a hurdle or barrier that has to be overcome. If certification is construed as a means of excluding people, practices or equipment, the addition of human factors as a further discipline included in certification processes would tend to increase the role of certification as a barrier. Certification would then exclude on human factors grounds additional people or practices formerly acceptable on all other grounds, without encouraging the inclusion of people or practices already legitimately excluded on other grounds.

It has been claimed (and was extensively discussed) that certification is a guarantee that a minimum or standard has been achieved, that requirements have been met or that a predetermined procedure for testing or verification has been followed, and that the essential purpose of certification relates to this defined minimum but does not go beyond it.

An important perspective in dealing with advanced aviation technologies concerns the generally high levels of safety of existing aviation systems. Although they are not perfect and may never become so since they are designed and operated by humans, existing aviation systems are nevertheless very safe, and compare favourably with many other kinds of system. Serious incidents are rare enough to be newsworthy. To what extent this safety record is wholly or partly attributable to certification processes is not easy to determine.

In many instances, the importance of certification and its effectiveness as an ultimate safeguard may depend on its legal status. In extreme cases, the certification of an aircraft or a person may be withdrawn. If it is, that aircraft does not fly and that person does not practise. A significant aspect of certification is therefore as a threat with legal backing. Many ascribe overriding importance to it for this reason.

Some Identifiable Human Factors Issues in Certification Processes

With any test or assessment procedure which has to be passed as a condition for acceptance, certification shares the characteristic that its existence inevitably changes the nature of that to which it is applied. In other words, certification changes what is being certified. Whatever is included in the certification process thereby acquires importance; whatever is excluded is diminished, often to the point of insignificance. In some contexts, certification can represent the culmination and ultimate judgment of all other processes in that if certification is not granted the other processes are nullified. This may discourage the expenditure of effort on any processes not included in certification, no matter how vital or significant they may be. It raises the issue of how to make certification sufficiently comprehensive to avoid this trap of concentrating resources exclusively on those aspects of the system subject to it. Considerable human factors experience and evidence can be applied to predict and explain how certification as a process may change whatever it is applied to, to advise on safeguards against this, and to try and ensure that certification is fully comprehensive and does not omit anything of vital significance.

If there is a tendency to think of the achievement of certification as an end in itself, an unacceptable corollary may be that the achievement of certification seems to remove any subsequent need for critical examination of what has been certified. Certification therefore could, in some circumstances, become an excuse for subsequent inactivity.

Human factors evidence can help to explain how certification has evolved and why current certification processes take their present form. The application of human factors to certification can be extended to consider the adaptability of system processes to novel situations, to estimate how recoverable they are if the system fails, and to judge the extent to which the need for certification originated in human variability, with the consequence that reductions in human variability should affect the nature of and need for certification processes. It is reasonable to expect that many existing human factors data, practices, methods and procedures should be applicable to certification processes, though this expectation requires verification.

There has been extensive discussion of the possible role of certification in the detection of resident pathogens in complex systems wherever certification is applied. The concept of resident pathogens has proven useful in explaining particular incidents where the causes have been latent and partly predictable in terms of retrospectively recognised deficiencies in system designs, procurements and procedures. Specific incidents in specific circumstances then trigger these resident pathogens. The concept has provided a rationale for ensuring that incidents of a particular type do not recur; but as a concept it may be much more useful for providing explanations of incidents that have occurred than for predicting incidents that will occur. When systems become very complex and include many advanced aviation technologies and their interactions, they must contain thousands of resident pathogens, many of which may never appear at all, and most very rarely and perhaps only after a long time. The potential number of pathogens may be too large to be dealt with comprehensively by any certification process which treats them individually instead of in categories. Therefore, an issue that arises is whether the concept of resident pathogens is as appropriate when applied to certification as it is when applied to incidents and their investigation.

An essential aspect of any human factors approach to certification would seem to be to determine how much existing certification processes actually add to safety, to check that no certification process impairs safety, and to recommend possible modifications to the certification process that could enhance safety still further. It may be that the existence of independent checks in the form of certification is what is important rather than the detailed nature of the certification procedure itself. If so, satisfactory checks could in principle take many forms, and their forms would not be critical. They could be as comprehensive as certification is intended to be; they could aim to be complete, which is at best a major undertaking and perhaps ultimately impossible to achieve; or they could be arbitrary in the sense that a series of checks would be made but the particular checks chosen would be selected at random or according to other criteria not known beforehand.

Regarding certification, a human factors issue is how much the existing certification processes actually contribute to safety and whether it is their detailed nature or the fact that they are unavoidable which has the main effect. Alternative explanations for safety may be in terms of the professionalism of those employed in aviation coupled with their training, which together are primarily responsible for the high standards achieved and the consequent high levels of safety, rather than the certification which may not necessarily add much. A corollary is that changes in certification might not lead to major system changes, but that large changes in the attitudes of those who work within systems to their profession or to the systems might have major consequences for safety.

The acceptability of certification processes by those who apply them and by those to whom they are applied would seem to be essential preconditions for their practicality and success. It is difficult to impose processes which people do not agree with and are not prepared to follow. This implies that in some contexts the notion of certification may have to be promulgated and positively advocated in terms of its benefits.

Certification procedures have not only to be devised and validated but they have to be taught. Whatever else the certification depends on, it should not be on the whim of the particular individual or group tasked with the conduct of the certification procedures. This implies a definition of the knowledge required for effective certification, appropriate training, verification that the knowledge is present and can be applied by each individual, and some means of establishing the reliability with which the certification procedures are followed. The issue of the teachability of certification therefore has to be addressed and this should rest on human factors learning principles.

There seemed to be general agreement that certification must not only be independent but be seen to be independent, and that certification processes should be open and available for inspection and part of the public record so that they are seen to be fair and it is possible to confirm by verification that they have all been followed correctly. This seems obviously sensible and desirable and a matter of common sense, but human factors evidence suggests that while it should be implemented as the main policy it is unlikely to prove sufficient. In many contexts, it has been found necessary in the interests of safety to provide some form of confidential incident reporting system whereby actions or omissions which occurred or could have occurred, or which were observed, or which the system did not prevent but could have been potentially dangerous, can be reported anonymously in the interests of safety and efficiency, without attributions of blame and possible loss of employment. Much of this evidence would not be forthcoming if it all had to be part of the public record. This issue is likely to arise in certification whether there is a human factors input to it or not. Precedents from other contexts suggest that certification is not likely to be an exception to this recognised need, and therefore some additional channels for gathering information outside formal public records may also be required.

Some Issues Concerning the Application of Human Factors as a Discipline

The world of certification does not owe human factors a living. The burden of proof of the value of human factors rests squarely on those within the discipline of human factors itself to prove its worth, its applicability, its independence, and its cost-effectiveness. It is reasonable for others to demand tangible evidence that human factors can be beneficial when applied to certification.

A problem encountered elsewhere when human factors is applied is likely to recur with certification. This concerns the general utilisation of what is already known and the practical application of existing evidence. One of the means by which human factors as a discipline can prove its worth is to show that existing evidence can be applied to good effect, but an educative process will be needed so that others know that this evidence exists and of what it consists.

Human factors as a discipline has the problem of great variability in the strength of the evidence on which its recommendations depend. For many traditional ergonomic applications, existing data of known high validity and generalisability could be applied to certification processes; but for other aspects of certification, particularly those concerned with higher cognitive functions, a much lower degree of certainty and confidence often applies to the human factors recommendations that can be made. Nevertheless, even the poorest of these recommendations remains a substantial improvement over the ill-informed options, guesses or random choices which are the main alternatives. Human factors specialists tend to view this problem as peculiar to their own discipline whereas it is often widespread in others. It does, however, mean that recommendations usually have to be accompanied by further information on the appropriate level of confidence in them. This in turn suggests that much of the advice must come from human factors specialists themselves who are most likely to possess this further essential information about the strength of the evidence underpinning their recommendations. Where the value of the data is not high but the data are not valueless either, the question arises of how to treat human factors recommendations in a certification context when they rest on somewhat tentative evidence.

A related point is that since certification is a real activity, it is essential to reach practical compromises whenever the requirements of different disciplines related to certification conflict. The origins of such conflictions have to be ascertained in the interests of reaching working compromises. For this reason also the human factors specialists involved need not only a knowledge of the human factors evidence but also a knowledge of its strengths, its generalisability, the conditions attached to it, and the extent to which it can be compromised without being invalidated. This point raises the issue of how well the human factors profession is equipped to furnish this sort of advice and of whether those giving it would need some special training in certification processes and in the applicability of human factors evidence to them.

A related point concerns the certification of the people who would apply human factors to certification, and the appropriate training for them. The efficacy of the human factors contributions to certification must not rely heavily on the characteristics of the individual specialist who makes those contributions. There are problems in the training of human factors specialists for certification work in defining required knowledge, testing that all possess it and can apply it, and ensuring uniformity of training and professional standards. Perhaps there is a requirement for an official compendium or textbook of human factors data applicable to certification. An additional contribution of the human factors specialist in relation to certification should be to ensure that what has been defined as comprising the certification processes is capable of being taught to the people who will carry them out so that certificators all reach an agreed required standard and have a comparable understanding of what they are doing. This is the familiar issue of teachability.

The role of the human factors specialist can also be considered in terms of the status of the human factors evidence and whether it should be advisory or mandatory. The ultimate responsibility for the certification must rest with the professional employed as a certification expert; yet human factors has a role equivalent to that of several other disciplines where it may be occasionally expected that a human factors requirement takes precedence over all others. However, human factors requirements are usually taken into account without overriding others. Certainly, those making human factors contributions must be perceived by others as maintaining their independence. A fundamental problem arises because someone has to pay for the human factors time, effort, and resources devoted to certification. It is natural that whichever organisation pays should claim some say in how the money is spent. This is a recurrent problem not confined to certification but arising in many other human factors and safety contexts whenever the findings are in potential contradiction with the policies, practices or wishes of those providing the funding.

There is also a problem in deciding how much effort and money should be devoted to the application of human factors to certification. One kind of policy is to select a percentage of the total certification budget for human factors work. This is better than no policy or than no human factors effort at all, but it is not the best means of insuring good high quality human factors contributions. Some certification procedures require far more human factors resources than others so that any chosen percentage will be too high for some applications and insufficient for others. Additionally, funding alone does not guarantee quality of effort or of its deployment.

Before the Workshop was held, there seemed to be a presumption that it might be difficult to gain acceptance for human factors in certification and to establish a constructive role for human factors within certification processes. This did not seem to be as contentious an issue as was expected, although it could be claimed that those who would accept an invitation to attend the Workshop were already predisposed towards human factors and might not in this respect be representative.

Under-Emphasised Issues

For part of the Workshop, the participants worked in six groups, considering respectively why, what, who, where, when and how human factors might be applied to the certification of advanced technologies. It is surprising perhaps that no group addressed the question of whether it should be applied. Although this seldom arose as a problem during the Workshop it does not mean that the question can be ignored or was ignored, since in a sense it is the most obvious question of all. But there did seem to be widespread agreement that there was a role for human factors in certification processes.

There was remarkably little discussion during the workshop about research, and particularly about any supporting laboratory work for the application of human factors to certification. Perhaps this is simply a reflection of contemporary funding priorities, or it may even be thought that laboratory studies do not offer a particularly profitable research approach to human factors and certification. However, the lack of suggested research at a meeting of this kind is almost revolutionary and a comparatively recent development, and this alone constitutes grounds for questioning whether the apparent absence of any identified need for supporting research really does offer the optimum approach.

Information System Certification: Purview, Perspective And Projections

Anthony Debons[1] & Esther H. Horne[2]

0117

[1]University of Pittsburgh
[2]Research Institute for Information Science and Engineering(RIISE)

Purview

Background

The objective of the present workshop on certification is to ensure that human safety is maintained in Air Transportation Systems. During the past twoto three decades, several air transportation events have led to loss of life, material resources and national prestige. The first NATO Advanced Study Institute, held in Bad Windsheim, West Germany in 1986, raised the question as to the role of information system failure in such instances. It also raised the question as to whether a methodology was available, or could be developed, by which a body of experts trained in analysis and information systems design could be marshaled to determine how the breakdown of the information exchange system led to the overall system failure. The participants of that institute detailed a number of methods that could be applied. But the how, when and who of the application of these methods in determining information system failure were not addressed in sufficient depth to recommend the structure of a program to achieve this objective.

Whether performed by humans, machines or a combination of both, data processing is a fundamental component of all artificial (human augmenting) information systems. The question of the role of automation in the performance of an artificial information system supporting air transportation was addressed at the second NATO Institute held in Acquafredda di Maratea, Italy in 1990. There is a paucity of incidents in which automation can be considered to have directly caused an air transportation event leading to loss of life and property. But it was clear from deliberations by participants that the human-machine interface presented a plethora of problems. Humans are not perfectly performing information systems. Basically, that is why artificial augmenting systems exist. The extent to which machinery can be interlaced with human motor and cognitive capacity in order to assure a degree of performance that obviates error and system breakdown remains a challenge to engineers, psychologists, system designers and organizational managers. The third NATO Institute held in Portugal in 1992 questioned the extent that our present data, information and knowledge about this interface can be verified and validated.

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

Fundamental issues of Present Workshop

The area of interest in this present workshop was certification of human factors as applicable to air transportation systems. Problems of definition and general focus are crucial for maintaining an "even keel" in the attempt to formulate a practical perspective for sponsors. Moreover, definition and focus are a matter of some importance to the work reported in this paper. Analysis of previous deliberations was undertaken to provide participants with an integrated view of theories, propositions and views on matters related to certification.

Several "givens" can establish the framework for the task of organizing the intellectual substance of previous deliberations:

1. No certification process can exist without some rule or standard on which the *process of certification can be undertaken.*

2. Certification and certainty are correlated to the extent that the *judgment* that is part of certification rests on the degree of verification and validation of whatever human factors concepts, propositions and principles apply to the specific case and which can be substantiated by theory, experimentation and field practice.

3. The *object* for which certification is executed requires identification if the certification process is to achieve relevance. For example, *certifying the procedures* applied in establishing operational status foraircraft engines is one thing, while *certifying the persons* in executing the procedures is another. Certifying the *human factors principles* that apply in the checking and which insure the mechanical status of the engine still another matter.

Methodology and Applications

In this paper we attempt to organize the deliberations of the NATO Advanced Study Institute held in Portugal in 1992 so as to serve the needs of the participants of the present workshop. The issues of verification and validation of human factors principles in the analysis and design of air transportation systems are considered relevant to the major concern of the present work shop, namely, certification.

We project the *cognitive need of the participants of present workshop to be synthesis.* Synthesis is one of several cognitive functions and is defined operationally by Bloom (1956):

> "...reorganize material to generate new patterns or structures...analyze trends, clusters, relationships, etc. and produce predictions, projections, reports...generate new or creative...outputs"

We plan to support and aid the cognitive function of synthesis by the execution of the following steps that are part of the knowledge counseling method (Debons, 1992):

1- reading the text of the Portugal proceedings in detail

2- Identifying major ideas, propositions, and theories for each of the papers presented at the Advanced Study Institute convened in Portugal in 1992.

3. Obtain from each director of the present workshop an assessment of the major concepts relevant to certification prior to the convening of the workshop.

4. Derive a mapping of the concepts (relationship of each item to each other; establish clusters, etc.).

5. Re-read proceedings in detail based on steps 1-4.

6. Generate proposals on certification in accordance with work shop goals and objectives.

We have applied our analysis to *three major fundamental dimensions of certification* based on which the data from the Portugal NATO ASI could be organized and studied.

- Capacities (performance)
- Measurement
- Information Flow

There were 53 formal presentations with a total of 679 pages of text. Each idea, thought, proposition, speculation, and thesis considered representative of the participants' position on verification and validations of complex human systems were aggregated to conform to these dimensions. These are included in Tables 1 a, b and c.

In addition, each proposition, theory, principle, and idea was cited on individual index cards, including the respective author's name and page number where the citation could be found. There were 122 citations so identified. These cards were then given to each of the work shop directors at a scheduled meeting. They were asked to examine the cards individually and, if they so desired, add new items on individual cards to the set. Based on Paul Stager's assertion "The quest is for certainty", the directors were asked to indicate for each item the extent of certainty that could be applied based on available Knowledge supporting the item in question (given experimental data, operational reports, etc.; that is, "very good", "good" or "poor". The data for two respondents based on this criteria are indicated in Tables 2 and 3. The third respondent chose to undertake the task, given the same criteria differently (Table 4).

Subsequently, certainty was given the added reference to utility. A high degree of certainty warranted high utility for certification; good certainty – marginal utility; and little or no certainty – poor utility. For example, experiments and other studies on dark adaptation have confirmed that human eye dark adaptation reaches an asymptote within 15 minutes following exposure to darkness, depending, of course, on the individual eye's previous exposure to light Thus the human factors principle on dark adaptation is substantiated by a high degree of certainty given the available data and, as such, provides high utility to the certification process. On the other hand, our level of understanding of human attention does not offer this level of certainty. The phenomena of attention is governed by a number of confounding factors. When attention is considered an important event element,the certification of human factors principles pertaining to attention could be considered, at best, marginal or questionable (poor) utility.

The directors were then asked to place the respective items (indicated on the index cards) in relation to each other This provided a map showing the relationship between the items, establishing vectors and item clusters (see Figures 1 a, b). One of the directors chose to show the relationship differently (Figure lc).

Table la
Verification/Validation Human Factors in Air Transportation System: Human Capacities (Performance)

Author(s)	**Citation(s)**
Dubois-Gaussin	Capacities influenced by mental imagery
Schand	Impact of stress
Sauveur	Eight(8) areas of importance to enhance human-machine capacities
Tonner-Kalmbach	Human response time as a factor
Rottmann & Wattler	Simulation of capacities
Wilkinson (*)	Details classification concept
Gibson	training of capacities
Jorna	Impact of automation (both broadens/increases criticality of human increasing task difficulty)
Sanders & Roelofsma	Need for general model for extent of human performance limits methodological issues
Logie	Working memory critical tools for engaging high level cognitive needs
Kolrep	Basic characteristics of skilled memory effect of change in task
Hancock	Shared performance: allocation of strategies
Foster	Resiliency (true vulnerability of technological social systems cannot always be predicted
Hollnagel	Cognitive reliability; knowledge intensive functions
Reason	cognitive errors in accidents
Hunt:	Pilot competency Flight crew Performance management

*Critical paper for certification

Table 1b
Verification/Validation: Human Factors in Air Transportation System: Human Capacities -- Measurement

Pitts, Kayten & Zalenchak:	Systems are evaluated late in development cycle, thus finding problems are costly.
Harwood:	Hard to develop criteria for V & V Continuously changing environment ecological demands performance criteria not obvious
Wood & Sarter:	Context bound approach exams the system as its functioning System design flaws abound
Stubler, Roth, & Muman:	There should be a framework for evaluating systems; the dynamic properties of systems cause problems
Rosness:	Not possible to identify and access all possible sequence of events or states which may occur in system
Reason:	No established means for assessing the cultural properties of an organization
Andrews:	A number of techniques for achieving V&V are available(task analysis;human reliability assessment
Jorna:	No matrix for workload
Smolensky & Hitchcock:	There is a measurement error in establishing mental work load
Jack:	Users as certifiers
Hancock:	Ecological approach: include total environment-consistent with information flow concept
Stager:	Ecological validity- representativeness of subjects, of variables and of setting
Lane	Use prototyping whenever possible
Toner & Kalmbach	detailed training evaluation
Smoker:	Quantification of risk
Pitts, Kayton, & Zalenchak:	Human performance measurement available
Shaffe:	Video documentation of processes (black box concept)

Table 1b (continued)

MEASUREMENT	
Westrum:	Obtain reaction from users to system functions
Svensson:	V & V is a continuous process
Chavez:	V & V a continuous process Difficult to measure value system Acknowledge Cultural Factors
Dujardin:	Provides 7 steps in understanding ATC process
Jack:	User interface model
Schaad:	Tests are needed for V & V
David:	Assess different approaches to determine which works for system Analysis *
Rosness:	Difficulty in assessing and communicating uncertainty
Plant:	Need to develop a V & V knowledge base system
Hollnagel:	Recommend point to point analysis (cause-effect) Evaluation of real systems not enough Dynamic simulations important
Baker:	Incident investigation is recommended

* Suggested critical paper for certification

Table 1c

Verification/Validation Human Factors in Air Transportation System

INFORMATION FLOW

Bangen:	Need detailed exposition of information flow
Pitts, Kayten, &Zalenchak:	Information transfer detailed
Folles & Volden:	Integration of information
Or:	Information Flow requirements
Zakharova:	Information Flow model
Rottmann & Wattler:	Impact of Information on workload
Wieringa & Stassen:	Continuous evaluation of system state(non-linear behavior of components
Clare:	Role of Human assimilation of information Yourdon Methodology (core)
Westrum:	Flow must be timely, detailed pertinent and honest Top management is advocate of information flow
Hunt:	Require command-control of processes in system

Table 2

Item Numeration		Item Desegnation	Rating for Certification
10		Desingers as certifiers	Good
11		Users as certifiers	Poor
12		How, what, where	Poor
13		Lawyers	Good
14		Users as designers	Poor
20		Differences in assumptions	Poor
21		Differences in knowledge systems	Poor
22		Modeling	Good
221		Test	Good
222		EATPUTr	Good
223		Other theoretical models	Good
23		Simulation	Good
24		Experimentation	Good
30		Legal analysis	Good
		Cost-benefit analysis	Very good
32		Physiological measures	Good
33		Human personality assessment	Poor
34		Data flow analysis	Good
35		Post accident analysis (failure) investigative methods	Very good
50		Human Factors (HF) engineering	Good
51		Cognitive engineering	Poor
511		Cognitive performance measure	Poor
512		DM/PS performance	Good
513		Knowledge of outcomes on performance	Good
514		Acceptance of low certainty outcomes	Good
515		Theory of working memory	Good
52		Knowledge engineering	Poor
53		Operations research	Poor
54		Workload assessment	Good
55		Psychomotor performance	Good
56		Work physiology	Good
57		Anthropometric	Very good
Summary:	A	Certifiers	Good 2; Poor 6
	B	Modeling/Experimentation	Good 6; Poor 2
	C	Analysis	Very good2; Good 3; Poor 1
	D	Cognitive engineering	Good 5; Poor 2
	E	Workload assessment	Very good1; Good 4; Poor 2
26		HF practitioner as certifier	Poor
27		Users as designers	Poor
7		Group interaction in teams	Good
6		Physiological measure	Good
5		Theory of working memeory	Poor
4		Application of cognitive science	Good
3		Knowledge engineering	Good
2			

Table 3

Item Cluster for Respondent 2

Item Numeration	Item Desegnation	Rating for Certification	
26	HF practitioner as certifier	Poor	A
27	Users as designers	Poor	
7	Group interaction in teams	Good	
6	Physiological measure	Good	
5	Theory of working memeory	Poor	B
4	Application of cognitive science	Good	
3	Knowledge engineering	Good	
2	Differences in knowledge systems	Good	
1	Theories of cognition	Good	
18	Application of current data	Good	
17	Modeling	Good	
16	Experimentation	Good	C
21	Simulation	Good	
20	Training	Good	
19	Selection	Good	
22	Automation	Poor	
23	Human-Computer interface	Good	
24	Research practices	Poor	
25	Quantitative/qualitative data	Poor	
8	Individual difference assessment	Poor	
9	Workload Assessment	Pooor	
10	Situation awareness assessment	Poor	
11	Human reliability assessment	Good	E
12	System assessment	Poor	
13	Data flow analysis	Good	
14	Cost benefit analysis	Good	
15	System reliability assessment	Good	
32	Cultural factors	Poor	
31	Organization/Management Philosophy	Poor	
30	Policies/Procedures	Good	F
29	Data Management	Poor	
28	R&D Implementation	Good	

Summary:	A	Certifiers	Poor 2
	B	Concept Assessment	Good 6; Poor 1
	C	Methods	Good 6;
	D	Practices	Good 1; Poor 4
	E	Assessments	Good 4; Poor 4
	F	Cultural factors	Good 2; Poor 3

Table 4

Item cluster for respondent

Item Designation	**Rating for certification**
Cost benefit analysis	Very Good
Data flow analysis	Very Good
Knowledge engineering	Very Good
Physical measures	Very Good
Users as designers	Good
Modelling	Good
Simulation	Good
T.E.S.T. Model	Good
Experiment	Good
Human reliability assessment	Good
Workload assessment	Good
Subjective assessment	Good
Working memory	Poor
Theoretical knowledge systems	Poor

Perspective

1. Tables 3, 4 and 5 present the data obtained on the mapping of concepts from the respective directors. Of the 16 items that were included on the original list of items provided to the directors, the respondents added 16 others with a total of 32 items (Tables 2 and 3). The third respondent maintained the 16 items included in the initial tabulation (Table 4).

2. The three respondents clustered the items basically into three groups; namely, *personnel, measurement and culture.* The central position of culture in the mapping scheme (the relationship of each item to each other) indicates it was considered a predominant variable in any consideration of certification. The present "degree of certainty" as to the human factors principles that could be applied to culture was judged as poor. Availability and application of human factors principles on policies, practices and the implementation of research was judged as good.

3. Generally, methods of data flow analysis, human reliability, physical measures, experimentation, modeling and simulation offer the best potential for the certification of Human Factors principles. We should add however, that these methods were not without the directors' misgivings, particularly as to their application to verification and validation, and thus, suspect as to their utility for certification.

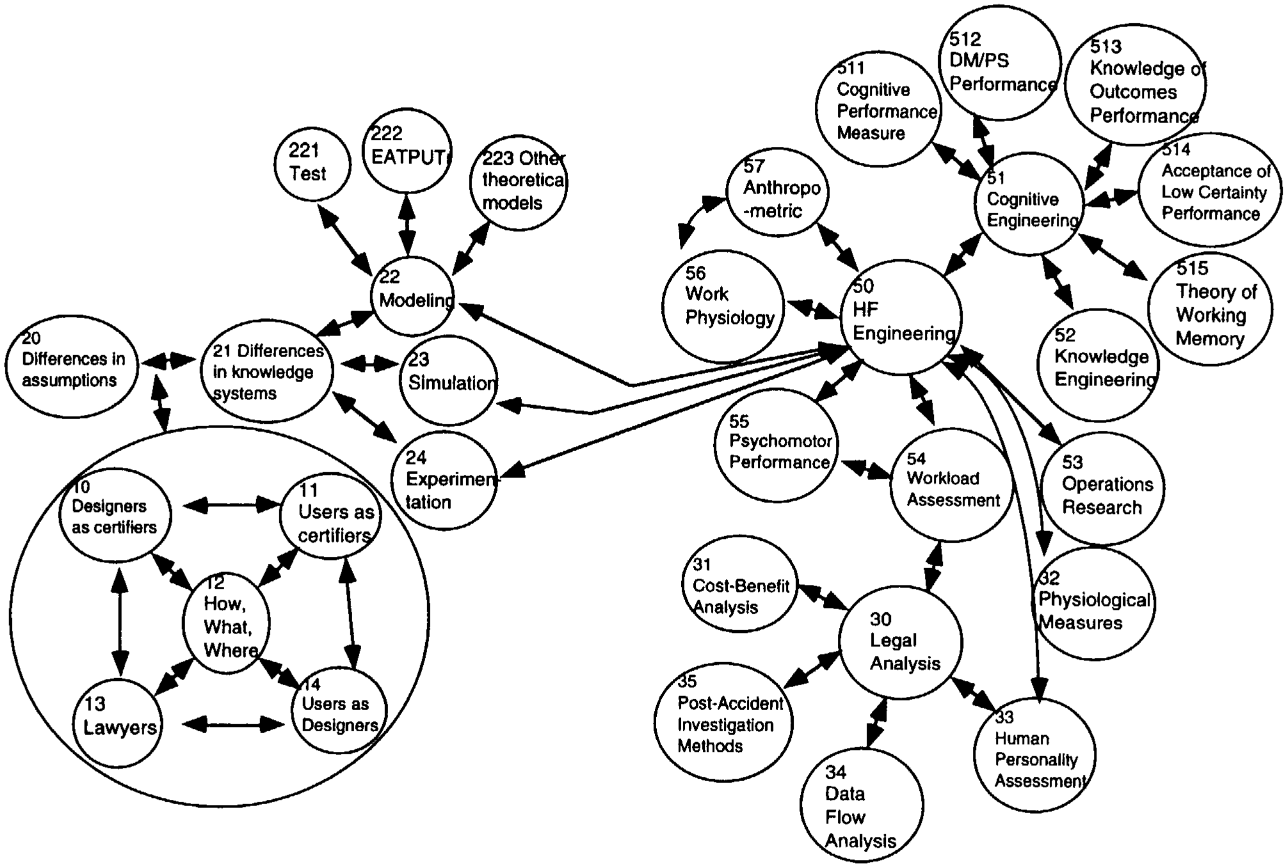

221 Test
222 EATPUT
223 Other theoretical models
22 Modeling
20 Differences in assumptions
21 Differences in knowledge systems
23 Simulation
24 Experimentation
10 Designers as certifiers
11 Users as certifiers
12 How, What, Where
13 Lawyers
14 Users as Designers
57 Anthropo -metric
56 Work Physiology
50 HF Engineering
55 Psychomotor Performance
54 Workload Assessment
511 Cognitive Performance Measure
512 DM/PS Performance
513 Knowledge of Outcomes Performance
514 Acceptance of Low Certainty Performance
515 Theory of Working Memory
51 Cognitive Engineering
52 Knowledge Engineering
53 Operations Research
32 Physiological Measures
31 Cost-Benefit Analysis
30 Legal Analysis
35 Post-Accident Investigation Methods
34 Data Flow Analysis
33 Human Personality Assessment

Figure 1a. Mapping of Concepts: Respondent 1

Projections

The central focus of the current workshop is the determination of the six critical interrogatives that constitute the basic framework of any data, information and knowledge (D-I-K) system; namely, the what, where, when, who, how and why. The D-I-K system is the lifeblood of any operational system. We can best understand the nature of a D-I-K system if we liken it metaphorically to the circulatory system in living systems (Miller, 1978). In this system, there are sensors that elicit (capture) data from the external world. The sensors provide the organism's awareness of this world. This awareness incorporates those processes that define the transmission component. It is the transmission component that moves the energy from the sensors to the brain. We identify this primarily as the central nervous system. The brain processes the data, which then makes possible the derivation of judgment that is the basis for problem solving and decision making. At this point, the motor system provides the means for the actions chosen to safeguard the integrity of the organism. This fundamentally open system is characterized by continuous feedback among the various components. As mentioned, the D-I-K system is fundamental to all humans and operational systems. But the technological (artificial) D-I-K system augments such functions for humans (Simon, 1981). When reference is made to the air transportation system, the concept of a D-I-K system is fundamental whether it pertains to humans within such systems or the technology that augments their native capacities. The model provides a framework to help understand the nature and function of symbols (data, language, etc.), the human-machine response, and transfer (movement) of such symbols throughout the system.

The present interest is centered on the certification of human factors principles that guide the design and use of advanced technologies that will govern the performance of an advanced air transportation system.

In his paper, Andrews (1993) makes clear the essential point governing certification:

> Are the techniques of validation and verification which have evolved, been developed, and are currently in application, able to identify all possible key sources or causes of error in systems and enable error resistant systems to be created or, where incidents do occur, provide for information for subsequent investigations, so that the courts can arrive at undeniably *safe and reliable* verdicts?

Andrews' thesis answers some of the interrogative:

What:	identify all possible key sources or causes of error in systems
Where:	location and conditions in which incidents occur
When:	the time that incidents that occur
Why:	enable creation of resistant systems – provide for information for subsequent investigations – courts can arrive at undeniably safe and reliable verdicts

Andrews does not address the "how" and "who" (courts are implied). It is to these two interrogative that we focus our discussion.

The proceedings on the verification and validation of human factors principles provide moderate assurance that a certification program can be formulated and implemented, given our present ability to establish standards for the analysis, design and performance of complex systems. Thus, suggested certification programs will be, by necessity, prototypical and judged for their evolutionary potential and continuous assessment.

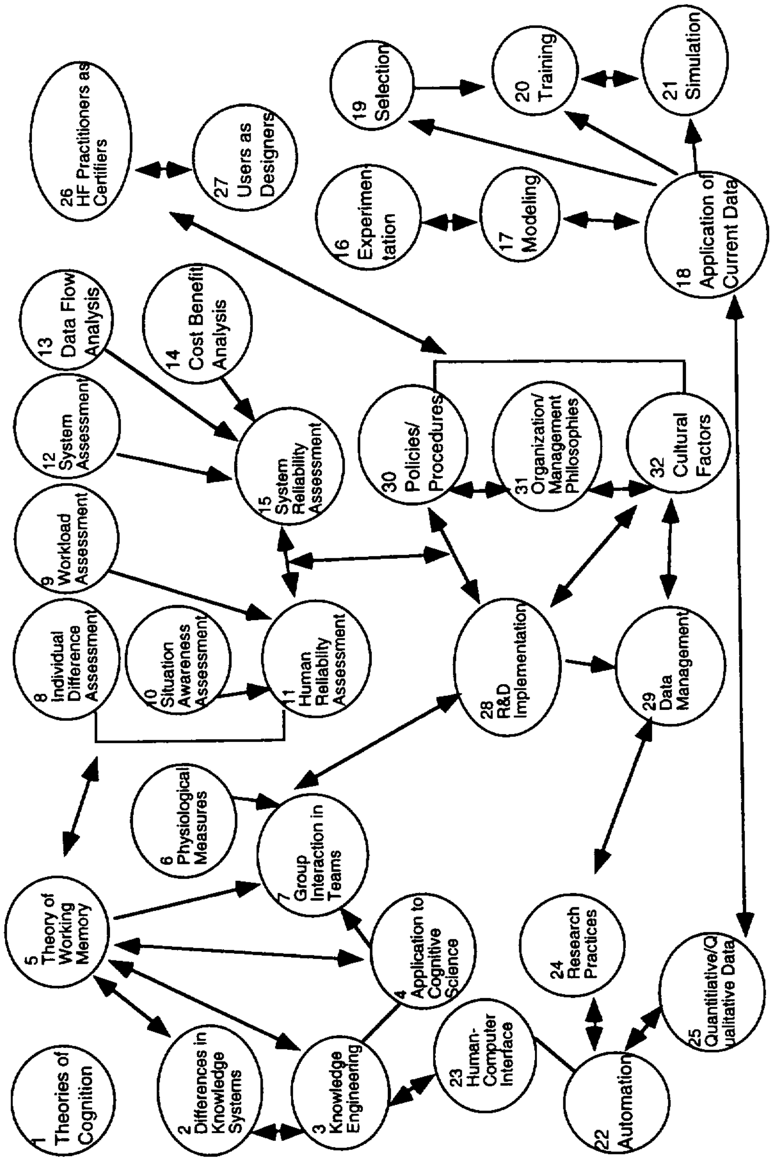


Figure 1b. Mapping of Concepts: Respondent 2

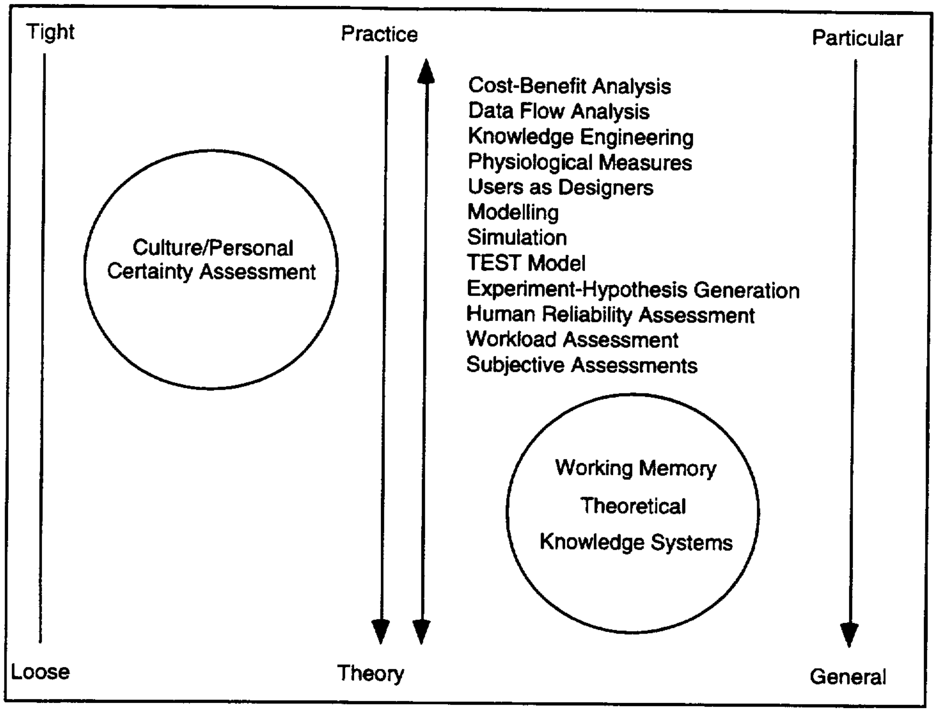


Figure 1c. Mapping of Concepts: Respondent 3.

There is precedence in Human Factors research for certification of human factors principles derived from standards. Standards make certification possible. The design of workplace arrangements and anthropometric considerations for the fabrication of clothing to meet the special demands of the environment are two instances in point, but there are others as well. As a matter of fact, the *Handbook of Human Engineering* (Van Cott, et al., 1972) conceivably is a representative compendium or compilation of such standards. The fact that standards often serve to stifle progress may be a central point for advancing that any certification program be dynamic and that the certification program drive the standards, not the converse. But there are other problems as well with standards as measures for verification and validation.

Meanwhile, certification is not forever. The underlying assumption that governs this workshop is that there will be advanced (changing) technologies for which the question of establishing the acceptable limits of both human and machine performance is the desired goal, Thus, a certification program needs to be dynamic. This means that certification requires individuals who are *trained* to deal with the eccentricities that such technologies provide and which blend with the human requirements. This narrows down the design of a certification system. The certification system can be envisaged as a supra data, information and knowledge (D-I-K) system consisting of components that are normally considered the components of such systems as we have detailed previously.

The three respondents of the present report have placed cultural factors as the center of a certification program. Examination of the literature on certification practiced in the physical and social sciences (engineering, psychology, psychiatry, medicine and law) offer a clue to the complexity of the certification issue. Germane are issues of education and training, socio-political-economic influences, testing philosophies, ethics in practice, etc. (Scheiber, 1991). Westrum (1993, in print) and others have stressed the importance of individual education and training for advanced concepts and skills, the structure of organizations and institutions, and the manner by which individuals are judged and rewarded in job performance when criteria for such reinforcements vary from culture to culture.

This paper cannot include the innumerable psycho-social technical statements and reports that are relevant and need to be addressed in the formulation and development of a certification program. What is predominant in these reports concerns the role and influence of professional groups in formulating and governing certification programs. These programs are often developed from field experiences. From these policies are established education, training, licensing, and ethical practices and many other aspects that govern the lives of professionals in their day to day work. Further, it is clear that although such programs are continuously changing as to requirements, they are not without extended debate and uncertainty. Certification is seen as more a matter of evolution than revolution.

Unquestionably, certification of human factors principle in the application of advanced technologies to complex systems must account for *technology transfer*. Technology transfer refers to the process of introducing new, advanced technologies to developing countries and cultures around the globe. Technology transfer can profit substantially by having a comprehensive process of certification that acknowledges basic differences in cultural responsiveness to technology, in general (Kaplan, 1992). Meanwhile, this raises a plethora of human factors questions for which there is a paucity of principles that can be applied.

Certification often assumes the status of legal reference; legality postulated on evidence and precedence. From the scientific perspective, evidence is a corollary to data, derived from observation and experimentation. Data are gathered based on certain rules for the conduct of scientific investigation. In the human factors domain, there is ample evidence to support the verification and validation of human performance *under certain conditions*. As is well known, affirmation of evidence is correlated to the extent of variance that is the property of all events, and the manner in which these events are treated (perceived) by humans. We are reminded of Heisenberg's uncertainty principle (1950) from quantum physics in its application to certification objectives. As said previously, certification is never forever, or absolute (p=l). Certification of principles on practices pertaining to hazards to life and property are vitally important. It is this concern that is paramount and that a certification program should be prepared to address.

Last, but certainly not least, is the importance of the Data, Information and Knowledge (D-I-K) system and the need for verification, validation and certification of the processes that underlay the system. Whereas Human Factors research stresses the architectural aspects of systems (hardware, software, human-machine interface), the point that is often missed is that performance of these aspects remains totally dependent on the data, information and knowledge flow between the various components of the system. Although the emphasis has been on the human component as a processing (cognitive) entity, it is also best to acknowledge that the cognitive response is a function of the energy (stimulus) available to the respondent. This energy is represented in the form of symbols, signals and other formulated technologies (language, number systems, etc.). Again, metaphorically, alike the measuring of blood pressure in determining the status of an individual's physical health at any point in time, methods are needed to determine the status of the data, information and knowledge flow throughout the various components of the air transportation system (EATPUT) (Debons & Horne, 1988) at any point in time (Leroux, 1992).

Recommendations

The sponsors and directors of this workshop have asked that the workshop attempt to provide some proposals for achieving certification of Human Factors applications in an Air Transportation System. Applying the Directors' guidance using the interrogative proposed, we recommend the following:

What should be certified?

The contents of the 1992 ASI conducted in Portugal do provide the basis for those elements of the Air Transportation System that can be certified. Specifically, those elements are the surveillance of the data, and the information and knowledge flow of the system. The documentation governing the flow is that which can represent a form of certification.

How should it be done?

Several steps are recommended:

1. Transcribe the technical substance of the ASI on verification and validation as part of manual that details the what, when and how the various conditions they address are to be certified. For example, how detected; periodicity conditions; mode error; data type; post stress; design flaw (Woods & Sarter, p. 133, 1992).

2. Establish a training program directed at establishing a corps of certifiers available and responsible to an independents, non profit, entity. The manual would be the basis for the training of such corps of certifiers.

3. Establish a central store for data collected by certifiers (after Rosneso, p.187, 1992/draft). The support for such a data base would come from the independent agency.

4. Initiate a research program, supported by the independent agency, that attempts to more specifically define the certification process. To the extent that the certification process would cross international boundaries, the research would attend to the role of cultural factors that could enhance or mitigate the benefits that are intended. Cultural factors that influence ATS operations are the least understood although their importance cannot be minimized (Kaplan, 1991).

5. The independent agency would convene an annual conference of certifiers. At this conference, critical issues pertaining to certification of human factors principles in Air Transportation System would be critically examined and studied. For example, discrepancies in data collection obtained from certifiers could arise. Thus, the evaluative process requires continuous surveillance and refinement (Baker, p. 281, 1992/draft).

Who should have the authority to do it ?

A specific foundation (independent agency) supported by the industrial-private interest who see the benefits of such a program to their organization (very similar perhaps to the New York Conference Board) should have the authority to certify. The sponsors of the present workshop would establish, guide and support the foundation.

References

Alexander, C. (1967). *Notes on the Synthesis of Form.* Harvard University Press, Cambridge, Massachusetts.

Andrews, C. J. A. (1993). Major Incidents, safe and reliable verdicts and the process of validation and verification. In J. A. Wise, V. D. Hopkin & P. Stager (Eds). *Verification and validation of complex systems: Additional human factors issues.*(pp.261 277).

Bloom, B.S. et al.(1956). Taxonomy of educational objectives: The classification of educational goals. *Handbook I: The Cognitive Domain.* New York: David McKay Co.

Debons, A., Horne, E., & Cronenweth,S. (1988). *Information Science: An Integrated View.* G. K. Hall. Boston, Massachusetts.

Heisenberg, W. (1950). *The Physical Principles of the Quantum Theory.*

Kaplan, M. (1991). Issues in cultural ergonomics. In J. A. Wise, V. D. Hopkin., & M. Smith (Eds.). *Automation and System Issues in Air Traffic Control.* Berlin: Springer-Verlag.

Kohler, W. (1947). *Gestalt Psychology.* New York. Liveright.

Leroux, M.(1993). Towards Cooperative tools for Air Traffic Controllers. In The Controller, the Automation and Automation, *Le Transponder*, No. 9. April, 1993. pp.40-47.

Miller, J. G. (1978).*Living Systems.* McGraw-Hill Book Company. New York.

Mintzberg, H. (1979). *The Structuring of Organizations.* Englewood Cliffs, NJ pp.35-64.

Scheiber, S. C. (1991). Certification and recertification. Special Issue: Social and Economic Influences on psychiatric education. *Psychiatric Quarterly*, Vol 62 (2), pp. 153-164.

Simon, H. A. (1981). *The Sciences of the Artificial.* 2nd ed. (Karl Taylor Compton Lecture, 1968). MIT Press, Cambridge.

Van Cott, H. P., & Kinkade, R. G. (Eds.). (1972). *Human engineering guide to system design.* Washington, DC: US Government Printing Office.

Appendices: The Group Papers and Participants

PRECEDING PAGE BLANK NOT FILMED

PAGE 372 INTENTIONALLY BLANK

The “What” Group Paper

What is already certified without human factors specific focus?

- The usability (system procedures) and safety of the system within a global environment when operated by the end-user according to an existing set of minimum requirements incorporating very little human factors design.
- The global procedures which give coherence to the general macro-system. The certification is accomplished according to international regulations. It is a prerequisite of system certification, but can change with new technologies.
- The proficiency of end-user.

For system certification, procedures are currently twofold: first, certification of the design, device, or machine, and second, licensing of any individual copy of the system. The same procedure applies for personnel: certification of a training course followed by licensing of an individual.

What should be certified with human factors focus?

Certifying the process

There are two models for dealing with the process:

1. Introducing HFE via contractual requirements by application of HFE standards and guidelines.

In reference to Taylor & Macleod (1994) and Small & Rouse (1994). We need to develop new guidelines or complete existing guidelines.

Advantages	Drawbacks
<Cheaper ?> quicker, less bureaucratic. Timely	Are we capable of writing non ambiguous guidelines for all domains related to human factors, especially cognitive activities?

2. Certification of the end-product. Global acceptance, final acceptance tests (a straw man), and ensuring the system complies with contractual requirements.

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PRECEDING PAGE BLANK NOT FILMED

PAGE 374 INTENTIONALLY BLANK

Advantages	Drawbacks
legal forced more appropriate for simple systems	poorly adapted to complex systems, more expensive, maybe impossible if things have to be changed.

Certifying HF engineers

HF engineers provide guidance on applying standards and guidelines, direct and monitor evaluation studies, interpret outcomes, provide design recommendations, and finally officially certify proper application of standards and guidelines.

It is implicitly assumed that contractors and governments must use these certified HF engineers.

Advantages	Drawbacks
Put the responsibility on the engineer. Possibly to deal with situations which are too complex to be captured in guidelines and standards (use of judgements). Could be complementary to developing new human factors guidelines.	Raise the question of how we are sure that certified human factors people would be employed in a contract.

References

Small, R. L., & Rouse, W. B. (1994). Certify for success: A methodology for human-centered certification of advanced aviation systems.In: J. A. Wise, V. D. Hopkin, & D. J. Garland (Eds.), *Human Factors Certification of Advanced Aviation Technologies.* Daytona Beach: Embry-Riddle Aeronautical University Press.

Taylor, R. M. & MacLeod, I. S. (1994). Quality assurance and risk management: perspectives on human factors certification of advanced aviation systems. In: J. A. Wise, V. D. Hopkin, & D. J. Garland (Eds.), *Human Factors Certification of Advanced Aviation Technologies.* Daytona Beach: Embry-Riddle Aeronautical University Press.

C-5.

The "Why" Group Paper: "Why" of Certification

Different Levels of Certification

At the outset, it is considered important to clearly differentiate the objective of the certification, if the "why" is to be established.

- Certification of the system design – its components, their function and interaction. Design is considered the ordering of components and function to achieve specific objectives.
- Certification of application and conformance to established human factors principles in both the design of the total system and in application of such principles in operating the systems.
- Certification of those who *exercise the process* of certification. This includes the general principles which underlie the process, particularly its application.

Basic Purpose

1. The Government must demand control in order to maintain popular approval for its performance at protecting the many from the potential mistakes of the few.
2. Certification of proper application of human factors principles is based on the determination of whether human actions which are initiated and driven by technology and human agents can maintain efficient and effective performance while also insuring individual safety and well being.
3. Certification provides the basis for accountability and corrective action (adaptability) based on responsibility according to standards. Certification is, in this sense, evolutionary in practice.
4. Due to changing demographics of an international community dependent on air transportation, it is important that general agreements are reached on requirements and that these agreements are applicable to new technologies that influence their design.
5. There is a need for a human factors database to acknowledge design changes from one system to another. This capacity allows for identifying and differentiating good or bad designs.

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

6. Certification can be considered a feedback mechanism for improving the quality of training. This feedback mechanism reflects assessed performance.

7. Certification provides the basis for assessing the effectiveness of human factors principles and leads to updating the human factors data bank.

8. Certification provides the basis for establishing the reliability of the entire system.

Why Not?

1. System certification programs presently exist. If experiences from such systems are ignored, others would be deprived of fruitful insights into systems design and performance.

2. Certification promotes increasing bureaucracy in analyzing and designing systems.

3. There is no base line for critical system performance information, particularly in reference to cost effectiveness.

4. Certification can become a punitive (practice) rather than a creative process.

5. Certification promotes dismay for a good established system.

The “Who” Group Paper

Who Should be Involved in the Certification of Complex Aviation Systems?

We realize, as have all other groups, that it is hard to consider “who” in isolation from “how,” “when,” “where” and so on. Nonetheless, we have constructed an overview of *who* with only some reference to what and how. It should be said that we had a mixed perspective – academic human factors and sociology, military ergonomics, international standards, industrial consultancy, aviation systems, etc., – and thus have not tried to delve too far into specifics, for we are in danger of presenting a very diffuse viewpoint.

We therefore emphasize the following at the outset:

1. Any system must attain its own assurance of integrity, within the certification process and in the development of other systems to which it is applied.
2. As a logical corollary to this, the process and especially associated documentation must be transparent and available to interested parties within the limits of commercial availability.
3. A process must be realistic and therefore must recognize and utilize existing institutions and frameworks within which to work.
4. Certification must be applied to the whole system, e.g., aircraft, crew, ATC, controllers, maintainers, and interactions between them. Thus hardware, software, training, procedures, and communications networks must be included, and mutual interactions between them must be explained.
5. Moreover any certification process must be robust enough for useful and efficient application to existing systems and those likely to emerge in coming decades (e.g., CNS/ATM – Communications/Navigation/Surveillance/Air Traffic Management).
6. In addition, the certification approach used, in both development and operation, must work within social, psychological, and technological dynamics.
7. Professional societies and their special subgroups have a pivotal role in developing a certification structure, providing appropriate analyses, test and evaluation tools and criteria, and in producing protocols for “user” traits with varying degrees of prototype simulation. Emphasis is on methods and analyses rather than the less universal and potentially more dated data and design guidelines and, in any case, are determined by current state of the art.

Who	*Notes*
Requirers	Impetus and legitimacy to process
International governing bodies – ICAO	Standards, overseeing, implementation

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

Authorities – FAA/CAA/DNA/etc. – ISO, etc. – National governments	Oversee, implement, and intervene Where necessary
Players – Suppliers: aircraft, maintenance, avionics – Communications (datalink and radio) – Customers: airlines, airports, passengers – Users: pilots, ATC, passengers, etc.	Nothing will work unless there is agreement among these on remit and skeleton system of certification. Internal pressure here also, e.g., customers or suppliers, users or customers, etc.
Guardians –professional bodies in human factors psychology, aviation, etc.	Responsible for setting standards of professional competence and training generally and specifically, and supplying people and knowledge into the system.
Influencers – Committee on human factors in complex systems; covered by authorities among Guardians (and players?) and witnesses (e.g., lawyers).	Specify analyses, test and evaluation and documentation procedures that should be followed by Doers for the suppliers, and advise the extent of specific criteria available.
Doers – Design and development team and their consultants, professionally qualified in appropriate disciplines.	Professionally trained and competent to design, develop, and apply all required analyses, T E documentation procedures to design of aircraft, ATC, communication technologies, socio-technical systems, etc.
Beauties and Beasts – Test workers of all aviation systems	Test users of aviation systems—pilots ATC, etc. – who are trained and educated to work with Doers (and/or relatives) to test nominally (beauties) or to limits (beasts).
Detectives – Independent body of professionally trained "testers" responsible to authorities.	Testers supplied by authorities where signaled to do so by users, customers, national bodies, etc. They are also professionally trained and competent.

The "When" Group Paper

A timeline for the certification process is presented in Figure 1. Shown are the major phases of the development process and corresponding certification events during each phase. It should be noted that the events are not purely sequential but may occur concurrently or with some overlap. Time sequencing is based on and coincides with current certification procedures that human factors activities should become a part of. The primary advantages and disadvantages of conducting efforts within each stage of the certification process are presented below.

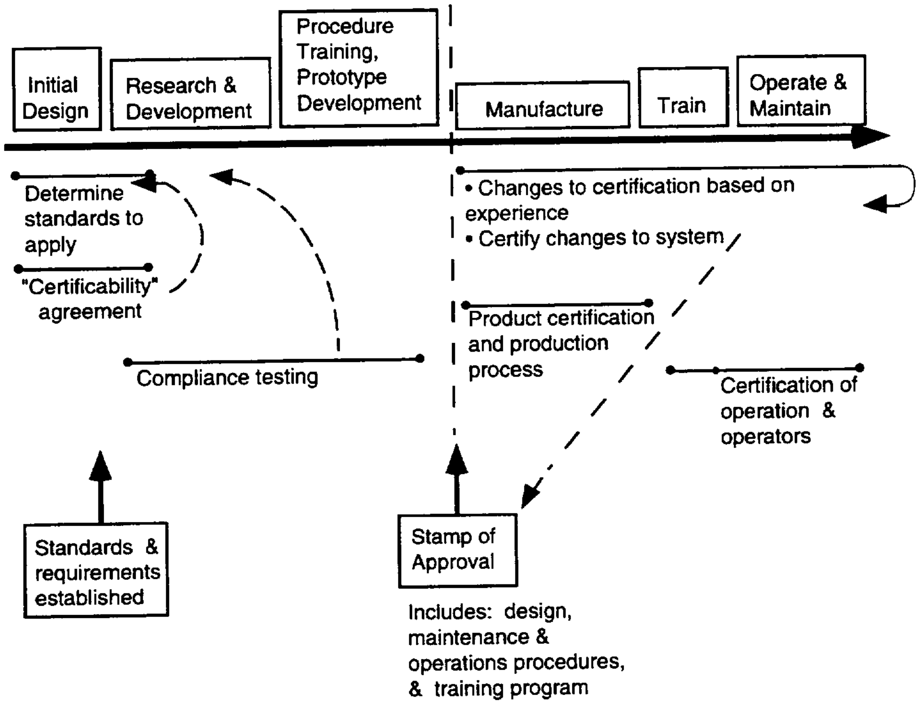

Initial Design
Research & Development
Procedure Training, Prototype Development
Manufacture
Train
Operate & Maintain
Determine standards to apply
"Certificability" agreement
Compliance testing
• Changes to certification based on experience
• Certify changes to system
Product certification and production process
Certification of operation & operators
Standards & requirements established
Stamp of Approval
Includes: design, maintenance & operations procedures, & training program

Figure 1.

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

DETERMINE STANDARDS [Time: Initial design phase]

Advantages

- Adapt standards and requirements to new project/system concept/technologies
- Filter out known human factors problems.
- Provide designers with standards to incorporate into design early enough to be cost effective
- Standard validity can be checked

Disadvantages

- Pressure on certifiers to use lower standards than they should

CERTIFICABILITY AGREEMENT [Time: Beginning with establishment of requirements standards until completion of initial design and early R&D on new technologies]

Advantages

- Provides company with reasonable assurances prior to major investment in system
- Encourages innovation as risk is managed
- Helps to interpret/apply/clarify ambiguous or general standards and requirements

Disadvantages

- Compromises independence of certifier (by involving them in design)
- Difficult to perform with limited available data at this stage

COMPLIANCE TESTING [Time: After establishment of certification conditions (agreement) until compliance is established for each component and the entire system]

Advantages

- Allows detection of human factors problems associated with multi-component system in a safer, cost effective manner prior to fielding a system
- Allows safe testing of system failure situations
- Allows safe development of procedures and training
- Provides confidence in design acceptability prior to building system
- Provides data for certification standards
- Allows modifications and alternatives to achieve certification

Disadvantages

- Time consuming and costly
- Testing may not be fully representative of actual use, system, or conditions
- May encourage over reliance on testing instead of analysis

STAMP OF APPROVAL [Time: After compliance with all requirements]

Advantages

- Certification provides formal approval prior to expense of mass production and fielding of system with possible compromise of safety
- Establishment of compliance with prescribed standards

Disadvantages

- May not have sufficient data/know ledge to provide definitive stamp of much value
- Cannot guarantee absence of human factors problems, particularly without actual operations and systems
- Legal liability
- Pressure on certifier to meet schedule deadlines

PRODUCT AND PRODUCTION CERTIFICATION [Time: Beginning of manufacturing cycle of each system]

Advantages

- Provides a check of human factors issues associated with manufacturing/product variances and evolution

Disadvantages

- Cost

CERTIFICATION OF OPERATIONS AND OPERATORS [Times: Beginning of operations and before each operator uses system individually]

Advantages

- Confirms proper use of prescribed methods and user performance level before individual operational use

Disadvantages

- Human factors input may not be necessary or cost effective on certain tasks

CERTIFICATION OF CHANGES TO CERTIFIED SYSTEM [Time: During each change cycle]

Advantages

- Enables improvement gained from practice
- Keeps certification current

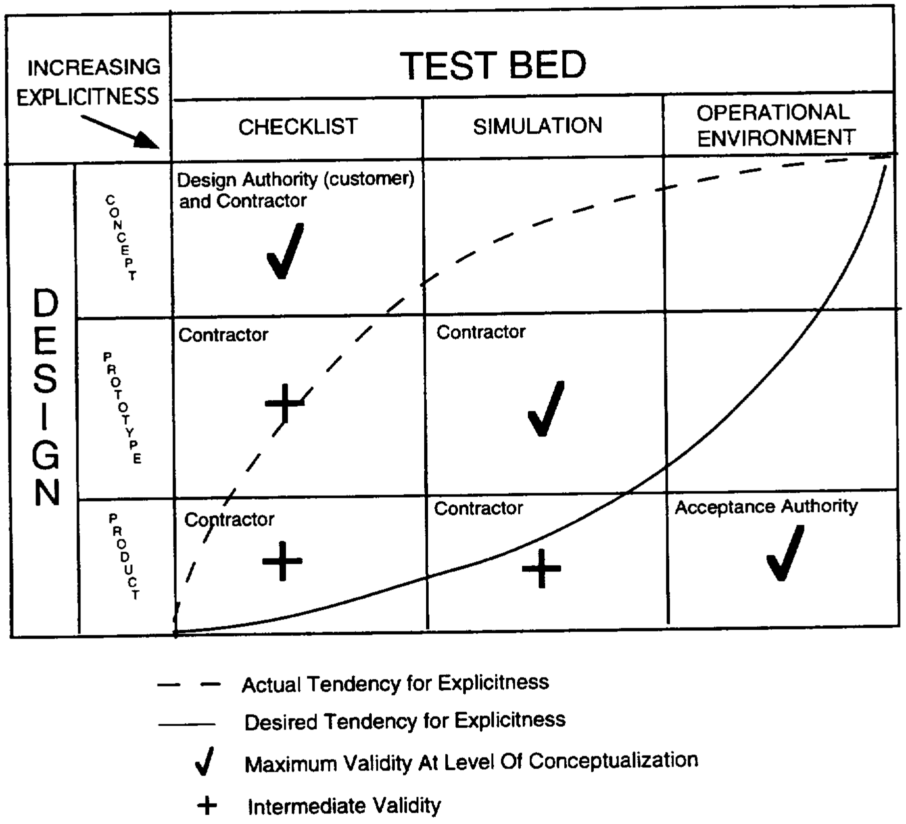
Disadvantages

- Slows change process
- Expensive to conduct and administer

The "Where" Group Paper

Introduction

We will illustrate our concept of "where" using two matrices. The first matrix (Figure 1) shows a generic relationship between the design and test processes and how the certification process may be described within this context. The second matrix (Figure 2) is less abstract and relates certification steps to common human factors areas.



INCREASING EXPLICITNESS		TEST BED		
		CHECKLIST	SIMULATION	OPERATIONAL ENVIRONMENT
DESIGN	CONCEPT	Design Authority (customer) and Contractor ✓		
	PROTOTYPE	Contractor +	Contractor ✓	
	PRODUCT	Contractor +	Contractor +	Acceptance Authority ✓

— — Actual Tendency for Explicitness
—— Desired Tendency for Explicitness
✓ Maximum Validity At Level Of Conceptualization
+ Intermediate Validity

Figure 1. Degree of Validity as a relationship between design stages and test beds.

As a consequence of our abstraction/reality basis we categorize the design process and test bed on opposite ends of the spectrum from abstract conceptualization to operational reality. The design process involves three broad steps: concept development, prototyping and production.

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PAGE 384 INTENTIONALLY BLANK

PRECEDING PAGE BLANK NOT FILMED

The test bed is a framework for evaluating the design process (through certification). At the conceptual stage evaluation proceeds according to broad rules (e.g. human factors checklists and expert judgment) and flows through various forms or levels of simulation fidelity and finally to the actual operational environment. Figure 1 illustrates the framework and indicates some possible results. The figure shows relationships between each of three levels of design and test in a design-test matrix, from the abstract to reality. Evaluation in the abstract (e.g. through checklists) is possible at all levels of design. Conversely, realistic evaluation in the operational environment is possible only with the actual design product. Evaluation at intermediate levels of realism via simulation can occur with both prototypes and products. Possible outcomes are indicated by cells with a check mark, while impossibilities are indicated by cells with a cross.

In Figure 1, maximum assessment validity at each level of design and test abstraction – reality is indicated by diagonal cells of the matrix. Other cells correspond to reduced levels of validity. Checklist assessment is the least valid evaluation method of the final product. Conversely, operational evaluation of the final product is the most valid assessment. All levels of design and test provide data for progressive acceptance and support the process of achieving certification. Final certification is completed after testing the product in the operational environment.

"Where" can be considered in terms of the implicitness or explicitness of assessment criteria. In the design-test matrix, the degree of implicitness or explicitness is indicated diagonally such that explicitness increases from top left to bottom right. Operational testing of a product provides the most explicit criteria for judging system performance and effectiveness.

There is tension governing the implicitness-explicitness characteristic. This tension is created by increasing automation and software demands, leading to less explicit "physicalistic" design/test criteria, and the needs of the human factors community for more explicit criteria.

The physical dimension of “awareness” is illustrated by the primacy of responsibility for each cell's activity and is divided between customer design authority (DA), contractor (C), and acceptance authority (AA).

Both customer and contractor are equally involved with concept design and testing. The acceptance authority maintains prime responsibility for operational testing of the product independent of the customer and contractor. Primary responsibility for simulation and prototype testing resides with the contractor.

Figure 2 presents a matrix identifying certification elements along the abstraction-reality dimension. Column headings identify possible areas for human factors certification. Row elements identify activities for human factors certification. Cell entries identify certification elements.

Four of the five possible areas identified for human factors certification relate to the system and systems personnel:

- Human-machine interface (HMI) system
- Procedures for operating the system
- Training of personnel to operate the system
- Personnel responsible for operating the system

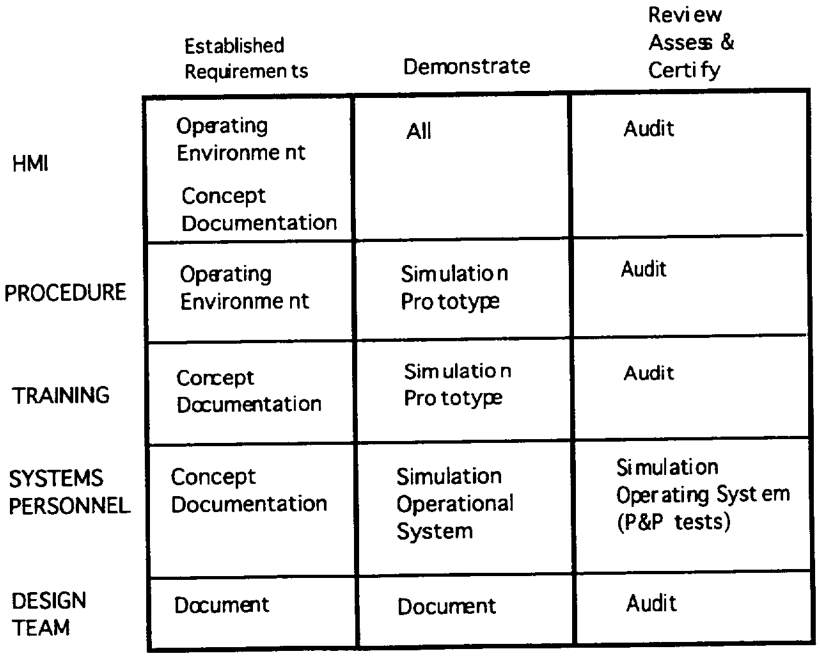
The fifth area, Design Team Certification, includes individuals who are responsible for system design, development, and testing.

Three possible activities are identified for the certification process:

- Establish requirements and criteria for certification
- Develop information to demonstrate that certification requirements and criteria are satisfied
- Review, assess and certify that components, systems, procedures, training programs, personnel, and so on have satisfied certification requirements and criteria

The physical or geographic locations certification activities may occur are listed below:

- Certifying organization which establishes certification requirements, reviews, assesses, and certifies
- Manufacturing facility, independent testing and evaluation facility, government facility, and commercial customer facility which develops information for demonstrating satisfaction of certification requirements and criteria



	Established Requirements	Demonstrate	Review Assess & Certify
HMI	Operating Environment Concept Documentation	All	Audit
PROCEDURE	Operating Environment	Simulation Prototype	Audit
TRAINING	Concept Documentation	Simulation Prototype	Audit
SYSTEMS PERSONNEL	Concept Documentation	Simulation Operational System	Simulation Operating System (P&P tests)
DESIGN TEAM	Document	Document	Audit

Figure 2. Matrix of Certification Elements and Possible Areas of Human Factors Certification

Each matrix cell in Figure 2 contains "where" certification will occur. We organized the matrix to show the various abstraction levels of the test bed: conceptual documentation, simulation prototype, operational environment, audit, and examinations. Each cell indicates a different level of abstraction. For example, conceptual documents and the operational environment are where requirements and criteria for certifying human-machine interfaces (HMI) should occur (upper left cell). Requirements definition and criteria development for HMI certification is an iterative process and refinement in more realistic operational environments is necessary for advanced aviation technologies, due largely to our general lack of understanding of their impact on system performance. In our example the subsequent phase in the certification process is to demonstrate that requirements for HMI have been met at all levels of abstraction of the test bed (upper middle cell). Information generated here will be analyzed and synthesized in order to review, assess and certify HMI (upper right cell).

The Pros and Cons of Where

We have discussed the pros and cons of where only in terms of levels of test bed abstraction and explicitness of the certification criteria. We believe the pros and cons of "where" along these two dimensions can be described in terms of a tension between confidence and cost. If "where" is at the abstract end of test beds and the implicit or internal end of certification criteria explicitness, the cost and confidence are low. If the test bed is the operational environment *and* the certification criteria are explicit, the cost *and* confidence are higher.

There is a desire, because of cost and perceived design efficiency, to accelerate toward test bed realism and explicit certification criteria early in the design process. The desire is countered by state-of-the-art human factors principles and emergent properties of intelligent automation that tend to push test bed realism and explicit criteria toward the end of design cycle and, worse, into the operational cycle. Herein lies the cost-confidence tension, the dilemma of human factors certification of advanced aviation technologies.

Conclusion

"Where" is an elusive concept embodied also within "who", "what", and "how". As we have explained, "where" is a multifaceted concept, containing meaning beyond physical geographic location.

The "How" Group Paper

Human Factors Certification

The assessment of the human factors within, a total system which includes both human and non human elements, compared to a standard in terms of obligatory and desriable elements, to fit the purpose of its design.

Static Analyses

Dynamic Performance

- Modeling
- Simulation
- Operational
- Normal
- Emergency
- Cognitive requirements

How Should Certification Be Done?

Options: positives versus negatives

1. Define certification

2. Organistaion/political
    - financial safety productivity
    - impartiality (non stake holder)
    - total ecological cost

3. Technical
    - systems
    - operations
    - personnel
    - certifiers

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

4. Publish standards and guidelines for the certification process

- standards and guidelines
- knobs and dials – good knowledge
- physiological – environment well known
- cognitive – too general and hard to interpret, incomplete and ambiguous
- team activity – poor knowledge, needed

Standards will evolve based upon the exposure and feedback of certification and must be creditable to stakeholders.

Cross cultural concerns for rationalisation and standardisation.

Publish procedures for applying for certification.

Publish the certification process.

Certification Organisation

1. Independent/neutral group/organisation, (without undue influence by stakeholders)

2. Small standing organisation
   - aplication standards
   - organise terms; stakeholders representation
   - publish certification procedures

3. Ad hoc certification groups
   - HF experts
   - Subject matter experts
   - engineering
   - maintenance

Information needed	**Information sources**	**Data**
standards	simulation	quantative
detailed requirements	multilevel (part task to high level)	qualitative
design rationale	normal	subjective/objective
	emergency	design
		safety verification

Figure 1. Information Matrix

Stakeholders (Significant Interested Parties)

- Producer/manufacturer
- Designer
- Programmer
- Users
  - organisations: people who want the system
  - operators: people who ... the system function
  - support people: other groups needed for the operation and maintenance of the system.

1. Elicit the purpose of the system and the stakeholders
   - (basis for validation)

2. Statement of the specific standards and guidelines to be used to certify the system

3 State the information to be provided to the certifier

4. certification team

A. Review required data
B. Conduct on-site assessment of systems run by producer
C. Formalised and recorded communication between producer and certification team
   i. Phased preliminary evaluation
   ii. Final total systems provisional certification
D. Field data of operations
   -Designers
   -Users
E. Final certification

Certification items	Advantages	Disadvantages
Publish standards and guidelines for certification process	explicit guidance provides existance for measurements focus for communication	understanding may require HF expertise in specific areas adequate standards do not exist in all relevant areas time and cost
Publish procedures for supplying for certification, and the certification process	facilitates entry into the process by the producer gives producer early notice of certification requirements the definition of responsibility	may adversly impact the process of developlment cost and time

Figure 2. Certification process and attendant advantages and disadvantages

certification organisation (Independent)	unbiased evaluation	may have less extensive experience with 'system' more personnel resources who certifies the certifiers may be politically unacceptable
small standing organisation to manage HF certification	modest cost continuity response speed expertise on regulations	creates new organisation may be difficult to estimate skill
ad hoc certification groups: HF experts subject matter experts engineering maintenence	expertise availability effective use of resources	delays in getting experts communication difficulty
data collection: identify needs and interests of significant interested parties producers/manufactures - designers - programmers users - organisations - operations - support	know who the players are know what they want promotes ownership of final system reduce conflicy in certification cycle improves the potential for user acceptance improve certification quality	political pressure time consuming perceived as reducing control by interested parties

Figure 2. (cont'd). Certification process and attendant advantages and disadvantages

elicit purpose of the system	basis of validation	
state information to be supplied to the certifiers - certification information needs optimise information resources data/answers	producers have a large amount of control in collecting data consistency in the data collection and reporting process provides data requirements at the beggining of the process data may be avialbale for future development	cost of collecting the data
certification team - review required data - conduct on site assessment of systems run by producers	speed certification process increase validity of data	validity of data may be unknown time and cost
formalised comunications between producers and certification team	impartiality lower cost to producer by reality testing	rigidity
evaluation process - phased preliminary evaluation	speed up overall evaluation process early feedback	longer certification process possibly greater cost
provisional system certification	permits initial fielding of systems data source builds confidence in users	premature acceptance
data from field operations	final total system 'operator in the loop' evaluation	time delay and cost
final certifcation	approval to use and market system safe effective system	may be residual problems

Figure 2. (cont'd). Certification process and attendant advantages and disadvantages

Participants

Dr. Rene R. Amalberti
Direction Générale de l'Aviation Civile
11 Bvd Hotel De Ville
Aulnay-Sous-Bois 93600
FRANCE
Tel.: (+33) 1 69 88 87 81
Fax.: (+33) 1 60 84 04 48

Dr. Rod Baldwin
Baldwin International Services
34 Dellegaass
Heffingen L-7651
LUXEMBOURG
Tel.: (+352) 872 11
Fax.: (+352) 879 785

Dr. Tom Bernard
University Of South Florida
College Of Public Health
13201 Bruce B. Downs Blvd
Tampa, FL 33612-3805
USA
Tel.: (+1) (813) 974-6629
Fax.: (+1) (813) 974-4718

Mr. Taylor Bracey
Embry-Riddle Aeronautical University
54 Seabreeze Dr.
Ormond Beach FL 32176
USA
Tel.: (+1) (904) 226-6375
Fax.: (+1) (904) 226-6459

Dr. Birgit Bukasa
Institute for Traffic Psychology
Austrian Road Safety Board
Ölzeltgasse 3
A-1030 Vienna
Austria
Tel.: (+43) (222) 717 70 170
Fax.: (+33) 62 25 95 99

M. Jean-Philippe Cottenceau
67 Rue Victor Boissel
5300 Laval
FRANCE
Tel.: (+43) 56 45 00
Fax.: (+43) 98 63 98

M. Victor Bernard Day
Eurocontrol
72 Rue de la Loi
1040 Brussesls
BELGIUM
Tel.: (+32) (2) 729 3583
Fax.: (+32) (2) 729 3976

Prof. Anthony Debons
RIISE
115 Edgeclift Drive
Carnegie, PA 15106
USA
Tel.: (+1) (412) 279-5020
Fax.: (+1) (412) 281-6170

Dr. Mica R. Endsley
Texas Tech University
Dept. of Industrial Engineering
Lubbock TX 79409
USA
Tel.: (+1) (806) 742 3543
Fax.: (+1) (806) 742 3411
E-Mail.: ajend@ttacs1.ttu.edu

Ms. Alyson Evans
Civil Aviation Authority
ATCEU
Bournemouth International Airport
Christchurch
Dorset BH23 GDF
UNITED KINGDOM
Tel.: (+44) 202 472 123
Fax.: (+44) 202 472 236

Ms. Karen Faris
Embry-Riddle Aeronautical University
Center For Aviation/Aerospace Research
600 So. Clyde Morris Blvd.
Daytona Beach FL 32114-3900
USA
Tel.: (+1) (904) 226-6381
Fax.: (+1) (904) 226-7050

Ms. Diane Farrow
Embry-Riddle Aeronautical University
Center For Aviation/Aerospace Research
600 So. Clyde Morris Blvd.
Daytona Beach FL 32114-3900
USA

Ms. Irene Gaillard
ARAMIIHS
31 Rue de Cosmonautes
ZI du Palays
31077 Toulouse Cedex
FRANCE

Mr. Vincent Galotti
ICAO
3 Bis Villa Emile-Bergerat
92522 Neuilly-Sur-Seine
Paris
FRANCE
Tel.: (+33) (1) 46 37 96 22
Fax.: (+33) (1) 46 24 09 14

Dr. Daniel J. Garland
Embry-Riddle Aeronautical University
Center For Aviation/Aerospace Research
600 So. Clyde Morris Blvd.
Daytona Beach FL 32114-3900
USA
Tel.: (+1) (904) 226-6790
Fax.: (+1) (904) 226-7050
E-Mail.: garland@db.erau.edu

Dr. Richard Gibson
Embry-Riddle Aeronautical University
Center For Aviation/Aerospace Research
600 So. Clyde Morris Blvd.
Daytona Beach FL 32114
USA
Tel.: (+1) (904) 226-6385
Fax.: (+1) (904) 226-7050
E-Mail.: gibson@db.erau.edu

Dr. Richard Gilson
University Of Central Florida
Psychology Dept/CAHFA
Phillips Hall
Orlando, FL 32816
USA
Tel.: (+1) (407) 823-2216
Fax.: (+1) (407) 823-5156

Mr. Patrick Guide
Embry-Riddle Aeronautical University
Center For Aviation/Aerospace Research
600 So. Clyde Morris Blvd.
Daytona Beach FL 32114-3900
USA
Tel.: (+1) (904) 226-7102
Fax.: (+1) (904) 226-7050
E-Mail.: guidep@db.erau.edu

Dr. Rune Haglund
Civil Aviation Administration
S-601 79 Norrköping
SWEDEN
Tel.: (+46) 11 19 24 14
Fax.: (+46) 11 19 26 40

Dr. Peter A. Hancock
University Of Minnesota
164 Norris Hall
172 Pillsbury Drive 5E
Minneapolis MN 55455
USA
Tel.: (+1) (612) 625-8527
Fax.: (+1) (612) 626 7496
E-Mail.: hancock@vx.acs.umn.edu

Dr. Lewis Hanes
2023 Wickford Rd.
Columbus, OH
USA
Tel.: (+1) (614) 487-8655
Fax.: (+1) (614) 487-8655

Dr. Kelly Harwood
Sterling Software
FAA/NASA-Ames Field Office
Main Stop 210-2
Moffet Field, CA 94568USA
Tel.: (+1) (415) 604-5051
Fax.: (+1) (415) 604-0173
E-mail:
Kelly_Harwood@qmgate.arc.nasa.gov

Mr. L.W. Hennessy
Embry-Riddle Aeronautical University
Center For Aviation/Aerospace Research
600 So. Clyde Morris Blvd.
Daytona Beach FL 32114-3900 USA
Tel.: (+1) (904) 226-7107
Fax.: (+1) (904) 226-7050
E-Mail.: marks@db.erau.edu

Mr. V. David Hopkin
Peatmoor
78 Crookham Rd
Church-Crookham
Fleet Haints GU13 0SA
UNITED KINGDOM
Tel.: (+44) 252 620 221
Fax.: (+44) 252 377 839

Dr. Esther Horne
RIISE
302 Fox Chapel Rd. #212
Pittsburgh PA 15238-2336
USA
Tel.: (+1) (412) 963 1900
Fax.: (+1) (412) 963 1926

Mr. Alistair Jackson
EUROCONTROL Experimental Centre
B.P. 15
91222 BRETIGNY-SUR-ORGE CEDEX
FRANCE
Tel.: (+33) (1) 69 88 75 44
Fax.: (+33) (1) 60 85 15 04
E-Mail.: jac@vm.eurocontrol.fr

M. Denis Javaux
Service de Psychologie du Travail
Bat B32
Fapse-Ulg
4000 Sart-Tilman
BELGIUM
Tel.: (+32) 41 56 20 209
Fax.: (+32) 41 56 29 44
E-Mail.: JAVAUX@vm1.ulg.ac.be

Mr. Hartmut Koelman
Eurocontrol
Rue De La Loi, 72
B-1040 Brussels
BELGIUM
Tel.: (+32) (2) 729 3952
Fax.: (+32) (2) 729 3783

Dr. Marcel Leroux
CENA
7 Edouard Belin
BP. 4005
31055 Toulouse
FRANCE
Tel.: (+33) 62 25 95 65
leroux@cenatls.cena.dgac.fr

Mr. Iain MacCleod
Aerosystems International
West Hendford
Yeovil Somerset BA20 2AL
UNITED KINGDOM
Tel.: (+33) 935-77326/34145

Dr. Carol Manning
FAA
Civil Aeromedical Institute
P.O. Box 25082 AAM-511
Oklahoma City OK 73125
USA
Tel.: (+1) (405) 954 6849
Fax.: (+1) (405) 954 4813

Dr. Michel Masson
FAPSE
B 32
5 bld du Rectorat
Sart-Tilman
Universite de Liege
BELGIUM
Tel.: (+32) 41 56 22 36
Fax.: (+32) 41 56 29 44

Dr. Andrew McClumpha
RAF/IAM Farnborough
Psychology Divison
Farnborough
Hants GU14 6SZ
UNITED KINGDOM
Tel.: (+44) (252) 377 83

M. Jean Paries
Bureau Enquetes Accidents
246 Rue Lecourbe
75732 PARIS CEDEX 15
FRANCE
Tel.: (+33) 1 40 43 40 49
Fax.: (+33) 1 45 58 01 23

Dr. William Rogers
BBN, Inc.
NASA Langley Research Center
M/S 152
20 West Taylor Rd.
Hampton, VA 23665-5225
USA
Tel.: (+1) (804) 864-2045
Fax.: (+1) (804) 864-7793
E-Mail.:
william_rogers.haib@qmgate.larc.nasa.gov

Ms. Laurence Rognin
L.A.A.S.
7 Avenue Du Colonel Roche
31077 Toulouse Cedex
FRANCE
Tel.: (+33) 61 33 62 41
Fax.: (+33) 61 33 64 11
E-Mail.: rognin@laas.fr

Mr. Ron Small
Search Technology, Inc
4898 South Old Peachtree Road
Suite 200
Norcross GA 30071-4707
USA
Tel.: (+1) (404) 441-1457 (1458x114voice)
Fax.: (+1) (404) 263 0802
E-Mail.: rons@searchtech.com

Dr. Mark Smolensky
Center for Aviation/Aerospace Research
Embry-Riddle Aeronautical University
600 So. Clyde Morris Blvd.
Daytona Beach FL 32114-3900
USA
Tel.: (+1) (904) 226-7103
Fax: (+1) (904) 226-7050

Dr. Paul Stager
Department of Psycholody
York University
4700 Keele Street
Toronto, Ontario M3J 1P3
CANADA
Tel.: (+1) (416) 736-5122
Fax.: (+1) (416) 489-1532
E-Mail.: pstager@vm1.yorku.ca

Dr. Earl Stein
Federal Aviation Administration
FAA Tech Center
ACD-350
Atlantic City Int'l Airport NJ 08405
USA
Tel.: (+1) (609) 485-6389
Fax.: (+1) (609) 485-6218
E-Mail.: steine@admin.tc.faa.gov

Dr. Andy Tattersall
University Of Wales, Cardiff
School Of Psychology
P.O. Box 901
Cardiff CF1 3YG
UNITED KINGDOM
Tel.: (+44) (222) 87 4000
Fax.: (+44) 222 874858
E-Mail.: tattersall@cardiff

Dr. Robert M. Taylor
Special Senses Division
RAF-IAM
Farnborough
Hants GU14 6SZ
UNITED KINGDOM
Tel.: (+44) (252) 394 120
Fax.: (+44) (252) 277 839

Dr. Ron Westrum
19017 Saxon Drive
Beverly Hills, MI 48025
USA
Tel.: (+1) (313) 647-1015
Fax.: (+1) (313) 487-8755

Ms. Florence Wilbaux
Direction Generale De L'Aviation Civile
246 Rue Lecourbe
75732 Paris Cedex 15
FRANCE
Tel.: (+33) 1 6043 6532
Fax.: (+33) 1 6063 6532

Dr. John R. Wilson
Dept. Of Manufacturing Engineering
University Of Nottingham
Nottingham
NG7 2RD
UNITED KINGDOM
Tel.: (+44) 602 476 769
Fax.: (+44) 602 514 000

Dr. John A. Wise
Center For Aviation/Aerospace Research
Embry-Riddle Aeronautical University
600 So. Clyde Morris Blvd.
Daytona Beach FL 32114-3900
USA
Tel.: (+1) (904) 226-6384
Fax.: (+1) (904) 676-5416
E-Mail.: wise@db.erau.edu

Mr. Mark A. Wise
University of Central Florida
Psychology Dept/CAHFA
Phillips Hall
Orlando FL 32816
USA
Tel.: (+1) (904) 226-7106
Fax.: (+1) (904) 226-7050
E-Mail.: wisem@db.erau.edu