JYVÄSKYLÄ STUDIES IN EDUCATION, PSYCHOLOGY AND SOCIAL RESEARCH 207

Narciso González Vega

Factors Affecting Simulator-training Effectiveness

Esitetään Jyväskylän yliopiston yhteiskuntatieteellisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Villa Ranan Blomstedt-salissa
joulukuun 5. päivänä 2002 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Social Sciences of the University of Jyväskylä,
in the Building Villa Rana, Blomstedt Hall, on December 5, 2002 at 12 o'clock noon.

UNIVERSITY OF JYVÄKYLÄ

JYVÄSKYLÄ 2002

Factors Affecting Simulator-training Effectiveness

JYVÄSKYLÄ STUDIES IN EDUCATION, PSYCHOLOGY AND SOCIAL RESEARCH 207

Narciso González Vega

Factors Affecting Simulator-training Effectiveness

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2002

Editors
Tapani Korhonen
Department of Psychology, University of Jyväskylä
Pekka Olsbo, Marja-Leena Tynkkynen
Publishing Unit, University Library of Jyväskylä

ISBN 951-39-1370-8 (nid.), 978-951-39-5144-3 (PDF)
ISSN 0075-4625

Jyväskylä University Printing House, Jyväskylä 2002

ABSTRACT

Narciso González Vega
Factors affecting simulator-training effectiveness
Jyväskylä: University of Jyväskylä, 2002, 162 p.
(Jyväskylä Studies in Education, Psychology and Social Research
ISSN 0075-4625; 207)
ISBN 951-39-1370-8 (nid.), 978-951-39-5144-3 (PDF)
Yhteenveto: Simulaattoriharjoittelun tehokkuuteen vaikuttavat tekijät
Diss.

The global aims of this research were: Firstly to attempt to improve the effectiveness of training provided by means of simulators. Secondly, to provide empirical evidence for these improvements. Thirdly, to attempt to use cognitive theory and simulator training concepts as guidance to improve skill acquisition. And, finally, to provide guidelines based on research results for future developments of training simulators. Three experiments were designed to examine some of the factors which can influence skill acquisition through simulator training. In Experiment 1, novice students performed the operation of a naval diesel generator on a machinery control room simulator according to part-task, critical part-task, and full-task training schemes. Trainees were transferred to a task of increased complexity. Results indicated that the part-task training scheme was more effective in the acquisition and transfer phases. With the same task setting, Experiment 2 provided evidence of the advantage of simulator training with augmented cueing strategies as compared to non-augmented cueing. Additionally, effectiveness was increased with the implementation of augmented cueing training on a PC-based low-fidelity simulator. Experiment 3 examined the relationship between simulator fidelity and training effectiveness. Even though no substantial differences were found between the high-fidelity and low-fidelity simulators, cost-effectiveness favoured the low-fidelity simulator. Results are used to provide suggestions for future developments of more cost-effective training simulators.

Keywords: simulator, training, skill acquisition, training effectiveness, naval, complex skill, transfer

Author's address
Narciso González Vega
Department of Psychology
Agora/Psykocenter
University of Jyväskylä
P.O. Box 35
FIN-40014 Jyväskylä
FINLAND
E-mail: gonzalez@psyka.jyu.fi

Supervisory group
Professor Pertti Saariluoma (main supervisor)
Department of Computer Science and Information Systems
University of Jyväskylä, Finland

Professor Heikki Lyytinen
Department of Psychology
University of Jyväskylä, Finland

Reviewers
Professor Peter G. A. M. Jorna
National Lucht en Ruintevaart Laboratorium (NLR)
National Aerospace Laboratory, The Netherlands

Professor Veikko Teikari
Department of Industrial Engineering and Management
Helsinki University of Technology, Finland

Opponent
Professor Peter G. A. M. Jorna
Nationaal Lucht-en Ruimtevaartlaboratorium (NLR)
National Aerospace Laboratory NLR, The Netherlands

ACKNOWLEDGMENTS

This dissertation has a long history which started in 1994. It has gone through several mountains and valleys, through the sea, air, and ground, fortunately, with the company and support of many people to whom I am truly thankful. For this, I would like to show my gratitude to Pertti Saariluoma and Heikki Lyytinen for their supervision. For their invaluably effortful revision work, I would like to thank Peter G. A. M. Jorna and Veikko Teikari. I again thank Peter G. A. M. Jorna for accepting the role of opponent. I would also like to thank Jane Erskine for her timely language corrections and comments, Paula Salmi and Heikki for their suggestions, Kenneth Eklund for his translation into Finnish, and Asko Tolvanen for his advice on statistical issues. I would also like to thank, Raija Mehto, Paavo Kuusimäki, Tarja Etelälahti, and Päivi Parikka for softening the terrain in the last steps. In general to all of those involved in the LKK project.

The seminal work of this research was funded by a research grant from the University of Salamanca, Spain (1994). The empirical work was carried out while I was working at SAINSEL, in Sevilla, Spain, funded by the EUCLID CEPA RTP 11.1 Project MASTER (Military Application of Simulation and Training concepts based on Empirical Research). The compilation, additional literature reviews, and writing of the final manuscript was funded by a grant from the University of Jyväskylä, Finland.

I am grateful to those who have shared their knowledge with me, Gerardo Prieto Adánez, Juan Carro Ramos, and José Antonio González del Campo, (Department of Basic Psychology, Psychobiology, and Methodology, University of Salamanca, Spain). I am thankful to Jesús Bermejo Muñoz, Eduardo Ruiz Pérez, Sergio Retamero Merino, and José Manuel Benítez García (developer of the NDGSTRT software tool) (SAINSEL, Spain); Celestino Sanz Segundo, Rafael Benítez Domínguez, Ricardo Molina, and all the participants in the experiments from the Department of Naval Systems and Machinery (University of Cádiz, Spain); Antonio Gramaje Cornejo (ISDEFE, Spain), Jan Moraal, Johan B. J. Riemersma, John C. G. M. van Rooij, Ivonne Barnard, and Danielle Verstegen (TNO, The netherlands), Jan Joris M. Roessingh, and Jelke van der Pal (NLR, The Netherlands), Hélène de Fontenilles (Thomson TTS, France), Michel Moscato (University of Ruen, France), Jacques Krawiès (Dassault Aviation, France), Eric Farmer (DERA, UK). I would like to appologise for not being able to name many other MASTER partners.

Finally, I would like to show my heartily thankfulness to my father Valeriano, to Toñi, Vicente, Javi, Judit, Hellevi snd Jussi, and Jouni.

To María Oliva and Faustino in memoriam

I dedicate this dissertation to Outi.

Jyväskylä, December 2002 Narciso González Vega

CONTENTS

1 INTRODUCTION

The working conditions and functions of workers have changed drastically in the last decades. Much of the manual work has been replaced by robotic and process control systems. In certain areas, job performance demands have changed from physical motor to more intellectual or cognitive exigencies (Norros, 1989; Rasmussen, 1993). Computer and information technology systems accompany us almost in every aspect of our daily lives not only at work but also at leisure time (i.e., computers, mobile telephones, Internet, etc.). Most of these systems require our participation to accomplish their functions and tasks. We refer to these as human-machine systems (HMS) or human-computer systems.

In the accomplishment of the functions and tasks of any complex HMS, we contribute to it by means of our more or less skilled performance through the human-machine interface (HMI). Constraints to our interaction with the system are determined by our information processing capacities and limits at the functional and neural levels. Paradoxically, even though systems are designed by humans, they impose information processing demands that exceed the capacities of the interacting human. Therefore, systems which do not take into account the capacities and limits of the human are prone to induce erroneous human performance (Wickens, 1992). Error-free performance is a general goal, very seldom achieved, by any HMS. When a HMS behaves erroneously, problems of different dimensions are likely to occur, e.g., shipwrecks, aircraft accidents, nuclear power plant emissions, and so on, irrespectively of whether the errors can be independently attributed to the human or to the machine (Miller and Swain, 1987; Reason, 1990).

Problems arising from this paradox can be remedied by improvements on the machine component of the system, the human user, or both. Improving the performance or the characteristics of the machine, would be provided by an engineering solution. Improving the performance of the operator would or should result from a more psychologically oriented or interdisciplinary solution. In this case, we are adopting a human factors solution to our problem. In brief, human factors is an applied discipline concerned with the design of

HMS that accommodate the limits of the human (see, Salvendy, 1987; Wickens, 1992).

Training the operators of a system so that their performance matches the system’s demands is one of the possible ways to remedy problems arising from the paradox mentioned above. The HMS under consideration in this research has been the machinery control system of a ship. Hence, this research deals with the application of simulation technology to the realm of training. In particular, with the training on the complex skills needed by the operators of ship machinery systems. Simulator training provides unique opportunities for training the operators of actual systems on tasks similar to the ones they can encounter at work. On many occasions, training can also be provided in tasks and situations, that hopefully, one never has to face, i.e., nuclear power plant accidents, plane crashes, ship-wrecks, etc. This can be accomplished without risking the health of the trainees and/or the integrity of the system and the environment. At the same time, simulator training can facilitate the training of operators at a more reasonable cost per training unit than if the actual system were used for training purposes.

Simulator training has been, so far, technology and cost driven (Andrews, Carroll, and Bell, 1995; Stedmon and Stone, 2001). Customers, i.e., civil and military training establishments, or companies providing simulator training for their employees, request a simulator system to a simulator development provider. In the best case scenario, this request is accompanied by a set of technical specifications in order to fulfil a set of training needs. The simulator provider offers a technical proposal at a certain price. If the price does not match the customer's budget, cheaper solutions can be provided by cutting down the technical characteristics in the original set of specifications. In this manner, so far, training simulators are technology and cost driven. Nevertheless, authors in the human factors field propose that training simulator technology should be more research driven than it is at present (Andrews, et al., 1995; Glaser, 1990; Salas and Cannon-Bowers, 2001; Stedmon and Stone, 2001).

Depending on the aims of the research, the course of simulator training research can start from different stages. Also, the aims of the research can be constrained or guided by the functional state in which the training simulator is. Consider, for example, that a new complex system or subsystem is being deployed (e.g., a new nuclear power plant, new cooling system in a power plant, a new ship) and the operators of that system need training to use it in an efficient and safe manner. Most likely, the whole range of analysis, design, development, validation/testing, including the iterative loops for refinements of the different activities of the training system development should be performed (e.g., Branson, Rayner, Cox, Furman, King, and Hannum, 1975; Goldstein, 1986; Patrick, 1980; Rasmussen, 1983,1993).

Moving to an alternative extreme situation, say that the training programme and the training simulator already exists and has been operative for some years. The results of the training simulator are, according to the users, satisfactory, but there is a need to consider the possibility of improving its efficacy (Benítez-Domínguez, 1996). Constraints preclude the manipulation of the training simulator as the software code is not open. The premises where the simulator is located cannot be changed, nor can the simulator be transported elsewhere. Additionally, changing other factors in the training programme negatively affect the training results of the actual trainees. This scenario means that the research was carried out in these simulator facilities before the actual trainees started their actual training programme. Thus, their participation in this research was not part of their current training programme. Another constraint has been the reduced research literature available in the domain of naval training simulation. Apart from being scarce, this has been diverse in its aims, training methodologies, tasks, and training evaluation procedures. In practice, this has meant that little direct guidance and few points of contrast have been found for the present research within this specific domain.

This was the context from which this research departed. Our global research aims were: Firstly, to improve the effectiveness of the training simulator while taking into consideration the conditions and constraints stated above. Secondly, to provide empirical evidence of these improvements. Thirdly, to attempt to use cognitive theory and simulator training concepts as guidance for simulator training improvement. And, finally, to provide research based guidelines for the development of training simulators as well as for future research in the interdisciplinary field of human factors.

In the following sections, models of skill acquisition, retention and transfer will be described. Theory and results from research on skill acquisition processes, related to prescriptive training theories should provide better training methods, techniques, and devices, after extensive research on these issues has been carried out. Aims and hypothesis of this research are stated as an epilogue to the report of the three experiments presented here. Conclusions and suggestion for future simulator development conclude this research.

1.1 Acquisition of skills

As stated earlier, most present-day complex HMS and subsystems, such as ship propulsion plant operation, power plant operation, ship handling, etc., consist of hardware, software and personnel. These components work together to accomplish some function or output goal. The accomplishment of that goal depends on a number of variables that represent system functions. Thus, functions are activities or tasks performed to achieve the goal of a system (Drury, Paramore, Van Cott, Grey, and Corlett, 1987), usually by means of the interaction between the machine and the operator via the HMI.

Despite the numerous existing definitions, there is a common agreement that a task is a condition that requires a set or unit of human/system behaviour that contributes to the accomplishment of a specific functional objective of the task (Drury et al., 1987). The advantage of defining tasks is that it lays the foundations on which task descriptions can be made according to a common format, irrespective of which system or function is being dealt with. One possible format to describe a task, is: input - operation/transformation - output. Tasks can be simple or complex. Depending on the level of detail required for the description of a task, a complex task may be subdivided into elements referring to identifiable subtasks or elements of that task and described according to the format presented above (Anderson, 1987; Miller and Swain, 1987).

Development of task taxonomies is an important issue in the training domain because it has been suggested that the characteristics of the to-be-learned tasks that are to be performed in complex systems should determine the content of the training programme (Colley and Beech, 1989a; Holding, 1987). A different position argues that the relatively general principles affecting learning may be independent of any task taxonomy. These basic principles of learning define the conditions under which learning may take place (see, Langley and Simon, 1981). In accordance with the conditions and constraints indicated above, in our research we approached the task description at the functional level, i.e., system start-up, and malfunction solving.

Until recently, most studies on skilled performance have focused on perceptual-motor skills. Nevertheless, the concept of skill is now seen as applying to the broad domain of human performance, (i.e., perceptual, motor, cognitive). In addition, the differences and/or similarities between the classical perceptual-motor versus intellectual skills dichotomy are also changing nowadays (see, Rosenbaum, Carlson, and Gilmore, 2001). A skill is considered to be a capacity to perform a task acquired or learned through instruction or training and practice, instead of being innate; it is a goal-directed behaviour, and uses feedback for error correction —unless there is not sufficient time for the execution of error detection and correction—; it is organised spatially and temporally; it takes place as a smooth, automated, and highly integrated pattern of behaviour (Anderson, 1980; Annet, 1991; Colley and Beech, 1989b; Fitts, 1962/1990; Leplat, 1989; Schneider and Shiffrin, 1977; Welford, 1976). Acquisition of skills and learning is enabled by the memory system. The human as an information processor perceives information, encodes, stores, performs operations, and retrieves it from the memory system when needed (Baddeley, 1997; Wickens, 1992).

Different authors propose different skill categories in their investigations. Anderson (1980) considers skills as either declarative or procedural. Holding (1987) describes skills along two dimensions: verbal vs. motor, and simple vs. complex. Welford (1976) identifies three types of skills corresponding to different stages in information processing activities. The skills involved in the different stages of human information processing activities are: perceptual skills, motor skills, and intellectual or cognitive skills (Colley and Beech, 1989b).

Glaser (1990) reviewed recent learning research from a cognitive perspective, and lamented that the learning processes and instructional or training implications show very few commonalties across task domains. However, he suggested that an integrated theory prescribing a mix of instructional approaches for specific training purposes should be developed.

Nevertheless, cognitive theory still has a long way to go before achieving that promising research objective. In return, this should have a remarkable impact on training research and practice (Morrison, 1991).

Integration of learning theory and training theory should influence the effectiveness and cost of training. The efficiency of the training delivery system should be increased as long as it is able to predict the speed with which skills are acquired, the extension and quality to which they are retained, and the amount and conditions to which they are transferred to the actual operational situation. It means that trainees would acquire skills needed to perform their tasks in less time, with a higher level of performance, and would retain and transfer them more efficiently to their actual jobs. If these goals could be achieved, then costs could be reduced by reducing time spent in training activities, and by selecting and designing optimal training devices (e.g., simulators). Characterising training devices as optimal, means that they produce the best results, in the shortest possible time, and at a low economical cost (e.g., Gonzalez, Carro, and Prieto, 1994, Hesketh, 1997).

Some trends in experimental cognitive psychology have attempted to characterise the regularities found in the learning processes. One of these concerns the mathematical description of the relationships between the amount of practice and performance. This approach is known as the power law of practice (Newel and Rosenbloom, 1981). Other approaches have attempted to integrate performance models with models of learning. These are widely known as skill acquisition models (e.g., Colley and Beech, 1989b; Masson 1990, Morrison, 1991; Patrick, 1991).

1.1.1 The power law of practice and performance

The importance of the power law of practice in the skill acquisition research emanates from the fact that it is a very general empirical phenomenon. The observation that power law functions of performance speed-up with practice is consistently found in many task domains, including proving geometry theorems (Neves and Anderson, 1981), evaluation of logic electronic circuits (Carlson, Sullivan, and Schneider, 1989), and air-traffic control (ATC) in the Kanfer-Ackerman task (Ackerman, 1988; Lee and Anderson, 2001). Because it applies to a wide variety of tasks and performance measures, it has also been termed the ubiquitous law of practice. See Newell and Rosenbloom (1981) for a comprehensive review on this topic. The power law of practice represents a quantitative empirical relationship between the time taken to perform a task (e.g., reaction time), and the number of practice trials (Newell and Rosenbloom, 1981). The impact of this phenomenon in the study of basic cognitive processes has been such that the evidence of a power law in task performance is considered a necessary condition to accept or reject theoretical postulates and hypotheses, e.g., Anderson, (1982, 1983, 1992), Anderson and Fincham (1994), Lee and Anderson, (2001), Logan (1988, 1992, 2002), Newell and Rosenbloom

(1981), Schneider and Shiffrin (1977) Shiffrin and Schneider (1977), to name a few of the most relevant authors.

The mathematical function interrelating amount of practice and performance appears to follow the form of a power law. If we plot the logarithm of the time to perform a task against the logarithm of the practice trial number it always produces an almost straight line.

This general learning equation was apparently first found by Snoddy (1926), in an experiment of mirror-tracing of visual mazes (Fitts, 1962, 1990; Newell and Rosenbloom, 1981). Later, Snoddy (1926) proposed the log-log law of practice, the equation of which as put forward by Newell and Rosenbloom (1981) is as follows:

$$\mathrm{Log}\ (\mathrm{T}) = \mathrm{Log}\ (B) - a\ \mathrm{Log}\ (N),$$

where:
T is the time required to perform the task,
B is the performance time in the first trial,
a is the slope of the line (i.e., the learning rate), and
N is the trial number.

The linear equation of the power law of practice (Newell and Rosenbloom, 1981) has the form:

$$\mathrm{T} = BN^{-a},$$

where:
T is the time required to perform the task,
B is the performance time in the first trial,
N is the trial number, and
a is the slope of the line (or, the learning rate).

Other versions of the power law, such as that by Logan (1988, 1992, 2002) include an asymptotic value which represents the limit of learning. The point where more learning would be difficult to achieve in terms of reaction times to perform a task. This power function has the form:

$$\mathrm{RT} = a + bN^{-c},$$

where:
RT is the reaction time,
a is the asymptote,
b is the difference between initial and asymptotic performance,
N is the amount of practice in sessions or trials per item, and
c is the slope (or, the learning rate).

The power law of practice applies to a great number of tasks (Newell and Rosenbloom, 1981). These tasks include perceptual-motor tasks, such as cigar manufacturing, and perceptual tasks, such as learning to read inverted text, memory tasks, problem solving, and so on.

Newell and Rosenbloom concluded that the power law of practice holds not only for speed measures of performance, but also for other measures of performance such as accuracy (Newell and Rosenbloom, 1981). Despite this conclusion, they also stated, that the power law effect is not as strong for accuracy measures as it is for speed measures of performance. They also demonstrated that the power law of practice is empirically more robust than other mathematical functions, such as exponential and hyperbolic functions. See, Newell and Rosenbloom, (1981) for detailed descriptions on supporting evidence for the power law of practice.

1.1.2 Learning models of skill acquisition

In its origin, the power law of practice was just an empirical finding with no theoretical basis. Nevertheless, it was striking enough to attract the attention of researchers. Attempts have been made to develop a learning theory which could explain these empirical findings. Strictly speaking, the models which are discussed below were not learning models at their inception, with the exception of Fitts' (1962/1990) description of the phases in complex skill learning. For example, Anderson (1980, 1982, 1992) aimed at developing a computational model of memory; Fisk and Schneider (1984), Schneider and Shiffrin (1977), Shiffrin and Schneider (1977) tried to explain automaticity by means of attentional mechanisms proposing two different modes of processing, i.e., controlled vs. automated. Logan (1988) proposed the instance theory of automatization to explain these phenomena as memory retrieval rather than attentional processing limitations. Some of these postulates have been extrapolated to the research on skill acquisition, and, in the human factors research, are accepted as skill acquisition models (e.g., Colley and Beech, 1989b; Masson 1990, Morrison, 1991; Patrick, 1991). In this direction, two different trends can be identified. Models of skill acquisition which suggest two distinguishable stages such as the ones proposed by Adams (1971), Schneider and Shiffrin (1977), Shiffrin and Schneider (1977), and Logan (1988). And, three-stage models of the skill acquisition process represented by the proposals of Fitts (1962/1990), Fitts and Posner (1967), Anderson (1980, 1982, 1992) and Neves and Anderson (1981).

Lewis (1979) tried to explain theoretically those performance results which satisfactorily fitted power law functions. In doing so, he argued that performance improves according to a power function because a task is a combination of many components of the task, and with practice, performance on each component improves exponentially. Even though a power function can be derived with proper and quite restrictive assumptions about the relative contribution of individual component processes to overall performance, e.g., the key stroke level components in the Ackerman-Kanfer Air Traffic Control (ATC) task (Lee and Anderson, 2001), Lewis' theory had little acceptance among psychologists.

Newell and Rosenbloom (1981) suggested that the power law is the result of chunking processes. In the psychological research field, chunking has two possible meanings. Firstly, the concept of chunk was developed by Miller (1956) in respect to the capacity of the human memory. A chunk is a unit of information held in working memory, and working memory is assumed to have a capacity of 7 ± 2 chunks (Miller, 1956). A chunk can consist of a letter, a syllable, a word, or other units. Though this view has been accepted for almost 50 years, more recent and controversial evidence suggests, that short-term memory is only capable of maintaining from three to five independent chunks (i.e., 4 chunks on average) (Cowan, 2001). Secondly, Newell and Rosenbloom (1981) proposed that the learning mechanisms which produce performance data with the shape of power functions are chunking processes. These chunking processes account for the combination of simple rules into larger ones. The chunking theory, as well as Lewis' explanation, depend on restrictive environmental conditions to produce a power function which fits performance data. Despite these restrictions, the chunking model usually behaves like power functions, or otherwise, closely approximates them (Newell and Rosenbloom, 1981).

Anderson (1982, 1983, 1992) and Neves and Anderson (1981) proposed a more developed computational learning system which also tried to explain the phenomenon of the power law of practice. Contrary to other explanations, their proposal and further developments of the model have reached a higher level of acceptance and sophistication (Colley and Beech, 1989a; Lee and Anderson, 2001; Masson, 1990).

As stated earlier, different researchers agree on the point that skills are acquired through learning processes and practice. Also, most current theories of skill acquisition agree on the nature of the learning processes involved in skill acquisition (Ackerman and Kyllonen, 1991). Even though learning is considered to be a continuous process (Newell and Rosenbloom, 1981), theorists have offered a description of the learning process as divided into phases or stages of skill acquisition. We shall consider the three major phases of skill acquisition as described by Anderson (1980, 1982, 1992), Fitts (1962/1990), and Fitts and Posner (1967). In our discussion, we try to relate the different models to show that the skill learning processes seem to follow similar pathways even though the theoretical constructs as well as the explanatory mechanisms proposed differ.

The first phase starts when the trainee faces the task for the first time. It has been termed the cognitive phase by Fitts (1962, 1990), Fitts and Posner (1967), and the encoding phase by Neves and Anderson (1981). After instructions about the task are given to the trainee, he/she begins to understand the basic task requirements. The trainee then formulates a general idea about what is required from him/her. At this stage, declarative knowledge is used by the trainee to recall the instructions and the rules applying to the task to be performed (Anderson, 1980). Declarative knowledge is represented in a way that generally allows conscious-mediated retrieval (Shiffrin and Schneider, 1977). Thus performance at this phase is slow, attentionally effortful, and error prone (Fisk, Ackerman, and Schneider, 1987; Schneider and Fisk, 1983).

The second phase of skill acquisition has been called the fixation phase by Fitts (1962/1990), associative stage by Fitts and Posner (1967), or proceduralisation by Neves and Anderson (1981). The mechanism which leads the acquisition of skills from the cognitive or declarative stage to the procedural one is called by Anderson (1982) the knowledge compilation process. Knowledge compilation follows the declarative stage and happens during the associative stage. The main element of performance improvements during this phase occurs due to an increase in the strength and efficiency of associations between stimulus conditions and response patterns (Fitts and Posner, 1967). Knowledge compilation at this stage is associated with converting declarative knowledge into production rules (Anderson 1992). Knowledge compilation does not eliminate declarative knowledge, rather it remains available as an alternative means for performing the task (Anderson, 1992). At this stage, accuracy and speed of performance are increased (Anderson, 1982; Colley and Beech, 1989b; Fitts and Posner, 1967).

The third phase is known as the autonomous phase (Fitts, 1962, 1990), composition (Neves and Anderson, 1981), or proceduralised knowledge (Anderson, 1982). In Anderson (1982), the proceduralisation and composition phases were assimilated into one, which was called proceduralisation. When the trainee has reached a skill level such that performance requires minimal attentional effort, and at the same time is fast and accurate, then the knowledge to perform the task has been automated (Shiffrin and Schneider, 1977). Fitts (1962/1990) also called this stage automation of performance. In contrast to declarative knowledge, procedural knowledge does not require conscious mediation. If performance on the task has been sufficiently proceduralised, so that declarative knowledge is no longer required to accomplish the task goal , declarative knowledge, relating to how to perform the task may be unavailable to consciousness with no decrease in task performance (Anderson, 1982).

1.1.3 Learning mechanisms in skill acquisition

The most relevant models of skill acquisition have been described in the two previous sections, including the quantitative, empirical model, or the power law of practice, and principally, a theoretical model developed to explain results which accommodate the power law functions. The explanatory model we have described is Fitts' (1962/1990) and the reformulation by Neves and Anderson (1981) of Fitts' model of skill acquisition, and later refinements of the model (Anderson, 1982, 1983, 1992). That description has focused chiefly on the stages of the skill acquisition process, or the topological aspects of the qualitative changes occurring during skill learning (i.e., error reduction, speeded performance, reduction in the use of attentional resources).

Attention and memory mechanisms have been posited to explain the automatisation of skills through extensive periods of practice. A two-process theory of visual detection, search and Attention was developed by Schneider and Shiffrin (1977) and Shiffrin and Schneider (1977). The theory proposes two qualitative different modalities of processing. These are automatic processing and controlled processing. Automatic processing is characterised by fast parallel information processing that can be carried out almost independently of other concurrent tasks. This is, it demands little or no attention. The requisite for automatic processing skill to develop is a consistent mapping between stimuli and responses. Controlled information processing is a slow serial mode of processing which happens under conscious control. Controlled information processing demands a high degree of attention. The conditions for controlled processing skill to develop are varied mappings between stimuli and responses. Conditions with this characteristic can be, for example, when there are no consistent rules to perform a task, or when a task is new to the person and the consistencies have not yet been learned (see, Fisk, Ackerman, and Schneider, 1987, Schneider and Shiffrin, 1977, Shiffrin and Schneider, 1977). In further developments of the theory of automatisation, Schneider (1985) and Schneider and Detweiler (1987) propose two learning processes, priority learning and association learning, to account for how automaticity develops. These processes have mainly been studied in the areas of visual search and memory search tasks. More recently, attentional processes, i.e., internal versus external focus of attention, have also been proposed to affect the use of automatic control processes in the acquisition and retention of motor skills (e.g., Wulf, McNevin, and Shea, 2001).

Instead of a reduction in attentional resources to perform a task as a result of practice, Logan (1988) proposed a mechanism based on memory retrieval to explain the development of automaticity. According to Logan's instance theory of automatisation, early in practice, task performance is based on the execution of a general algorithm (e.g., children learning to add a single digit). With practice or experience, trainees learn specific solutions to specific problems. This problem-solution mapping (stimulus-response association) is considered an instance of the task. Instance theory assumes that each encounter with a problem lays down a separate trace in memory. When trainees encounter the same problem again, they can respond with the solution directly retrieved from memory or with the one computed by the algorithm. The development of automaticity is thus considered to be a transition from performance based on the execution of an initial algorithm (i.e., non-automatic performance), to performance based on direct memory retrieval (i.e., automatic performance) (Logan, 1988).

This transition is explained by Logan as a race between the algorithm and the memory retrieval process. Actually, the race is established between the algorithm and the fastest instance retrieved from memory. At some point in training, performance will depend on memory retrieval rather than the algorithm. This can happen as a result of a strategic decision of the trainee to entirely rely on memory and abandon the algorithm. Alternatively, as a consequence of the statistical characteristics of the race, while the distribution of the finishing time of the algorithm across training remains constant, the finishing time for the retrieval process decreases. The instance theory of automatisation predicts a power-function speed-up, not only for the mean of the distribution of reaction times, like most theories of skill acquisition, but also for the standard deviation (Logan, 1988). This theory was tested in five experiments using lexical decision and alphabet arithmetic tasks which provided support for its hypothesis (Logan, 1988). The prediction of a power law function by the instance theory of automaticity was also supported by analysing data sets from two previous experiments using an alphabet arithmetic task (Compton and Logan, 1991) and dot-counting task (Lassaline and Logan, 1993) (see, Logan, 1992; Logan and Stadler, 1991). In contradiction, with pseudo-arithmetic and alphabet arithmetic tasks, Rickard (1997) did not find supporting evidence for the predictions of the instance theory of automatisation. With less emphasis on the power law of practice, Logan has extended the instance theory to more complex and integrated views of cognitive processing such as the recent instance theory of attention and memory (see, Logan and Gordon, 2001; Logan, 2002).

Neves and Anderson (1981) were interested in how students learn to use postulates and theorems in geometry tasks. They considered how postulates are encoded in memory, how procedures are created after these encoded postulates, and how the use of procedures speeded up with practice. The research method was implemented on the computer by means of a production system program which performed the tasks. They found that the production system provided a good model of the behaviour of all subjects at all levels of skill. Consequently, they proposed their model of skill acquisition and the mechanisms responsible for it.

The learning mechanisms they postulated were: encoding, proceduralisation, and composition. In a later review of the theory of the Adaptive Control of Thought (ACT*) (Anderson, 1992), three learning processes are also postulated. These processes are, knowledge encoding, knowledge compilation, and strengthening of production rules and declarative facts.

Encoding is the mechanism by which perceived information is stored in the memory system. Information is encoded declaratively as a set of facts in a semantic network (Neves and Anderson, 1981). Once information has been stored in memory, it is considered as knowledge and the person can retrieve and use it for different purposes. This set of facts is then used when needed by general interpretative procedures (IF-THEN rules, or pairs of condition-action statements) to guide behaviour. This way of encoding information represents the procedures which guide behavior as data. Data can be used for reasoning about problems, and to plan actions to change procedures if needed (i.e., accommodate a procedure to solve a problem faster than usual). A second way of representing knowledge about procedures is encoding them as production systems which can be executed without interpretative operations (Neves and Anderson, 1981). This encoding process was called procedural encoding by Neves and Anderson (1981). Application of knowledge in this format is faster than declarative knowledge, because the condition and the action required in that situation are directly connected.

Proceduralisation is the mechanism by which knowledge represented in a declarative format is converted into a procedural format. Neves and Anderson (1981) assumed that every time a production matches a semantic network structure in long-term memory that must be retrieved into working memory, the proceduralisation mechanism creates a new production which has the semantic network structure incorporated in it. This new production held in working memory does not need to access long-term memory to retrieve knowledge from it. Therefore, use of procedural knowledge is faster than use of declarative knowledge. The proceduralisation mechanism converts progressively declarative knowledge into procedural knowledge by reducing the amount of data that has to be retrieved from long-term memory (Neves and Anderson, 1981). For this procedural version of a first declarative one to become a unitary procedural representation, another mechanism is needed which would account for this transformation. This mechanism is the composition process (Neves and Anderson, 1981).

Composition was assumed to occur concurrently with proceduralisation, but it is also thought of as a process which continues after proceduralisation has been completed. Neves and Anderson proposed that composition operates by combining pairs of productions which are executed sequentially into a single production, i.e., chunking. Thus, it is expected that combining procedures into a larger procedure will reduce the time to execute a procedure.

Anderson (1992) attempted to explain phenomena associated with the concept of automaticity from the framework of the ACT* theory of skill acquisition. In this paper, he proposed three mechanisms to explain transitions from stage to stage in the skill acquisition process. These mechanisms were somehow different from those described earlier.

Encoding is the mechanism that plays its role at the starting point of task performance. If the task has never been performed before by the trainee, no productions would be available at that moment. Hence, knowledge is encoded directly from experience and encoding processes are declarative at this point.

From declarative knowledge, subjects create production rules. Production rules are specific for a particular task. The learning process of constructing task-specific productions was named knowledge compilation. The mechanisms of knowledge compilation and proceduralisation have similar functions. They eliminate the need to retrieve knowledge from long-term memory and reduce the number of smaller productions that have to be used to perform the task.

The third learning process involves the strengthening of declarative knowledge and production rules. Strength is considered to be the determinant of how active a declarative fact or a production rule is. The level of activation of different declarative facts and procedures determines which one is selected to accomplish the task. The strength of a production or declarative fact is assumed to increase a constant unit each time they are used to perform the task. The implication of this strengthening process is that, the stronger a production is, the more chances it has to be selected to activate the data to which some productions match (Anderson, 1992). Thus, productions compete between themselves for the activation of the data to which they match. Hence, the strongest production is the one selected to perform the task.

In relation to the different phases of skill acquisition, those models also propose the existence of different types of knowledge which is stored in the human memory system and retrieved when needed. Anderson (1980, 1992) makes a distinction between the knowledge underlying recall of the facts we know and the knowledge required to perform various intellectual tasks. These two types of knowledge are referred to as knowing that, or declarative knowledge, and knowing how, or procedural knowledge. An example of declarative knowledge is a fact such as "Helsinki is the capital of Finland".

Although, most declarative knowledge can be verbalised, its form does not need to be verbal. In fact, Neves and Anderson (1981) proposed that all incoming information is encoded declaratively as a set of facts in a semantic network. Sometimes, procedural knowledge, on the other hand, cannot be verbalised (Anderson, 1980). An example of procedural knowledge is knowing how to speak in one's mother tongue. Most people often forget the rules governing the use of their own language (i.e., the grammar), or simply cannot report what they do to speak to another person. The way of encoding procedural knowledge is production systems (Anderson, 1976, 1981). Productions consist of a pair of condition-action statements or IF-THEN rules, and production systems consist of groups of productions (Masson, 1990).

Rasmussen (1983) discussed some distinctions in defining the different categories of human performance. Performance is supposed to be based on the internal representations or mental models of all factors surrounding the task to be performed, and the task itself. Rasmussen (1983) made a distinction similar to that of Fitts' (1962/1990) when describing the categories of human performance. The categories of human performance and the descriptions proposed by Rasmussen (1983) are detailed below.

Skill-based performance represents sensory motor performance during activities, and, following a statement of an intention, takes place without deliberate control as smooth, automated, and highly integrated patterns of behaviour. Input for this type of performance is sensory information, and output is a motor response. The flexibility of skilled performance is due to the ability to compose, from a large repertoire of sub-routines, the sets suited for specific purposes. Examples of this kind of performance can be seen in very simple tasks such as picking up an object, or more complex ones, such as ship handling.

The next level Rasmussen (1983) proposes is rule-based performance. Activity in this case is controlled by a stored rule or procedure which may have been learned through previous experience, communicated by another person as know-how information, or directly applied from an user's manual for a device such as video. In this case, the person is usually able to report which rules or knowledge are being applied in the performance of the task.

When no knowledge of the first two types is available for an unfamiliar complex task, performance is controlled by a more general and flexible conceptual information or knowledge-based representation of the task (Rasmussen, 1983). Knowledge-based representations include information about the goal to be achieved, the means to accomplish it, the functional relationships between the elements, and the predicted results of the action considered. Attentional resources must be devoted to this process in order to integrate incoming information in a coherent manner. This information is represented by a mental model, schemata, or knowledge structure (Johnson-Laird, 1983; Rasmussen, 1983; Rouse and Morris, 1986).

Until now, we have described the models and mechanisms of the memory system that allow the storage of information into memory (i.e., learning or skill acquisition) and the types of knowledge contents stored within. Learning is considered to occur when the information encoded into working memory, is transferred to more permanent format into long-term memory (Atkinson and Shiffrin, 1968). Performance on the other hand is considered to be dependent on the maintenance in working memory of information retrieved from long-term memory. This widely accepted conceptualisation of the functioning of the memory system, nevertheless, introduces practical and theoretical problems when processing of a large amount of information is required. The problem arises due to the fundamental properties of working memory. This is, its dependence on time, i.e., chunks cannot be held in working memory longer than about 10-20 seconds, and its demands of attentional resources, i.e., chunks cannot be maintained in working memory without attention (Wickens, 1992).

According to this view, the task we used in our experiments would be impossible to perform in the initial phase of training. The task involved many component sub-tasks, probably not cued by previously performed steps, nor by the task environment. This is contrary to the view of Schneider and Detweiler (1987), and probably, the whole amount of information, i.e., knowledge about the functional relations among sub-systems, plus the ordered sub-tasks required to complete the procedures assigned to the trainees could not be chunked in any way since they were not familiar with the system. The required knowledge is highly specific of this domain (Ericsson and Kintsch, 1995). On the other hand, some experts do not seem to have this problem in their areas of expertise including taxi drivers and professional chess players (Ericsson and Kintsch, 1995; Ericsson, Krampe, and Tesch-Römer, 1993; Ericsson and Lehmann, 1996; Kalakoski and Saariluoma, 2001; Saariluoma and Laine, 2001; Saariluoma and Kalakoski, 1997).

Ericsson and Kintsch (1995) proposed a theory that without altering the current views of the memory system and its functions, elegantly provides clarification on the issue of exceptional performers which cannot be accounted for by the traditional view of the limited capacity of working memory (e.g., Baddeley, 1986). They argue that subjects can acquire domain-specific memory skills that allow them to expand their effective working memory capacity for a particular activity. Thus, what has been considered long-term memory can adopt the functions of long-term working memory. Originally, Chase and Ericsson (1982) proposed that expanded memory capacity, i.e., skilled memory, could be achieved under certain conditions. First, subjects must be able to store information rapidly in long-term memory. This implies that they already have a large body of knowledge and patterns for that information. Second, the activity must be very familiar to the experts so that they can anticipate future information retrieval demands. Evidence of the necessity of extensive (10 years) and deliberate practice is provided by Ericsson, et al. (1993). Under these two conditions, selective storage of information in long-term memory is possible. Third, subjects must associate the encoded information with the appropriate retrieval cues. By means of the association between the encoded information and its retrieval cue, it is possible to activate a particular retrieval cue in short-term memory which reinstates the encoding conditions to retrieve the required information from long-term memory. Sets of retrieval cues organised in a stable structure are conceptualised as retrieval structures (Chase and Ericsson, 1982; Ericsson and Kintsch, 1995).

Two types of mechanisms can account for the development of long-term working memory skill. The acquisition or generation of new retrieval structures in long-term memory and the use of retrieval structures. In sum, individuals can acquire memory skill to accommodate expanded demands of working memory in a specific task domain. Extended working memory capacity is acquired over an extended period of time and in response to task relevant training. Retrieval structures clearly differ among individuals as well as among task domains (Ericsson and Kintsch, 1995).

1.1.4 Retention of skills

Most of the skills learned through simulator training fall within the category of complex skills (Schneider, 1985). These types of skill are easily forgotten (Hurlock and Montague, 1982). One of the factors affecting skill forgetting is the non-utilisation period, - the time elapsed between simulator training and job placement. In the case of ship engineers, if this period is too long, then it is likely that they must relearn those skills on the job. The goal of a training programme, and training simulators by extension, consists of the trainee performing at the best possible level on the actual job or task for which the training programme was developed. The training process provides the trainee with the opportunities to acquire new skills, retain them in memory, and transfer them to the actual job situation. The psychological processes of skill learning/acquisition, retention, and transfer, are usually disclosed for the purpose of research on the variables affecting them, and the effects they might have on the results of training. Nevertheless, the interrelationships between these processes are very tight, and should not be separated in the training programme development process; rather, factors affecting acquisition, retention, and transfer processes should be integrated in the system to achieve the best training result.

Retention performance measures evaluate what information has been learned during training, and the dynamics of that stored information between the time of original learning and the time when it is used, in the absence of practice. The dynamics of the information stored in memory refer, broadly, to the processes of remembering and forgetting that information during the time of the retention interval.

There are models which try to predict the decay of memory representation strength. They study forgetting and describe strength as mathematical functions of memory decay (i.e., exponential function, power function, and exponential power function) (Sticha et al., 1990). Other approaches are interested in the variables affecting skill retention in the absence of practice. Thus, some findings which directly relate to the training process will be described in the next sections.

Hurlock and Montague (1982) contend that complexities of Navy operations make it difficult to develop and maintain the skills and job proficiency of enlisted personnel. Hence, fleet readiness could be seriously affected by personnel skill deterioration over non-utilisation periods. Research into the factors affecting Navy skill retention can provide understanding as well as guidelines to prevent skill deterioration and its likely disastrous effects on naval and other human-machine systems (Hurlock and Montague, 1982).

Sticha et al. (1990) reported some conclusions on the factors affecting retention of skills in military training. Hurlock and Montague (1982) also provided findings and conclusions regarding the implications of skill retention for Navy task training. The results reported by Hurlock and Montague (1982), and Sticha et al. (1990) are very similar. This indicates the generality of these results regardless of the type of task (i.e., military in general or Navy in particular). The following conclusions include the effects of training conditions, type of task, trainee's ability variables, job conditions, and retraining factors or refresher training on later skill retention. Motor skill retention is not so relevant in most naval systems as is the retention of cognitive skills. Despite this, the retention of motor skills is addressed in the following discussion due to their contribution to task performance, usually in the context of the human-machine interaction.

Trainees can reach different skill levels at the end of the training programme. Skill acquisition level is an important predictor of retention. Increasing the number of repetitions of a task (i.e., over-learning) during training will enhance training performance, and retention at a later time (Driskell, Willis, and Copper, 1992; Hurlock and Montague, 1982; Sticha et al., 1990). Repetitions are generally effective when they apply to both practice trials and test trials. However, repeated tests do not improve retention if the job, or task is performed with aiding material (Sticha et al., 1990).

Active practice and spaced practice improve retention. Active practice means that the trainee is motivated to fulfil the training tasks, and is searching to achieve the best results from his/her training in knowledge and future performance. Spaced practice refers to the way in which training activities are organised in time, and is opposed to massed practice in the psychological research literature. Spaced practice is spread out in time, allowing longer intervals between practice trials according to different schedules. Massed practice, on the contrary, concentrates practice in time so that the time spent in training to a certain skill level is reduced (Bahrick, 1979;Montague and Knirk, 1993; Shebilske, Goettl, Corrington, and Day, 1999).

Retention is not necessarily improved by the use of mnemonic techniques (Hagman and Rose, 1983). Mnemonic techniques are a series of strategies which explicitly detail the organisational structure of the to-be-learned task (Wickens, 1992). One such technique is chunking. Chunking in this context has a completely different meaning from that described in relation to the mechanisms of skill acquisition proposed by Newell and Rosenbloom (1981). While in Newell and Rosenbloom (1981) it is a mechanism assumed to produce skill acquisition, in this context, as seen below, it is a strategy for increasing the working memory capacity. A chunk is a unit (e.g., a letter, a syllable, a word) of information held in working memory. Working memory has a capacity of 7 ± 2 chunks (Miller, 1956). Chunking consists, broadly, of expanding working memory capacity by grouping chunks into bigger chunks. Another mnemonic technique is the method of loci. This method works by providing a visual framework of a fixed order. Every piece of information can be associated to a point in the visual framework (Baddeley, 1997). Despite their utility in experimental research, these mnemonic techniques do not necessarily increase retention in tasks performed at work (Hagman and Rose, 1983). More recent analyses of the capacity of working memory postulate an average capacity of 4 chunks (Cowan, 2001)

Procedural tasks (e.g., ship machinery operation), which are the most important in naval settings, are the most difficult to learn and the most easily forgotten (Hurlock and Montague, 1982). They are much more quickly forgotten than continuous control tasks or remember fact tasks. The number of steps in a procedure (e.g., safety procedure) seem to accurately predict the forgetting of that procedure (Hurlock and Montague, 1982; Sticha et al., 1990). Over-learning can help to reduce memory load in complex procedural tasks and increase retention (Hurlock and Montague, 1982)

Knowledge of results or corrective feedback is a key factor for reducing errors and for maintaining and improving skills over time, i.e., affects the amount and quality of skill acquisition, and interacts with other factors (Van Matre, Pennypacker, Hartman, Brett, and Ward, 1981). This factor is especially critical in procedural tasks when a high level of performance is needed, e.g., in system damage control operations. Nevertheless, other authors propose that feedback should be removed progressively from the training task context in order to improve transfer of training to the job situation (Schmidt, 1991; Schmidt and Bjork, 1992; Hesketh, 1997).

The trainees' general ability affects acquisition/learning performance of a task more than it affects retention. This means that trainees with higher general ability will learn how to perform a task faster than trainees with lower general ability (Hurlock and Montague, 1982; Sticha et al., 1990). However, if both groups are trained to the same level of performance they will show equal performance in a retention test after a period of no practice on the task (Shendel and Hagman, 1991).

In naval tasks, as well as in other domains, skill deterioration is a primary function of the length of the non-utilisation period. Job conditions in the US Navy, among others, often involve more than one period of non-utilisation between skill training and skill utilisation on-the-job (Hurlock and Montague, 1982). Retraining can quickly dissipate the effects of forgetting due to the non-utilisation periods (Hurlock and Montague, 1982).

Motor skills play a role in the performance of a great variety of tasks in different areas (e.g., car driving, ship handling, aircraft piloting, etc.). Motor skills share the main components of the description of skill given above, but they focus on the muscular movements required to perform the task. We shall outline some of the conclusions from the review on motor skill retention by Shendel and Hagman (1991).

The conclusions reached in their review are divided into three categories, which include: task variables, procedural variables, and learner or trainee variables. Retention decreases as a function of time. The longer the time between the end of training and the retention test, the more it is forgotten by the trainee. One of the variables affecting this forgetting function is the level of original learning (Montague and Knirk, 1993; Shendel and Hagman, 1991). As we have seen in the previous section, trainees should be trained to the highest possible level to reduce forgetting.

Categories of motor responses include continuous, discrete, or procedural responses. Continuous responses are those in this classification which are retained better over time. Discrete and procedural responses, however, cannot be retained so well without regular practice (Shendel and Hagman, 1991). This finding has also been reported above when discussing procedural tasks.

The organisation or structure of a task is important for retention if the response is not well learned. This could happen at the beginning of training, or in the case of a complex task. On the contrary, well-learned responses do not need the task to be highly organised (Hurlock and Montague, 1982; Shendel and Hagman, 1991). This could be the case at the end of reasonable practice or in the training of simple tasks.

Compatibility of display-control relationships can enhance motor learning, retention, and transfer. Training in high-compatibility equipment requires less time to reach and maintain performance at criterion level than training on incompatible display-control equipment. Moving the cursor on a computer screen (display) with a mouse (control) is an example of display-control compatibility, because moving the mouse to the right makes the cursor move to the right. A case of incompatible control-display could be if moving the mouse to the right makes the cursor go to the left. Thus, compatible control-display training equipment is preferred for the implementation of the training programme (Hurlock and Montague, 1982, Wickens, 1992).

Display specificity of external cues is important early in the training process. The cues that the trainee receives from the task display are used to guide his/her performance early in training. Apparently, later in training, external cues do not provide more information than that provided by the trainee's internal sources (e.g., Hesketh, 1997; Shendel and Hagman, 1991).

Therefore, external cues are not considered to be relevant factors of motor skill retention.

Over-training and over-learning is the procedure in which training trials are repeated after criterion performance level has been reached. Repetition of training has been found to increase both skill learning level, and retention. As we have seen in the previous section, the learning level is an important determinant of later skill retention, and learning level is enhanced by repetition of training or over-training (Driskell, Willis, and Copper,1992; Hurlock and Montague, 1982).

Spaced training and massed training have different effects on motor and verbal tasks. Massed training produces better acquisition in verbal tasks, but retention is better when between trial intervals are increased (spaced training). In motor tasks acquisition and retention exhibit similar performance levels under both massed and spaced training schedules (Hurlock and Montague, 1982).

Mental practice might increase acquisition and performance of motor skills. Nevertheless, research has not demonstrated conclusively its effects in enhancing retention of motor skills.

Knowledge of results (KR) consists of information, provided externally to the trainee, about the discrepancy between the trainee's actual response and the criterion response (Holding, 1987). Knowledge of results is critical during the early training of motor skills. Still, trainees who perform better during acquisition when they receive KR in more trials or more accurate KR usually perform worse in retention tests, when KR has been withdrawn, than those trainees who have received less useful KR, or KR has progressively been withdrawn during training (Holding, 1987; Shendel and Hagman, 1991; Schmidt and Bjork,1992).

Response-produced feedback designates the sensory consequences of motor responses associated with them (i.e., sights, sounds, feelings, proprioceptive sensations). Increases in the number of feedback channels and the quality of response-produced feedback during training facilitate skill acquisition as well as skill retention.

Augmented feedback is provided during training by means of extra cues (i.e., augmented feedback cues) or extra information, neither of which is intrinsic to the task being performed. Augmented feedback cues are associated to the correct responses, so the trainee can anticipate the correct response and avoid errors. This is opposed to response-produced feedback, in that it is inherent to the responses being learned during training. When augmented feedback is used in the training programme, acquisition performance may increase. Nevertheless, performance is usually hindered if augmented feedback is removed during retention tests (Lintern, 1980; Lintern, Thomley-Yates, Nelson, and Roscoe, 1987).

Training, generally speaking, consists of study trials and test trials. In study trials, the trainees are presented with the information or movement they must learn. In test trials, trainees attempt to recall or execute the response from the representation in memory. The use of test trials during training hinders acquisition performance, but improves retention of both motor and verbal skills (Hurlock and Montague, 1982).

Variability of practice refers to the way the training of more than one task is scheduled. The two more popular experimental preparations for this concern have been the blocked schedule and the random schedule. In the blocked condition, training is completed for each task before training passes on to another task. In the random condition, training trials of the different tasks may be intermixed in different randomised or pseudo-randomised conditions. Variable practice is provided by the randomised conditions. The findings have been that while variable practice delays skill acquisition, it increases performance on retention and transfer tests in motor, visual search, and verbal tasks (González-Vega and González del Campo, 1993; Lee and Magill, 1983; Magill and Hall, 1990; Schmidt and Bjork, 1992)

Hurlock and Montague (1982) propose, that when the time from the end of training to the time when the acquired skills have to be used is too long, then the consequences of forgetting these skills may be disastrous. Hence, refresher-training devices have been devised to overcome the negative effects of the retention interval. Retention intervals in some situations, such as military or industrial, may last months between the time when trainees have completed training and the moment when they have to perform the tasks for which they have been trained (Hurlock and Montague, 1982; Montague and Knirk, 1993). In these cases, refresher training should be considered as a possible alternative in the implementation of the training programme.

Regarding trainee's variables which affect acquisition and retention, general ability level has been one of the most conclusive. Results suggest that general ability affects skill acquisition level, while it does not directly affect retention (Ericsson and Kintsch, 1995). Hence, trainees of higher initial ability level tend to achieve higher levels of acquisition performance than trainees of lower ability level. Nevertheless, the forgetting rate seems to be equal in both groups, though they reach different levels of forgetting because acquisition level is higher in higher ability trainees. Consequently, if both high and low ability trainees were trained to the same level of performance, then they should show the same level of retention (Ericsson and Kintsch, 1995; Shendel and Hagman, 1991).

Findings on the conditions affecting skill retention have been outlined above. Some of these conditions which enhance retention can be directly incorporated into the training programme and the simulator used to deliver training. Nevertheless, in most cases more research is needed to arrive at conclusive results about their effects on skill retention. In particular, scheduling of practice, knowledge of results, and augmented feedback have not produced conclusive results. Rather, they have shown transient effects on acquisition performance which are not separable from more permanent learning effects, such as retention, and transfer measures (Shendel and Hagman, 1991).

If the factors which improve skill retention are incorporated into the training programme, then trainees will show less forgetting when they face the actual job tasks. Therefore, the goal of trainees' retaining the skills acquired through training is more likely to be accomplished. At the same time, if trainees show a high level of skill retention on the tasks mastered during training, transfer to the actual job will be facilitated. Apart from other factors affecting transfer, retention of the skills learned during training enables the trainees to retrieve and apply that knowledge to the new situation. If trainees were not able to recall previously learned skills, then transfer should be minimal. Training of transferable skills is posited by Hesketh (1997) as one of the possible ways to ameliorate skill loss and employability in present-day social unemployment conditions (see also, Frese and Altmann, 1989; Hesketh, 1997; Ivancic and Hesketh, 2000; Schmidt, 1991; Schmidt and Bjork, 1992).

It has been shown how skill acquisition is inferred from empirical measures of acquisition, retention, and transfer, each one having its own implications. Following this logic, the next sections will define the concept of transfer, describe experimental paradigms developed to assess transfer, and report different measures devised to extract useful transfer meanings. Theories of skill transfer will be dealt with in section 1.2.3. There, the relationships among these theories and the assumptions underlying simulator development can be best understood.

1.1.5 Transfer of skills

Transfer of skill refers to how skills acquired/learned when practising a task, are applied to perform another task (Adams, 1989; Holding, 1987; Osgood, 1949; Thorndike, 1903). For example, skill transfer can occur from performance on a training simulator to performance on the actual system or when a driver buys a new car, from driving the old to driving the new one. Skill transfer also happens between tasks or tasks variations. For instance, from driving in Spain to driving in Finland, from driving in summer to winter and vice-versa, when changing from one job position to another, etc.

Skill transfer is a critical issue for most training programmes because the effect of the training programme will be assessed as the degree to which skills acquired in the training context benefit performance in the actual job (e.g., Salas and Cannon-Bowers, 2001). When assessing transfer, three results can be obtained: Positive transfer, if practice in the training situation is transferred to the operational one. Negative transfer, if practice in the training context impairs performance in the actual system. And, zero transfer, if training has no effect on performance in the operational situation (Adams, 1989; Holding, 1987). Hence, training programmes pursue high and positive transfer to the operational system, since negative and zero transfer are considered as training programme failures.

Transfer is an empirical measure. Some experimental paradigms have been devised which seek different goals in assessing transfer. The paradigm most commonly used to assess transfer compares performance in the training programme with on the job performance; this is called the forward transfer paradigm. Other methods for assessing transfer, such as system trainer to training simulator transfer, can also be used to assess the effectiveness of training. Some of those transfer paradigms and measures will be described in the remaining of this section.

Paradigms of transfer deal with the experimental preparations which seek the assessment of transfer of learning from a task, context, device, etc., to other different conditions. In the psychological and human factors research literature, three main paradigms of transfer have been proposed: Forward transfer, backward transfer, and quasi-transfer paradigms (see e.g., Adams, 1989; Goettl and Shute, 1996; Holding, 1987; Lintern, et al., 1987). Typical transfer experiments use two groups to test planned hypotheses, the experimental group, and the control group. The experimental group is trained to perform the first task (Task A), and then transferred to perform the second task (Task B). The control group only performs task B. Thus, differences in performance between the experimental and control groups are due to transfer from task A to task B, given that the control group has not been trained in task A.

Forward transfer consists of the assessment of the transfer from one task to the next. This is the classical paradigm used to assess transfer from a training device to the on-the-job performance situation. Thus, this paradigm is used to assess cost-effectiveness of simulator training, for instance.

Backward transfer consists of assessing the degree to which skill performance on actual equipment transfers to operational equipment. If positive transfer is found, the results using this paradigm indicate that positive forward transfer is likely to occur from the training device to the actual system. In this paradigm, skilled personnel are used to test transfer from the operational equipment to the training device. It can be useful in assessing the potential effectiveness of training devices under development. Nevertheless, it does not provide an estimation of the amount of forward transfer likely to occur.

Quasi-transfer intends to assess the extent to which training in one training device configuration transfers to another configuration. This paradigm can be useful in providing measures for selecting cost-effective configurations of training devices. The rationale behind this suggestion is that, for example, a less expensive configuration of a training device which produces positive transfer to a more expensive configuration, could be targeted to be implemented in the training device instead of a more expensive configuration. Of course, the training effectiveness of both configurations should be well established before deciding on which configuration is to be implemented in the training device.

Quantitative measures of transfer provide the possibility of further analysis to select, develop, and implement cost-effective training devices for a training programme. A basic index of transfer is percent transfer of training. This formula calculates the percentage of improvement in performance in task B as a result of training in task A. This calculation is carried out for the initial period of performance in task B (Adams, 1989; Holding, 1987). The mathematical expression of percent transfer of training is shown in the formula below; this formula is adequate for calculating transfer of training when lower scores indicate improvement of performance such as better speed and error scores (e.g., reaction time, number of trials to learn).

$$\text{Percent transfer} = \frac{(\text{control group}) - (\text{experimental group})}{\text{control group}} \times 100$$

The formula expressed below is adequate when lower performance scores indicate degraded performance, such as, accuracy measures (e.g., correct responses).

$$\text{Percent transfer} = \frac{(\text{experimental group}) - (\text{control group})}{\text{control group}} \times 100$$

However, as Holding (1987) noted, these percentages of transfer may not remain constant as training on task B progresses. Thus, for assessing training effectiveness, other sensitive measures may be more appropriate. In particular, for training devices which require cost-effectiveness assessment, the formulas expressed above are useless.

Williams and Flexman (1949) evaluated a training device in terms of the number of trials needed to reach criterion. The question under hypothesis was whether the experimental group required fewer training trials in order to learn in the aircraft after training to criterion in the training device than the control group. They found that the experimental group needed fewer training trials on the aircraft than the control group. The percent transfer of training was modest, and they also computed an efficiency ratio, which is an index of how much practice is needed on the training device to achieve time savings in the actual equipment (Adams, 1989; Holding, 1987; Roscoe, 1971).

This index is called the incremental transfer effectiveness ratio (Roscoe, 1971), or cumulative transfer effectiveness ratio (CTER) (Holding, 1987). Despite the different terms used by different authors, the ratio is identical. The cumulative transfer effectiveness ratio expression is shown below.

$$\text{CTER} = \frac{(\text{hours needed control}) - (\text{hours needed after training})}{\text{simulator hours}} \times 100$$

In this formula, hours are one of the possible measures that can be used to calculate the cumulative transfer effectiveness ratio. Other measures can be trials, money, or whatever is best for deciding on the implementation of a training programme. One aspect that has to be taken into account in calculating incremental transfer effectiveness ratios for training devices is that increments in time devoted to training devices for delivering training produce diminishing returns. This means that a point would be reached where increments in time spent on the training device would not produce cost-effective training. This is because, the training time needed in the operational system being equal, any excess in the time spent on the training device over which performance does not improve is useless, and costs unnecessary money.

1.1.6 Training by means of simulators

Learning theories are descriptive and specify the conditions under which knowledge and skills can be acquired (Bailey, 1982). Training theories are prescriptive because they specify effective and efficient ways to obtain knowledge and skills (Bailey, 1982). Thus, training is the systematic organisation of the learning process. Its goal is to enable the trainee to acquire (learn) the knowledge and skills needed to perform new tasks, and to transfer those to the actual job situation in the most cost-effective manner.

Any training programme must ensure that the knowledge and skills required to perform proficiently at a certain job, are provided to trainees. Thus, the outcome of the training programme should be the acquisition, retention and transfer of skills by the trainees. The training programme may deal with different types of tasks for which different training methods may be required.

The terms instruction and training are basically used interchangeably in the human factors literature. Both refer to the problems of what kind of information should be provided to the learner/trainee, when it should be administered, how the learning/practice experiences should be arranged, and by what means the instruction/training programme should be supported (i.e., instructor, computer-based, simulator, on-the-job).

The differences between instruction and training are very small. Both share the final goals of the learning/acquisition of knowledge and skills, and the transfer of these to somewhere outside the learning/instructional/training setting. This somewhere outside of the instructional/training setting may be the society outside the classroom in a school, or the job that a professional must execute. The term instruction is used more for the organised learning process in educational environments, while training is widely used in the field of human-machine systems (Adams, 1989).

A more constrained position about the differences between instruction and training is the one suggested by Annett (1991), and Holding (1987). They distinguish instruction and training by referring their arguments to the paradigms used in the processes of skill acquisition research. Thus, they distinguish two main research paradigms, practice and instruction, each having their own characteristics. In practice (which could be replaced by training in an applied situation) experiments, the trainee makes repeated attempts to perform the task (Annett, 1991). Within instruction, the experimenter (or trainer) supplies the learner with verbal instruction, advice, and correction (Annett, 1991; Fitts, 1962, 1990; Holding, 1987).

1.1.7 Defining training simulators

A broad conceptual definition of simulators can be as follows: Simulators are devices that attempt to duplicate as closely as possible the characteristics and environment of real systems run by one or more operators (Flexman and Stark, 1987). The purpose of simulators is to provide appropriate information and response capabilities to trainees so that they can practice and learn the skills they will perform in the operation of the real systems.

Two remarks must be made about this vague definition. First, the generality of this description means that simulators intend to resemble a complete operational system such as a ship, for instance. Other training devices may be devoted only to training tasks on a sub-system which is a part of the ship system. Second, despite the intention of closely duplicating the actual system, simulators do not need to be exact replicas of the real system. Hence, simulators can resemble the system more or less accurately, but they are still simulators (Adams, 1989).

The general characteristics defining training simulators were proposed previously. But, they do not help much in distinguishing simulators from other training devices. However, distinctions based on their functions and characteristics discriminate better between simulators and other devices.

Two main functions can be observed in training simulators. Firstly, they present information similar to that supplied by the operational system which requires skilled operators. Simulators store, process, and display information that reflects the functional characteristics of the system. They also store, process and display information about the effects of pertinent environmental events, and the effects of control inputs performed by the operator (Flexman and Stark, 1987). Secondly, simulators incorporate special instructional features that propitiate and increase their capacity to support practice on the tasks and the acquisition of skills. The purpose of these instructional features is to facilitate the trainees' knowledge and skill transfer to the operational system which is being simulated (Adams, 1989; Caro, 1988; Flexman and Stark, 1987).

According to Flexman and Stark (1987), training simulators posses six basic characteristics that differentiate them from other training devices. These characteristics are: synthetic, data storage and processing capacity, the dynamics of the system, the information displayed in response to control inputs, the support of training on the whole task, and the instructional control of the information provided to facilitate learning.

Simulators are constructed only to provide task information rather than supporting actual operational functions. In-flight simulators, for instance, are bound by the same parameters as a normal aircraft, but they have synthetic control systems that can be adjusted to represent different types of aircraft (Flexman and Stark, 1987).

Simulators store data that represent the dynamics of the system being simulated, and the task-relevant portions of the environment in which the system must perform. Data storage and processing are functions usually performed by a computer controlling the simulator (Flexman and Stark, 1987).

The responses of a system to a control input or to some external factors may be important variables for the trainee learning to understand, control, and employ the real system in order to perform the assigned tasks. The simulator uses its stored and processed information to simulate the dynamic responses of the actual system, which is provided to the trainee in the training session (Flexman and Stark, 1987).

Simulators incorporate controls and displays that are as real as possible in order to support trainee skill acquisition. All the elements of the actual system interface are represented, because they are expected to enhance learning of accurate perceptions and their use in the operational system (Flexman and Stark, 1987).

Simulators are developed to support full-task (FT) rather than part-task (PT) training. Therefore, simulators allow the trainees to practice the tasks under conditions similar to those of the actual tasks, such as workload, stress, and time pressure. The flexibility of training simulators permits them to be used to provide trainees with complementary practice on difficult or critical task elements. The training simulator forms the context in which previously acquired skills or sub-skills are integrated to perform the whole task (Flexman and Stark, 1987).

Another characteristic of training simulators is their ability to control the information that supports training. Information is provided to trainees so that, through practice, they can develop the skills necessary to perform their operational activities. The purpose of instructional control is the facilitation and enhancement of skill acquisition and transfer of knowledge and skills to the operational situation. Instructional control is commonly accomplished through the intervention of an instructor. The instructor can participate directly in the conditions of the training process or the role of the instructor can be implemented in a computer program that controls the progress of training for maximum training effectiveness. See, Flexman and Stark, (1987) for a comprehensive discussion on this.

In conclusion, training simulators are devices that attempt to duplicate the characteristics and the environment or context of the operational systems which they represent. Simulators support two main functions: Firstly, they present information similar to that of the real system; and store, process, and display information about the effects of environmental conditions as well as the effects of control responses executed by the operators. Secondly, training simulators provide training experiences by means of special instructional features to facilitate retention and transfer of training to the operational system.

1.1.8 Other training devices

As stated earlier, there are other training devices that share some of the characteristics of training simulators but have different labels. Two of the more common training devices are the procedure trainer and the system trainer.

The procedure trainer is one of the lower level devices in the category of training equipment. It is designed to provide training in the basic procedural steps in system operation. The trainees are presented with enough information so that they can observe and learn those basic procedures. Procedure trainers contain sufficient information to allow fundamental procedural control inputs to be reflected in the displays directly related to the control (Flexman and Stark, 1987). An example of this category could be a trainer for learning the starting up procedures of a power plant. Similar examples can be found in the naval context, for instance, in the use of a procedures trainer for the use of automatic radar plotting aids.

System trainers are more advanced than procedure trainers in the extension of the simulation they provide. System trainers offer broader simulation than procedure trainers, but they deal only with the tasks associated with a particular sub-system, e.g., fresh water generation plant aboard ship. Usually, training in some sub-systems, such as a radar system, involves practice in difficult or critical tasks, and therefore the achievement of its training goals are time consuming. Thus, a system trainer can be developed to provide practice on those perceptual tasks without the need for expensive training simulators which support most functions of a system such as a ship or aircraft.

Operator training and assessment is currently being carried out in a training setting supported by different training media. Although these devices are generally efficient in accomplishing their goals, some questions, such as the efficiency in promoting transfer of skills, still remain unsatisfactorily answered (Barnett and Ceci, 2002; Salas and Cannon-Bowers, 2001). A new concept in training delivery has appeared to tackle the deficiencies of other training devices. The development of computer hardware and software capabilities has enabled this development in simulator training.

The concept is embedded training (Caro, 1988; Thomson and Spears, 1990). Embedded training means the use of sophisticated simulators for training operators which are embedded within the operational system (Thomson and Spears, 1990). It means that hardware and software capabilities of the actual operational system are used to deliver training and test the operator's performance while the operational system performs its own tasks (Caro, 1988). Embedded training does not necessarily share the same hardware and software equipment of the operational system; rather, it can be built into or added to the operational system (Thomson and Spears, 1990). In this sense, embedded training may be a procedure, if it is implemented on the operational system, or it can be a device, if it is implemented on an independent machine.

The purposes of embedded training are twofold: first, to provide opportunities for the trainee to practice his/her performance on-the-job situation and second, to provide special and unique learning opportunities. Embedded training has some advantages over other traditional approaches to training. Firstly, it allows immediate on-the-job experience on the operational system. Secondly, it provides self-paced practice in critical, hard to master skills within the operational system without putting it at risk. Thirdly, by simulating actual workstation equipment, the embedded training component can provide the trainee with realistic, system specific cues, responses, and displays (Thomson and Spears, 1990).

Applications of embedded training will increase as cockpit designs incorporate more programmable control and display functions. Through programming, these systems can simulate operational events such as hostile electronic countermeasures, and the effects of own-ship weapons (Caro, 1988).

Though embedded training is a relatively new training concept, it shows promise for solving some problems in training, such as how to train performance in outstandingly difficult tasks, very uncommon situations, and complex tasks, and to ensure transfer to the actual system. Still, research on this topic is scarce, and conclusions about its effectiveness are lacking (Caro, 1988; Stedmon and Stone, 2001).

1.1.9 Simulator functions and instructional features

In previous sections, training simulators and other training devices have been described. To accomplish the training objectives established for training simulators, these make use of special training capabilities. These special training capabilities allow simulators to perform supplementary functions that facilitate and enhance the results of the training programme (Flexman and Stark, 1987). Generally speaking, what are called supplementary functions by Flexman and Stark (1987), are termed instructional features by Sticha et al. (1990). Even though the terminology is different, we consider that instructional features are enclosed within supplementary functions, in the sense that special simulator functions are carried out by, or implemented as, instructional features. The main difference, as will be shown, is that while supplementary functions may be performed either by the instructor or the simulator, instructional features are implemented as simulator functions. This is, the simulator performs the functions which are enabled by the instructional features.

Supplementary functions will therefore be detailed in the next sections including: briefing and demonstration, provision of practice, performance analysis, enhancement of learning, performance assessment, and practice under adverse conditions. Also, the associated instructional features for each function will be described. A summary of simulator functions and instructional features is provided in Table 1.

1.1.9.1 Briefing and demonstration functions

As an instructional feature, the briefing is the pre-training activity which serves to prepare the trainee for particular training objectives. It may include a review of a trainee’s past performance or an audio/visual description of the next exercise (Sticha et al., 1990). Briefing information can be administered through a cathode-ray tube display, or by means of sound recordings synchronised with the automatic demonstration (Semple, Cotton, and Sullivan, 1981).

Demonstration may be administered in different ways. The instructor can provide demonstrations from his/her normal position in the system, or they can be provided from the controls on the instructor’s station, or from the trainee’s controls.

TABLE 1
Summary of Simulator Functions and their Corresponding Instructional Features

Simulator Functions				
Briefing & Demonstration	Practice	Performance Assessment & Analysis	Learning Enhancement	System Malfunction & Failure
Instructional Features				
Briefing Utility	Initial Conditions & Scenario Control	Instructor Operating Station	Simulator Record/Replay	Malfunction Control
Automated Simulator Demonstration	Automated Adaptive Training	Automatic Performance Measurement	Closed-Circuit Television	
	Automated Controllers	Hardcopy/Printout	Automated Performance Alert	
	Computer Controlled Adversaries	Remote Graphics Display/Replay	Automated Cueing & Coaching	
	Reposition	Data Storage & Analysis	Computer-Managed Instruction	
	Freeze	Procedures Monitoring		

Automated demonstrations serve the function of providing a model of the desired performance by allowing the instructor to pre-record and replay a certain activity. The defining characteristic of this feature is that it is permanently stored in the simulator. It is assumed that demonstrations are more useful for training when the trainee is learning a new and difficult skill (Sticha et al., 1990).

1.1.9.2 Practice functions

The ultimate goal of simulator training is to provide the trainees with the knowledge and skills necessary to efficiently perform their jobs. Practice methods consist of trainees’ performing the tasks in the training programme so that knowledge and skills are acquired through the practice on these tasks.

Thus, practice is, or should be, the main ingredient in most training programmes because practice methods are effective in providing knowledge and skill acquisition throughout training (Holding, 1987).

Theoretical and empirical research suggest that practice methods are effective if other factors are also taken into account (Cannon-Bowers, Tannenbaum, Salas, and Converse, 1991; Schneider, 1985). Some of these factors include the characteristics of the task, the information provided to trainees about the task goals, the information provided during task performance. This is, information after action or feedback and knowledge of results. Furthermore, practice methods should be assisted by other training methods such as verbal instruction, demonstrations, and guidance, to ensure the achievement of the training goals. The latter can be served by some of the instructional features of the simulator as described here.

The basic purpose of simulators is to provide the controls and display information to trainees when practising their goal tasks. It means that simulators support practice on the tasks being trained (Flexman and Stark, 1987). Simulators are valuable devices because they are able to reproduce and present information about the conditions of the system and the environment necessary for the acquisition of the skills for which the practice session has been designed. It is important at the early phases of training so that the trainee has the opportunity to experience, and recognise, the effects of control inputs on the performance of the system. This facilitates the acquisition of the skills necessary to perform the tasks.

According to the instructional features classification, different instructional features can contribute to the provision of practice, as follows.

To arrange a practice session in advance, the initial values of a variety of environmental and system dynamic conditions and parameters must be pre-stated. The initial conditions feature enables the simulator to pre-select and store these parameters in order to set these conditions rapidly (Sticha et al., 1990). This feature may be contained within the scenario control feature. Scenario control enables the instructor to configure and control the simulator so that simulated events occur in accordance with a specific training scenario. Training scenarios are highly structured sequences of events intended to provide the trainees with practice on various tasks, such as landing manoeuvres (Sticha et al., 1990). Scenario control may involve pre-specified practice events, or it may be controlled by adaptive training algorithms.

Automated adaptive training is an approach to training in which the difficulty of a task is adapted to the skill level of the trainee. Training starts at relatively low difficulty levels that increase as trainee performance on the task improves. Automated adaptive training permits the instructor to select the adaptive variables. These variables are then used according to some instructional sequencing algorithms by the simulator. The sequencing algorithm is based upon trainee's performance on the previous trial/s. For example, the algorithm may state, IF performance on the previous trial is correct, THEN increase difficulty on the next trial. Even though adaptive training algorithms can be used in scenario control, as seen before, automated adaptive training solely pursues the adjustment of task difficulty to trainee skill level to enhance skill acquisition rate, not the control of the training scenario.

The simulator supports practice, and provides quite accurate information about the controls and displays of the system in response to control inputs. In presenting information as accurately as possible, it may be necessary to present information about controller information to the operator (Sticha et al., 1990). This feature is especially important in flight training, as information from the controller may be vital for mastering the tasks being trained. This feature may be fully automated by computer-based voice recognition (i.e., simple requests from the trainee) and voice synthesis to provide adequate controller responses. It can also be performed by the instructor. In this case, the computer calculates the information that has to be provided to the trainee, and the instructor provides it.

Simulation of adversary weapons, aircraft, ship, etc., is important in tactical training. Computer-controlled adversaries are therefore computer models that allow for the simulation of enemy system performance. Computer adversaries can be fully automated, or can be under the control of the instructor. Again, the accuracy of task information is important in promoting skill acquisition, and hence, adequate performance in the operational situation.

Reposition permits the instructor to position the simulated system in a state which is relevant to the training scenario. For example, positioning the simulated aircraft close to the landing track, instead of flying for minutes before that situation occurs. This feature allows for time saving in training, and is also useful for practicing extremely difficult tasks, because useless practice can be avoided (e.g., practicing regular flight after that skill has been achieved) (Sticha et al., 1990).

The freeze feature involves the capability to stop all or selected parts of the training scenario for training purposes. Activities can be frozen by the instructor or automatically stopped by the simulator under certain conditions. Freeze possibilities range from stopping the whole system to freezing only a part of it, while other components continue functioning. The freeze feature is frequently used to train procedural components of a task (Sticha et al., 1990).

1.1.9.3 Performance assessment and analysis functions

Learning and performance are interrelated, although they are different concepts with different meanings. The relationship comes from the fact that learning is inferred from the operator's performance measurements, either on-the-job or in the training situation, on condition that a valid and reliable relationship has been established between performance and learning on that task (Flexman and Stark, 1987). Usually, performance measurement is difficult in the operational task performance. Yet it is even more difficult to relate operational task performance to skill acquisition level, due to uncontrolled variables affecting performance.

Since most events represented on the simulator involve a mathematical expression or a measurable input to or output from the simulator computer, these events are suitable for measurement and analysis. As simulators provide a great amount of information, effective analysis is dependent on the adequate selection of the relevant information to assess aspects of performance and learning processes. Thus, this simulator function allows for the filtering of relevant information related to performance and learning measures, and avoids redundant or irrelevant information.

Performance assessment carries out three functions in the training programme. First, it rates the trainee's performance with respect to the task demands for which that person is being trained. Second, performance assessment determines when the trainee is ready to progress from one training stage to the following one. Third, it identifies and diagnoses the causes of incorrect performance, or performance which appears to reflect hindered learning progress.

The instructor operating station feature carries out the function of supplying information to the instructor about a trainee's current performance during a simulated mission. This information may be displayed alphanumerically or in graphic form. This feature monitors discrete performance, and allows the instructor to assess a trainee's performance and detect possible problems in the training progress.

Automatic performance measurement calculates quantitative measures of a trainee's performance. It thus assesses a trainee's progress, and provides information for diagnosing trainee performance problems. This information is not used as direct feedback to the trainee, but is used by the instructor to evaluate the trainee. This information may also be used as input to other instructional features.

The printing feature provides a permanent record on paper of the performance measurement data provided by the automatic performance measurement feature. Recorded data may be used for debriefing trainees, to monitor trainees' performance, or for course evaluation.

Remote graphics display/replay supplies a graphic or symbolic display of trainee's performance. It also provides information to the instructor about current performance status (Semple, Cotton, and Sullivan, 1981). Similarly, it can be used to provide detailed post-training performance feedback to the trainee (Sticha et al., 1990).

Data storage and analysis serves to store, analyse, and retrieve data from trainees, groups of trainees, or the simulator itself. Individual data can be used for briefing in pre-training and group data can be used by instructors to evaluate the training programme.

The procedures monitoring feature allows the instructor to monitor trainee performance of normal and emergency procedural tasks. It monitors continuous performance. The procedures monitoring feature is similar to the instructor operating station, but this feature monitors discrete responding.

1.1.9.4 Learning enhancement function

Simulators possess a number of characteristics that make them effective and efficient in enhancing trainees' skill learning. This is accomplished by means of the instructional features that simulators incorporate. The capability of instructional features in enhancing learning comes from their ability to present immediate clearly defined performance feedback to the trainee, knowledge of results, augmented feedback, standardised presentation of task conditions, and the ability to implement adaptive training according to the trainee's skill acquisition progress. In the following section, some instructional features which are defined explicitly as implementing learning enhancement functions will be described.

The record/replay feature allows the instructor to record a trainee's performance during a simulation run, and replay it later for its review with the trainee. This feature can provide the trainee with knowledge of results and feedback information about his performance. It is most useful when trainees are learning a new difficult skill or when detailed performance feedback is required (Sticha et al., 1990).

The purpose of the closed-circuit television feature is to monitor and record the observable behaviour of the trainee while performing the task at the simulator. After the performance of the trainee has been recorded, it is replayed to him/her during the debriefing session. It provides performance feedback, and the opportunity to review and rehearse one's activities.

Performance alerts are visual or auditory signals presented to the trainee or the instructor when performance tolerances have been exceeded. The purpose is to enhance monitoring abilities of both the trainee and the instructor (Sticha et al., 1990). Thus, it provides feedback and performance cues so that learning can be enhanced by the use of this feature.

The automated cueing and coaching feature is activated when performance tolerances have been surpassed. It is similar to the automated performance alert, but in this case the feature provides the trainee with a coaching message that asks the trainee to take corrective actions. The coaching message can replace the alert signal or can be added to it. This feature is especially useful in self-administered training (Sticha et al., 1990).

Training functions can be managed by the computer. For example, it can assess which training objectives have been achieved and make appropriate assignments for new exercises (Semple, Cotton, and Sullivan, 1981). Arranging training experiences according to achieved goals should enhance both performance and skill learning progress.

1.1.9.5 System malfunction and failure function

Most human-machine systems have reached a degree of complexity and reliability that system failures are rare. The occurrences of failures can have disastrous consequences because of the increased reliance placed on the system.

Thus, the skills required to execute corrective actions in response to system failures are highly important. It is almost impossible to train these skills in the operational situation for practical, economic, and safety reasons. Therefore, simulations of system malfunctions must or should be implemented for training in the context of simulators.

The malfunction control feature serves the function of simulating system malfunction and failures in a training scenario. It therefore provides training in emergency procedures which are of extreme relevance to performance in the operational system. This feature permits the instructor to insert simulated malfunctions in a given training scenario. Malfunctions and failures can be inserted either manually or automatically.

1.2 Improving simulator training effectiveness

Certainly, the number of factors which might determine simulator training effectiveness can be countless and vary widely across different viewpoints and domains of application (see e.g., Hays, Jacobs, Prince, and Salas, 1992). Training effectiveness is considered here as a function of two dimensions: economic and skill level.

It seems that within the human factors, and within the realm of training practice and research, there are two opposed trends guiding the design and development of training programmes. One posits that training should progress from easy to difficult training activities (see e.g., Carlson, et al., 1992; Bainbridge, 1993). The other proposes that training should be made so that similar conditions, including the difficulty of the tasks, should be presented during training (see e.g., Schmidt and Bjork, 1992; Schmidt, 1991).

1.2.1 Verbal instruction

Verbal instruction, apart from other training methods such as simulator training, plays an important role in most training programmes in different domains, e.g., industrial process control systems, military systems, etc. A training session currently starts by briefing the trainees. The briefing is an act in which the trainer/simulator provides the essential information required by the trainees to perform the tasks, and gives precise instructions about what the trainees are expected to do in that training session. The briefing can be administered to trainees in different ways on the simulator. Depending on the facilities the simulator includes, briefing can be delivered verbally by the instructor. Also, the instructor can be coached by the simulator in maintaining the same structure and content of the briefing session. In addition, the briefing can be automatically administered by the simulator. When provided automatically by the simulator it is considered to be an instructional feature (Flexman and Stark, 1987).

Verbal instruction consists of telling the trainee what to do by means of speeches, lectures, discussions, and also, written materials (Holding, 1987; O'Hara, 1990). It can be used to provide the trainee with the instructions for performing the task, for outlining the task goals, for procuring knowledge of results about his/her performance etc. Verbal methods are efficient and economical when used for these purposes. However, in many cases the translation of verbal information into actions makes this method not recommendable for the whole training programme (Holding, 1987).

Verbal instruction could be used to provide information about declarative knowledge (Glaser, 1990), for instance, general principles such as understanding control of loss of altitude, speed, etc. Thus, at the beginning of training it could be more useful to provide conceptual knowledge than to practice the task immediately. This conceptual knowledge, in turn, has the disadvantage of not being specific to the context, and, consequently, additional training may be required to link it to the specific situation (Fredericksen and White, 1989).

One factor related directly to verbal instruction for general principles is how the ability of trainees affects the application of that knowledge to practical situations (Holding, 1987). As Holding suggested, application of specific rules to the actual situation may not be automatic. Therefore, verbal and practice materials could be intermixed in different ways to produce better transfer (Fredericksen and White, 1989).

Another factor which affects the effectiveness of verbal instruction is the complexity involved in the information that has to be transmitted. The more complex the information, the less effective this method becomes. Thus, as a rule of thumb, verbal instruction must be arranged in the simplest possible manner (Holding, 1987).

1.2.2 Part-task (PT) training sequences

The aims of PT are two-fold: first, it intends to reduce training costs by developing cheaper training devices than full-scale system simulators (Eberts and Brock, 1987; Wightman and Lintern, 1985), and second, to improve skill acquisition efficiency by promoting faster learning. PT training versus full-task (FT) training is an important research topic for several reasons. Research results on the relative benefits of training by means of FT or PT schedules yield contradictory results. Some researchers favour training according to a FT schedule since it avoids the problem of integrating component skills into a whole (Gopher, Weil, and Siegel, 1989). Other researchers have provided evidence of the beneficial effects of PT training schedules over FT schedules (Frederiksen and White, 1989; Wightman and Lintern, 1985). Additionally, the tasks to which these training regimes have been applied are commonly perceptual motor tasks (i.e., manual vehicle control, aircraft piloting, etc.)

(Wightman and Lintern, 1985). Little research has been devoted to cognitive procedural tasks such as those usually required in naval and other work environments (i.e., power plant operation, machinery control etc).

1.2.3 Visual cueing

The essential point of this method is that instead of trainees being told what to do, they are shown by different means what they are expected to do. Two ways of showing the trainees what to do can be visual demonstrations, or physical guidance (Holding, 1987).

Visual demonstrations can adopt the form of imitation, in which the trainee observes another performer on the task or at work (Shebilske, Regian, Arthur, and Jordan, 1992). Also, demonstrations can be carried out using films in which task performance is shown (e.g., Ivancic and Hesketh, 2000). Presenting the trainee with additional visual cues is another form of showing the trainee what to do. Generally speaking, these methods are usually supplemented by practice on the task itself, and not used in a separate fashion, as stated above.

The trainee learns by observing and imitating another's performance. The utility of this is not completely understood, but it is effective when compared to trainees who have not observed the performance on the task (Holding, 1987; Shebilske, et al., 1992). Social variables such as the model being present when trainees perform the task, the status of the model (i.e., teacher or peer), and skill level of the model performing the task affect the results of this kind of training. Thus, if the model is present when trainees perform the observed task, the trainees' performance on the task is better than that of trainees without the model being present. Furthermore, observing an unskilled teacher is less effective than observing an unskilled peer, and this is less effective than observing a skilled model of either kind (e.g., Bandura, 1977, 1982, 1986).

Imitation may be useful in the early stages of skill acquisition, when attentional resources are needed to perform the task and performance is error prone and slow. Observation provides the trainee with a standard to which he/she can compare his/her performance, and reduce the number of alternative actions (Holding, 1987). As we can see, imitation is not used alone, but is supplemented by practice in almost every case.

Visual cues are used to show the trainee what to do in some detail. Cues are additional information given to trainees so that they can associate cues, external to the task, to the responses required by the task, and thus be able to anticipate the correct response. Visual guidance may be used to facilitate errorless performance in simple movements, but it can detriment performance when cues are withdrawn by distracting the trainee's attention to response-produced feedback. Visual cues can be used early in training, but as training progresses, cues should be withdrawn progressively to avoid degradation of established performance and facilitate retention and transfer (e.g., Lintern, 1991a).

Physical guidance is especially important in motor learning. Physical guidance can be executed by stopping an ongoing movement, and by physically forcing a particular movement. Stopping a movement prevents the trainee from performing incorrect alternatives. Forcing a movement provides information about which is the correct response (Holding, 1987).

The relative effectiveness of response prevention and response forcing depend on the type of movements which are being trained. For serial movements such as typing, response prevention appears to be more useful than forcing responses (Holding, 1987). The availability of knowledge of alternative responses also affects the effectiveness of these methods. Thus, a trainee trained by forcing responses has no possibility of becoming acquainted with the different possible alternatives to perform the task. Hence, information about the different alternative movements should be inserted into the training programme (Holding, 1987).

1.2.4 Training simulator fidelity

Training simulators are developed with the aim that training on simulators will transfer to the operational situation for which the training programme has been developed. As will be seen, theoretical underpinnings of the training simulator transfer research have been based on the relationships between identity, degree of similarity, or fidelity functions as simulator characteristics, and the operational setting characteristics (Lintern, 1991b). Transfer of training and simulator fidelity, therefore, are intimately related in both research and applied contexts.

Training simulators are effective devices for facilitating knowledge and skill transfer from the simulator to the operational system (Adams, 1989; Caro, 1988; Flexman and Stark, 1987; Roessingh, 2002). Transfer effectiveness has been assumed to be due to simulator fidelity. The concept of realism or fidelity was, and is, even nowadays, based on Thorndike's theory of identical elements (Thorndike, 1903) and the theory of common elements (Thorndike, 1931). The condition of identity can happen, for example, in the case that both the training simulator task and the operational task have the same objectives, elements, or approaches. In this case, according to Thorndike (1903), training on one task should transfer to the other task.

Thorndike's theory suggests that skill transfer from the simulator to the operational system will occur if the simulator and the operational system share common elements (Caro, 1988; Thorndike, 1931). Thus, transfer will increase as the number of elements in common between the simulator and the system increase.

Osgood (1949) developed the concept of transfer surface based on an extension of Thorndike's (1931) theory. Osgood's theory relates training transfer to stimulus-response similarities between the task at the simulator and the task at the operational setting. Osgood's theory suggests that a correspondence of one-to-one between the elements or features of the simulator and the simulated system will produce a high and positive skill transfer (Masson, 1990).

More recently, from Anderson's (1987) ACT* model, Singley and Anderson (1989), and Speelman and Kirsner (2001), skill transfer from one task to another is predicted by the number of productions shared by the production systems that governs performance on both tasks. Anderson's ACT* model has also inherited characteristics of Thorndike's (1903) identical elements theory of transfer (see, Anderson, 1987; Robins, 1996; Singley and Anderson, 1989; Speelman and Kirsner, 2001).

The apprenticeship, i.e., learn by doing the actual task, mode of training in different sectors, such as arts, handicrafts, intellectual, and so on, has been the rule for centuries. Moreover, taking into consideration the theories discussed above, it is easy to understand why high simulator fidelity has been the main goal in developing training simulators both past and present. This is to say, that since skill transfer is the chief purpose of simulator training, the best way to achieve high skill transfer to the operational system is by designing and developing high-fidelity (HF) training devices. Despite this traditional and apparently reasonable logic, the problem is not so simple, as will be seen in the next sections.

1.2.4.1 Dimensions of simulator fidelity

Defining simulator fidelity is a difficult task. Generally speaking, simulator fidelity refers to the degree to which a training simulator represents the operational system and its environment (Flexman and Stark, 1987). Despite the clarity of the definition of simulator fidelity that Flexman and Stark provided, the landscape around this concept is a lot more obscure. Definitions of fidelity have been posited since the 1950's, but, after revisions by various authors, the only agreement is that there is no consensus among them concerning the nature of simulator fidelity (Hays, 1980; Semple, Hennessy, Sanders, Cross, Beith, and McCauley, 1981; Sticha et al., 1990).

In their review, Sticha et al. (1990) propose that fidelity is composed of two dimensions: realism and comprehensiveness. This definition tries to capture all the relevant factors which affect fidelity with the purpose of accurately assessing simulator fidelity. Based on that ground, the relationship between simulator fidelity and training effectiveness could be tested more effectively. This would therefore allow for savings in time and cost in both simulator development and training development processes.

Realism is one of the two dimensions proposed to characterise simulator fidelity. Realism is conceptualised as the measured similarity between the training device attributes and the corresponding attributes of the simulated system. Realism is affected by three types of attributes: (1) The configuration of the static displays and controls (e.g., the cockpit configuration, remote control console); (2) the dynamic responses of all non-static components (e.g., simulator motion, external visual images, process parameters). (3) The sensory stimuli generated by the training simulator (Sticha et al., 1990).

The realism of the sensory stimuli provided by the simulator is related to the realism of both static displays and controls, and dynamic responses. Nevertheless, the relationship between static and dynamic stimuli may be hindered by other variables. For example, dynamic display responses may be highly realistic. But, if perceptual variables such as display resolution, brightness, etc., are inadequate, then dynamic display responses may be unable to provide the adequate visual stimuli to perform the task. Therefore, if realism is low in one or various of the attributes which affect this dimension of fidelity, then fidelity is decreased (Sticha et al., 1990).

The other dimension of simulator fidelity is its comprehensiveness. Comprehensiveness is defined as the range of potential training applications that the simulator may have (Sticha et al., 1990). For instance, full mission simulators are more comprehensive than system trainers, instrument trainers, part-task simulators, and so on.

The attributes which determine simulator comprehensiveness are the following: (1) Dynamic response range. This attribute refers to the range over which simulator components are able to respond in a realistic manner; (2) range of simulated operational tasks that can be performed in the simulator; (3) the range of simulated operational conditions which can be simulated (e.g., simulator may include adverse operational conditions, equipment malfunctions, etc.); (4) The range of sensory stimuli that can be generated by the simulator. This refers to the range of sensory stimuli provided by the operational system which is also provided by the simulator, e.g., visual information, movement, vibration, g-force.

It has been suggested that realism and comprehensiveness are separable dimensions of fidelity. Nevertheless, both realism and comprehensiveness are constituent parts of simulator fidelity. Training simulators and other training devices can be assessed independently on both dimensions. However, in practice it would be incorrect to assess simulator comprehensiveness without taking into account the realism of its components. The reason for this argument is as follows: given a simulator which is highly comprehensive, but the realism of its attributes is so low that it makes effective training on critical tasks impossible, then the high level of comprehensiveness is useless (Alessi, 1988).

Obviously, in practice, this situation is highly unlikely to occur. However, little is known about the relationships between realism, comprehensiveness, training effectiveness, and cost of training devices. Despite some attempts in this direction (see, Alessi, 1988; Hays et al., 1992) Sticha et al., 1990), more research on this subject should prove its value in selecting cost-effective training methods and simulators.

1.2.4.2 Simulator fidelity and training effectiveness

HF simulators, especially flight simulators, have been developed from assumptions that are in agreement with the theoretical proposals reported above and the apprenticeship tradition. They have been used mainly to reduce the costs of on-the-job training, and to provide training in tasks that when performed in reality would cost more money and perhaps lives (Adams, 1989; Caro, 1988; Orlansky and String, 1978; Sticha et al., 1990). Nowadays, budgetary arguments are accompanied by other factors when determining the use of simulators for training. Two of the factors considered important for the adoption of simulator training are: their ability to increase skill level over that achievable on the operational system, and to provide training on tasks that cannot be performed on the actual equipment (i.e., accidents and system failures) (Adams, 1989).

Many experiments on relatively HF flight simulators have been conducted until now. Although percent transfer values have been relatively low, the cost of training on these simulators, compared to aircraft training, has made them profitable (Alessi, 1988; Diehl and Ryan, 1977; Lintern, 1991a; Orlansky and String, 1977).

Other experiments, however, have shown that effective transfer of training in some tasks can be accomplished with the use of low-fidelity (LF) simulators. Studies on procedures training have shown that LF devices, such as a photographic 'mock-up' of a cockpit, produced as much transfer to the aircraft as a HF simulator (Dougherty, Houston, and Nicklas, 1957; Prophet and Boyd, 1970).

In part based on experimental results such as those cited above, in part because increasing fidelity costs more money, i.e., this can make simulators low cost-effective for training, and in part because a better use of technology (for example, PT simulators, computer based training, etc.) can produce cost-effective training, the emphasis is nowadays placed on low-fidelity (i.e., low-cost) effective training devices (Alessi, 1988; Salas and Cannon-Bowers, 2001; Sticha et al., 1990; Van Matre, Ellis, Montague, and Wulfeck, 1993).

2 GOALS AND HYPOTHESIS OF THIS STUDY

The present research aims to improve the effectiveness of simulator training in the area of the prerequisite skills required by novice trainees in order to operate a ship machinery control system. The skills mainly addressed in the following experiments can be characterised as procedural (Anderson, 1982, 1987; O'Hara, 1990; Rasmussen, 1993). The strategies adopted to improve training effectiveness in terms of skill acquisition speed-up and its relationship to training costs are based on psychological and applied training research. One general hypothesis is that the use of research derived principles for simulator training, in particular, is useful as well as desirable. This also applies to other training strategies such as on-the-job training. Of course, common sense indicates this, but this hypothesis needs constant and cyclical efforts to obtain empirical support. Training theory must be applicable in practice and the results obtained after that application should be used for further theoretical developments. Thus, the cycle can roll once more.

It is postulated here, that simulator training technology per se does not ensure the most effective training results that could be expected. Nevertheless, simulator technology which facilitates the implementation of training principles derived from research would improve the effectiveness of training. Three factors that can affect the effectiveness of training provided by means of simulators are investigated in this research: Part-Task training strategies (Experiment 1), augmented cueing (Experiment 2), and simulator fidelity (Experiment 3). Consequently, three hypothesis stem from these factors:

Part-Task training is hypothesised to produce higher level of skill achievement than full-task training schedules, in terms of task performance accuracy and speeded task performance. Additionally, training provided through a part-task schedule that addresses a critical component subtask (malfunction diagnosis), is assumed to facilitate the acquisition of more effective performance strategies than simple part-task and full-task schedules. These hypotheses are tested through the transfer from one task to a variant of it with increased complexity in the same simulator.

Augmented cueing training implemented on a low-fidelity simulator is postulated to produce higher skill acquisition levels than training with the absence of this strategy, in terms of task performance accuracy and speed. In addition, training on a low-fidelity simulator which implements this strategy is hypothesised to facilitate the development of more effective performance strategies than training without this feature. These hypotheses are tested through the transfer from one task to a variant of it with increased complexity in the same low-fidelity simulator. This training strategy cannot be implemented in the high-fidelity simulator employed in Experiment 3.

It is hypothesised that training by means of a low-fidelity simulator which incorporates a training strategy not available in the high-fidelity simulator, would eliminate the interference provoked by the change from task performance on one HMI (i.e., workstation computer), to another HMI (i.e., remote control console and local control panels). It is also postulated that the skill level facilitated by both simulator-fidelity configurations differs between the low-fidelity and the high-fidelity simulators. It is expected that the low-fidelity simulator would produce higher skill level in terms of task performance accuracy and speed of performance. These hypotheses are tested through the transfer from one task to a variant of it with increased complexity and the transfer from performance on the low-fidelity simulator to the high-fidelity simulator.

The experiments testing the hypotheses postulated above are reported in the following sections.

3 EXPERIMENT 1: PART-TASK TRAINING

3.1 Introduction

Training through PT versus FT strategies is important for three main reasons: In the first place, research results on the relative benefits of training by means of FT or PT have yielded contradictory results. Some researchers favour training according to a FT schedule since it avoids the problem of integrating component skills into a whole (Gopher, Weil, and Siegel, 1989). Others have provided evidence of the beneficial effects of PT training schedules over FT (Frederiksen and White, 1989). In the second place, the tasks to which these training regimes have been applied are commonly perceptual motor tasks (i.e., manual vehicle control, aircraft piloting, etc.). Little research has been devoted to cognitive tasks such as those usually required in naval environments (i.e., machinery control, radar operation etc). Finally, if PT training is more beneficial than FT training, then training could be achieved more efficiently in economic and performance attainment terms (Eberts and Brock, 1987). Therefore, this experiment aimed at obtaining empirical evidence of the relative value of training schedules based either on PT or FT strategies.

The machinery aboard ship are very complex systems. From the different ship machinery sub-systems, a diesel generator was selected as feasible to implement the experiments in this research. Consequently, the task employed was the operation of a diesel generator. It is worth noting that modern ships possessing the sort of machinery control equipment (direct manipulation interactive process control software) are very similar. Differences can be found in the type of engines and the HMI that different ships incorporate. The structure and functions of the system are alike. The generators aboard ship provide the electricity supply so that many other pieces of equipment can work.

Operating a diesel generator involves the execution of tasks both in normal and emergency conditions. Emergency conditions should not appear on an isolated diesel generator system unless the generator is working. In practice, as derived from pilot studies, the complete isolation of the diesel generator was not possible. Alarms pertaining to sub-systems shared by other sub-systems could eventually be triggered, e.g., alarm of low level in the fresh water generation plant, due to continuous filling-up of the fresh water expansion tank when it has a leakage. The fresh water generation plant is not represented in the diesel generator screen and is a sub-system which serves the cooling of other generators as well as the main engine among other devices (Benítez-Domínguez, 1996).

Operation of a diesel generator can be fractionated (Wightman and Lintern, 1985) into two main subtasks:

- starting up the system to its normal working condition, and
- supervising its behaviour and solving the malfunctions if they occur.

Performing both fragments of the task requires the execution of procedures. This feature of the task is more obvious in the start-up procedure for which a number of steps have to be performed in a sequential order as shown in Table 5. Supervision of the system's behaviour and solving its malfunctions also require the execution of procedures, (see Table 6), if emergencies threaten the system itself and the devices (i.e., electrical consumers) being served by it. The operation of the diesel generator under normal conditions chiefly involves executing the procedure to start it up. In this case, memory skills to recall the adequate start-up procedure are the main skill components involved in the execution of the correct procedure. The supervision of the system behaviour normally occurs in between the execution of the start-up procedure and the execution of the malfunction solving procedure. This should be the case if trainees were fast enough in starting-up the generator. In practice, as observed in pilot studies, if the trainees did not perform the start-up procedure fast enough, alarms appeared before the finalisation of the previous procedure (e.g., start-up). In this situation, time-sharing skills or task change must be applied (see e.g., Ackerman, Wickens, and Schneider, 1984; Eyrolle and Cellier, 2000; Tsang, Velazquez, and Vidulich, 1996; Wikman, Nieminen, and Summala, 1998, for related research on time-sharing skills and tasks). That is, both procedures can be performed in an intermixed way (i.e., time-sharing), or a change can be executed from the start-up to the malfunction solving procedure (i.e., task change), and back to the start-up procedure. In accordance with the experience of the subject matter experts, malfunctions are less frequent than one might think. Nevertheless, for this reason it is very important to provide training on such situations. Subject matter experts also consider these to be a crucial training goal. When emergencies occur, the procedures to be executed in order to restore the normal functioning of the system rely heavily on perceptual and decision-making skills.

Motor knowledge and skills are required to operate the different devices on the HMI (i.e., valves, electrical pump, etc.) thus contributing to the integral task performance of the operators. Motor skills contribute to the performance of procedures in both normal and emergency conditions. Despite its procedural nature, the operation of the diesel generator can be defined as a complex task (Schneider, 1985) or a multi-component task (Lane, 1987), at least in the sense that skills of different characteristics contribute to its performance. Acquiring the skills to master the operation of the system requires the inexperienced trainees to acquire a set of concepts, rules, procedures and strategies. The concepts, procedures, and rules are provided as instructions before practising the tasks. Performance strategies should be developed by trainees as training progresses. The main strategies involved in the operation of the system are: time sharing, and diagnostic strategies. The latter are conceptualised as backward elimination of non-malfunctioning components. This should be the ideal strategy to identify and repair malfunctioning devices, or the strategy commonly used by expert operators when they do not have a direct explanation for an alarm, i.e., in some cases a particular alarm can indicate to the expert which device is failing without further examinations.

As reported above, system malfunctions can interrupt the performance of other procedures (i.e., start-up). The analysis of the operation of the diesel generator provided direction about how to decompose the task into parts in order to operationalise the part-task training strategy. Additionally, the malfunction solving process is considered more critical than just starting up the system. Thus, the strategy chosen to decompose the task into its components is based upon the time-sharing or simultaneousness dimension as suggested by Gopher, et al. (1989) and Wightman and Lintern (1985). That is, the operation of the diesel generator was sub-divided according to the fractionation scheme (Goettl and Shute, 1996; Gopher, et al., 1989; Wightman and Lintern, 1985).

Two decompositions were devised that made up two different PT training strategies. One was based on the fractionation scheme and the other on the criticality of its sub-components at a lower level. Specifically, one sub-component of the malfunction-solving task is considered critical to correct task performance according to the subject matter experts. That is, the diagnosis of the malfunction or the identification of the malfunctioning device/s within the system (i.e., diesel generator). The existence of malfunctions is indicated by sound and optical alarms. The alarms do not necessarily identify the malfunctioning devices. Instead, they show the result of a malfunction. This is why diagnosing malfunctions is considered to be a critical sub-component of the malfunction-solving subtask.

These decompositions of the task are shown below and constitute the training strategies investigated in this experiment:

- FT strategy: Operation of the diesel generator
- Fractionated segments PT strategy: start-up the system and supervise/solve malfunctions
- CPT strategy: critical task component or malfunction diagnosis and identification followed by the fractionated segments PT strategy

With this three training strategies a between-subject experiment was designed which assessed the effects of the 3 training strategies by means of a TOT paradigm.

In order to be able to compare the training effects of the three training strategies, the same transfer task was used for the three experimental groups. The task will consist of the FT (i.e., operation of the diesel generator) in which the subjects must solve malfunctions not practised in the acquisition phase of the experiment. The complexity of the transfer task is comparatively higher than the acquisition task. Four malfunctions, not experienced earlier, interrupted the execution of the start-up procedure more often. Thus, task switching was more intensive. Other research has employed increased task complexity (e.g., Speelman and Kirsner, 2001) or increased task difficulty (e.g., Doane, Sohn, and Schreiber, 1999) to assess skill transfer. Naturally, assessing skill transfer from one task to a variant of it could be replaced by the transfer to another training simulator in a quasi-transfer experiment (cf. Experiment 3). It is assumed that transfer mainly applies to the strategies developed to diagnose malfunctioning devices, whether these strategies are effective or not (see e.g., Doane, Sohn, and Schreiber, 1999). Nevertheless, the actual simulator set-up does not really allow for a verification of this assumption. Future research could address this issue.

This experiment tries to identify the effects of 3 training strategies on the operation of the diesel generator. The two PT training strategies are hypothesised to produce higher skill level in terms of performance accuracy than the FT strategy. This effect should be observed not only after the acquisition phase but also on the transfer to the FT. Furthermore, the PT and the CPT strategies should produce faster task performance than the FT. Additionally, if the CPT strategy facilitates the development of appropriate strategies to diagnose malfunctioning components, this group should perform more accurately and faster than the FT and PT groups in the transfer phase.

Two independent variables are contrasted in this experiment. The training strategy used to train novice trainees on the operation of a diesel generator by means of a machinery control room simulator, and the effect of practice upon task performance. The training strategy is manipulated at three levels and is treated as a between-subjects factor (i.e., three experimental groups. The training strategies are described as follows:

- FT strategy (FT): Operation of the diesel generator as described above (see Tables 5 and 6)
- Fractionated sub-tasks strategy or part-task (PT): the training schedule proceeds from starting up the system, followed by supervising and solving malfunctions, followed by training on the FT.
- CPT strategy: the training schedule proceeds from critical task component or malfunction identification, followed by PT training on the fractionated sub-tasks, and training on the FT.
- The practice variable was treated as a within-subjects variable considering:
- Practice block of runs.

- Trainees' performance on the task was assessed during the experiment by means of the following dependent variables:
- Number of runs needed to reach performance criterion. The criterion is achieved when 3 runs are consecutively performed correctly
- Performance accuracy or error rate in each run considering the steps to complete the procedures
- Average time to complete the task/s or performance time
- Performance accuracy in the pre-test and post-test.

3.2 Method

3.2.1 Participants

In the first stage of this study, thirty trainees, inexperienced on the operation of a naval diesel generator, volunteered to participate in the experiment. Nevertheless, when the experiment started only twenty one finally registered to participate. Trainees were extracted from the same population. All were young adults (male = 17; female = 4) studying in their second year of the Nautical Sciences: Naval Machinery Systems degree. They were required to participate in 4 experimental sessions of approximate 2 hours. The first experimental session was run in groups of two or three trainees (7 trainees/experimental group). The first session of the experiment was previously assigned for each trainee. Due to limitations on the availability of trainees, subsequent sessions could not take place on consecutive days for most of the trainees. Thus, the remaining sessions were arranged taking these constraints into account. As will be shown in the results section, the effects of these changing conditions were controlled by including these variables -number of trainees per session and time between sessions- as covariates in the analyses of covariance. Trainees were randomly assigned to each training group, the FT, PT, and CPT. No significant differences existed among groups in their background experience with computers, mouse experience, proficiency in English, and previous studies (see Table 2 below for a summary). Therefore, it can be concluded that trainees were homogeneously distributed across training groups.

TABLE 2

Trainee Sample: Kruskall-Walis Comparison of Mean Ranks among Experimental Groups on Previous Knowledge/Experience (rank values range from 1-5 on all variables)

	Full-Task		Part-Task		Critical Part-Task		Kruskall-Walis	
Background Experience	*M*	*SD*	*M*	*SD*	*M*	*SD*	χ^2 (2)	*p*
Computer Experience	2.14	0.90	2.29	0.76	2.71	0.76	1.94	0.38
Mouse Experience	2.14	0.90	2.29	0.76	2.86	0.38	2.86	0.24
Knowledge of English	1.57	0.53	1.14	0.69	2.00	0.58	4.37	0.11
Previous Studies	1.86	1.57	2.14	1.95	3.00	1.53	1.57	0.46

3.2.2 Instruments and materials

Before describing the instruments and material used in this experiment, a brief contextualisation of the equipment used in the 3 experiments reported here is provided. The MCRS facility includes a lecture room with 4 Unix workstations networked through a LAN to a server which runs the MCRS, a video projection screen, and 4 Macintosh computers with multimedia educational material. See Figure 1 for an overview of this room. Accesible from this room is the HF MCRS. This is divided into three areas, the local control panels, the remote control console, and the instructor operating station. Connected through a LAN network to the server, this runs the MCRS. From the MCRS, exclusively the diesel generator was operated by the trainees. See Figure 2 for an overview from the point of view of the remote control console, and section 5.2.2 for a description of the elments used in Experiment 3. The ship machinery system model can be operated from the workstations in the same way as from the HF simulator, if the local control panels and the additional controls are not used. The workstations do not incorporate any of the additional displays and controls found in the HF MCRS. In Experiment 2 (augmented cueing), two PC computers were used. These runned the NDGSTRT, a software tool developed for Experiment 2 and the LF group in the acquisition phase of Experiment 3. From the ship machinery system, this tool only simulates the diesel generator with similar behaviour to the diesel generator in the MCRS, as evaluated by the subject matter experts.

This experiment (i.e., Experiment 1) was run by two experimenters on three workstations in the simulator facility (see Figure 1). The three workstations were separted by panels so that each trainee could only see his/her own display. Although the MCRS model does not allow for separate simulation of different sub-systems, the diesel generator was the main component in use, all other sub-systems were kept with constant parameters as much as possible. Printed material included a questionnaire to record trainees' data on their previous knowledge and experience, the instructions to provide familiarisation and operational knowledge of the diesel generator simulator, as well as the test used to evaluate trainees' understanding of the dependencies between components in the diesel generator before and after the experiment.

FIGURE 1 Equipment and environment used in Experiment 1. The three Unix workstations at the bottom of the picture were used. Connected by a LAN network to the server, these run the MCRS. From the MCRS, exclusively the diesel generator was operated by the trainees.

FIGURE 2 Overview of the HF MCRS equipment and environment used in Experiment 3 by the HF group throughout the experiment. The LF group used this equipment only in the transfer phase. From the remote control console, on the left, only the CRT display and the keyboard in front of it were used. The local control panels are located in the contiguous room to the left of the console (visible through the window).

Additional printed material consisted of the printed reports of trainees' performances on the tasks and the expert performance templates to analyse their own performance.

3.2.3 Experimental design

As shown in Table 3, the experiment was designed as a transfer of training with three independent experimental groups or conditions defined by the training strategy provided to them. Seven trainees were randomly assigned to each experimental group.

TABLE 3
Experimental Design

Group	Acquisition Phase			Transfer Phase
Full-Task	Full-Task			Full-Task
Part-Task	Part-Task	Full-Task		Full-Task
Critical Part-Task	Critical Task (Malfunction Identification)	Part-Task	Full-Task	Full-Task

The tasks in the experimental design for the acquisition and transfer phases practised by each experimental group during the experiment are specified in Table 4.

TABLE 4
Specification of the Acquisition and Transfer Tasks

Critical Task	Part-Task	Full-Task	Transfer Full-Task
Fractionated identification of 2 malfunctioning devices	Fractionated: - Start-up segment - Supervision and malfunction-solving of 2 malfunctioning devices	Continuous performance of the full-task with 2 malfunctioning devices	Continuous performance of the full-task with 4 malfunctioning devices not practised before

3.2.4 Procedure

Trainees in the three groups underwent approximately the same amount practice, i.e., about 390 minutes, irrespectively of the experimental manipulations. Prior to the experiment, the trainees were administered a test aimed at evaluating their understanding of the functional dependencies of the system's components. This test intended to evaluate if knowledge not directly addressed by training is acquired by trainees and can be made apparent after training. The test was adapted from Benítez-Domínguez (1996). From the original test, which included each main sub-system in the ship machinery system, only the components directly related to the diesel generator were represented in the test. The same test was administered about one month after the experiment was completed. From a diagram representing the components of the diesel generator (i.e., numbered boxes) the participants were asked to link with directional lines each of the listed elements/components considering their functional dependency. The pre- and post-test task is illustrated and explained in Figure 3.

Unir con lineas dirigidas todos los elementos enumerados atendiendo a la dependencia de su funcionamiento.
(Link with directional lines each of the listed elements/components considering their functional dependency)

1.- Motor auxiliar (auxiliary engine or diesel generator)
2.- Alternador (alternator)
3.- Cuadro eléctrico (electrical switchboard)
4.- Compresor de aire de accionamiento eléctrico (electrical air compressor)
5.- Botella de aire comprimido para arranque (start compressed-air bottle)
6.- Bomba acoplada de lubricación (shaft lubricating pump)
7.- Bomba acoplada de agua dulce (shaft fresh water pump)
8.- Bomba acoplada de agua salada de circulación (shaft circulating sea water pump)
9.- Bomba acoplada de combustible (shaft diesel-oil pump)
10.- Sistema de circulación de agua salada (fresh water circulating system)
11.- Sistema de refrigeración de agua dulce (fresh water cooling system)
12.- Sistema de lubricacion (lubricating system)
13.- Sistema de combustible (diesel-oil system)
14.- Sistema de arranque (starting system)

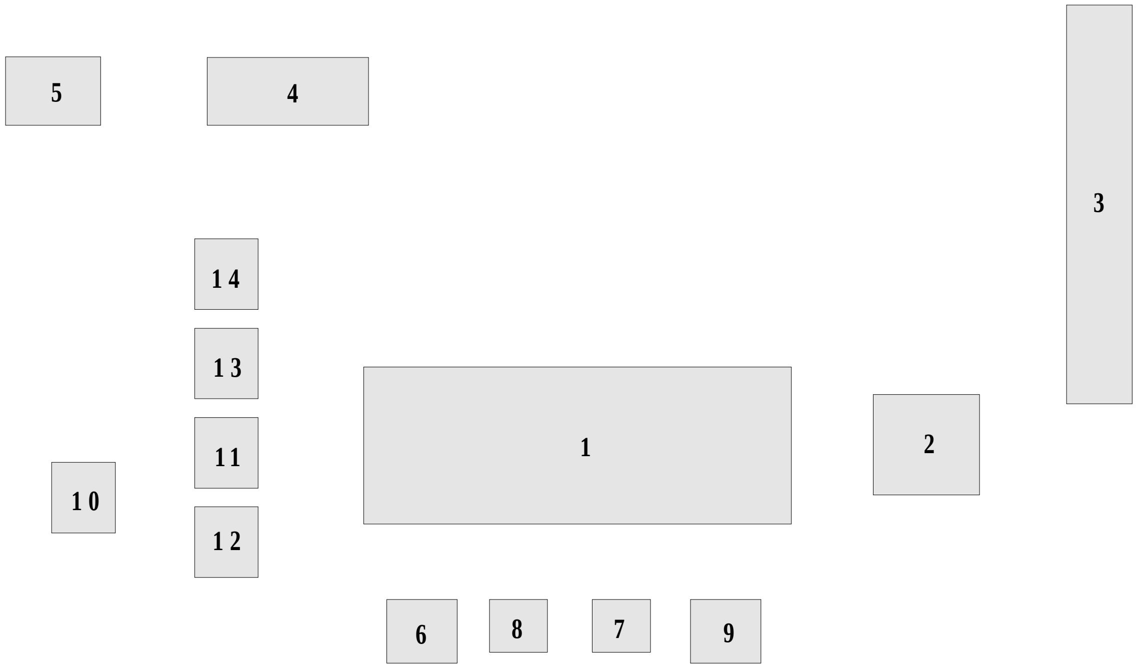

FIGURE 3 Pre- and post-test task as presented to the trainees before and after Experiment 1. Above the numbered grey boxes is explained what each of them represents. Adapted from Gonzalez and Sanz (1998).

Figure 4 represents the expected correct directional links that should have been drawn by the trainees between the system's components in the pre- and post-test task. This test was collectively administered both before and after the experiment.

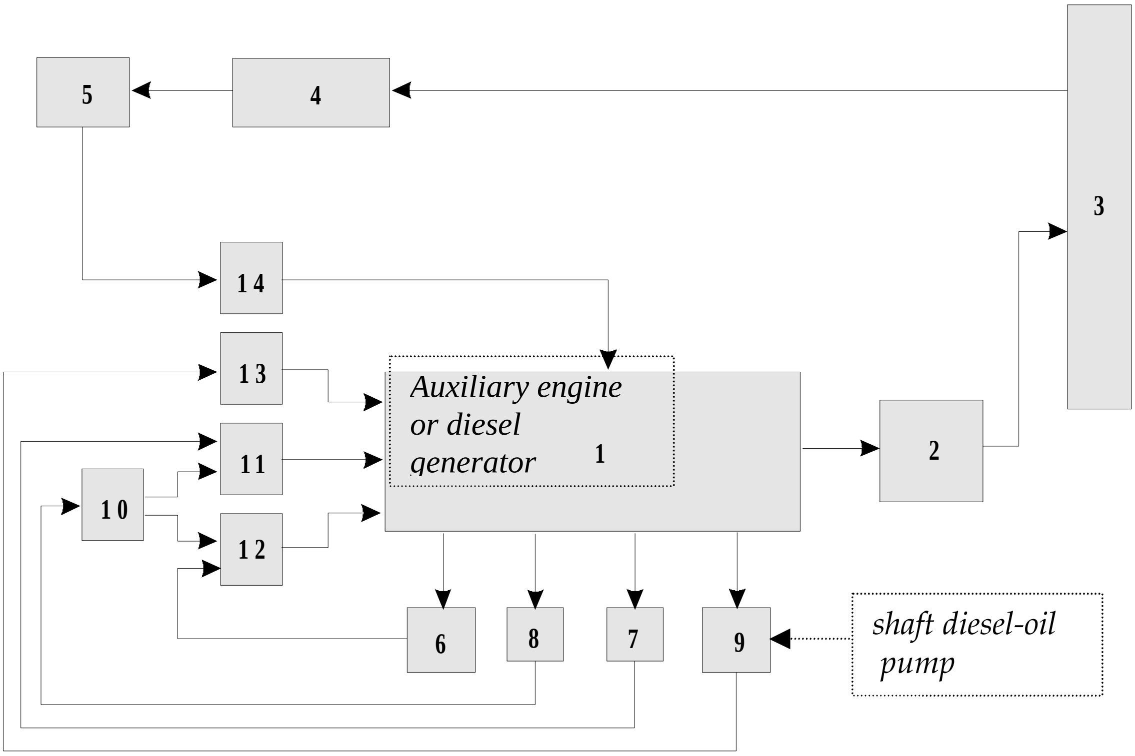

FIGURE 4 Expected correct directional links between the components of the diesel generator. Example: The shaft diesel-oil pump depends on (is moved by) the auxiliary engine or diesel generator. The dashed boxes and arrow have been added to illustrate what the boxes 1 and 9 represent. If the shaft diesel-oil pump is moved by the diesel generator, Then the arrow should start from the diesel generator (box 1) and end on the shaft diesel-oil pump (box 9). Adapted from Gonzalez and Sanz (1998).

When arriving at the first experimental session, the trainees were required to fill in a questionnaire on their previous experience (see Table 2). After they had completed this questionnaire, they continued with the experiment itself. One of the experimenters explained the interactive machinery control system and its graphical interface to the trainees. The next instructions included the knowledge to operate the system (i.e., the procedures) and the sort of malfunctions that can occur with their corresponding effects. Trainees were then informed that no other malfunctions than the ones present in the malfunction lists could occur on the diesel generator. The possible malfunctions related to the diesel generator were represented by the simulator in two separate menu lists or pages which were accessed by pressing the "MALFUNC. LIST" key on the keyboard. These were: the menu list "4000 Diesel Generator 1 – TBCH/DO/LO" and the "4100 Diesel Generator 1 - FW/SW/MISC". The first contained the possible malfunctions in the sub-systems: turbocharger (TBCH), diesel oil (DO), and lubricant oil (LO). The second contained the possible malfunctions in the sub-systems: fresh water (FW), sea water (SW), and other devices or miscellaneous (MISC). Moving between the two menu lists was achieved by pressing the “down” an “up” keys on the keyboard. Trainees were familiarised with the operation of the system in two practice runs after they were provided with knowledge of the system. The first of these practice runs consisted of the execution of the start-up sub-task, see Table 5, and the second consisted of the execution of the malfunction-solving task, see Table 6.

TABLE 5
Start-up Procedure for the Diesel Generator 1

Step	DG 1 Start-up Procedure
1	CHECK IF LO level in LO Sump is lower than 52 %, IF it is lower refill opening the LO Sump make up valve, IF NOT, continue in Step 3
2	WHEN the level in LO Sump reaches 52 %, close the LO Sump make up valve
3	CHECK IF FW level in Exp. Tank is lower than 52 %, IF it is lower refill opening the Exp. Tank make up valve, IF NOT, continue in Step 5
4	WHEN the level in Exp. Tank reaches 52 %, close the Exp. Tank make up valve
5	Open LO filter 1 valve
6	Start the LO electrical priming pump
7	CHECK IF LO pressure after the pump increases, IF it does continue on Step 8, IF NOT go back to Step 6
8	Open SW suction valve
9	Open SW discharge valve
10	Open DO filter 1 valve
11	Open DO shut off valve to DO pump
12	Set the FW temperature control Set point at 60°
13	CHECK IF there is any TRIP indicator lit on the ENGINE CONTROL panel, IF THERE IS press RESET, IF THERE IS Not, continue in Step 14
14	Press LOCAL mode on the ENGINE CONTROL
15	Start the engine, pressing START on the ENGINE CONTROL
16	CHECK IF engine rpm or “N” increase, IF they do continue on Step 17, IF NOT, go back to Step 15
17	Set the LO electrical priming pump on AUTO mode on the Pump Ctr. panel

TABLE 6

Malfunction-solving Procedure for the Diesel Generator 1

Step	DG 1 Malfunction-solving Procedure
1	Detect alarm
2	WHEN the alarm buzzer sounds, turn off the acoustic alarm by pressing the ALARM SILENCE key on the keyboard
3	WHEN there is any blinking optical alarm lit, e.g., $\boxed{\text{H}}$, acknowledge it by clicking with the left trackball button on it
4	To identify the malfunctioning device check and compare differential pressure increments, pressures and/or temperatures related to the alarm
5	IF the malfunctioning device is one in a pair of filters, open the alternative one, close the malfunctioning one, and continue in (2) in Step 6, IF the malfunctioning device is NOT any of this type of devices, continue in (1) in Step 6
6	(1) IF the engine is running stop it by pressing START on the ENGINE CONTROL panel, IF the engine is NOT running, (2) open the malfunctions list by pressing the MALFUNC. LIST key on the keyboard
7	Open the list of the DG 1 subsystem
8	Open the page "4000 Diesel Generator 1 - TBCH/DO/LO", IF it is needed to go to another page use the keys down ∨ or up ∧ to browse malfunction pages
9	Repair (i .e., reset) the malfunction/s which is/are thought to be provoking the alarm by clicking with the right trackball button on the variable number, e.g., on M5001 not on the descriptive text, e.g. DG 1 TURBOCHARGER DIRTY
10	Check if the malfunction/s was/were correctly repaired, IF it/they was/were, click EXIT to close the malfunctions list and continue in Step 11, IF NOT, go back to Step 4
11	Check IF there are any other active alarm/s, IF there are go back to Step 1 or 2 or 3 or 4 depending on the state of the alarm, If NOT, START the engine
12	CHECK IF engine rpm or "N" increase, IF they do continue on Step 13, IF NOT, go back to Step 11
13	IF necessary, Set the LO electrical priming pump on AUTO mode on the Pump Ctr. panel

Practice on these two familiarisation runs was aided with the indications of the experimenters when the trainees became blocked (i.e., between steps of the procedure) or lost (i.e., not finding the appropriate device to act upon) among the different devices on the HMI. Trainees in the three groups followed the same familiarisation practice. During this first session the trainees were encouraged to ask each question they needed in order to understand the procedures and the dynamics of the system. They were also informed that during the following sessions they would not obtain support from the experimenters. Trainees were strongly advised not to share with other colleagues the contents of the experimental sessions in order to prevent other trainees from learning through their partners. They were also discouraged from searching for and/or studying material related to ship machinery systems.

In the beginning of the second session or acquisition phase of this experiment, the trainees were allowed to rehearse the procedures and the information concerning the operation of the simulator until they considered they remembered how to perform the tasks without it. The rehearsal period lasted 5-10 minutes approximately. After this, the trainees were not allowed to consult this information. This rehearsal period was also provided in the beginning of the third session or transfer phase. During the acquisition phase, the trainees performed the operation of the diesel generator with two malfunctioning devices during twelve runs. The time allowed to perform this task was limited to 6 minutes, but trainees could stop the run whenever they considered they had completed the task. The FT group practised the task according to the FT strategy. The PT group performed the task following the PT strategy, which began with the start up procedure during 3 minutes after which knowledge of performance was provided. The malfunction-solving task, during 3 minutes, followed the start up and knowledge of performance was provided immediately after. In the first three runs, CPT trainees only performed the malfunction identification sub-task (i.e., sub-component of the malfunction-solving task). Trainees in this group were required to diagnose the malfunction basing their decisions on the supervision of the system's behaviour. They had to evaluate the system functioning through the data shown on the display (i.e., pressure, temperature parameter values, etc). Trainees were not required to repair the malfunctioning components, i.e., they did not operate the diesel generator. Six minutes were allowed on each of these first three runs for the trainees to decide which were the malfunctioning components and to explain their diagnostic process, i.e., how they arrived at their decisions. Knowledge of performance was provided after each of these runs indicating whether their diagnostic processes were sufficient or not to identify the chosen components as malfunctioning. No reference was made concerning the accuracy of their final choice. In the remaining of the acquisition phase, the trainees in the CPT group performed the task in the same way as the PT group. The CPT trainees greatly differed from those of the FT and PT on the first three runs. They did not operate the diesel generator and the knowledge of performance provided was focussed on the adequacy of the diagnostic process, not on the accuracy of its result.

In the third practice session or transfer phase, each group of trainees performed the operation of the diesel generator in the same way. This time, the task included the start-up of the system plus the solution of four malfunctions not experienced earlier. Practice was provided during twelve runs with available time to perform the tasks limited to 8 minutes.

Knowledge of performance was provided after each run in the familiarisation runs (first session), acquisition and transfer phases. This was carried out in order to facilitate the reduction of errors and to maintain performance improvement (Van Matre, Pennypacker, Hartman, Brett, and Ward, 1981) across runs. Knowledge of performance was comprised of the performance record resulting from the trainee's performance in that run in comparison to a performance template taken from an expert machinery control room operator performing the tasks. As stated above, knowledge of performance was different for the CPT group on the first three runs of the acquisitions phase. The expert performance templates were previously printed and supplemented with the translation into Spanish of the simulator messages. Trainees were required to correct their own results according to the instructions given. Trainees were allowed to correct and examine their performance record for approximately 3 minutes. The experimenters provided guidance to trainees when they were not adequately correcting their performance records. Response-produced feedback was displayed on the screen after the trainees performed an action (i.e., black to colour change of the device symbol and vice-versa). When trainees chose to repair a certain device on the malfunctions list, response-produced feedback was also provided displaying a message on the bottom left corner of the screen indicating the result of that action. Three messages were available, one indicated that the action was correct, and two messages indicated an incorrect action, either due to an incorrect choice (i.e., false alarm) or to the system being in running state. To avoid possible personal and system damage, as in real systems, most of the components cannot or should not be repaired while the system is running.

Trainees were debriefed after they had finished the transfer phase. Debriefing consisted of the explanations offered to the issues that were still unresolved after training, and the questions that the trainees asked about the system. The three groups practice the tasks approximately during the same amount of time (i.e., 6.5 hours). Especial attention was devoted to provide understanding of the behaviour of the system and the interactions between sub-systems. After the trainees had finished the experimental sessions, they were again encouraged not to share the knowledge gained throughout the experiment with their colleagues. This was done in order to prevent other potential trainees from acquiring knowledge outside of the experimental context.

3.3 Results

In the following analyses, results are considered non-significant when the probability of rejecting the null hypothesis being true is greater than 10% (i.e., $p > 0.10$). Both parametric and no-parametric statistical tests were used. Data were transformed for parametric tests. Performance time was logarithmically transformed and errors were square root transformed. For the analyses of covariance, rank scaled data were square root transformed.

3.3.1 Training strategy effects on performance criterion

The number of training runs required to achieve performance criterion (cf. section 3.1.) were qualitatively and quantitatively analysed. The qualitative analyses consisted in comparisons of the percentage of trainees reaching performance criterion in each experimental group and the number of runs required. These results are graphically presented in Figure 6. In the quantitative analyses, the FT group was compared against the PT and CPT groups by means of Mann-Whitney *U* tests. The skill training process can be described as an orderly progression of performance improvement as depicted in Figure 5.

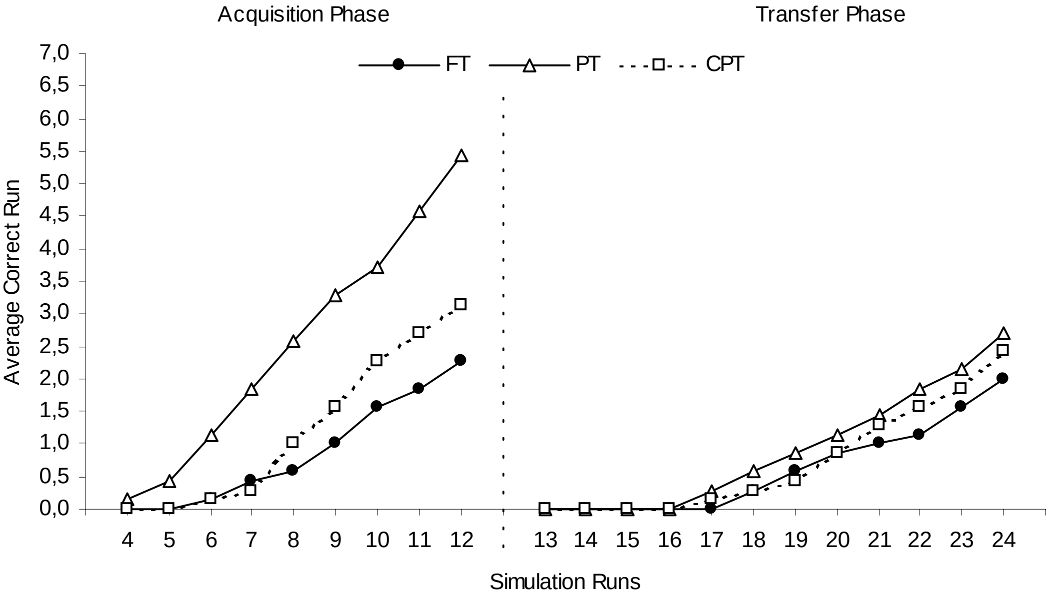

FIGURE 5 Average cumulative frequency of correct runs in the acquisition and transfer phases as a function of simulation run and FT vs. PT vs. CPT training groups. FT = Full-Task group; PT = Part-Task group; CPT = Critical Part-Task group.

Acquisition phase

In the acquisition phase of training, 4 trainees in the FT group did not reach the criterion within the 12 available runs. In the PT group, 1 trainee did not reach the criterion. Similarly to the FT group, 4 trainees in the CPT group did not reach performance criterion. These results are graphically presented in Figure 6.

The Mann-Whitney *U* test revealed a significant difference between the FT and PT groups, $U = 9.00$, $p < 0.05$, mean ranks being 9.71 and 5.29 respectively. The PT group achieved performance criterion faster than the FT group. A similar comparison between the FT and CPT groups did not show differences between them, $U = 23.00$, $p > 0.90$, mean ranks being 7.71 and 7.29. The FT and the CPT trainees required approximately the same amount of practice on the task to reach the criterion in the acquisition phase.

Transfer phase

The transfer task was more difficult than the acquisition task. This can be inferred from the results concerning the number of runs required by trainees to achieve performance criterion. In the FT group, 5 trainees did not reach performance criterion within the available training runs. Within the PT group, 4 trainees did not achieve performance criterion. Similarly to the PT group, 4 trainees in the CPT did not achieve performance criterion. A graphical depiction of the results is shown in Figure 6.

These results reveal that 9 and 13 out of 21 trainees in the acquisition and transfer phases, respectively, did not reach performance criterion. Despite this phenomenon, the PT group was faster in achieving the criterion than the FT and CPT groups in both acquisition and transfer phases. This was more apparent in the acquisition phase.

The Mann-Whitney test did not show a significant difference between the FT and PT, $U = 19.50$, $p > 0.54$, mean ranks being 8.21 and 6.79 respectively. The comparison between the FT and CPT groups also did not show differences, $U = 21.50$, $p > 0.71$, mean ranks being 7.93 and 7.07 respectively. The three experimental groups required approximately the same number of practice runs to reach the criterion in the transfer phase.

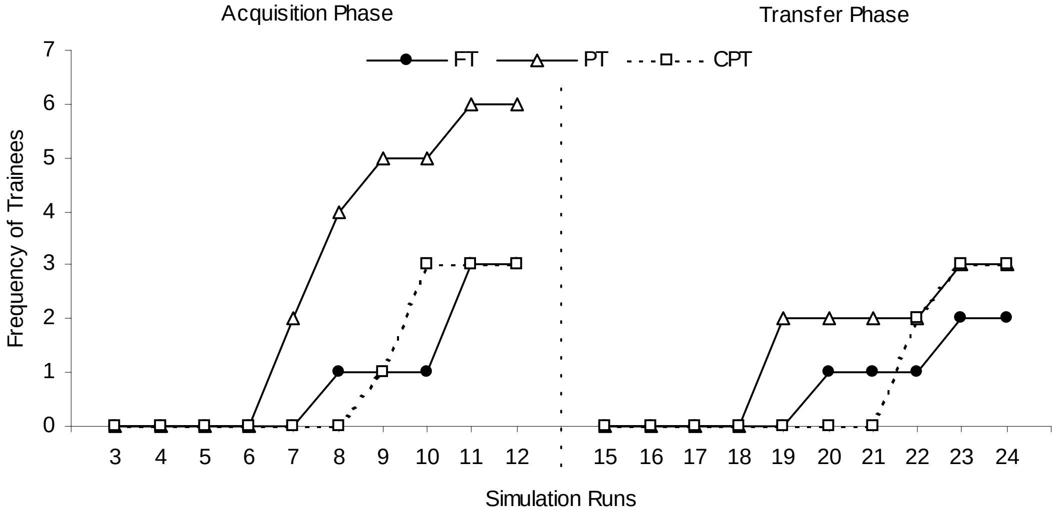

FIGURE 6 Acquisition and transfer phases: cumulative frequency of trainees reaching performance criterion in a particular training simulation run as a function of training group. FT = Full-Task; PT = Part Task; CPT = Critical Part-task.

3.3.2 Training progress as a function of training strategy

Two dependent variables were used to analyse the effects of the training strategies and the effects of practice. These were the percentage of errors made on each run, and the time spent performing the task on each run. The statistical tests were split-plot factorial analyses of variance (ANOVA) with four covariates. The covariates included in the analyses were the time elapsed between the first and second practice sessions, number of trainees in each session, knowledge of English, and previous studies. The ANOVAs were 3 (group FT, PT and CPT) × 3 (blocks of 3 runs) for the acquisition phase with the covariates stated earlier. The first factor was a between-subjects and the second a within-subjects. As the CPT group only performed the malfunction-finding task on the first three runs, the first block of three runs was eliminated from the analysis for the FT and PT groups as well. In the transfer phase, the ANOVA analysis, based on the percentage of errors made in each run was a 3 (FT vs. PT vs. CPT groups) × 4 (blocks of 3 runs) with the same covariates as in the acquisition phase except for the time elapsed between sessions. In the transfer phase this was the time elapsed between the second and third sessions. For the performance time variable, the ANOVA was a 3 (FT vs. PT vs. CPT groups) × 3 (blocks of 3 runs) with the same covariates. This decision was made because each group failed to perform the task correctly within the available time in the first three runs of the transfer phase.

3.3.2.1 Training strategy by practice effects on performance errors

Acquisition phase

- Predictive effects of the covariates

The regression analysis did not show a significant effect of the covariates. That is, the covariates used in the analysis did not predict trainees' performance on the task. In other words, the time elapsed between the first and second practice sessions, number of trainees (2 or 3) per session, knowledge of English, and trainees' previous studies do not seem to affect their performance on the task.

- Between-subjects effects, training strategy

As illustrated in Figure 7, the ANOVA showed significant effects of the training strategies, $F(2, 14) = 4.63$, $MSE = 25.42$, $p = 0.029$. The FT group made more errors throughout the acquisition phase than the PT group, and more than the CPT group in the third and fourth blocks.

- Within-subjects effects, practice or blocks of runs and training strategies:

The practice factor revealed significant differences, *F*(2, 36) = 35.77, *MSE* = 30.36, *p* = 0.000. All training groups reduced their performance errors during the acquisition phase. The interaction training group × block also showed significant differences *F*(4, 36) = 6.05, *MSE* = 5.14, *p* = 0.001. This result indicates that the reduction of errors across training blocks was different for the three groups. The FT group reduced performance errors almost constantly through the third and fourth blocks. The PT group had a sharp reduction of errors from the second to the third block and stabilised from the third to the fourth block. The critical part task group started with the highest error rate of all groups, that was drastically reduced on the third block, error reduction continued from the third to the fourth block. Nevertheless, CPT trainees made more errors than PT and less than FT trainees on the fourth block.

- Univariate tests

Critical differences were computed for the within-subjects effects by means of univariate F tests taking the fourth block as the reference category. The block effect was significant on the second and third blocks, *F*(1, 18) = 47.08, *MSE* = 1.15, *p* = 0.000, and *F*(1, 18) = 11.67, *MSE* = 0.54, *p* = 0.003, respectively. The interaction, group × block was also significant on the second and third blocks, *F*(2, 18) = 5.94, *MSE* = 1.15, *p* = 0.01, and *F*(2, 18) = 6.26, *MSE* = 0.54, *p* = 0.009, respectively.

Transfer phase

- Predictive effects of the covariates

The regression analysis did not show a significant effect of the covariates. That is, the time elapsed between the second and third practice sessions, number of trainees (2 or 3) per session, knowledge of English, and trainees' previous studies do not seem to affect their performance on the task.

- Between-subjects effects, training strategy

The ANOVA did not show a significant effect of the training strategies, *F*(2, 14) = 1.25, *MSE* = 9.26, *p* = 0.32. Nevertheless, the full-task group made more errors throughout the transfer phase than the PT and CPT groups as illustrated in Figure7.

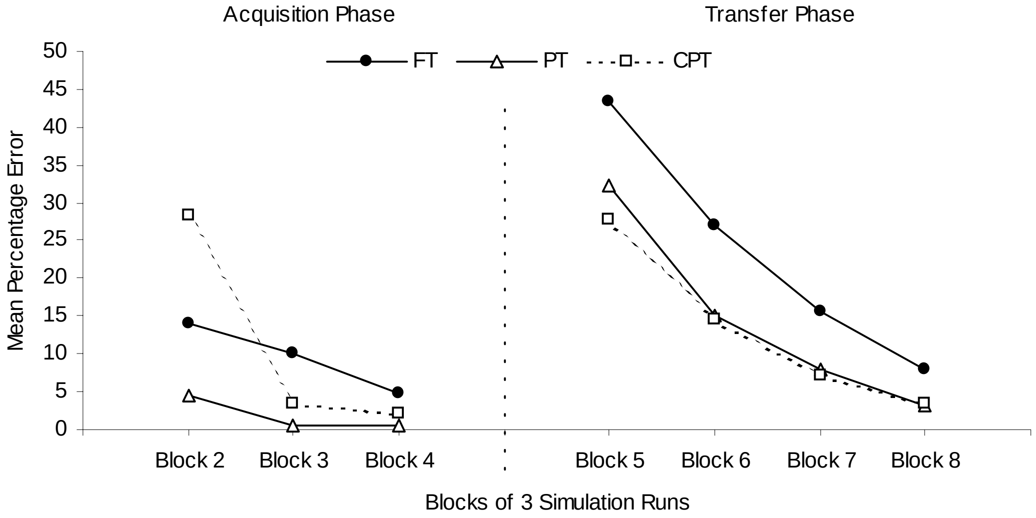

FIGURE 7 Acquisition and transfer phases: Mean percentage of errors as a function of training group and practice block of 3 simulation runs. FT = Full-Task; PT = Part Task; CPT = Critical Part-task.

- Within-subjects effects, practice or block of runs and training strategies:

The practice factor revealed significant differences, *F*(3, 54) = 28.49, *MSE* = 52.18, *p* = 0.000. All training groups reduced their performance errors during the transfer phase. The interaction training group × block was not significant *F*(6, 54) = 0.07, *MSE* = 0.13, *p* = 0.999. This result indicates that the reduction of errors across training blocks was similar for the three groups.

- Univariate tests

Critical differences were computed for the within-subjects effects by means of univariate *F* tests taking the fourth block as the reference category. The block effect was only significant on the first block, *F*(1, 18) = 63.67, *MSE* = 2.43, *p* = 0.000. The interaction, group × block was not significant. These results are depicted in Figure 8.

3.3.2.2 Training strategy by practice effects on performance time

Acquisition phase

- Predictive effects of the covariates

The covariates do not seem to predict trainees' performance time on the task.

- Between-subjects effects, training strategy

The ANOVA showed a significant effect of the training strategies, *F*(2, 14) = 10.78, *MSE* = 0.87, *p* = 0.001. The FT and CPT groups required more time to perform the task than the PT group throughout the acquisition phase. Figure 8 shows these results.

- Within-subjects effects, practice or blocks of runs and training strategies

The practice factor revealed a significant effect, *F*(2, 36) = 38.99, *MSE* = 0.75, *p* = 0.000. All training groups reduced their performance time during the acquisition phase. The interaction training group × block also showed a significant effect *F*(4, 36) = 3.66, *MSE* = 0.07, *p* = 0.013. This result indicates that the reduction in performance time with practice followed different patterns across training groups. Performance time reduction in the FT group was slower than in the PT across training blocks. Trainees in the FT group were also slower than CPT trainees on the last two blocks of runs. The CPT group had a sharper reduction in performance time than the FT and PT groups. Nevertheless, CPT trainees were only slightly faster than FT trainees in the acquisition phase.

- Univariate tests

Critical differences were computed for the within-subjects effects by means of univariate *F* tests taking the fourth block as the reference category. The block effect was only significant on the second block, *F*(1, 18) = 50.18, *MSE* = 0.03, *p* = 0.000. The third block was not significant. The interaction, group × block was significant on the second block, *F*(2, 18) = 3.77, *MSE* = 0.03, *p* = 0.043, and marginally significant on the third block *F*(2, 18) = 3.28, *MSE* = 0.009, *p* = 0.061.

Transfer phase

- Predictive effects of the covariates

Similarly to the acquisition phase, the regression analysis did not show significant effects of the covariates. The covariates did not predict trainees' performance time on the transfer task.

- Between-subjects effects, training strategy

The ANOVA did not reveal a significant effect of the training strategies in the transfer phase, *F*(2, 14) = 0.98, *MSE* = 0.02, *p* = 0.40. On average, the FT group required more time to perform the task than the PT and CPT groups throughout the transfer phase, though this difference was not significant.

- Within-subjects effects, practice or blocks of runs and training strategies

The practice factor revealed a significant effect, *F*(2, 36) = 14.98, *MSE* = 0.07, *p* = 0.000. All training groups reduced their performance time during the acquisition phase. The interaction training group × block did not show a significant effect *F*(4, 36) = 1.03, *MSE* = 0.00, p = 0.41. This indicates that performance in the transfer phase was not affected by the training strategy followed in the acquisition phase of training. The FT group was also slower than the CPT on the last two blocks of runs. The CPT group had a sharper reduction in performance time than the FT and PT. Nevertheless, in the acquisition phase, CPT trainees were only slightly faster than FT trainees. These results are graphically presented in Figure 8.

- Univariate tests

Critical differences were computed for the within-subjects effects by means of univariate *F* tests taking the fourth block as the reference category. The block effect was only significant on the second block, *F*(1, 18) = 17.84, *MSE* = 0.01, *p* = 0.001. The third block was not significant. The critical differences test did not reveal significant differences for the interaction group × block.

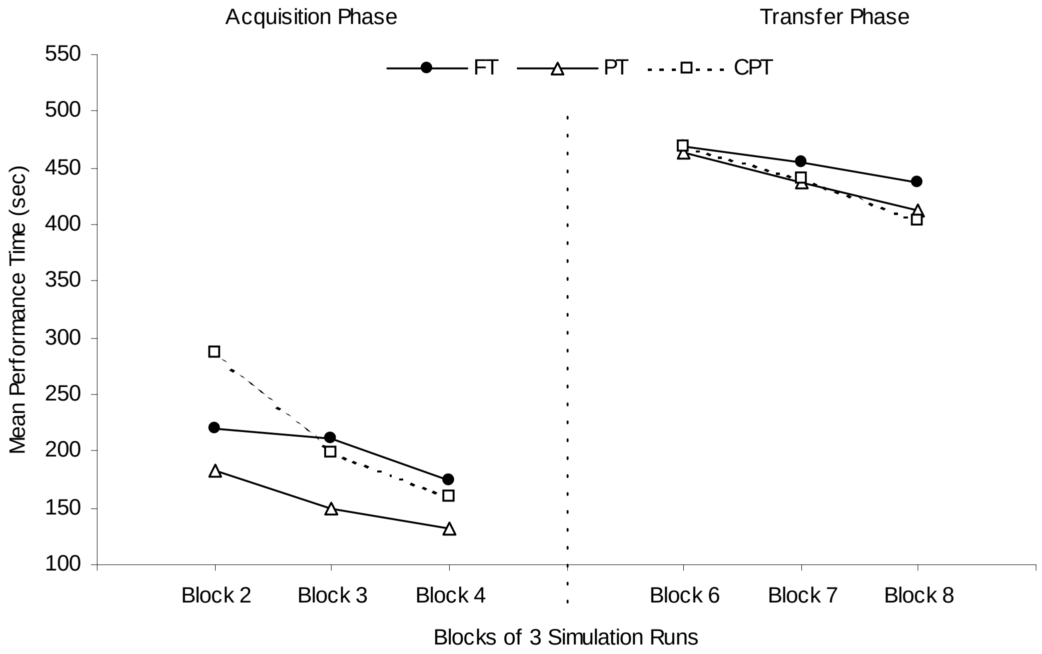

FIGURE 8 Acquisition and transfer phases: Mean performance time (sec) as a function of training group and practice block of 3 simulation runs. FT = Full-Task; PT = Part Task; CPT = Critical Part-task.

3.3.3 Transfer effects of the part-task training strategy

Results reported earlier indicated that a certain amount of interference is produced by the change from the acquisition task to the transfer task. The interest now is placed on the possibility that different training strategies produce different amounts of interference. Direct exploration of this issue, apart from the qualitative evidence illustrated through the figures, is not possible as the tasks varied in their degree of complexity. Also, the means and standard deviations of the distributions of performance measures of both conditions are different. In order to extract some conclusion with a small probability of rejecting the null hypothesis, i.e., that the three training strategies produce similar interference effects from the acquisition phase to the transfer phase, being true, the percentage of errors variable was standardised into z scores (mean = 0; standard deviation = 1) for the FT, PT, and CPT groups, for the blocks of runs 3 and 4 of the acquisition phase, and the blocks of runs 6 and 7 in the transfer phase. With these data, a split-plot factorial ANOVA was performed. The analysis was a 3 (FT vs. PT vs. CPT groups) × 2 (acquisition vs. transfer phases) × 2 (blocks 3-4 vs blocks 6-7) with groups as the between-subjects factor, and acquisition/transfer phases and blocks of runs as within-subjects factors. The results approached significance on the effect of training group $F(2, 18) = 2.48$, $MSE = 1.40$, $p = 0.11$. The other main effects as well as the interactions did not reach significance (all Fs < 1). These results are consistent with the results favouring the PT and CPT groups in the transfer phase (see Figure 7).

3.3.4 Understanding the functional dependencies between sub-systems

Trainees understanding of the dependencies between the components of the system was analysed by means of a 3 (FT, PT, and CPT groups) × 2 (pre-test and post-test) split-plot factorial ANOVA with the time elapsed between the end of training and the post-test stage as a covariate. The first factor was a between-subjects and the second a within-subjects factor. The results indicate that the covariate did not affect the results of the post-test. As shown in Figure 9, the difference between training groups was significant, $F(2, 17) = 7.90$, $MSE = 16.25$, $p < 0.004$. PT trainees made less errors than FT and CPT trainees in the pre-test and in the post-test. This difference is more obvious in the post-test. The CPT group made less errors than the FT in the pre-test as well as in the post-test. Despite this effect of the training strategies, the difference between pre- and post-test was not significant. The interaction training group by testing stage was also not significant . These effects indicate that trainees' performance on the post-test did not improve as a result of training, regardless of the training strategy that they followed in the acquisition phase.

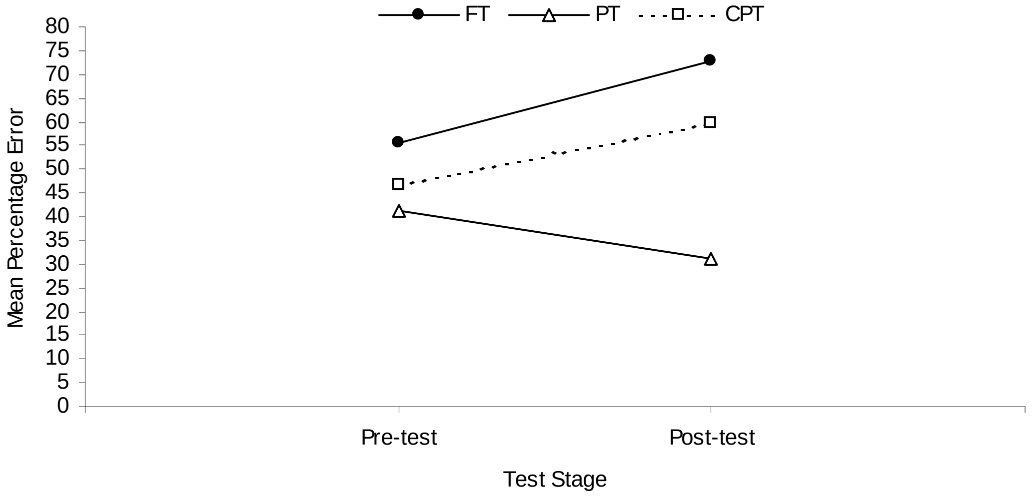

FIGURE 9 Mean percentage of errors in the pre-/post-test task as a function of training group. FT = Full-Task group; PT = Part Task group; CPT = Critical Part-task group.

3.4 Discussion

Firstly, the qualitative analysis provided evidence of the difficulty of the task in the acquisition phase as 42.86% of the trainees did not reach performance criterion (i.e., correct performance on three consecutive runs within the available time). Also, the task difficulty increased in the transfer phase as 61.9% of all trainees did not achieve the criterion. Despite these negative results, the qualitative analysis indicated that the PT training strategy resulted in a higher proportion of trainees achieving criterion than those in the FT and CPT in the acquisition phase. This was also supported by the quantitative analysis (Mann-Whitney). The PT trainees required fewer runs to achieve the criterion than the FT and CPT trainees. No differences were found between the FT and CPT groups. The lack of significant differences between the FT and CPT strategies can be attributed to the distinctive training provided to the CPT trainees on the first three runs of the acquisition phase, i.e., malfunction diagnosing only, and its associated knowledge of performance. These did not favour the CPT trainees in the operation of the diesel generator when the CPT group followed the PT training schedule (second block) as much as it did in the PT group. The highest error rate and performance time shown by the CPT group on the second block of runs in the acquisition phase (See Figures 7 and 8), can be explained by this lack of practice on the operation of the system for the first three runs.

In the transfer phase, the proportion of PT and CPT trainees achieving performance criterion was higher than that of FT trainees. Nevertheless, this result was not supported by the quantitative analysis. On average, trainees in the FT, PT and CPT required the same number of runs to achieve the criterion.

The results indicate that task fractionation strategies produced higher skill levels than approaching the training from strategies based on the whole task in the acquisition phase. Regarding performance measures in each group, it is important to highlight how the number of errors made by the CPT group on blocks 3 and 4 in the acquisition phase, were in between those of the FT and PT (see Figure 7). In the transfer phase (Figure 7), the number of errors made by the CPT and PT groups were similar, and contrasted with the higher error level of the FT group. Performance time for the FT and CPT groups was similar in the third and fourth blocks of the acquisition phase (Figure 8) and longer than for the PT group. In the transfer phase, performance time for the PT and CPT groups was similar and both groups performed the task faster than the FT group in blocks 7 and 8 (See Figure 8). These results also suggest that training on the critical diagnosis of malfunctions in the first stage of training (first three runs) delayed skill acquisition to perform the start-up and malfunction solving segments of the task. In the transfer phase, this effect was reversed to a performance level of the CPT trainees more similar to the PT than to the FT trainees.

The issue that several trainees did not reach the criterion can be attributed to the defined performance criterion itself, to insufficient training time, or to the degree of difficulty required to understand the performance records and templates provided to trainees after each training run. The performance criterion to be achieved by trainees in this experiment, was established so that no single mistake was allowed in the evaluations of performance, regardless of whether trainees had attained the task goal or not, i.e., the diesel generator in running order without alarms. This indicates that some procedural mistakes which are not fatal for the integrity of the system determined erroneous performance on several runs. Thus, the performance criterion established may have been too strict. Though this effect should be taken into account in future experiments, it does not obscure the effects of the different training strategies as will be shown later. The available time was set in order to match the availability of the simulator facility and was kept similar across training groups. Each trainee followed training for about 390 minutes (6.5 hours) which was considered sufficient by the experimenters. Additionally, overextended training time should likely equalise the results of training across training strategies, hence, not providing meaningful information (Speelman and Kirsner, 2001). Knowledge of performance was difficult to understand in the early stages of training as informed by trainees and demonstrated by their incapacity to correct their own performance records. This difficulty was reduced gradually throughout training. The low knowledge of English of the trainees could account for these difficulties though special attention was devoted to this issue in the first practice session. Also, printed performance templates were supplemented with translations of the simulator messages into the trainees' mother language (Spanish).

The results of the parametric analyses yielded results in line with the qualitative analyses. The covariates included in the analysis did not predict performance measured either by accuracy or by performance time. This means that previous knowledge, amount of trainees per session, and time elapsed between practice sessions are independent of the performance exhibited by trainees on the procedural tasks studied in this experiment. The analyses of the training process, i.e., training strategy by practice runs demonstrated that the PT training strategy produced higher skill achievement than the FT and CPT strategies in the acquisition phase. In the case of the CPT group this was probably due to the critical task practice in the first three runs of the acquisition phase, which resulted in the highest error rate in the second block of runs. In the third and fourth blocks, the CPT group outperformed the FT group, but, made more errors and performed the task more slowly than the PT group. The highest benefit of the PT strategy was evidenced not only by performance accuracy, but also by the performance time that trainees' employed on the tasks. Similar performance patterns can be observed on performance accuracy and performance time across training groups. See Figures 7 and 8 for comparison of these patterns.

In the transfer phase, despite the lack of statistical significance, performance differences mainly continued favouring both the PT and CPT groups against the FT group. Similarly to the qualitative analysis regarding performance criterion, the CPT group shifted from closer similarity to the FT group in the acquisition phase, to closer similarity to the PT group in the transfer phase. This was more outstanding in performance assessed by means of performance time. Lack of statistical significance could have been due to the continuous provision of knowledge of performance to trainees. The high difficulty of the transfer task, as evidenced by the high error rates across training groups, could have diminished the statistical value of the differences between training strategies.

The analysis considering trainees’ understanding of the functional dependencies between system’s components did not show improvement across training groups after training. This is contradictory to the results obtained after analysing the training process, i.e., the effects of training strategy by blocks of runs, since skill improvement is obvious throughout the training period. Explanations for these divergent results can be considered in relation to the nature of the task employed to assess understanding of the system and the nature of the trained tasks.

Alternatively, the graphical representation of the system both in the simulator and the pre-test/post-test task could account for these discrepancies. The tasks trained in this experiment mainly focussed on the dynamic processes involved in the functioning of the system. The pre-test/post-test chiefly relates to the structural organisation of the system. This issue was not directly addressed in the training tasks. The graphical representation of the system in both tasks was different. While in the pre-test/post-test task the system’s components were represented as a set of separate boxes, in the simulator, the components are represented by distinctive symbols and colours, and linked by directional lines indicating the flow direction of its fluids. This difference alone may have acted as a barrier to the transfer from one mental representation to the other. Also, the labels shown in the simulator display were different from the separate list of labels presented in the pre-test/post-test task. Hence, it can be concluded that the task employed to assess the understanding of the dependencies between the system's components was not valid. Other assessment tasks should be devised which demonstrate high internal validity and are directly related to the contents and aims of training.

4 EXPERIMENT 2: LOW FIDELITY SIMULATOR WITH AUGMENTED CUEING

4.1 Introduction

Empirical research on training effectiveness and its relation to simulator fidelity has been mainly developed within aircraft systems (Hays et al. 1992). This has usually been associated to aircraft pilot training. Within ground vehicle systems the amount of research devoted to these issues is far more restricted. In both domains, fidelity (i.e., objective or functional and perceptual fidelity or realism) is mainly related to the visual and motion systems and the engineering or technical requirements associated with their implementation. Unfortunately, within naval systems and other related systems such as industrial plants, very restricted research has been found which addresses the effects of simulator fidelity and training effectiveness. In naval systems, cognitive procedural tasks such as radar operation, machinery control, etc., are considered to be the most relevant (Hurlock and Montague, 1982) as opposed to motor tasks which are important relative to vehicle handling.

One empirical research addressed the effects of ship motion on motor skills such as tracking, key pressing, and tracing. From these tasks, only key pressing was unaffected by the movements of the ship motion simulator (McLeod, Poulton, du Ross, and Lewis, 1980). In relation to training, McLeod et al. (1980) proposed that motion should be included in the training device if this factor is critical to the acquisition of motor skills involved in ship mobility functions. On the contrary, if fidelity is not very critical for training goals, such as perceptual-cognitive skills training for performing procedural tasks, which can be best trained on PT simulators, simulator motion should not be incorporated (Coleman, 1988; McLeod et al., 1980; Wetzel, Van Kekerix, and Wulfeck, 1987).

According to these researchers, HF simulators should not be necessary for training perceptual-cognitive or procedural skills as no motor skills are the goal of training. If inexperienced trainees were being trained on these types of skill, HF simulators would be even less necessary.

During interviews with experts and users (MASTER, 1996) carried out in the Naval, Ground and Air Forces in Spain as well as in other countries, most of the interviewed personnel advocated for the highest possible fidelity. This was so explicit that some instructors in the Air Force declared that they prefer real flying than any kind of simulator though they valued the training advantages of simulators, i.e., economy and risk avoidance. Specifically, within the Combat Information Centre (CIC) simulator, the surveyed instructors indicated that they did not need a motion system for that simulator, but other fidelity characteristics of the simulator were highly valued. Namely, the perceptual fidelity of the CIC consoles. The external appearance or realism of the consoles, the arrangement of controls and displays on them, the spatial distribution, and so on, are factors that are considered to be very important for the training of CIC operators. The identical elements (Thorndike, 1903), transfer surface (Osgood, 1949), production rules (i.e., If-Then rules) shared between tasks (Anderson, 1987) or hypotheses of transfer of training based on the similarity between tasks or the operational environments (i.e., devices) are considered to be a theoretical anchor for these opinions.

These subject matter expert opinions can be better understood if we take into account that the training objectives at which this simulator aims, are the familiarisation of the CIC operators with the equipment they will use aboard ship. Additionally, most of the tasks the operators will perform in the real environment will be mainly trained on the actual equipment. Another important factor that ensure expert opinions remain anchored to their preference for the highest fidelity of the training device is the absence of any formal evaluation of the transfer effects of this type of simulator to the real system. In fact, formal evaluation of the training effects of neither HF nor LF simulators has ever been performed.

Even though the opinions of the experts should be taken into account to acquire any training device, formal evaluation of the training and transfer effects of these devices should help to validate those decisions and thus arrive at more profitable outcomes in terms of economical (value for money) and performance achievement of the trainees.

From these issues, it can be extracted that human factors research findings are quite distant from actual training practices. It should be desirable that the results of this research could help to bridge the existing gaps between training research and practice. The present experiment aims at these goals and is motivated by the following points which summarise the discussions provided above:

- The vast majority of empirical research addressing the relationship simulator fidelity and training effectiveness has been carried out in the aircraft systems domain. In the naval systems, this is almost lacking.

- The type of relevant tasks currently investigated in the aircraft systems area have been perceptual motor (e.g., manual tracking, aircraft piloting), while in the naval domain the most relevant are complex procedural tasks.
- Formal evaluation of the transfer effectiveness of HF simulators in the naval domain is missing.
- Consequently, little guidance is provided from the research and interviews with experts and trainers with respect to the training effectiveness of naval simulators and thus which fidelity parameters should be incorporated to achieve a certain level of effectiveness (Gonzalez, 1996b,c).

Similarly to Experiment 1, the research constraints stated above, and the availability of naval simulators guided our selection of a Machinery Control Room Simulator (MCRS) and its associated system management task to execute the present experiment. For a description of the nature of the task involved in this experiment see section 3.1. in Experiment 1.

The HMI on which the ship engineer operates the diesel generator and the rest of the machinery subsystems is the cornerstone of this experiment. The fidelity characteristics that will be altered on the perceptual fidelity of the HMI are discussed below.

The Naval Diesel Generator Simulator Training and Research Tool (NDGSTRT) was developed specifically for this experiment and Experiment 3. It is similar to the diesel generator sub-system implemented by the MCRS in the HMI contents and functions. It has been implemented on a Personal Computer (PC) instead of a Unix workstation. The differences between these two simulators are summarised in Table 7. The NDGSTRT was developed because the experimental manipulation proposed for this experiment could not be implemented on the MCRS.

TABLE 7
Differences between the MCRS and the NDGSTRT Implementing the Diesel Generator

	MCRS	NDGSTRT
Type of computer	Workstation	PC
Operating system	UNIX	Windows 95
Screen size	17"	14"
Input devices	Integrated trackball & keyboard	Separated mouse & keyboard
Display content	Interactive process diagram	Same + augmented cueing
Control functions	Operate equipment	Operate equipment
Response-produced feedback	Black to colour and vice-versa change, value data update	Same + value data check indication
Knowledge of performance	Printer action log	On-Screen action log
Performance data recording	No	Yes
Simulation completeness	Exhaustive	Reduced to training goals
Interaction between subsystems	Highly interactive	Isolated subsystems
System behaviour	Based on highly accurate software models	Based on dynamic changes of system states/contingencies (i.e., If-Then rules)
Cost	High	Low

As stated above, the MCRS workstation is identical to that of the remote control console which is also similar to those aboard ship. The remainder of the control equipment on the remote control console is not considered here because beginner trainees are not trained on the fully realistic simulator. They are only familiarised with the system on the workstations.

The fidelity parameters that can be considered are the differences in screen size, input control devices, and the presence/absence of augmented visual cueing or visual guidance (Holding, 1987). The presence of augmented cueing also represents a departure from fidelity of the simulator with respect to the real systems because the actual systems do not incorporate this feature. In the present experiment, the utilisation of the HF simulator was avoided because it was not available at the time this experiment was run. Nevertheless, a comparison of the effects of training of both training devices is performed within Experiment 3. The rationale for incorporating the augmented cueing training strategy is provided below in relation to the characteristics of the HMI.

The HMI display is presented in Figure 10, the environment is shown in Figure 1. The input devices of the HMI are a trackball and a keyboard. Most of the control inputs are performed by means of these devices on the remote control console and on the MCRS workstation employed in Experiments 1 and 3.

The operation of ship machinery systems, in general, and of a diesel generator, in particular, is currently performed on interactive process diagrams displayed on a computer screen. These diagrams display the symbols of the devices that can be operated through the HMI and data showing the state and the behaviour of the system. Figure 10 presents the interactive process diagram of the diesel generator used in this experiment.

The HMI display of this simulator as well as on board ship, is far from optimal for beginner trainees. This has been ascertained by questioning the trainees during the initial pilot studies executed to investigate the feasibility of the machinery control room simulator to carry out the experiments in this research. Some problems are associated with the poor discriminability between symbols which can and cannot be manipulated by the operator (i.e., the trainee). Display clutter exacerbates the difficulty in the search for target devices, i.e., valves, pump, pressure and temperature data, etc. on which the trainees must perform an action or check for particular values.

Many different manipulations could be performed on the HMI to avoid these problems during the early training of inexperienced trainees. This is, to improve the readability of the HMI in terms of the perceptual characteristics of the symbols on the display representing the actual equipment and its working conditions. These manipulations should affect the legibility of the visual information provided by the interactive process diagram. Changes to improve the actual HMI display characteristics could cause additional problems during the current training of the experimental participants, i.e., interference with the acquisition of the required skills and with the operation of the actual equipment.

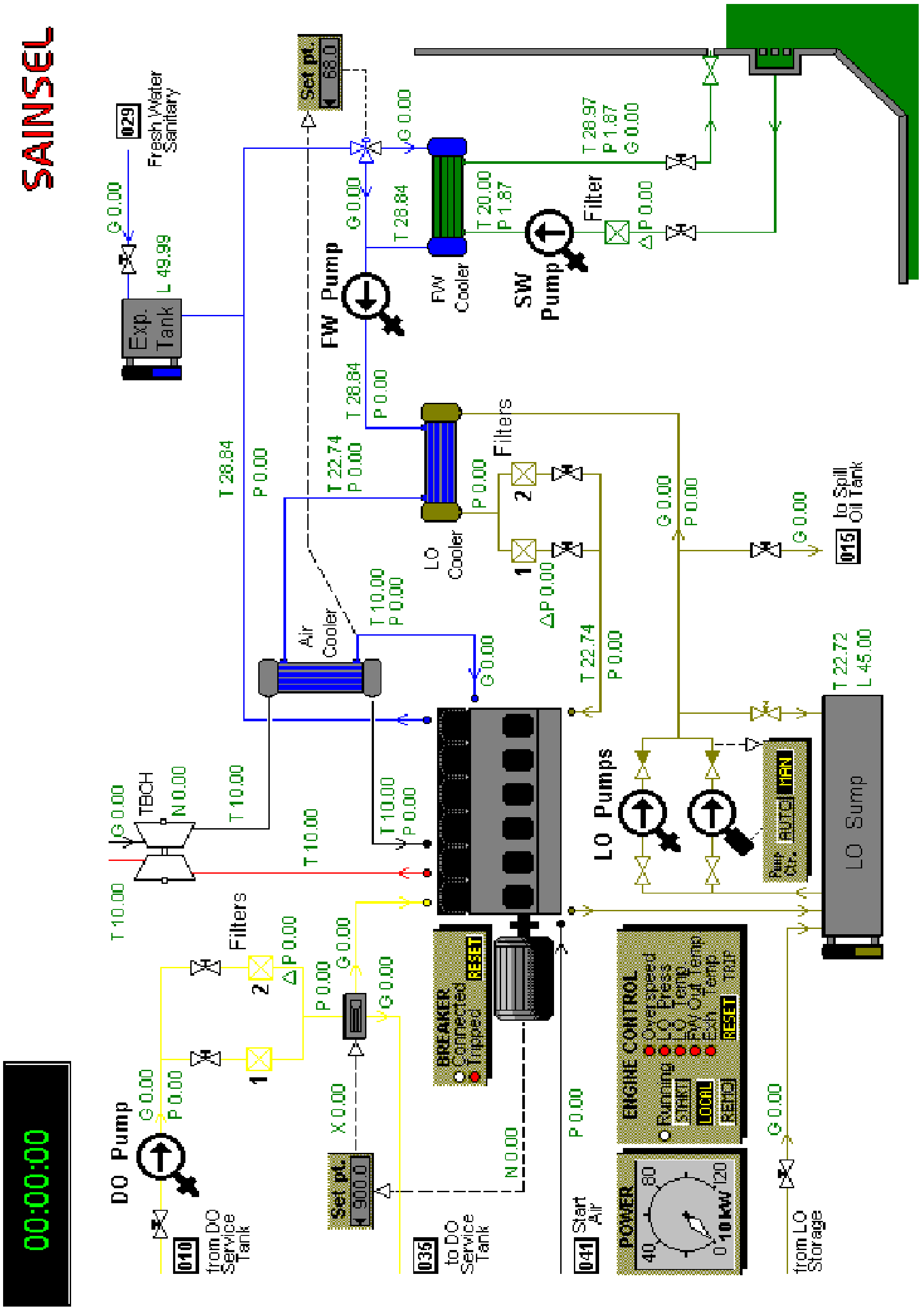

FIGURE 10 Interactive process control diagram of the NDGSTRT. On the PC computer screen, the background appears as light grey (MS Windows standard).

Carlson, et al. (1992) found that procedural guidance produced effective learning in comparison to discovery learning (i.e., variable template, fixed template). Taking into consideration the task and the issues discussed above, two different training strategies have been devised. One consists of incorporating augmented cues (e.g., Holding,1987) to help or guide the trainees in identifying the target devices and relevant value data to identify the malfunctioning device as suggested by the results of the pilot studies and Experiment 1. The other does not incorporate augmented cueing. The two training strategies were implemented on the NDGSTRT.

- NDGSTRT implementing the augmented cueing strategy
- NDGSTRT without augmented cueing strategy.

With these two training conditions, a between subject experiment was designed which assessed the effects of the 2 experimental conditions by means of a ToT paradigm. The acquisition and transfer tasks are those used in Experiment 1.

From the above discussions, it is hypothesised that training on procedural tasks which incorporates augmented cueing on the NDGSTRT could result in faster and higher training achievement compared to the training provided on the NDGSTRT without augmented cueing. That is, the augmented cueing training strategy could produce better training results than that of the non-augmented cueing. The null hypothesis predicts that both training strategies would produce similar training effects.

Other beneficial effects on training results could be produced by the augmented cueing strategy against non-augmented cueing. This is unclear as no reference on this subject has been found. Perhaps, more effective task performance strategies could be promoted by cue augmentation. Nevertheless, this statement is only tentative and must await for the analysis of the results.

The fidelity parameters on which the PC-based low-fidelity (i.e., low-cost) or LF and the high-fidelity or HF simulators differ are not assumed to produce distinct acquisition and transfer effects. This assumption is investigated in Experiment 3. If no differences are found between the group trained without augmented cues on the NDGSTRT and the HF simulator used in Experiment 3, the LF could be considered to be more cost-effective than the HF. Additionally, in the HF simulator used in Experiment 3, trainees had to walk from the local control panels room to the remote control console. These two rooms are separated by a door, and the operating positions are approximately 10.5 meters apart from each other. Also, the experimental procedure differed between these two experiments: in the present procedure, two trainees followed training at the same time but did not interact with each other, in Experiment 3, trainees participated in individual sessions. Therefore, the interpretation of this comparison should be considered with caution.

Two independent variables will be contrasted in the present experiment. The training strategy used to train inexperienced trainees on the operation of a ship diesel generator and the effect of practice upon task performance. The training strategy includes two levels and will be treated as a between-subject variable. The two training strategies are the augmented cueing vs. non-augmented cueing implemented on the LF NDGSTRT.

With these variables, two independent (between-subject) experimental groups were formed:

- Augmented cueing (AC)
- Non-augmented cueing (NAC)

The practice variable was treated as a within-subject variable considering:

- Training runs,
- Practice blocks, i.e., blocks of three runs, and,
- Acquisition and transfer phases which serve to compare how training strategies affect skill acquisition on both training groups.

An important aim of any training programme is that trainees attain a certain level of proficiency in task performance. This is commonly known as performance criterion. In this experiment, performance criterion is defined as the achievement of task goals in three consecutive runs. The task goals are achieved when trainees bring the system to the normal working condition and without active alarms . Additionally, trainees must achieve the task goals within the available time period, without making false attempts to repair the possible malfunctions, and executing each action/step required by the procedures.

The dependent variables on which trainees were assessed during training are thus considered in two ways. Firstly, dependent variables were analysed in relation to the attainment of performance criterion and the result of training as a whole. Secondly, the same variables were taken in relation to the training process or how performance level improves throughout training. By measuring trainees' performance in these ways, not only is an indication of the efficiency of training obtained, but also an understanding of the training process. The dependent or performance variables measured during this experiment are as follows:

- Runs: Number of runs needed to reach performance criterion (RCR). That is, three consecutive runs achieving task goals.
- Steps: Mean number of steps or actions executed in each training run (STP)
- Performance time: Mean performance time dedicated to perform the operation of the system (PFT)
- Errors: Mean percentage of errors in each training run (ERR)
- Attempts: Mean number of false attempts to correct malfunctions in each run (ATT)
- Checks: Mean number of data value checks performed in each run (CHK)

Differently from Experiment 1, the performance criterion was set so that errors which are not crucial to the integrity of the system do not prevent trainees from achieving performance criterion within the available number of training runs. This decision was adopted taking into consideration the opinion of the subject matter experts as well as the results of Experiment 1. In this experiment, no errors of any type were allowed to grant trainees the achievement of performance criterion. It was pointed out then that a very strict performance criterion could have influenced the small number of trainees attaining the criterion within the allowed training time. In this sense, an attempt has been made to prevent the possible misleading effects of a very strict performance criterion.

4.2 Method

4.2.1 Participants

Sixteen university students volunteered to participate in this experiment. All were extracted from the same population. They were all young adults studying in their third year of the Nautical and Navigation Sciences degree (male = 11; female = 5). Trainees were inexperienced in the operation of ship machinery control systems. Therefore it is assumed that their knowledge of the system is similar. Trainees in both groups were required to participate in 3 experimental sessions, in groups of two. The experimental sessions lasted about 2 hours. These were arranged so that the three experimental sessions could take place on three consecutive days and at approximately the same time of day for each pair of trainees. Trainees were allowed to decide when to start the experimental sessions. After they had chosen the date for the first session, the following two sessions were fixed for the next two consecutive days. Trainees were randomly assigned to both experimental groups (AC vs. NAC). The two groups do not differ substantially in computer experience, mouse experience, knowledge of English, and previous studies (see Table 8 for an overview).

TABLE 8
Trainee Sample: Mann-Whitney U Comparison of Mean Ranks between Experimental Groups on Previous Knowledge/Experience Variables (rank values range from 1-5 on all variables)

	Augmented cueing		Non-augmented cueing		M-W *U* test	
Background experience	*M*	*SD*	*M*	*SD*	*U*	*p*
Computer experience	2.75	0.71	2.13	0.99	20.5	0.14
Mouse experience	2.38	1.19	2.13	0.99	26.5	0.50
Knowledge of English	1.88	0.35	1.88	0.64	31.5	0.95
Previous studies	1.88	0.64	1.75	0.71	29.5	0.68

4.2.2 Instruments and materials

The experiment was implemented on two personal computers running the NDGSTRT software tool of which the characteristics were described earlier (cf. section 4.1.). Printed materials included a questionnaire to record the trainees' previous knowledge/experience, and the instructions to provide familiarisation as well as operational knowledge of the system. Operational knowledge included the procedures to operate the diesel generator and explanations of the effects of the possible malfunctions.

4.2.3 Experimental design

The experiment was designed as a transfer of training (ToT) with two independent experimental groups or conditions. Trainees were randomly assigned to each experimental condition, eight to the experimental group or augmented cueing (AC), and eight to the control group or non-augmented cueing (NAC). Table 9 shows the organisation of the experimental conditions by practice sessions.

TABLE 9
Experimental Design

Group	Acquisition Phase	Transfer Phase
AC	Augmented Cueing	Non-Augmented Cueing
NAC	Non-Augmented Cueing	Non-augmented Cueing

Note. AC = Augmented cueing group; NC = Non-augmented cueing group.

The tasks on which the experimental groups were trained during the acquisition and transfer phases were the same across groups and different across training phases. These were the same as the ones used in Experiment 1:

- Acquisition phase: Operation of the diesel generator with two malfunctioning devices.
- Transfer phase: Operation of the diesel generator with four malfunctioning devices not practised before.

4.2.4 Procedure

On arrival at the first session, the trainees were questioned about their personal data and previous experience. After they had filled in this questionnaire (cf. section 4.2.1. for an overview of the trainees' group statistic test), they were introduced to the experiment itself. The experimenter provided the instructions to familiarise the trainees with the interactive machinery control system and its graphical HMI. The next instructions included the knowledge to operate the system (i.e., the procedures) and the sort of malfunctions that can occur with their corresponding effects. Trainees were then informed that no other malfunctions than those present in the malfunction menu could occur on the diesel generator simulated by the NDGSTRT. These are the same as those implemented on the simulation equipment used in Experiment 1 and Experiment 3. Following the provision of knowledge, trainees were familiarised with the operation of the system with two practice runs. The first of these practice runs consisted of the execution of the separate start-up procedure, and the second with the execution of the malfunction-solving procedure. Practice on these two familiarisation runs was aided with the experimenter indications when the trainees became blocked (i.e., between steps of the procedure) or lost (i.e., not finding the appropriate device to act upon) among the different devices on the HMI. During the first session, the trainees were encouraged to ask each question necessary to understand the procedures and the dynamics of the system. They were also informed that during the next two sessions, they were not supposed to ask any further questions. Trainees were strongly advised not to share with other colleagues the contents of the experimental sessions in order to prevent other trainees from learning through their partners. They were also discouraged from looking for and/or studying material related to ship machinery systems.

In the second session or acquisition phase of this experiment, the trainees performed the operation of the diesel generator with two malfunctioning devices during twelve runs. The time allowed to perform this task was limited to 6 minutes, however, trainees could stop the simulation run whenever they considered they had completed the task. The AC group was provided with augmented cueing on 50% of the runs. That is, 6 runs in 12 were preceded by augmented cueing which demonstrated the performance on the task by highlighting the devices on which they had to act in the adequate sequence. The augmented cueing strategy intermittently, i.e., three times, highlighted the target devices reversing the colour of the background squared area in which any target device or parameter data value is embedded. This was done in the same order as in the procedures presented in Tables 5 and 6 and one device at a time. The malfunctions included in the malfunctions menu were neither shown nor highlighted. AC trainees performed the task in the normal fashion immediately after they were provided with augmented cueing. The augmented cueing was removed between the fifth and eighth simulation runs so that trainees did not become too reliant upon it. The pattern adopted for removal of the augmented cues was as is shown in Table 10. The NAC group performed the same task during the acquisition phase without augmented cueing.

TABLE 10
Pattern of Removal of the Augmented Cueing Strategy During the Acquisition Phase for the Augmented Cueing Group. Rows Present Simulation Runs and Whether Augmented Cueing (AC) Was Provided or Not in Each Run

	Augmented cueing group											
Run	1	2	3	4	5	6	7	8	9	10	11	12
AC	Yes	Yes	Yes	Yes	Yes	No	Yes	No	No	No	No	No

In the third practice session or transfer phase, both AC and NAC trainees performed the operation of the diesel generator in the same way. This time, the task included the solution of four malfunctions not experienced earlier. Training was provided during twelve runs with available time to perform the tasks limited to 7 minutes.

In order to facilitate the reduction of errors and maintain performance improvement (Van Matre, Pennypacker, Hartman, Brett, and Ward, 1981) across runs, knowledge of performance was provided on the computer screen after each run to both AC and NAC, during the practice runs (first session), acquisition, and transfer phases. This consisted of a performance template taken from an expert machinery control room operator performing the tasks, and the performance record resulting from trainee's performance in that run. The expert template was located along the left side of the screen and the actual performance record on the right side. As the screen was not large enough to show the whole performance records, subjects had to scroll down and up to check the complete lists. Trainees were required to correct their own results, marking a check box if a certain step/action was incorrect according to the instructions given. When the steps were correct, they merely left the check box blank. Trainees were allowed to correct and examine their performance for approximately 3 minutes. The experimenter provided guidance to trainees when they were not correcting their performance records correctly. Response-produced feedback was provided after the trainees chose to repair a certain device on the malfunctions list, the NDGSTRT displayed a message on the bottom left corner of the screen indicating the result of that action. This was either correct or one of two possible false attempt messages, i.e., attempt to repair active component or attempt to repair passive malfunction.

Trainees were debriefed after they had completed the transfer phase. Debriefing mainly consisted of the explanation of trainees' questions which, after training, were still unresolved. Special attention was devoted to provide understanding of the behaviour of the system and the interactions between sub-systems. After trainees had performed the experimental sessions they were again encouraged not to share the knowledge gained throughout the experiment with their colleagues. This was done in order to prevent other trainees from acquiring knowledge outside of the experimental context.

4.3 Results

Statistical tests are considered non-significant when $p > 0.10$, unless otherwise noted. Data were transformed in order to normalise distributions and homogenise variances which are prerequisite assumptions of ANOVA analyses and other parametric tests. Performance time was logarithmically transformed, runs, steps, errors, attempts and checks, were square root transformed. For significant interactions, univariate *F* tests were computed to investigate which means were significantly different.

The strategy of analysis adopted in this experiment aims at two objectives: first, the investigation of the level of achievement attained by trainees after training (i.e., includes the acquisition and transfer phases) in the AC and NAC groups is considered. This is investigated by analysing trainees' achievement of performance criterion as well as the results of training as a whole. Second, the investigation of the effects of the augmented and non-augmented cueing strategies on the skill acquisition process, which is analysed through the effects of practice (sequence of training runs) upon task performance. With this analysis the possibility that different performance strategies are developed as a function of training conditions can be examined.

The training results were thus analysed according to the following scheme: Firstly, measures concerning the achievement of performance criterion were analysed for both the acquisition and transfer phases. Secondly, a whole set of performance measures were averaged across training runs and analysed for the acquisition and transfer phases. Finally, the practice or training effects were analysed as indicated by the previous analyses for the acquisition and transfer phases. In this case, the training runs were grouped in blocks of three. The progress of performance improvement for the AC vs. NAC training groups is shown in Figure 11.

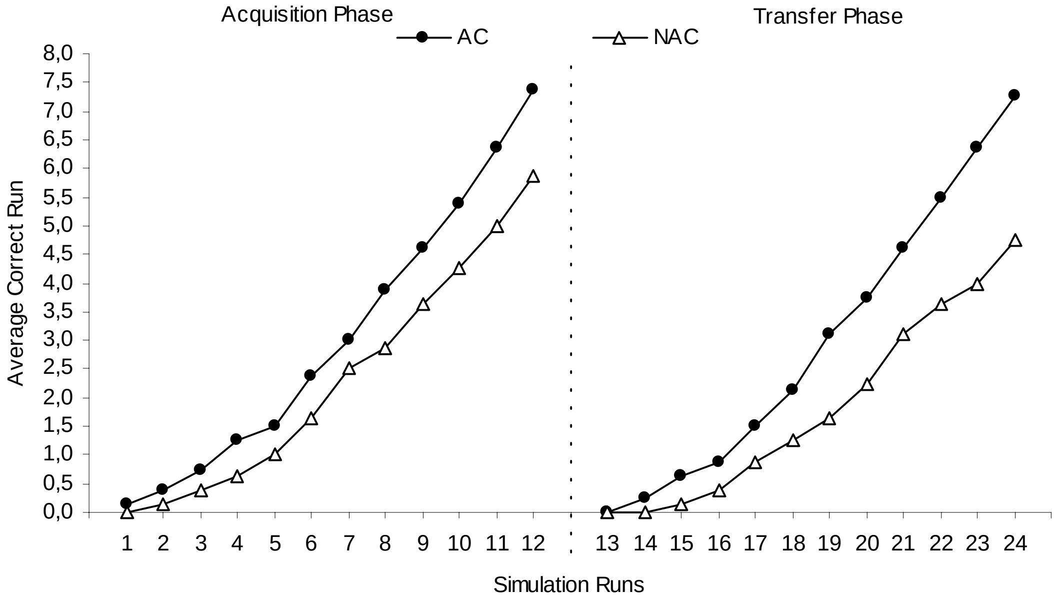

FIGURE 11 Acquisition and transfer phases: Cumulative average frequency of correct simulation runs as a function of AC vs. NAC training group. AC = Augmented Cueing group; NAC = Non-Augmented Cueing group.

4.3.1 Overall training results as a function of augmented cueing

The main independent variable has been training with or without augmented cues. The following qualitative and quantitative analysis are aimed at obtaining support to accept or to reject the null hypothesis. This is, no different training effects can be found between training groups using augmented cueing techniques (AC) or training without them (NAC).

4.3.1.1 Augmented cueing effects on performance criterion

Analysis of trainees' performance until they achieved performance criterion was examined by means of a qualitative analysis, and by means of non-parametric Mann-Whitney *U* tests. The qualitative analysis evaluated the percentage of trainees achieving the criterion. The non-parametric test analysed the number of the run on which trainees achieved the criterion.

Acquisition phase

After the acquisition phase, a qualitative analysis was first carried out to investigate how many trainees in the AC and NAC groups reached performance criterion and how many runs they required to attain this. An important result shows that while all of the AC trainees achieved performance criterion before the limited number of training runs (12), only 6 out of 8 NAC trainees did so, the remaining 2 did not reach performance criterion. These results are graphically summarised in Figure 12. This shows a relative advantage of AC over NAC because of those NAC trainees who did not reach performance criterion.

The Mann-Whitney *U* test performed on the runs on which trainees reached the criterion did not reveal significant differences between AC and NAC, $U = 20$, $p = 0.24$. Mean ranks obtained by AC and NAC were 7 and 10 runs, respectively. This result also points in the direction of a relative advantage of AC over NAC but not strongly enough to exhibit statistical significance.

Transfer phase

After the transfer phase, a similar pattern of results was found. In the AC group, all of the trainees reached performance criterion compared to only 5 out of 8 trainees in the NAC group. Three trainees in the NAC group did not achieve performance criterion. These results are graphically illustrated in Figure 12.

The results show that AC trainees were better than the NAC group in both the acquisition and transfer phases. Each trainee in the AC group reached performance criterion within the available training time while about a third of the NAC trainees did not.

With this qualitative analysis some support has been obtained for the alternative hypothesis, i.e., training which incorporates augmented cueing (AC) on the NDGSTRT can result in faster and higher training achievement in comparison to the training provided on the NDGSTRT without augmented cueing (NAC).

In the transfer phase, the Mann-Whitney *U* test did not show significant differences between AC and NAC, $U = 22$, $p = 0.33$. Mean ranks obtained by AC and NAC were 7.3 and 9.8 runs, respectively. This result also points in the direction of those obtained in the acquisition phase, a relative advantage of AC over NAC.

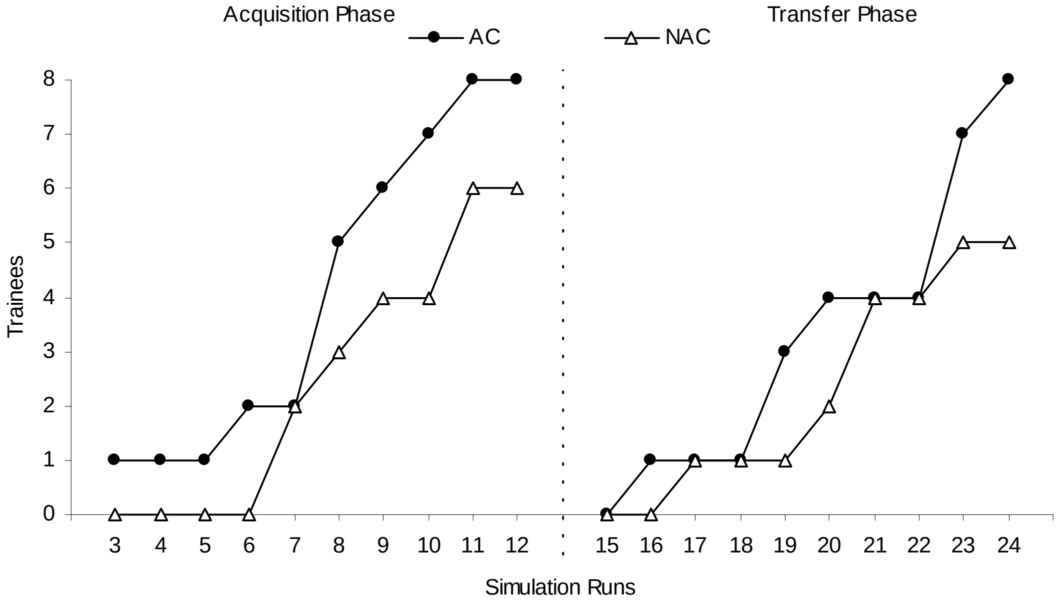

FIGURE 12 Acquisition and transfer phases: Cumulative frequency of trainees reaching performance criterion on a particular training run as a function of training group. AC = Augmented Cueing; NAC = Non-Augmented Cueing.

4.3.1.2 Overall augmented cueing effects on several performance measures

In the following sections, the results of the multivariate analysis of variance and the examination of the correlations between the dependent variables collected during training are reported.

4.3.1.2.1 Augmented cueing multivariate effects

To analyse the effects of the two training strategies, a multivariate analysis of variance (MANOVA) was performed which included all the dependent measures collected during training. These were:

- RCR or number of runs to achieve performance criterion,
- STP or average number of steps/actions performed in each run,
- PFT or average elapsed time during each training run,
- ERR or average percentage of errors made in each run,
- ATT or average number of false attempts to repair malfunctions, and,
- CHK or average number of parameter data checks performed to diagnose the malfunctioning devices on each training run.

Acquisition phase

No significant differences were found between AC and NAC in the multivariate analysis of variance for the acquisition phase. Given that the qualitative analysis performed earlier showed some differences between AC and NAC, this lack of significant differences could be provoked by the small sample size, and/or by noise produced by some performance variables which tend to stabilise as training progresses. This noise could be masking the effects of training strategies on other measures when including each variable in the analysis. Some of the noise-producing variables could be STP and PFT which tend to become stable as training progresses. Figure 13 illustrates these results.

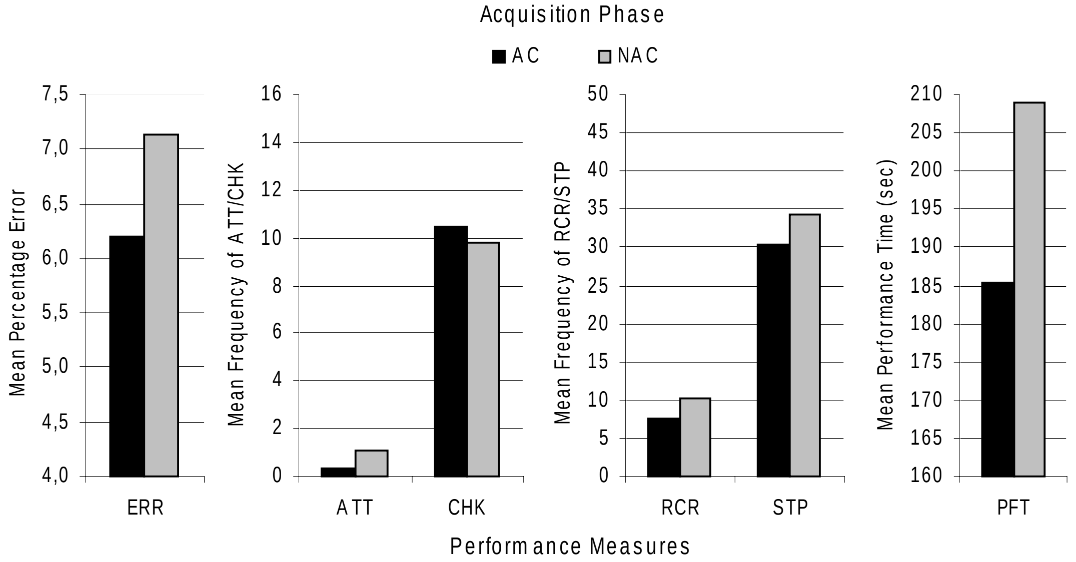

FIGURE 13 Mean acquisition performance on the dependent variables as a function of training group. AC = Augmented Cueing group; NAC = Non-Augmented Cueing group; ERR = percentage of errors; ATT = false attempts to repair malfunctioning devices; CHK = parameter value checks; RCR = runs required to reach performance criterion; STP = steps or actions performed in each simulation run; PFT = performance time spent on the task.

Another possibility could be that knowledge of performance provided to both groups of trainees facilitates training in both AC and NAC in a way which eliminates the effects of the experimental manipulations as it is confounded with the experimental conditions (Van Matre, et al., 1981). To investigate these possibilities, individual *F* tests were performed for the dependent variables. Both groups marginally differed in the number of runs required to achieve the criterion, $F(1, 14) = 2.88$, $MSE = 0.24$, $p = 0.11$. This concurs with the tendencies, in the same direction, found in the non-parametric analysis reported above. The results also showed a marginally significant effect of AC vs. NAC on the number of steps performed in each run, $F(1, 14) = 2.93$, $MSE = 0.17$, $p = 0.11$. The AC group required fewer steps to complete the task than the NAC group. AC and NAC groups differed on the number of attempts executed on each run in order to solve the malfunctions, *F*(1,14) = 3.09, *MSE* = 0,28, *p* = 0.10. The means of the AC and NAC groups are graphically represented in Figure 13.

Transfer phase

Similar analyses were carried out for the results of the transfer phase. The MANOVA did not show differences between AC and NAC. Independent F tests only showed a marginally significant difference between AC and NAC on the number of attempts trainees executed in each run, *F*(1, 14) = 2.84, *MSE* = 0.26, *p* = 0.11. AC made fewer false attempts to repair malfunctioning devices than NAC. This result is consistent with the acquisition phase, and can be interpreted as a difference in the strategies adopted by AC trainees as opposed to NAC in order to solve the problems of malfunctioning devices. The greater number of false attempts indicate an erratic trial and error strategy to repair malfunctioning devices. It must be noted that trainees in both groups were strongly advised not to use this type of erratic strategy. Instead, they should compare the values displayed on the interactive diagram to analyse the behaviour of the system and thus detect the malfunctioning devices before trying to repair them. The means of the AC and NAC groups for each dependent measure are graphically illustrated in Figure 14.

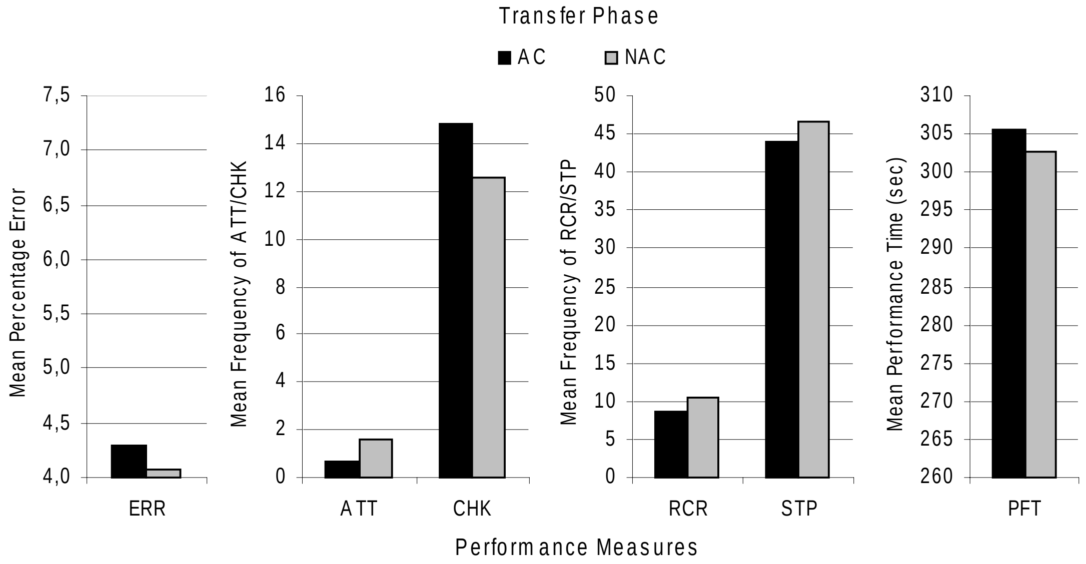

FIGURE 14 Mean transfer performance on the dependent variables as a function of training group. AC = Augmented Cueing group; NAC = Non-Augmented Cueing group; ERR = percentage of errors; ATT = false attempts to repair malfunctioning devices; CHK = parameter value checks; RCR = runs required to reach performance criterion; STP = steps or actions performed in each simulation run; PFT = performance time spent on the task.

4.3.1.3 Relationships between dependent variables

The MANOVAs did not show significant differences between the AC vs. NAC groups. To evaluate the possibility that the homogeneity of variances assumption was violated, Bartlett's-Box F tests were carried out for the acquisition and transfer phases. Neither in the acquisition phase nor in the transfer phase was the homogeneity of variances assumption violated. Thus, the possibility that knowledge of performance was playing an equalising role on AC and NAC assumes an increased value.

The degree of association between the dependent measures collected during training until trainees achieved performance criterion was analysed. This should demonstrate the consistency of the different measures to evaluate the effects of the training strategies across trainees until they reached the criterion. Additionally, this could show differential task performance strategies produced by the different training conditions if the correlation patterns differ between AC and NAC. The correlation tests also helped to decide which factors should be included in the ANOVAs for the test of practice effects.

4.3.1.3.1 Within-groups relationships

The results of the correlation tests were firstly examined for qualitative differences between AC and NAC. Descriptions of the following results are reported in these terms.

Acquisition phase

For AC, most of the correlations between performance measures were positive and low, $r = 0.48$. Some other correlations were higher and either positive or negative, $r = \pm 0.71$. The weakest association was found in the STP-CHK relation. Low association was also found in ERR-ATT. The relations RTC-STP, RTC-PFT, and RTC-ERR were also weak.

Some differences were observed for NAC: most of the relations were positive and strongly associated, $r = 0.82$, the weakest relationships, $r = -0.43$ were found in ATT-CHK, ERR-CHK, PFT-CHK, STP-CHK, and RTC-CHK, each of these relations was negative.

The main differences between AC and NAC are in the weaker correlations. While in AC, weaker associations are observed in the relations between ERR, and ATT, PFT, RTC, STP; in NAC, the weaker relations are observed between CHK and the other measures. On the one hand, this indicates that NAC trainees' performance was less associated to value data checks than it was in AC. On the other hand, in AC, ERR was less associated with other measures than it was in NAC.

Transfer phase

In the transfer phase, some of the correlations between performance measures in AC were moderately higher than in the acquisition phase, $r = 0.43$, and either positive or negative. Some other correlations were lower, $r = 0.41$, and positive or negative. The weakest association was found in the STP-CHK relation. Low associations were also found in the relations PFT-ATT, RTC-PFT, ERR-CHK, ATT-CHK, and RTC-PFT. The relations RTC-ERR and STP-ERR were higher than those observed previously.

The most relevant differences of NAC compared to AC in the transfer phase were the higher and negative associations found in ERR-CHK and ATT-CHK. Other associations became more similar to AC than in the acquisition phase. Nevertheless, the direction of the differences remained almost constant.

These results indicate that both training groups maintained similar task performance strategies in the acquisition and transfer phases. The strategy adopted by NAC trainees was closer to a trial and error than it was in the AC, which based its performance on value data checks to diagnose the problems. A summary of the results concerning the relationships between dependent measures for AC and NAC in the acquisition and transfer phases is presented in Table 11. The differences in the strength of the associations between AC and NAC are represented in the columns 'Diff'. The direction of these associations are indicated as positive or negative Pearson's (r) values.

TABLE 11
Within-group Correlations between Dependent Variables/Measures

	Acquisition Phase			Transfer Phase		
Correlated Dependent Variables	AC *r*	Diff.	NAC *r*	AC *r*	Diff.	NAC *r*
Runs to Criterion - Errors	0.11	<	0.93	0.44	>	0.17
Runs to Criterion - Attempts	0.82	=	0.88	0.80	=	0.83
Runs to Criterion - Checks	-0.71	>	-0.43	-0.58	>	-0.27
Steps - Errors	0.28	<	0.81	0.80	=	0.80
Steps - Attempts	0.42	<	0.95	0.93	>	0.70
Steps - Checks	0.03	<	-0.36	0.06	<	-0.47
Performance Time - Errors	0.36	<	0.83	0.18	<	0.37
Performance Time - Attempts	0.16	<	0.88	0.03	<	0.58
Performance Time - Checks	0.48	>	-0.36	0.49	>	-0.12
Errors - Attempts	-0.03	<	0.84	0.78	>	0.64
Errors - Checks	0.27	=	-0.24	-0.19	<	-0.47
Attempts - Checks	-0.56	>	-0.26	-0.19	<	-0.47

Note. AC = Augmented Cueing group; NAC = Non-Augmented Cueing group; Diff. = differences in the strength of the associations between AC and NAC.

4.3.1.3.2 Between-groups relationships

The correlations between the dependent variables were also computed across both training conditions. The significance of the correlation coefficients was tested in order to identify which of these were reliable.

Acquisition phase

The correlation analysis of the dependent variables demonstrated that most of the measures of performance are positive and highly associated. The significance test indicated that most of these are statistically significant. Nevertheless, the associations between STP-CHK, PFT-CHK, and ERR-CHK were not significant. Note that performance strategies have been proposed to depend on erratic malfunction solving, or on the check-up of relevant parameter values indicating system behaviour. Due to the difference of the correlation between these variables in AC vs. NAC in the acquisition phase, their correlations have not exhibited significance. Additionally, the relation ATT-CHK was stronger in AC than in NAC (see Table 11).

Transfer phase

In the transfer phase, the correlation coefficients also demonstrated positive and strong associations between the dependent variables. In this case, only eight of the fifteen possible relations were significant. The relationships between the average number of checks performed on each training run and the rest of the dependent variables was only significant in the relations RTC-CHK and ATT-CHK. During the transfer phase, the strength of the relationships between the dependent variables became weaker than in the acquisition phase.

4.3.2 Training progress as a function of augmented cueing

With the analysis of the practice or learning effects, the focus is placed on the training process and its dynamics. The possibility that AC vs. NAC groups developed different task performance strategies can be examined as well as the effects of practice.

As the sample size was not too large, separate analysis of performance on the dependent variables would probably not produce statistically significant differences in the ANOVA as they did not in the MANOVA. A possible solution to surmount this inconvenience can be to increase the number of factors in the analyses. Increasing the number of factors in the tests also increases the number of observations in each cell. This procedure could expose otherwise indiscernible differences between AC and NAC.

Analysing the effects of the blocks of runs on each of the dependent variables at a time, results in a split-plot factorial 2 (AC vs. NAC training group) × 4 (blocks of 3 runs) ANOVA with the first factor being a between-subjects, and the second a within-subjects. Adding another within-subjects factor results in a 2 × 4 × 2 ANOVA with the first factor being a between-subjects and the two second factors within-subjects. With this procedure, the effect of another factor sums to the total sum of squares of the analyses and increases its power by adding more degrees of freedom to the terms of the F ratio. The number of factors in the analyses concerning the effects of practice were thus increased by adding another within-subjects factor to the practice factor (blocks of runs).

The factors added to the practice factor in the following analysis were three of the bivariate combinations investigated in the previous analyses. The number of bivariate relationships were fifteen, but not all possible combinations should be included in different ANOVAs. This is so because it could cause significant differences to appear only due to chance. Therefore, reducing the number of analyses to a minimum reduces the probability that differences between the independent variables could happen solely by chance. The factors that were included in the following ANOVAs were selected on the basis of the results of the previous MANOVAs plus the inspection of the correlations between the dependent variables. Based on these grounds, three factors were selected. These factors included the number of steps performed on each run (STP) in combination with the number of attempts (ATT). These two factors showed marginally significant differences between AC and NAC in the acquisition phase. The other two factors included the combination ATT-CHK and ERR-ATT on which AC and NAC differed during acquisition and transfer phases. With this strategy, it is expected that interactions between both training strategies and the practice variable can be detected if they really occur. The results of these analyses are reported in the following sections.

4.3.2.1 Augmented cueing by block by number of steps-attempts results

The ANOVA was a split-plot factorial with 2 (AC vs. NAC training group) × 4 (blocks of 3 runs) × 2 (STP and ATT) factors, being the first between-subjects factor with two levels AC and NAC, and the latter two within-subjects with four levels, the four blocks of 3 runs, and two levels, measures of STP and ATT in each of the four blocks.

Acquisition phase

- Between-subjects effects or AC vs. NAC effects

The ANOVA revealed a marginally significant effect of training group, $F(1, 14) = 3.18$, $MSE = 22.59$, $p = 0.096$). The AC group performed fewer steps and fewer false attempts than the NAC group. These results are consistent with the mean differences found in the MANOVA and the strength of the association of STP-ATT in the acquisition phase (AC $r = 0.43$, NAC $r = 0.96$).

- Within-subjects effects or block and STP-ATT effects

The practice or block effect showed significant differences, *F*(3, 42) = 9.83, *MSE* = 15.18, *p* = 0.000. The number of steps and attempts performed decreased over training. The interaction group × block did not show significant differences. The STP-ATT effect was significant, *F*(1, 14) = 147.97, *MSE* = 537.01, *p* = 0.000. Both steps and attempts decreased over the acquisition period in both groups. The group × STP-ATT interaction did not reach significance level. This lack of significance of the interaction could be due to the fact that the number of steps required to perform the tasks tend to be stable after performance criterion has been achieved. Thus, masking the interaction. The block × STP-ATT interaction was statistically significant, *F*(3, 42) = 14.81, *MSE* = 16.11, *p* = 0.000. The interaction of second order group × block × STP-ATT did not indicate significant differences.

- Univariate tests

Comparisons of critical differences were carried out for the block effect, and the interaction block × STP-ATT. These indicated that only mean differences were different in the first block and means in the following blocks were not different, *F*(1, 14) = 18.53, *MSE* = 2.35, *p* = 0.001. The interaction block × STP-ATT was only significant in the first block, *F*(1, 14) = 25.48, *MSE* = 1.76, *p* = 0.000, and marginally significant in the second block *F*(1, 14) = 3.62, *MSE* = 0.90, *p* = 0.08, indicating that after the second acquisition block, only limited benefit of training is obtained. This was probably due to the proportion of trainees that reached performance criterion before the end of block 3. It is likely that after criterion has been achieved, performance stabilises as illustrated in Figure 15.

Transfer phase

- Between-subjects effects or AC vs. NAC effects

The AC and NAC groups, contrasting with the acquisition phase, did not differ significantly in the transfer phase. They performed a similar number of STP and ATT. These results are consistent with the lack of differences in means found in the MANOVA (i.e., exclusively considering the runs required to reach performance criterion) and the strength of the association of STP-ATT in the transfer phase which were more similar (AC *r* = 0.93, NAC *r* = 0.70) than in the acquisition phase.

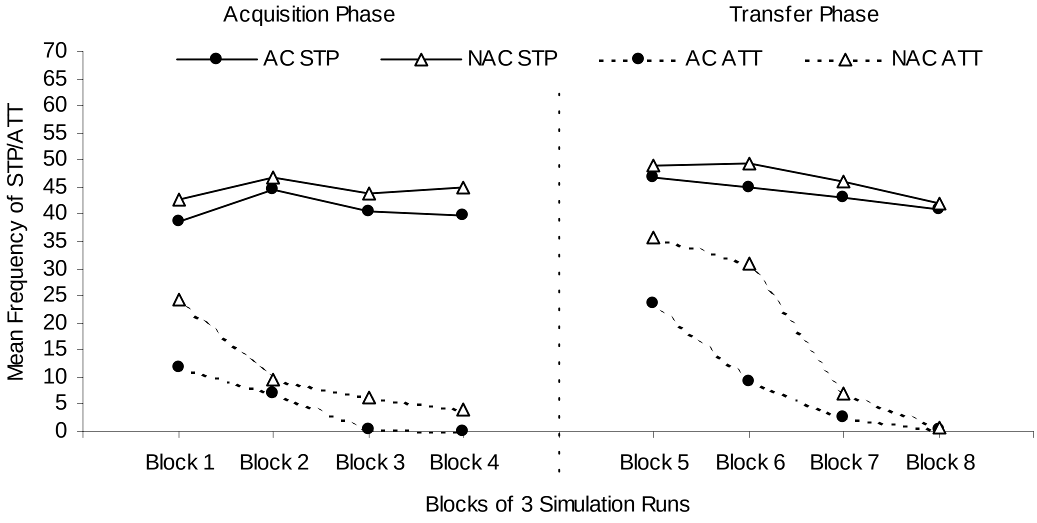

FIGURE 15 Acquisition and transfer phases: perfromance measures as a function of group and training block of 3 simulation runs. AC STP = Augmented Cueing group on the dependent variable STP or number of steps performed; NAC STP = Non-Augmented Cueing group on the dependent variable STP or number of steps performed; AC ATT = Augmented Cueing group on the dependent variable ATT or false attempts to repair malfunctioning devices; NAC ATT = Non-Augmented Cueing group on the dependent variable ATT or false attempts to repair malfunctioning devices.

- Within-subjects effects or block and STP-ATT effects

The practice or block effect showed significant differences, *F*(3, 42) = 21.92, *MSE* = 43.42, *p* = 0.000. The number of steps and attempts performed decreased over training. The group × block interaction did not reveal significant differences. The STP-ATT effect was significant, *F*(1, 14) = 123.11, *MSE* = 409.53, *p* = 0.000. Both steps and attempts decreased over the transfer period in both groups. The group × STP-ATT interaction did not reach significance. The block × STP-ATT interaction was statistically significant, *F*(3, 42) = 30.26, *MSE* = 28.94, *p* = 0.000. The second order interaction, group × block × STP-ATT, similarly to the acquisition phase, did not indicate significant differences.

- Univariate tests

In the transfer phase, similar comparisons of critical differences were carried out for the block effect, and the interaction block × STP-ATT. The results were similar to those found in the acquisition phase. Only the first block of runs was statistically significant for the block effect and the interaction block × STP-ATT. The only difference with respect to the acquisition phase is that the marginal significance of block 2 in the block × STP-ATT interaction disappeared in this phase of training. Results are presented in Figure 15.

4.3.2.2 Augmented cueing by block by number of error-attempts results

The ANOVA was similar to the previous analyses, a split-plot factorial with 2 (AC vs. NAC) × 4 (block) × 2 (ERR-ATT) factors. The first was a between-subjects factor, and the further two, within-subjects.

Acquisition phase

- Between-subjects effects

No significant differences were found between AC and NAC. Nevertheless, as Figure 16 shows, AC made fewer attempts than NAC, and also committed fewer errors after the second block of runs. This is consistent with the difference between-group means obtained in the MANOVA, and the degree of association of ERR-ATT in the acquisition phase (AC r = -0.03; NAC r = 0.84).

- Within-subjects effects

The practice or block effect produced significant differences, $F(3, 42) = 42.29$, $MSE = 102.06$, $p = 0.000$. The number of steps and attempts performed decreased over training. Similarly to the previous ANOVA, the interaction group × block did not show significant differences. The effect of ERR-ATT was significant, $F(1, 14) = 63.83$, $MSE = 188.50$, $p = 0.000$. Both errors and attempts decreased over the acquisition period in both groups. The group × ERR-ATT interaction was marginally significant $F(1, 14) = 3.80$, $MSE = 11.21$, $p = 0.072$. As Figure 16 shows, the reduction of errors and attempts was greater for AC than NAC. The block × ERR-ATT interaction was statistically significant, $F(3, 42) = 6.01$, $MSE = 6.09$, $p = 0.002$. The interaction group × block × ERR-ATT did not produce significant differences.

- Univariate tests

Critical differences were tested for the block effect. These demonstrated that performance in the first, $F(1, 14) = 76.47$, $MSE = 3.75$, $p = 0.000$, and second blocks $F(1, 14) = 10.32$, $MSE = 1.92$, $p < 0.006$, improved significantly in both AC and NAC groups, while no improvement was obvious in the others. The results of the tests for the interaction block × ERR-ATT were similar. The first, $F(1, 14) = 8.46$, $MSE = 1.54$, $p = 0.01$, and second blocks, $F(1, 14) = 5.18$, $MSE = 0.68$, $p = 0.04$ revealed significance on the reduction of errors and attempts. This indicates that after the number of steps to perform the tasks become stable, the reduction of errors continue to improve for some additional time. See Figure 16 for these comparisons.

Transfer phase

- Between-subjects effects

No significant differences were found between AC and NAC. Nevertheless, as Figure 16 shows, AC made fewer attempts than NAC, and also committed fewer errors since the beginning of transfer. The difference between AC and NAC on the number of errors made, disappeared in the last transfer block but was still apparent in the number of attempts in the last block. Differences between group means obtained in the MANOVA also indicate these effects. The degree of association between ERR-ATT in the transfer phase (AC $r = 0.78$, NAC $r = 0.64$) tended to the similarity even though it was stronger for AC than NAC compared to the acquisition phase where the association was very weak in AC and strong in NAC.

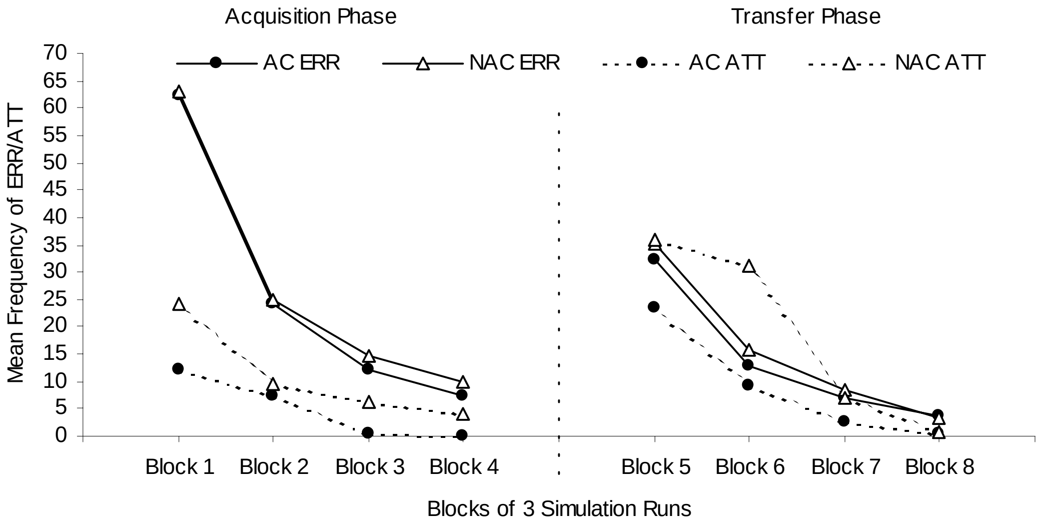

FIGURE 16 Acquisition and transfer phases: performance measures as a function of group and training block of 3 simulation runs. AC ERR = Augmented Cueing group on the dependent variable ERR or number of errors committed; NAC ERR = Non-Augmented Cueing group on the dependent variable ERR or number of errors committed; AC ATT = Augmented Cueing group on the dependent variable ATT or false attempts to repair malfunctioning devices; NAC ATT = Non-Augmented Cueing group on the dependent variable ATT or false attempts to repair malfunctioning devices.

- Within-subjects effects

The effect of block produced significant differences, $F(3, 42) = 44.40$, $MSE = 113.56$, $p = 0.000$. The number of steps and attempts performed decreased over training. The interaction group × block did not show significant differences. No significant differences were found on ERR-ATT. Both errors and attempts decreased quite similarly. The group × ERR-ATT interaction was significant at the 10% level $F(1, 14) = 4.49$, $MSE = 9.16$, $p = 0.052$. As depicted in Figure 16, the reduction of errors and attempts was greater for AC than NAC, especially in the number of attempts. The block × ERR-ATT interaction was significant, $F(3, 42) = 3.40$, $MSE = 3.99$, $p = 0.026$. The interaction group × block × ERR-ATT did not produce differences.

- Univariate tests

Similar tests were carried out for the block effect in the transfer phase. Only mean differences were found in the first block, $F(1, 14) = 147.90$, $MSE = 2.29$, $p < 0.000$. The results of the tests for the interaction block × ERR-ATT indicated that performance on ERR-ATT marginally improved in the second block, $F(1, 14) = 3.11$, $MSE = 1.31$, $p = 0.10$, and significantly in the third block, $F(1, 14) = 6.44$, $MSE = 0.50$, $p = 0.02$. That is, errors and attempts were significantly reduced in the second and third blocks in AC and NAC (see Figures 14 and 16 for comparisons).

4.3.2.3 Augmented cueing by block by number of attempts-checks results

A split-plot factorial ANOVA with 2 (AC vs. NAC groups) × 4 (blocks of 3 runs) × 2 (ATT-CHK) factors. The first was a between-subjects factor with two levels, and the remaining two within-subjects with four levels and two levels respectively.

Acquisition phase

- Between-subjects effects

Groups AC and NAC did not differ significantly during the acquisition phase. Their performance was quite similar on both number of attempts and number of checks. Despite this statistical similarity, Figure 17 illustrates a small difference in the number of checks until the fourth block. The AC group performed more checks than the NAC. AC consistently made fewer false attempts than NAC. Similar results were found in the previous MANOVA (see Figure 13). The analyses of the ATT-CHK correlation revealed that checks and attempts were more negatively associated in AC ($r = -0.56$) than in NAC ($r = -0.26$). This indicates that NAC false attempts to repair malfunctioning devices were less associated to the execution of value data checks than was the case for AC. A difference in problem solving strategies can be indicated by this relationship. The direction of the ATT-CHK relationship can be best understood in Figure 17.

- Within-subjects effects

Performance improved across blocks of runs. This effect was significant *F*(3, 42) = 6.28, *MSE* = 9.77, *p* = 0.001. The interaction group × block was not significant. ATT-CHK effect was statistically significant, *F*(1, 14) = 145.48, *MSE* = 1614.01, *p* = 0.000, the association ATT-CHK remained consistently across groups and blocks. The interaction group × ATT-CHK was not significant. The interaction block × ATT-CHK was significant, *F*(3, 42) = 13.52, *MSE* = 24.58, *p* = 0.000, the number of attempts systematically decreased across training blocks while checks tended to increase. The interaction group × block × ATT-CHK was not significant.

- Univariate tests

Comparisons of critical differences were carried out for the block effect. This only revealed mean differences in the first block, *F*(1, 14) = 11.50, *MSE* = 2.21, *p* = 0.004. The interaction block × ATT-CHK was only significant in the first block, *F*(1, 14) = 23.30, *MSE* = 2.93, *p* = 0.000. This indicates that after the second practice block, only limited training effects can be observed. Figure 17 shows these results.

Transfer phase

- Between-subjects effects

Similarly to the acquisition phase, groups AC and NAC did not differ significantly during this phase of training. Despite the lack of statistical differences, Figure 17 shows a difference on the number of checks and false attempts between AC and NAC. In the previous MANOVA (see Figure 14), the number of checks until trainees reached performance criterion was almost equal, but there was a marginally significant difference on the number of false attempts made. As indicated in the analyses of the ATT-CHK correlation in the acquisition phase, during transfer, these analyses demonstrated that checks and attempts were more associated in NAC (r = -0.47) than in AC (r = -0.19). This difference in problem solving strategies almost remained constant after the acquisition phase, though NAC tended in this direction through the transfer phase. This effect can be attributed to the fact that AC quickly stabilised the amount of checks performed in order to solve the malfunctions, thus reducing the strength of the association ATT-CHK.

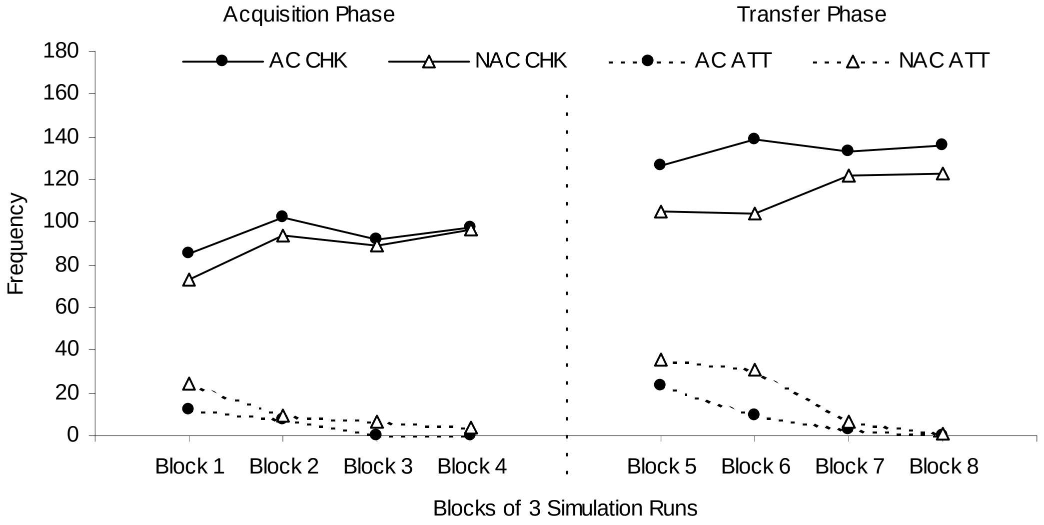

FIGURE 17 Acquisition and transfer phases: Perfromance measures as a function of group and training block of 3 simulation runs. AC CHK = Augmented Cueing group on the dependent variable CHK or number of parameter value checks performed; NAC CHK = Non-Augmented Cueing group on the dependent variable CHK or number of parameter value checks performed; AC ATT = Augmented Cueing group on the dependent variable ATT or false attempts to repair malfunctioning devices; NAC ATT = Non-Augmented Cueing group on the dependent variable ATT or false attempts to repair malfunctioning devices.

- Within-subjects effects

Figure 17 shows that attempts declined, i.e., performance improved across blocks of runs. This effect was significant *F*(3, 42) = 11.87, *MSE* = 26.83, *p* = 0.000. The interaction group × block was not significant. The ATT-CHK effect was statistically significant, *F*(1, 14) = 109.70, *MSE* = 2022.71, *p* = 0.000. The interaction group × ATT-CHK was not significant. The interaction block × ATT-CHK was significant, *F*(3, 42) = 33.78, *MSE* = 46.12, *p* = 0.000. Contrary to the acquisition phase and the other ANOVAs previously conducted, the interaction group × block × ATT-CHK was significant *F*(3, 42) = 2.65, *MSE* = 3.61, *p* = 0.061. This interaction is clearly shown in Figure 17. AC performed more checks and made fewer false attempts than NAC. These results were constant across training blocks.

- Univariate tests

As well as in the acquisition phase, critical differences were tested for the block effect. This also revealed a difference in the first, *F*(1, 14) = 29.94, *MSE* = 2.64, *p* = 0.000. The Block × ATT-CHK interaction was only significant in the first block of the transfer phase, *F*(1, 14) = 141.67, *MSE* = 0.96, *p* = 0.000. The interaction group × block × ATT-CHK was only marginally significant in the first and third blocks, *F*(1, 14) = 3.62, *MSE* = 0.96, *p* = 0.08, and *F*(1, 14) = 3.11, *MSE* = 1.43, *p* =

0.10, respectively. This indicates that after the second practice block, only limited advantage is gained with practice unless performance criterion has not yet been reached , as happened to some of the trainees in the NAC group. In this case, training would still continue to improve performance.

4.3.3 Transfer effects of the augmented cueing training strategy

Similarly to Experiment 1, the analysis of the transfer/interference effect of the AC vs. NAC training strategies was carried in the same way. The percentage of errors variable was standardised into *z* scores (mean = 0; standard deviation = 1) for the AC, and NAC groups, for the blocks of runs 3 and 4 of the acquisition phase, and the blocks of runs 6 and 7 of the transfer phase. The analysis (split-plot factorial ANOVA) was a 2 (AC vs. NAC groups) × 2 (acquisition vs. transfer phases) × 2 (blocks 3-4 vs. blocks 6-7) with groups as the between-subjects factor, and acquisition/transfer phases and blocks of runs as within-subjects factors. The results did not show significant differences on the effect of training group $F(1, 14) = 0.55$, $MSE = 0.34$, $p = 0.47$. The other main effects as well as the interactions did not reach significance (all Fs < 1). The disturbances in performance as a result of the task change from the acquisition to the transfer phase was similar for both AC and NAC groups of trainees.

4.4 Discussion

From a qualitative standpoint, the results of this experiment supported the hypothesis posited earlier. Training with the augmented cueing strategy produced higher and faster performance improvement compared to non-augmented cueing. In the acquisition phase, this was demonstrated by the achievement of performance criterion of 100% of AC trainees, compared to only 75% of NAC trainees. In the transfer phase, the percentage of NAC trainees that did not reach performance criterion increased to 37.5%, while 100% of AC trainees achieved criterion. Also, trainees in AC achieved performance criterion faster than NAC. The non-parametric tests showed a similar tendency in the number of runs required by trainees to achieve the criterion. In the F test computed on the squared root transformed number of runs that trainees required to achieve criterion, a marginally significant difference also favoured the AC group.

On the measures collected during the experiment for each trainee, the AC and NAC training strategies did not produce differences, though the examination of the means of AC vs. NAC showed better performance of AC on each measure during the acquisition phase. These differences disappeared on the time elapsed required by trainees to reach criterion and the number of checks made by them to identify the malfunctioning devices. Though not significant, means on the other variables still showed better performance of AC over NAC.

Three possible reasons which could account for the discrepancy between the qualitative analysis and the MANOVA were indicated earlier. First, a small sample size which does not provide for enough statistical power, and thus, no differences between training groups. The second, pointed in the direction of the likelihood that dependent variables, which after a certain level of proficiency has been achieved remain constant, such as the number of steps required to perform the tasks or the time required to perform them. The other possible explanation for the discrepancy between the qualitative and the quantitative analysis was that knowledge of performance systematically provided to trainees in AC and NAC after each training run produced stronger training effects than the cue augmentation, thus dissipating the effect of the experimental factor.

The group trained with augmented cues during the acquisition phase required fewer steps to reach performance criterion than that trained without augmented cues. AC trainees also performed fewer false attempts to repair malfunctions than did the NAC trainees. In the transfer phase, these differences between AC and NAC only remained constant for the number of false attempts to correct the malfunctioning devices until they reached performance criterion. These results support the hypothesis that augmented cueing produces the highest performance level as compared to non-augmented cueing. The degree of association between dependent variables showed that AC developed an adequate problem solving strategy. This consisted of the verification of the system state to diagnose the failure, after the malfunctioning device was thought to be correctly identified, the reparation was attempted. This strategy is inferred from the performance scripts which show this sequence of actions. NAC trainees on the other hand tried to repair the malfunctions with fewer system state verification steps than AC. This is also indicated by the smaller number of checks they performed -compared to AC- during the acquisition and transfer phases. This strategy points in the direction of one that is trial and error, even though both groups were strongly advised to verify the system state before attempting reparations.

Augmented cueing not only affected the amount of training achievement as indicated above, but also its quality. The analysis of the training process indicated that both training groups improved performance as training progressed, both in the acquisition and the transfer phases. The analysis of the effects of practice supported the hypothesis favouring augmented cueing training in the acquisition phase, even though augmented cueing was only provided in one half of the practice runs for AC. Therefore, it can be concluded that removing augmented cues does not necessarily hinder subsequent performance on procedural tasks as suggested by other authors for the retention of skills (Lintern, 1980).

The degree of the association between errors and attempts was greater for NAC than for AC in the acquisition phase. This means that many of the errors made by NAC were false attempts. In AC, many of the errors were of other types, such as unnecessary repetitions, and order of steps. The differences in the proportion of attempts and other errors between AC and NAC almost disappeared in the last block of the transfer phase. However, NAC still made more attempts than AC.

The difference found between AC and NAC in the association ATT-CHK in the acquisition and transfer phases was probably due to the fast stabilisation of performed checks in AC. In the acquisition phase, AC showed a stronger association than NAC. In the transfer phase, there was an inversion of this effect. On average and similarly to the acquisition phase, AC performed more checks than NAC. This is probably the best indicator of the performance strategies used by both groups. It is interesting to note that after developing a certain performance strategy, both groups tended to maintain this across the whole training process. AC developed a form of systematic diagnostic skill. NAC mainly performed the tasks on a trial and error basis. Therefore, we can propose that augmented cueing favoured the development and maintenance of appropriate diagnostic skills, while non-augmented cueing did not.

Future research should address some of the unanswered questions from this experiment. The main question should focus on the differential effects of augmented cueing strategies versus knowledge of performance by means of an experimental design which separates these effects. The schedule used in this experiment to remove augmented cues did not show adverse effects on AC trainees. Other schedules could be tested that without negatively affecting performance could speed-up training achievement. Response-produced feedback was present systematically in the diesel generator (i.e., NDGSTRT) as well as in Experiment 1. This consisted of two types, black to colour change and vice-versa of the device symbols, and feedback on the results of malfunction repairing attempts. Whether this is an effect which sums to the effects of other variables such as knowledge of performance could also be examined in future research. Additionally, the study of acquisition and development of strategic skills as a function of training strategies should also be considered in future research in the area of procedural task training. If, after the results of Experiment 3 are analysed, no skill achievement differences can be found between trainees trained on the NDGSTRT and the HF simulator; future experimental research should investigate whether these effects are artefacts or valid conclusions.

5 EXPERIMENT 3: LOW VS. HIGH-FIDELITY TRAINING SIMULATOR

5.1 Introduction

This research aims at the optimisation of training device costs and training effectiveness by means of empirical research that can be used for decision making within a systematic approach to training. An important cost-driving factor in training simulators is their degree of fidelity. Usually, a high level of fidelity is associated with highly expensive technological developments. However, the relationship between simulator fidelity and training effectiveness is not well understood for the moment (Alessi, 1988). Mainly, because the adoption of simulator training has been technology driven more than research driven. The aim of this research is to match the skills to be trained with appropriate levels of fidelity. This implies that the highest level of fidelity may be less cost-effective than other lower fidelity simulator configurations for the effective training of certain skills. Nevertheless, these suggestions need to be empirically validated in order to provide useful support for decision-making concerning the training programme development.

In the naval area, the type of simulator predominantly demanded by customers is that of the machinery control room. The requirements for a machinery control simulator typically specified are HF which are also the most expensive. However, the type of skills trained on this kind of simulator are primarily procedural. In fact, most tasks of importance aboard naval platforms are largely procedural e.g., machinery control, watch-standing, radar and sonar operation, etc (Hurlock and Montague, 1982; O'Hara, 1990).

The decisions leading to the procurement and acquisition of any training device should be supported by empirical data. In the context of machinery control room simulators, little research evidence has been found that either supports or refutes decisions leading to the acquisition of HF training simulators. This experiment primarily aims to bridge this knowledge gap by providing scientific and quantitative evidence of the relative effectiveness of a LF simulator as compared to a HF. The two simulators used (see section 4.1. for a detailed description) differ on several fidelity dimensions:

First, the software models driving the simulation are different: the LF simulator basically consists of a representation of the functional contingencies which can be observed in the sub-systems composing the diesel generator and is not affected by other sub-systems. The HF simulator is driven by an exhaustive software model in which all the sub-systems of the ship machinery interact dynamically and the sub-systems cannot be run in isolation. Objective fidelity for the HF simulator is clearly superior.

Second, the LF simulator is implemented on a PC computer with separate keyboard and mouse, scoring low on face validity. The HF simulator is implemented on a realistic remote control console placed in a separate room, and the local control panels of the different sub-systems which are located in a room contiguous to the remote control cubicle.

Because the LF simulator is clearly cheaper than the HF simulator with superior fidelity, the cost-effectiveness ratio of the LF simulator would prevail over that of the HF training device when similar or better training results can be obtained with the LF simulator.

As a second aim of the study, the LF simulator may be used to solve an actual training problem with respect to the poor transfer found by the instructors between the HMI in the remote control console (using a CRT monitor displaying direct manipulation interactive process diagrams) and the HMI interface in the local control panels (with push buttons and analogical displays). A specific training strategy will be adopted to solve this problem. Note that the HF simulator cannot be reconfigured to surmount this problem.

Applied research on training effectiveness and its relation to simulator fidelity has been mainly developed within the aircraft system domain. Within the ground system domain, the amount of research devoted to these issues is far more restricted. In both domains, fidelity is mainly related to the visual and motion systems which largely allow for motor skill training. Contrasting with this result, within the naval system domain, very restricted research has been found that addresses the effects of simulator fidelity and training effectiveness. Also, motor tasks are far less important than in the air and ground domains. As Hurlock and Montague (1982) stressed, in naval systems, procedural tasks such as sonar operation, target acquisition, weapon assignment, machinery control, and so on, are the most relevant (see also, Cowen, 1993, 1994). Thus, the visual and motion systems are not so important as they are in the aircraft and ground vehicle simulators.

One research example related to fidelity in the naval domain was the study carried out by McLeod, Poulton, du Ross, and Lewis (1980). They analysed the effects of ship motion on motor skills such as tracking, key pressing, and tracing. From these tasks, only the key pressing task was unaffected by the movements of the ship motion simulator. They concluded that motion should be implemented in the training device if this factor is critical for acquiring motor skills involved in ship mobility functions. However, when motion is not critical for the addressed training goals, such as procedural skills training, motion cueing should not be incorporated (McLeod, et al., 1980).

Other fidelity issues, and more interesting for the purpose of this experiment, relate to simulation models and HMIs. From a prevalent objective fidelity point of view, accurate simulation models (as compared to the system's behaviour using purely technical measures) are assumed to be crucial for training. For the consoles, it was found during interviews with experts and users that most of them advocated for the highest possible fidelity (which may be interpreted both as objective as well as face fidelity). Specifically, within the Combat Information Centre (CIC) simulator investigated, the instructors indicated that, apart from motion cueing, HF of the simulator interfaces and control room environment (consoles and other equipment) were highly valued, noticeable sonar, radar, weapon control. That is, the external appearance, their arrangement of controls and displays, their spatial distribution and so forth were factors considered to be very important for the training of pre-operational or inexperienced CIC trainees (Gonzalez, 1996a,b,c).

Although opinions of the experts must certainly be taken into account to acquire training devices, formal evaluations of the training and transfer effects of these devices could help to validate decisions and thus arrive at more profitable outcomes in terms of economical (value for money) and performance achievement of the trainees. In the case of procedural training in the naval domain, it appears that the present insight from a human factors and training perspective is quite distant from actual training practices. The results of this study can help to bridge the existing gaps between them.

This experiment tries to determine whether a HF (high objective fidelity for simulator software model and high face and objective fidelity for consoles) simulator is necessary to provide training on system management tasks (or diesel generator operations in particular), or alternatively, that a LF simulator can be used to produce the same training effects.

The HF and the LF simulators are able to simulate a number of malfunctions on the diesel generator. This feature allowed us to implement the acquisition and transfer tasks as differing in the number and type of malfunctions to be solved in both training phases. The malfunctions and which type were presented to the trainees during the experiment, were selected after discussions about their relevance and feasibility (i.e., independence of other sub-systems, and timely occurrence within the available time in each run) with the subject matter experts.

Much effort is dedicated to provide trainees with knowledge directed to understand the systems they will operate by means of classroom teaching before HF simulators are used to provide familiarisation with the system. Training on the skills necessary to perform the operative tasks is mainly provided on-the-job. The actual training is provided as follows: Firstly, trainees are provided with conceptual knowledge about naval engines in general; secondly, they are provided with familiarisation knowledge of the machinery aboard ship; and thirdly, with knowledge and simulator practice to operate the different sub-systems of the ship machinery system. The first step makes use of classroom teaching on the basic concepts of naval engines. The second is based on classroom teaching supported by overhead projections of the process diagrams representing each machinery sub-system. The instruction offered in this step is oriented to provide the trainees with the knowledge on the principles and procedures to operate the machinery system through the interactive software. The third step makes use of the interactive process diagrams first projected from the computer screen and explained by the instructor followed by practice on the work-stations running the interactive process software. In the next course, the trainees are presented with the representation of the actual machinery control systems aboard ship. These are represented by the local control panels of the machinery sub-systems and the remote control console (replica of a real one) which constitute the HF simulator in this experiment.

The instructors have found that the transfer of training from the system diagrams to the local control panels is difficult for the trainees. Trainees seem not to be able to transfer easily their previous training to the operation on the local control panels. This problem cannot be solved by means of the actual HF simulator. Therefore, a strategy has been devised which can help to solve this by means of the LF simulator (i.e., the NDGSTRT, described in section 4.1.). This strategy consists of coupling the actions and results of the trainees' inputs on the process diagrams to their corresponding actions and results on an interactive pictorial representation of the local control panels.

It should be noted that the fractionation of the task into sub-tasks which share the time to be performed (Gopher et al., 1989; Wightman and Lintern, 1985), as described above, was used as the main independent variable in Experiment 1. That is, the FT vs. PT training strategy factor was implemented according to this fractionation manipulation. Similarly to Experiments 1 and 2, the PT strategy was adopted to provide familiarisation on the HF and LF simulators taking part in this experiment. For the remainder of the experimental runs (i.e., acquisition and transfer phases), the FT strategy was adopted.

Considering the effects of practice, it is assumed that all trainees will improve their performance on the task as a result of practice. However, not only has the level of fidelity been manipulated, but also an instructional feature to address the problem of transfer from one HMI to another as mentioned above. In the LF simulator (but not on the HF simulator), it is possible to provide for an association of the buttons on the local control panels and the elements on the diagram on the remote control console. By doing so, the design confounds the effects of instruction and the effects of fidelity. This may be considered as natural because the instructional possibilities of the LF simulator are inherently better than those of the HF simulator. Besides, too few subjects were available to enable the addition of an additional condition without this instructional strategy. Assuming equal results on the basis of fidelity differences, but better results due to the additional instruction, the hypothesis to be tested is that

- the LF simulator produces better training effects (shorter practice time, fewer practice trials, fewer errors) than the HF simulator.

Two independent variables were contrasted in the present experiment. The level of fidelity of the simulators used to train inexperienced trainees on the operation of a diesel generator aboard ship, and the effect of practice upon task performance. The level of fidelity was treated as a between-subject factor. Two levels of fidelity formed the two experimental groups.

- HF simulator, represented by the machinery control room simulator
- LF simulator, represented by the NDGSTRT

The practice variable was a within-subjects factor implemented as:

- Number of practice run, or
- Practice block (3 blocks of 4 runs) number

Trainees' performance on the task was assessed during the experiment by means of the following dependent variables:

- Runs to criterion: The number of runs required by trainees to reach performance criterion (RCR). Performance criterion was reached when trainees performed 3 consecutive runs achieving the task's goals (i.e., the system is running without active alarms and no attempts to correct either active components or passive malfunctions)
- Errors: This is, the performance accuracy or error rate in each run (ERR), considering the steps to complete the procedures
- Blanks: The number of actions not performed within the available time (BLK) that should however be performed to achieve the task's goals
- Steps taken in each run: The number of actions executed in each run (STP)
- Performance time: Average time to complete the task in each run (PFT)
- Attempts: The number of false attempts made to correct malfunctions (ATT)

5.2 Method

5.2.1 Participants

18 university students took part in this experiment. The trainees were young adults (male = 15; female = 3) studying in their second year of the Nautical and Navigation Sciences degree. It is assumed that their previous knowledge of the system involved in this experiment is similar. All trainees were inexperienced

in the operation of ship machinery systems. They were required to participate in 4 experimental sessions lasting approximately 2 hours. Trainees in both groups participated in individual sessions. Trainees were randomly assigned to one of two groups, the HF and LF. The two groups do not differ significantly in background on computer experience, previous studies, mouse experience, and proficiency in English on their speciality (see Table 12 for an overview). Only mouse experience and proficiency in English reveal two possibly relevant trends in favour of the HF group.

TABLE 12

Trainee Sample: Mann-Whitney U Comparison of Mean Ranks between Experimental Groups on Previous Knowledge/Experience Variables (rank values range from 1-5 on all variables)

	Low Fidelity		High Fidelity		M-W *U* test	
Background Experience	*M*	*SD*	*M*	*SD*	*U*	*p*
Computer Experience	2.67	0.50	2.89	0.33	31.5	0.28
Mouse Experience	2.56	0.53	2.89	0.33	27.0	0.13
Knowledge of English	1.11	1.27	2.11	1.27	24.0	0.13
Previous Studies	1.22	0.44	1.44	0.53	31.5	0.33

5.2.2 Instruments and materials

The following sections describe the components of the HF simulator, the LF simulator and other informational material employed in this experiment.

The HF simulator equipment used to implement this experiment was the machinery control room simulator. From this simulator, only the following components were used in the experiment:

- The local control panel of the diesel generator which incorporates:
  - analogical revolutions per minute (RPM) indicator of the diesel engine,
  - analogical speed drop adjustable indicator,
  - speed drop decrease and increase push buttons,
  - diesel engine start/stop button,
  - diesel engine trip reset/indication push button,
  - local and remote operation mode selection push buttons,
  - diesel-oil filters 1 and 2 inlet valves open/close push buttons,
  - lubricant-oil filters 1 and 2 inlet valves open/close push buttons,
  - lubricant-oil electrical pump auto/manual operation mode selection push button, and
  - lubricant-oil electrical pump start/stop push button.
- The local control panel of the diesel generator auxiliary systems which incorporates:
  - exhaust temperature analogical indicator,
  - fresh water outlet temperature analogical indicator,
  - lubricant oil inlet pressure analogical indicator,
  - lubricant-oil sump low-level alarm light indicator,
  - fresh water low-level alarm light indicator,

  - sea water cooling inlet and outlet valves open/close push buttons,
  - diesel-oil shut off valve open/close push button,
  - lubricant-oil make up valve open/close push button,
  - lubricant-oil discharge valve open/close push button, and
  - fresh water make up valve open/close push button
- only the console based on a Unix workstation with its keyboard and trackball on the remote control console took part in this experiment. This displays the interactive process diagram of the diesel generator. The other components of the simulation facility were not directly involved.

The use of both the local control panels and remote control console was mandatory as malfunction solving is not possible on the local control panels, as it is not possible either in the real ship equipped with this kind of machinery control systems. Of course, when actual pieces of equipment must be repaired or changed, these operations are carried out on the actual equipment (e.g., repairing or changing a valve).

The LF simulator was implemented on a software tool or the NDGSTRT developed to simulate the same processes involved in the operation of the diesel generator simulated by the HF simulator. The functionality and the graphic appearance of this LF simulator was kept as similar as possible to the HF simulator to enable controlled comparisons between the training effects of both simulators. This software research tool was developed with MS Visual C ++.

The differences between the high and LF simulators are as follows: The LF basically consists of a representation of the functional contingencies which can be observed in the sub-systems composing the diesel generator and does not receive input from other sub-systems or is not affected by them (e.g., fresh water generation plant).

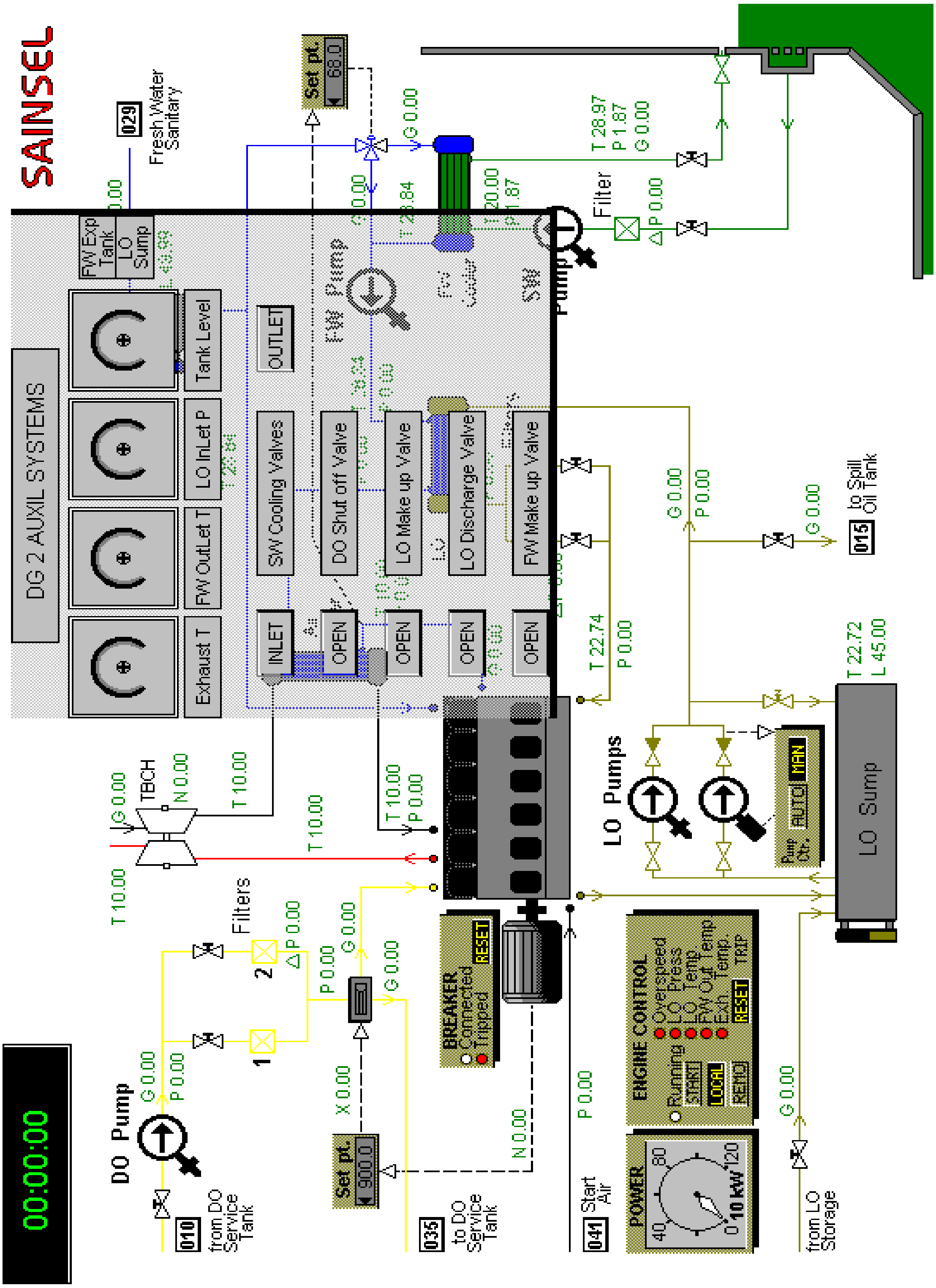

FIGURE 18 Example of the interface on the LF simulator. The overlaying grey rectangle represents the pop-up mimic panel of the DG 2 Auxiliary Systems panel. The remainder represents the interactive process control diagram.On the computer screen, the background appears as light grey (MS Windows standard).

These have been implemented as a set of IF-THEN rules capturing those contingencies. Its behaviour was validated by the subject matter experts. The HF is driven by an exhaustive software model in which all the sub-systems of the ship machinery interact dynamically and the sub-systems cannot be run in isolation. The controls and instruments of the LF simulator are (functionally correct, but not realistic) implemented on a PC computer with separate keyboard and mouse. The HF is implemented in a realistic remote control console placed in a room, and the local control panels of the different sub-systems which are located in a room contiguous to the remote control.

The NDGSTRT drove the experimental runs, provided knowledge of performance to the trainees after each run, and recorded the experimental raw data. The graphic interface is shown in Figure 18.

As the trainees did not possess any knowledge of the diesel generator system, information was provided before the practice on the tasks began. The instructions were specifically developed for the experiment. The materials provided include information to operate the HF and LF simulators as well as specific instructions to perform the task, i.e., the operation of the diesel generator. The experimental tasks and the instructions were similar to the ones used in Experiments 1 and 2. Nevertheless, the tasks were performed on different interfaces.

5.2.3 Experimental design

The design in this experiment was a quasi-transfer, composed of the experimental and control groups. The experimental group followed training on the LF simulator and was transferred to the HF simulator during a fixed period of time/runs on the selected tasks, while the control group was trained on the HF simulator and performed the selected transfer task for a fixed period of time/runs on the HF simulator.

Some previous experience could affect their performance on the experimental task. Knowledge of the English language, the technical English jargon used by the simulators, and their previous experience with computers was surveyed by means of a questionnaire at the beginning of the experimental sessions of each trainee. This information could eventually help to understand results not explained by the experimental manipulations.

TABLE 13
Experimental Design

	Acquisition Phase			Transfer Phase
Group	Runs 1-4	Runs 5-8	Runs 9-12	Runs 13-24
LF	PC process diagrams → Pop-up mimic local control panels	Pop-up mimic local control panels → PC process diagrams	Pop-up mimic local control panels and PC process diagrams	Local Control Panels and Remote Control Console
HF		Local Control Panels and Remote Control Console		Local Control Panels and Remote Control Console

Note. LF = Low-Fidelity group; HF = High-Fidelity group. The symbol → indicates that the effect of an action performed on one element on the PC process diagram was highlighted in its associated element in the pop-up mimic control panel, or, that the effect of one action performed on one element in the pop-up mimic control panel was highlighted in its associated element in the PC process diagram. The absence of the symbol → means that no corresponding effect was explicitly highlighted, such as in the runs 9-12 for the LF group.

5.2.4 Procedure

During the acquisition phase, nine trainees were trained on the HF simulator and the other nine on the LF simulator. Both groups were transferred to the HF simulator (see Table 13).

At the beginning of the first experimental session the 18 trainees were required to complete the questionnaire to record their previous knowledge (cf. section 5.2.1.) After completing the questionnaire, the experimenter provided the instructions concerning the system components and fluid colour codes, the knowledge to operate it, and functional knowledge of the behaviour of the system. This basic knowledge was followed by two practice runs guided by the experimenter on which the trainees became familiar with the system and its operation. These practice or familiarisation runs consisted of one run on which the trainees performed the start-up procedure, and another in which they performed the malfunction-solving procedure. Trainees were instructed to perform the tasks as fast and accurately as possible. The experimental tasks, i.e., start-up and malfunction-solving procedures differed from those used in Experiments 1 and 2 in that most of the control actions were performed on the control panels.

TABLE 14

Start-up Procedure for the Diesel Generator 2 in the HF and LF Simulators

Step	Panel	DG 2 Sart-up Procedure
CHK	Auxiliary systems	CHECK IF LO level in LO Sump is lower than 52 %
1		IF it is lower refill opening the LO Sump make up valve, IF NOT, continue in Step 3
CHK		CHECK WHEN the level in LO Sump reaches 52 %,
2		Close the LO Sump make up valve
CHK		CHECK IF FW level in Exp. Tank is lower than 52 %,
3		IF it is lower refill opening the Exp. Tank make up valve, IF NOT, continue in Step 5
CHK		CHECK WHEN the level in Exp. Tank reaches 52 %,
4		Close the Exp. Tank make up valve
5	Local	Open LO filter 1 valve
6		Start the LO electrical priming pump
CHK	Auxiliary systems	CHECK IF LO pressure after the pump increases, IF it does continue on Step 7, IF NOT go back to Step 6
7		Open SW suction valve
8		Open SW discharge valve
9	Local	Open DO filter 1 valve
10	Auxiliary systems	Open DO shut off valve to DO pump
11	Local	CHECK IF there is any TRIP indicator lit on the ENGINE CONTROL panel, IF THERE IS press RESET, IF THERE IS Not, continue in Step 14
12		Press LOCAL mode on the ENGINE CONTROL
13		Start the engine, pressing START on the ENGINE CONTROL
CHK		CHECK IF engine rpm or “N” increase, IF they do continue on Step14, IF NOT, go back to Step 13
14		Set the LO electrical priming pump on AUTO mode on the Pump Ctr. panel
15	Remote console	Set the FW temperature control Set point at 60°

The arrangements devised for the experimental group (LF) were intended to resemble the course of training as it is provided by the actual training programme (i.e., practice on the interactive diagrams followed by practice on the HF replica of a machinery control room). The experimental group commenced the acquisition phase performing 1/3 of the total number of runs (runs 1-4) on the NDGSTRT interactive process diagram. Their actions on the elements of the diagram produced a pop-up onset of the corresponding local panel on which the adequate push button was highlighted. These trainees then proceeded to the performance on the pictorial local control panels and its associated highlight of the corresponding diagram element on the runs 5-8 (4 out of 12 runs). They concluded the acquisition phase performing the remaining runs (runs 9-12) on the pictorial local panels without highlighting the associated diagram element. The start-up procedure and other actions of the malfunction solving procedure were executed on the local control panels. See Table 14 for the start-up and Table 15 for the malfunction solving procedures respectively. Trainees invoked exclusively the interactive diagrams to supervise the performance of the system and repair the malfunctions. Seven minutes were allowed for trainees to perform each run in the acquisition phase. In the transfer phase, these trainees performed the transfer tasks on 12 runs. The transfer task was the operation of the diesel generator under four malfunctioning conditions that were not experienced before. Eight minutes were allowed to perform each run in the transfer phase.

TABLE 15
Malfunction-solving Procedure for the Diesel Generator 2 in the HF and LF Simulators

Step	Panel	DG 2 Malfunction-solving Procedure
CHK	Remote console	Detect alarm
1		WHEN the alarm buzzer sounds, turn off the acoustic alarm by pressing the ALARM SILENCE key on the keyboard
2		WHEN there is any blinking optical alarm lit, e.g., H , acknowledge it by clicking with the left trackball button on it
CHK		To identify the malfunctioning device check and compare differential pressure increments, pressures and/or temperatures related to the alarm
3		IF the malfunctioning device is one in a pair of filters, open the alternative one, close the malfunctioning one, and continue in (2) in Step 4, IF the malfunctioning device is NOT any of this type of devices, continue in (1) in Step 4
4		(1) IF the engine is running stop it by pressing START on the ENGINE CONTROL panel, IF the engine is NOT running, (2) open the malfunctions list by pressing the MALFUNC. LIST key on the keyboard
5		Open the list of the DG 1 subsystem
6		Open the page "4000 Diesel Generator 1 - TBCH/DO/LO", IF it is needed to go to another page use the keys down ∨ or up ∧ to browse malfunction pages
7		Repair (i .e., reset) the malfunction/s which is/are thought to be provoking the alarm by clicking with the right trackball button on the variable number, e.g., on M5001 not on the descriptive text, e.g., DG 1 TURBOCHARGER DIRTY
CHK		Check if the malfunction/s was/were correctly repaired, IF it/they was/were, click EXIT to close the malfunctions list and continue in Step 8, IF NOT, go back to Step 4
8		Check IF there are any other active alarm/s, IF there are go back to the required Step, depending on the state of the alarm, If NOT, START the engine
CHK		CHECK IF engine rpm or "N" increase, IF they do continue on Step 9, IF NOT, go back to Step 11
9		IF necessary, Set the LO electrical priming pump on AUTO mode on the Pump Ctr. Panel

The control group performed the same tasks as the experimental group, but using the HF simulator and always using both the local control panels and the remote control console. No highlighted associations of the buttons on the local control panels and the elements on the diagram on the remote control console were provided.

After each run, all trainees received report of their performance. Note that the experimental group was given a performance report on the computer screen, whereas the control group obtained this report on a print-out. Trainees were then required to analyse the reports and evaluate their performance comparing it to the expert performance guide.

5.3 Results

For statistical reasons, the dependent variables were transformed in order to meet the requirements of parametric tests (i.e., normal distribution and homogeneity of variances). The variables runs to criterion (RCR), steps taken in each run (STP), errors (ERR), attempts (ATT), and Blanks (BLK) were square root transformed. Performance time (PFT) was logarithmically transformed. The test means of these variables were back transformed for graphical presentations. Differences are considered significant when the probability of rejecting the null hypothesis as being true (i.e., α) is less than 0.10.

5.3.1 Overall training effects as a function of fidelity

Analysis of trainees' performance as a function of the experimental conditions was addressed in two steps. First, a qualitative analysis of the percentage of LF and HF trainees reaching performance criterion within the available runs was approached. This qualitative analysis was further refined with a non-parametric statistical test (i.e., Mann-Whitney *U*). Second, a multivariate analysis of variance tested the overall effects of the training conditions on trainees' performance as assessed by the different dependent variables collected in this experiment. Similar but independent analyses were computed for the acquisition and transfer phases. Similarly to Experiments 1 and 2 training produced constant performance improment in both training groups LF and HF, this is depicted in Figure 19 for the acquisition and transfer phases.

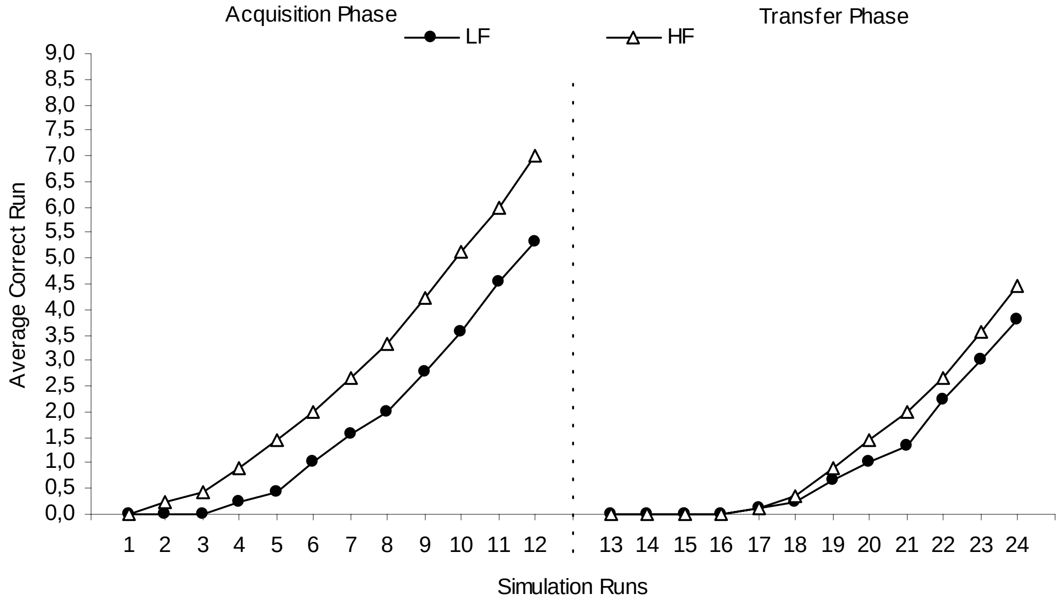

FIGURE 19 Acquisition and transfer phases: Cumulative average frequency of correct simulation runs as a function of LF vs. HF training group. LF = Low Fidelity group; HF = High Fidelity group.

5.3.1.1 Fidelity effects on performance criterion

In the qualitative analysis the percentage of trainees reaching performance criterion in a particular run is considered as a dependent variable. In the quantitative Mann-Whitney test the dependent variable used to compare LF against HF was the number of the run on which trainees achieved the criterion of performance (i.e., three consecutive runs without errors). These results are reported below.

Acquisition phase

After the acquisition phase, all nine HF trainees achieved performance criterion. Only seven (out of nine) LF trainees did so. Besides, all HF trainees reached performance criterion on the 11th run at the latest. Only six LF trainees reached performance criterion on the 11th run or before. These results are graphically presented in Figure 20.

The Mann-Whitney *U* test revealed a significant effect of fidelity at the 10% level, $U = 20.50$, $p = 0.08$. LF mean rank was 11.7 runs, while HF obtained a mean rank of 7.3 runs, indicating that on average, in the acquisition phase, HF trainees reached the criterion faster than LF.

Transfer phase

During the transfer phase, only 14 (out of 18) trainees achieved the criterion within the permitted 12 runs. Nevertheless, HF trainees were faster in achieving performance criterion than LF trainees, five out of nine HF trainees reached the criterion on the 23rd run or before, while only three LF trainees did so. Figure 20 illustrates these results.

In the transfer phase, the Mann-Whitney *U* test did not show a significant effect of fidelity, $U = 29.00$, $p = 0.34$. LF mean rank was 10.8 runs, while HF obtained a mean rank of 8.2 runs, indicating that on average, LF and HF did not differ substantially regarding how fast they achieved performance criterion.

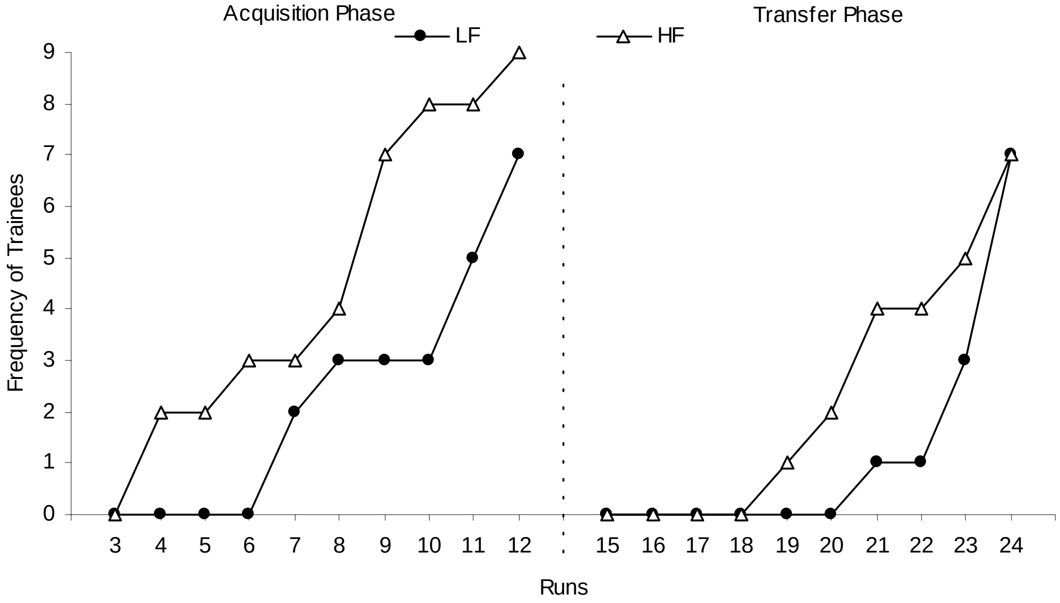

FIGURE 20 Acquisition and transfer phases: Cumulative frequency of trainees reaching performance criterion on a particular training run as a function of training group. LF = Low Fidelity group; HF = High Fidelity group.

5.3.1.2 Training effects of the interface on performance criterion

In this section, partial results of Experiment 2 and the present experiment are analysed in order to provide answers to the unsolved questions generated by Experiment 2. The group trained without augmented cues in the acquisition phase of Experiment 2 or NAC group was compared against the HF group of this experiment. These analyses should provide information about the relative importance of highly accurate software models of the behaviour of the system or models which represent this behaviour in a more simple way. Whether highly realistic interfaces are required or not will also be addressed in the discussion.

Acquisition phase

In the acquisition phase, the Mann-Whitney U test did not show a significant effect of fidelity in terms of the different interfaces, $U = 23.00$, $p = 0.24$. HF mean rank was 7.6 runs, while NAC obtained a mean rank of 10.6 runs, indicating that on average HF and NAC did not differ substantially on how fast they achieved performance criterion.

Transfer phase

In the transfer phase, HF and NAC both obtained an equal mean rank of 9.0 runs to reach the criterion of performance.

5.3.2 Overall fidelity effects on several performance measures

A multiple analysis of variance (MANOVA) was performed to assess the global effects of training as a result of the experimental manipulations (LF vs. HF). The different dependent variables collected throughout the training period were included in the MANOVA. During the acquisition phase, the HF group scored better on RCR, STP, ERR, and BLK (as shown in Figure 21 and Table 16). During the transfer phase, no significant differences were observed between LF and HF on any of the dependent variables (see Figure 22 for an overview and Table 16 for the test results)

TABLE 16
MANOVA Test Results for the Acquisition and Transfer Phases

	Acquisition Phase		Transfer Phase	
Dependent Variable	$F(1, 16)$	p	$F(1, 16)$	p
Frequency of Runs to Criterion (RCR)	4.70	< 0.05	1.31	0.27
Frequency of Steps per Run (STP)	8.13	< 0.01	2.32	0.15
Percentage of Errors per Run (ERR)	8.89	< 0.01	0.40	0.54
Frequency of Blank Steps (BLK)	4.49	< 0.05	0.26	0.62
Performance Time (PFT)	0.70	0.42	0.01	0.92
Frequency of False Attempts (ATT)	0.38	0.55	0.01	0.93

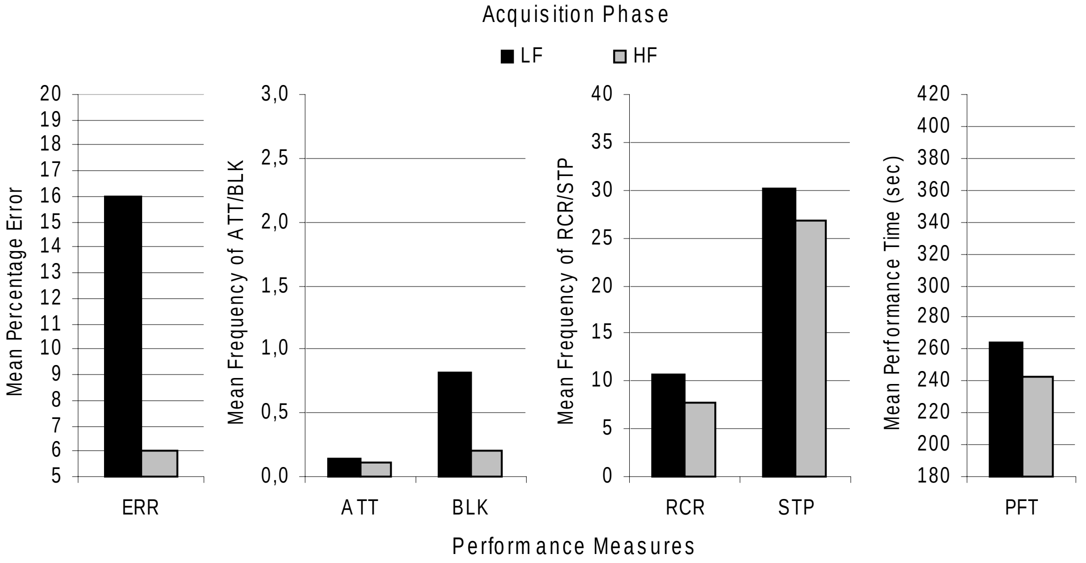

FIGURE 21 Mean acquisition performance on the dependent variables as a function of training group. LF = Low Fidelity group; HF = High Fidelity group; ERR = percentage of errors; ATT = false attempts to repair malfunctioning devices; BLK = Blank steps or not performed; RCR = runs required to reach performance criterion; STP = steps or actions performed in each simulation run; PFT = performance time spent on the task.

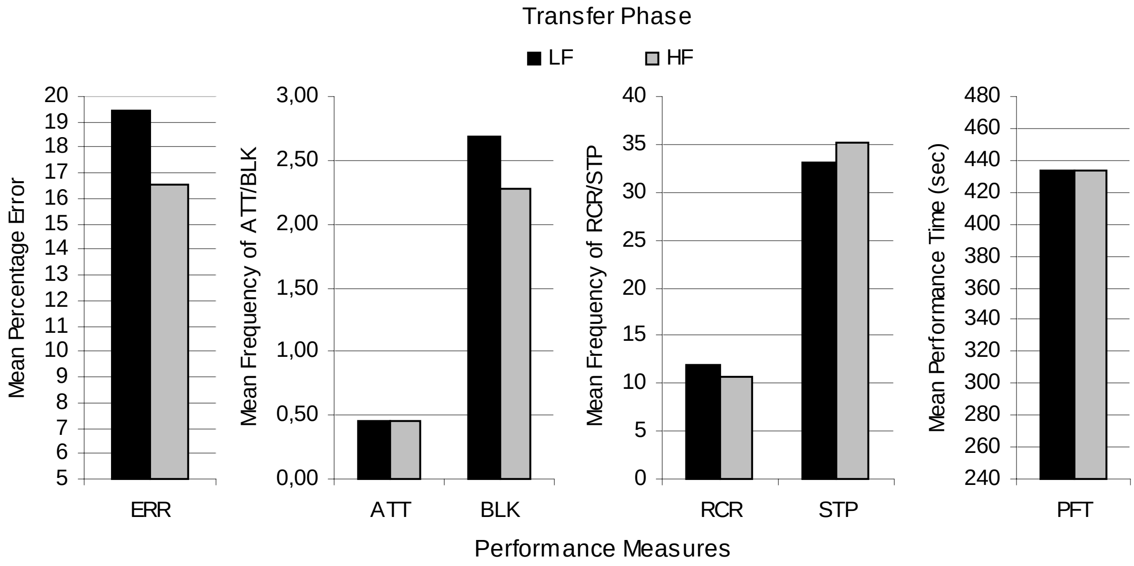

FIGURE 22 Mean transfer performance on the dependent variables as a function of training group. LF = Low Fidelity group; HF = High Fidelity group; ERR = percentage of errors; ATT = false attempts to repair malfunctioning devices; BLK = Blank steps or not performed; RCR = runs required to reach performance criterion; STP = steps or actions performed in each simulation run; PFT = performance time spent on the task.

5.3.3 Training progress as a function of fidelity

In the following sections, the analyses of the results include the effects of practice, as well as the fidelity factor, thus providing information on the training process as an addition to the global training results.

5.3.3.1 Fidelity effects on performance accuracy

A split-plot factorial analysis of variance was computed to analyse the effects of LF vs. HF on the percentage of errors trainees made in the acquisition and transfer phases. The analysis was a 2 (LF vs. HF groups) × 3 (blocks of 4 runs), with LF and HF being the between-subjects factor with two levels, and block, the within-subjects factor with three levels.

Acquisition phase

- Between-subjects effects or LF against HF:

As illustrated in Figure 23, the HF group consistently made fewer errors than the LF group throughout the acquisition phase ($F(1, 16) = 8.89$, $MSE = 1.21$, $p = 0.01$).

- Within-subjects practice effects:

The effect of practice revealed a significant effect, $F(2, 32) = 119.52$, $MSE = 84.60$ $p = 0.000$. Trainees in both the LF and HF groups constantly reduced their performance errors along the acquisition phase. The interaction group × block was not significant, $F(2, 32) = 2.04$, $MSE = 1.45$, $p = 0.15$. The error reduction rate across training blocks was almost similar in both LF and HF groups.

- Univariate tests:

Critical differences tests were computed for the block effect. Results indicated that mean differences were only significant for the first block as compared to the last, $F(1, 16) = 203.56$, $MSE = 0.83$ $p = 0.000$. The reduction of errors was not significant between the second and third blocks. The interaction group × block was significant on the second block of runs, $F(1, 16) = 4.68$, $MSE = 0.59$, $p = 0.05$, showing that the reduction of errors was sharper in LF than HF, but HF reduced errors faster than LF between the first and second blocks. It is likely that the changing conditions in LF during the acquisition phase delayed error reduction in this group as compared to HF which followed a persistent training condition.

Transfer phase

- Between-subjects effects or LF against HF:

As illustrated in Figure 23, both LF and HF groups constantly reduced errors throughout the transfer phase, without revealing a significant group effect, *F*(1, 16) = 0.40, *MSE* = 1.29, *p* = 0.54.

- Within-subjects practice effects:

The practice effect, similarly to the acquisition phase, revealed a significant effect, *F*(2, 32) = 104.77, *MSE* = 84.27 , *p* = 0.000. Error reduction was consistent in both LF and HF groups. The interaction group × block showed a significant effect, *F*(2, 32) = 2.58, *MSE* = 2.08 *p* = 0.09. The error reduction rate across training blocks was almost similar in both LF and HF groups until the second block. From the second to the third block, error reduction was sharper in LF than in HF.

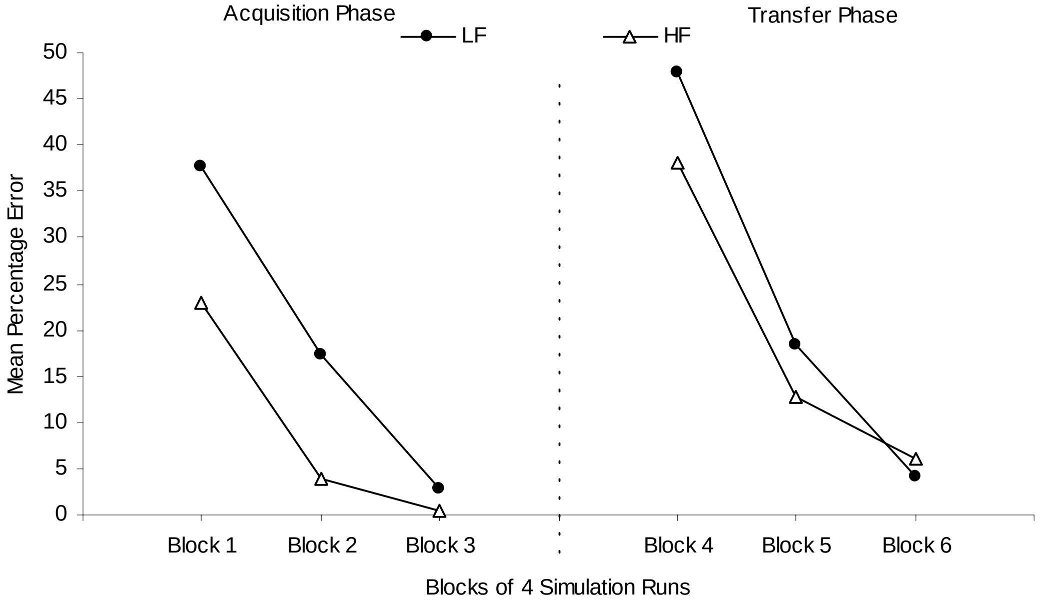

FIGURE 23 Acquisition and transfer phases: Mean percentage error as a function of LF and HF and training blocks of 4 runs. LF = Low Fidelity group; HF = High Fidelity group.

- Univariate tests:

Critical differences tests were computed for the block effect and the interaction group × block. Results indicated that mean differences were significant for the first block, *F*(2, 32) = 182.16, *MSE* = 0.91, *p* = 0.000, and significant in the second, *F*(2, 32) = 3.78, *MSE* = 0.70 *p* = 0.07 in comparison to the last block. The interaction group × block was significant in the first block of runs, *F*(1, 16) =

3.47, *MSE* = 0.91, *p* = 0.08, showing that the reduction of errors between the first and last block was sharper in LF than in HF. While LF maintained an almost constant reduction of errors, this slowed down in HF between the second and third blocks. The interaction, in the second block, had no significant effect.

5.3.3.2 Fidelity effects on blank steps

Similarly to the previous analysis on the effect of fidelity on performance accuracy, a split-plot factorial analysis of variance was computed to analyse the effects of LF vs. HF on the blank steps that trainees did not perform in the acquisition and transfer phases. The analysis was a 2 (LF vs. HF) × 3 (blocks of 4 runs), with LF and HF being the between-subjects factor with two levels and block the within-subjects factor with three levels.

Acquisition phase

- Between-subjects effects or LF against HF:

As illustrated in Figure 24, the analysis of variance showed a significant effect of fidelity, *F*(1, 16) = 4.49, *MSE* = 2.61, *p* = 0.05. Throughout the acquisition phase, HF avoided leaving actions unperformed (i.e., blank steps) to a greater extent than LF.

- Within-subjects practice effects:

The effect of practice was significant, *F*(2, 32) = 24.79, *MSE* = 6.52, *p* = 0.000. Trainees in both LF and HF groups increasingly avoided blank steps as the acquisition training phase progressed. The interaction group × block approached a significant level, *F*(2, 32) = 2.42, *MSE* = 0.64, *p* = 0.11.

- Univariate tests:

Critical differences tests were computed for the block effect. Results indicated that mean differences were only significant for the first block as compared to the last, *F*(1, 16) = 53.99, *MSE* = 0.24, *p* = 0.001. The reduction of errors was not significant between the second and third blocks. The interaction group × block was significant (10% level) on the second block of runs as compared to the third, *F*(1, 16) = 4.29,*MSE* = 0.29, *p* = 0.06, showing that the reduction of blank steps was faster in LF than HF, but HF reduced blank steps faster than LF between the first and second blocks.

Transfer Phase

- Between-subjects effects or LF against HF:

As illustrated in Figure 24, the analysis of variance did not reveal a significant effect of fidelity, $F(1, 16) = 0.26$, $MSE = 0.25$, $p = 0.62$. Both LF and HF groups reduced the number of blank steps in the transfer phase to the same extent.

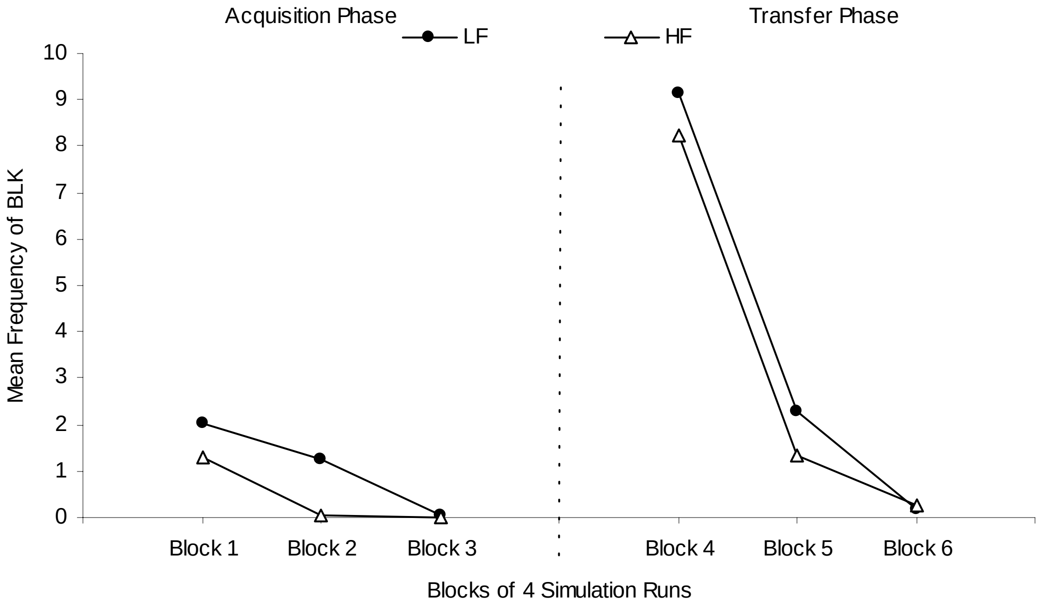

FIGURE 24. Acquisition and transfer phases: Mean frequency of non-performed steps (BLK) as a function of LF and HF and training blocks of 4 runs. LF = Low Fidelity group; HF = High Fidelity group.

- Within-subjects practice effects:

The practice effect, similarly to the acquisition phase, revealed a significant effect, $F(2, 32) = 156.69$, $MSE = 28.69$, $p = 0.000$. Reduction of BLK was continuous in both LF and HF groups. The interaction group × block did not reveal a significant effect. Both LF and HF groups reduced the number of BLK at a comparable rate as shown in Figure 24.

- Univariate tests:

Critical differences were computed for the block effect. Results indicated that mean differences were significant in the first and second blocks as compared to the last block, $F(1, 16) = 410.20$, $MSE = 0.14$, $p = 0.000$; and $F(1, 16) = 6.78$, $MSE = 0.23$, $p = 0.02$ respectively.

5.3.3.3 Fidelity effects on the number of performed steps

A split-plot factorial analysis of variance was computed to analyse the effects of LF vs. HF on the number of performed steps (STP) in the acquisition and transfer phases. The analysis was a 2 (LF vs. HF) × 3 (block of 4 runs), with LF and HF being the between-subjects factor with two levels and block the within-subjects factor with three levels. The dependent variable was the number of actions that trainees' performed in each run in order to achieve the task goals or STP.

Acquisition phase

- Between-subjects effects, LF vs. HF:

The analysis of variance showed a significant effect of fidelity, $F(1, 16) = 8.13$, $MSE = 1.30$, $p = 0.012$. HF required fewer steps to perform the task. Figure 25 illustrates this result.

- Within-subjects practice effects:

The practice effect did not show a significant effect, $F(2, 32) = 1.75$, $MSE = 0.09$, $p = 0.19$. The HF group required fewer steps than LF to perform the task across the acquisition phase. The interaction group × block revealed a significant effect, $F(2, 32) = 7.48$, $MSE = 0.37$, $p = 0.002$. While HF increased the number of steps from the first to the second block, LF reduced these in the second block. From the second to the third block, performance as evaluated by STP, was also different between LF and HF. STP remained almost constant in HF, while LF increased STP in the third block. These results are graphically presented in Figure 25.

- Univariate tests:

The interaction group × block was analysed for critical differences. The interaction was significant (at the 10% level) on the first block of runs, $F(1, 16) = 3.93$, $MSE = 0.07$, $p = 0.07$, showing that LF reduced STP from the first to the third block while HF slightly increased them. The interaction was also significant on the second block, $F(1, 16) = 14.38$, $MSE = 0.03$, $p = 0.002$, LF increased STP from the second to the third block, while in HF no substantial difference was appreciated from the second to the third block.

Transfer phase

- Between-subjects effects, LF vs. HF:

The analysis of variance did not reveal a significant effect of fidelity in the transfer phase, $F(1, 16) = 2.32$, $MSE = 0.49$, $p = 0.15$.

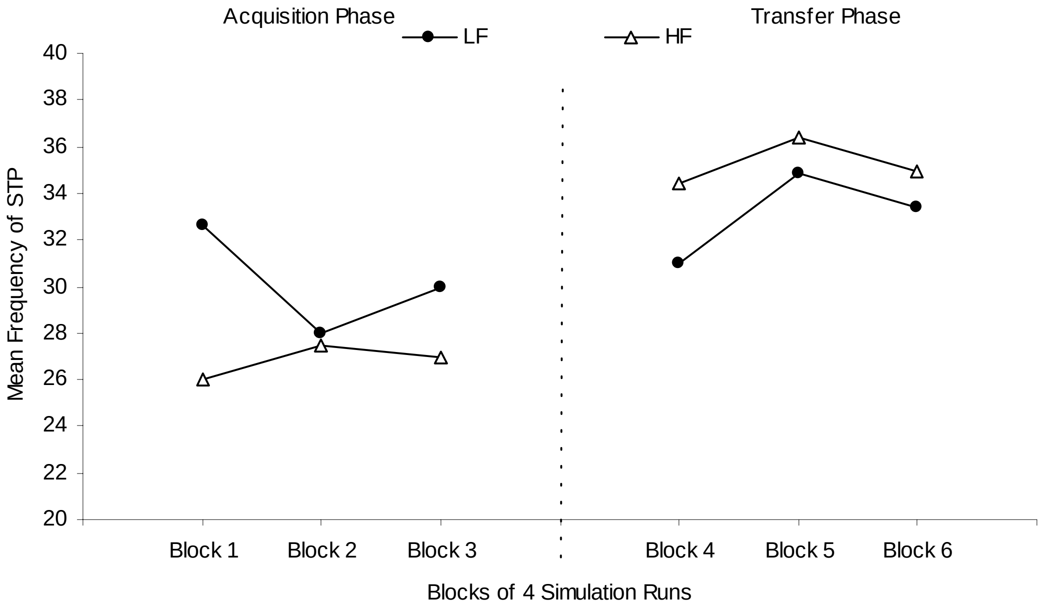

FIGURE 25 <Acquisition and transfer phases: Mean frequency of steps (STP) as a function of LF and HF and training blocks of 4 runs. LF = Low Fidelity group; HF = High Fidelity group.

- Within-subjects practice effects:

The practice effect was significant, *F*(2, 32) = 3.32, *MSE* = 0.28, *p* = 0.05. As Figure 25 illustrates, LF executed less STP to perform the task than HF. The interaction group × block did not show significant differences between LF and HF across training runs, *F*(2, 32) = 0.56, *MSE* = 0.05, *p* = 0.58.

- Univariate tests:

The practice effect was analysed for critical differences. There was only a significant difference between Block 4 and Block 5, *F*(1, 16) = 6.61, *MSE* = 0.06, *p* = 0.02. Both LF and HF reduced the number of steps required to achieve the task goals between the second and third blocks.

5.3.4 Transfer effects of the low vs. high-fidelity simulator configuration

The analysis of the transfer/interference effect of the LF vs. HF simulator configurations was performed in the same way as in Experiments 1 and 2. The percentage of errors variable was standardised into *z* scores (mean = 0; standard deviation = 1) for the LF, and HF groups, for the blocks of runs 2 and 3 of the acquisition phase, and the blocks of runs 5 and 6 of the transfer phase. The analysis (split-plot factorial ANOVA) was a 2 (LF vs. HF groups) × 2 (acquisition vs. transfer phases) × 2 (blocks 2-3 vs. blocks 5-6) with groups as the between-subjects factor, and acquisition/transfer phases and blocks of runs as within-subjects factors. The results showed significant differences on the effect of LF vs. HF group $F(1, 16) = 4.50$, $MSE = 2.81$, $p = 0.05$. The other main effects as well as the interactions did not reach statistical significance. Performance in the early stages (block 5 and 6) of the transfer phase was more negatively affected in the LF group than in the HF group. This result is consistent with the previous ANOVAs, see Figure 23.

5.4 Discussion

Considering the global effect, i.e., achievement of performance criterion by trainees, of the two simulator fidelity conditions studied in this experiment, no evidence has been found which supports the hypothesis of this experiment. In the acquisition phase, HF trainees showed better performance than LF. That is, HF trainees achieved performance criterion earlier than LF trainees. Not only was this result supported by the percentage of trainees in each group achieving performance criterion on a certain training run, but also by the analysis of the number of runs required by both groups until they reached the criterion. Nevertheless, when comparing the LF interface used in Experiment 2 with the HF used in the present experiment, no differences where found between them. The interface of the LF simulator employed in Experiment 2, i.e., non-augmented cues group or NAC, only utilised the interactive control diagrams. The HF simulator, employed in this experiment, mainly used the local control panels. Despite this difference, the advantage of HF over LF needs further consideration which will be dealt with in the following discussion of the whole set of results.

During and after the transfer phase, LF did not differ from HF on the percentage of trainees achieving the criterion in a certain run, nor in the number of runs needed by trainees to achieve the criterion. The comparison between the LF interface employed in Experiment 2 and the HF employed in the present experiment did not show differences in the transfer phase as it did not in the acquisition phase. The interface (HF vs. LF in the NAC group in Experiment 2) per se does not seem to differentially influence task performance.

Differences between LF and HF trainees in particular, rather than global performance measures, were only obvious in the acquisition phase. HF trainees made fewer errors, performed fewer steps, and left fewer unperformed actions than LF trainees. Both groups required similar amounts of time to complete runs, and made similar attempts to correct either passive malfunctions or active components.

The advantage of HF trainees in the acquisition phase was reduced in the transfer phase, during which no statistical differences were found between LF and HF in any of the performance measures collected in this experiment. However, the HF group still made fewer errors in the first two blocks of the transfer phase, probably indicating that the skill development process was better in HF than in LF. Alternatively, the consistency of the interface between the acquisition and transfer phases (i.e., HF) may have favoured skill transfer in the HF group.

Trainees in both groups improved their performance on the operation of the diesel generator as a function of practice. This indicates that training was effective independent of the experimental conditions addressing the fidelity of the training device. This is suggested by the ANOVA results concerning the interactions between fidelity (i.e., LF vs. HF) and the practice variable. The only difference between LF and HF was found in the reduction of errors in the transfer phase. This difference indicated that LF reduced errors faster than HF from the second to the third block as observed in Figure 23.

As stated above, the advantage of the HF group over the LF group in the acquisition phase needs further discussion. Two differences existed between LF and HF in the acquisition phase. On the one hand, LF followed a changing training sequence implemented in order to adjust the training progress to resemble the conditions of the actual training programme, which in turn - according to the subject matter experts- provokes interference when trainees are transferred to the HF simulator. HF instead, followed an invariable sequence. On the other hand, the LF group started training operating the diesel generator on the interactive diagram, while HF always performed most of the task components on the local control panels (i.e., only supervision of the system's behaviour and malfunction solving operations were exclusively performed on the interactive diagram).

Due to these differences, which were confounded with the fidelity variable in this experiment, it cannot be concluded that the disadvantage of LF in the acquisition phase was exclusively due to the LF of the training device. Reasons to explain this effect could be attributed to the interface (i.e., local control panels or interactive diagram) and/or the nature of the changing conditions in the LF simulator. In principle, the interactive diagram seems perceptually more complex than the local control panels. This complexity could have increased the task demands in comparison to the control panels (HF), and thus delaying skill acquisition in LF. Additionally, the changing training sequence probably required more attentional resources than the invariant sequence followed by HF trainees, thus, augmenting the task demands.

Trainees did not differ in their previous knowledge and experience as evaluated before the experimental runs. Despite this, HF tended to be different from LF in their previous knowledge of the English language (See Table 12). To test the possibility that this difference could be affecting the differences found between HF and LF in the acquisition phase, post-hoc analyses were computed on the dependent variables using trainees' knowledge of English as a covariate. Results indicated that knowledge of English was a good predictor of the number of false attempts made in the acquisition phase (p = 0.07). LF made more false attempts than HF (See Figure 21). This effect did not reach significance in the transfer phase. This might indicate that LF trainees learned the necessary English feedback messages to perform the tasks without being affected by their proficiency level in the transfer phase.

Therefore, the possible contribution of these factors (i.e., variability of practice, interface complexity, and proficiency in the English language) to an increment in task difficulty in the LF group could have interacted with the fidelity factor or could have been independent. The independence of fidelity and the factors increasing task demands was indirectly supported by the comparison between the NAC group in Experiment 2 and the HF group in the present one. No difference was found between them in the acquisition or transfer phases.

The transfer effects of an LF simulator as compared to those of an HF were small. Therefore, from a cost-effective point of view, the LF simulator is advisable though the skill development process appeared to be better when using the HF simulator.

Future research should address the training of ship machinery control tasks (procedural) on simulators differing in their degree of fidelity (low vs. high) which maintain the training schedule as constant. Also, different training schedules should be tested on simulators with similar/equal fidelity characteristics. With these experimental conditions, some answers could be provided to bridge the gaps remaining after this experiment. In any of these situations, the trainee sample size should be large enough to provide for robust results.

6 GENERAL DISCUSSION AND FUTURE SIMULATOR TRAINING

The following are conclusions extracted from the results reported above and should be understood in relation to the questions put forth in the experimental hypotheses tested in these experiments. Other research efforts and conclusions are also mentioned where applicable or related to the results presented earlier. Additionally, questions not addressed in this research, which can be of interest for future research and simulator training development and practice, are also proposed throughout the discussion.

The goal of training by means of simulators consists of the trainee performing at the best possible level on the actual job or task for which the training was developed. The training process provides the trainee with the opportunities to acquire new skills, retain them in memory, and transfer them to the actual job situation. The psychological processes of skill learning/acquisition, retention, and transfer, are usually disclosed for the purpose of research on the variables affecting them, and the effects they might have on the results of training. Nevertheless, the interrelationships between these processes are very tight, and should not be separated in the training system development process; rather, factors affecting acquisition, retention, and transfer processes should be integrated in the training system to achieve the best training result.

Simulators and training strategies

Retention performance measures evaluate what information has been learned during training, and the dynamics of that stored information between the time of original learning and the time when it is used, in the absence of practice. Transfer tasks not only evaluate retention but also the generalisation of the knowledge and skill acquired in one condition or task to another which has different characteristics. As skill retention is usually measured in the context of the same task, this does not ensure that skill generalisation will occur as a result of training. It is proposed here that as conditions and tasks usually change from the training situation to the real on-the-job situation, the result of training should be best evaluated through transfer measures. Specially, because the real situations are usually more complex and difficult than the training tasks. This is one of the reasons why in the experiments reported earlier, the transfer task was comparatively more complex than the acquisition task. For examples of increased complexity in transfer tasks see e.g., Doane, et al. (1999) and Speelman and Kirsner (2001), for different tasks within the same domain see e.g., Hinds, Patterson, and Pfeffer (2001).

Training provided through PT fractionation of the procedural diesel operation task resulted in more accurate and faster performance than training on the FT regime. This was statistically significant during the acquisition phase and qualitatively evident in the transfer phase. This result is opposed to the proposals of other authors for whom FT training is more advantageous than PT (Gopher et al., 1989; Speelman and Kirsner, 2001; Wightman and Lintern, 1985). Therefore, it is suggested that PT training be made available in simulators that address procedural training if these training regimes do not provoke interference with performance on the integrated tasks.

The CPT strategy employed in Experiment 1, albeit somewhat retarding skill development in the acquisition phase, did not hinder performance on the transfer task as compared to the PT and FT strategies. Other PT variations could be considered, but careful attention should be paid to the task to which those training strategies are applied. Availability of different PT strategies should be made available which are easy to implement on the simulator and easy to use by training developers and trainers. PT variations can be specified as training goals or sub-goals by subject matter trainers in co-operation with simulator developers and human factors experts. Within the training system development approach (see e.g., Goldstein, 1980, 1986; Patrick, 1991; Perez and Seidel, 1990; TRADOC, 1987) the results of these activities could contribute to training goals, training tasks, and training device specification.

Basically, the PT schedule used in Experiment 1 was a fractionation of the FT into its start-up and malfunction solving procedures. These were sequentially arranged in this order. This procedure did not hinder skill transfer to the FT in the PT group. Other groupings or organisations of PTs could be studied in order to develop more effective PT training schedules. Starting the acquisition phase with a critical component of the malfunction solving sub-task retarded the development of skills in the CPT group, though this effect dissipated with practice. Other training regimes, such as backward chaining (Wightman and Lintern, 1985), could also be arranged which could eliminate the adverse effects in the initial stages of training and improve skill development over the levels obtained with other strategies.

The augmented cueing strategy used in experiment 2 produced faster and higher performance achievement in both acquisition and transfer phases. It is thus proposed that simulators, which address the training of procedural tasks, incorporate augmented cueing or visual guidance techniques (Carlson, et al., 1992). This should focus not only on the procedures but also on the adequate performance strategies, which can be promoted by augmented cueing in the problem solving areas. Augmented cueing should be incorporated regardless of the degree of fidelity possessed by the simulator. Furthermore, in the absence of contradictory empirical support, the schedule used in this experiment to fade augmented cueing away could be used to prevent trainees from becoming dependent on the augmented cues.

The order in which the sequence of training tasks is organised should be carefully analysed when different interfaces exist in simulators and the actual systems. In this research, trying to solve the problem of low transfer from interactive control diagrams to control panels revealed possible implications for the scheduling of training. I followed the same order in which trainers arrange practice in their normal courses. In the acquisition phase, performance improvement of LF trainees was hindered by the training strategy provided to them. Perhaps, the order in which the two interfaces were presented was not adequate. Or perhaps, the visual connection made apparent to trainees distracted them, or imposed excessive processing demands that prevented them from taking advantage of this experimental preparation. Therefore, it is proposed that when different interfaces must be used in the course of training, the least demanding should precede the most demanding ones. Similar reasoning can be applied to training on the simulator followed by the actual system or on-the-job training. Another strategy, already mentioned, could be the backward chaining strategy proposed by Wightman and Lintern (1985).

Knowledge of performance (KP)

Knowledge of Perforamnce is not always consistently and contingently provided by instructors, is not very detailed, or is provided after long training runs on HF simulators thus reducing its effectiveness (e.g., Jaspers, 1980a, 1980b). KP was consistently provided throughout the training process and across training groups. It was also provided in the most detailed manner allowed by the simulator. Previous studies have demonstrated the beneficial effects of KP (Holding, 1987; Shlechter, Meliza, Burnside, and Bessemer, 1992). Some authors argue that providing the trainees with extensive information about the reasons for correct or incorrect performance results, has not proved more effective than providing minimal information (Schimmel, 1988). In our experiments, the effects of KP were confounded with the training strategies and practice effects. Thus, future research should evaluate its relative effect upon skill learning.

The relative value of KP, as indicated by trainees' comments, increased with practice. At the beginning of training, it was very difficult and bore little use for the trainees, but, as training progressed, its utility increased substantially. Hence, we propose that the KP provided by the simulator must be made easily comprehensible to the trainees from the very outset of training. Removing the provision of KP should also be taken into account to give trainees the opportunity to rely on their own evaluation of performance. The schedules for KP removal should be empirically validated and/or made flexible to adjust to the characteristics of the tasks and the trainees. It should also be made flexible so that different degrees of completeness can be adjusted to task demands and goals, as well as to trainees' preferences and abilities in line with Schimmel (1988) and Schmidt (1991).

In the experiments reported earlier, flexibility of KP was managed by the trainees themselves. As observed by the experimenters, as training progressed, trainees gradually spent less time evaluating their own performance. Two possible explanations can be offered for this behaviour: First, as trainees gained experience with the interpretation of their performance log, less time for self-evaluation was required. And second, as a result of accuracy improvement in task performance, fewer errors required correction and less memory adjustment should be made in preparation for the next simulation run (see, VanLehn, 1996). Still, other questions, not dealt with in our research, such as those proposed by Hesketh (1997) could be of interest for future research. For example, training without knowledge of results, not focusing on the learning process or how well the trainees are performing. In a sense, disrupting the automatisation of skills, and extending the period of analytic processing or knowledge based processing (Rasmussen, 1983), should enable transfer to occur (see e.g., Hesketh, 1997; Reeves and Weisberg, 1994).

I propose that simulators incorporate the means to provide KP in the most consistent and detailed way possible, regardless of their fidelity level. This could be a script of trainees’ performance presented on the simulator screen, printer format, or replay of performance with detailed explanations. For research and evaluation purposes, knowledge of performance should be recordable. In practice, this means, that the same KP information that is provided to the trainees throughout the practice session should be recorded, whether it is later used for particular purposes or not. When the simulator cannot automatically provide KP, e.g., the conditions were not considered during the specification and development phases, the trainer(s) should support this function. This implies that the handling of the recorded training session should be made easy and flexible to the trainers. Recorded sessions could contain flags, inserted by the trainers during the training session, marking particular situations or events for later debriefings with the trainees.

Provided that trainees are requested to evaluate their own performance on the task, this activity could be recorded on video by means of think aloud protocols and/or by means of the computer software. The question here, is to obtain as much and as detailed information as possible. Thus, further analyses of how the trainees evaluate their own performance could be carried out. The aim of this strategy is to capture those aspects of the information processing carried out by the trainees during these activities that cannot be registered otherwise. In addition, how the time spent on this activity is reduced with practice could also be evaluated. When planning for these aspects of the research, also the resources in terms of personnel, technical equipment, software capabilities and development, time requirements, and economical resources, should be taken into account. After all of these considerations are carefully evaluated, perhaps the practical implications of these analyses could likely justify the resources invested in them. Future research could take this issue into account.

Response-produced feedback

In the machinery control system used in these experiments, response-produced feedback, consisted of two types, black (i.e., inactive) to colour (i.e., active) change and vice-versa of the device symbols, and feedback on the results of malfunction repairing attempts. Throughout training, the experimenters observed that simulator response-produced feedback to trainees' attempts to repair malfunctioning devices was, sometimes, not understood, misinterpreted, or unattended. Response-produced feedback messages were displayed on the status bar, in the lower left corner of the screen, quite far from the central point of the screen. Predominantly, the messages indicating incorrect actions seemed meaningless to the trainees, at least, in the early stages of the training process. Increasing the salience of the feedback should optimise these issues. And, if the feedback is provided as a text message, length and meaning should be carefully designed to enable fast and easy interpretation. Optimising response-produced feedback should increase performance improvement in accuracy as well as in speed measures of performance.

As stated above, response-produced feedback seemed to be effective after a certain amount of practice. Studies on perceptual motor skills suggest that feedback is effective in facilitating skill acquisition. Thus, even in the absence of empirical support, we propose that whenever possible, response-produced feedback should be incorporated in simulators regardless of their fidelity. Lintern (1991b) argued that augmented feedback should be provided when the trainees are not performing correctly, but not when trainees are performing correctly. In the first case, feedback could direct trainees' attention to the correct informational invariants to guide correct task performance. In the second case, feedback could divert attention to artificial information (i.e., not present in the actual task context) or mask the natural informational invariants of the task that should guide correct performance. For the time being, as the response-produced feedback provided by the simulator is similar to the kind of feedback actual systems provide, we propose that it should probably be optimised in its informational value, but kept as an intrinsic part of the task.

Despite the fact that Lintern's suggestions were made in the context of manual tracking tasks, specifically, in relation to aircraft pilot training, Lintern's (1991b) ideas could be considered in other task domains. In our case, the performance of the procedures required more cognitive skills than motor. The behaviour of the machinery control room simulator did not reflect the actual time scale in which the actual system behaves. The events are driven by the mathematical software model of the system. This was done in the HF simulator by setting parameter values at lower or upper values than would happen in the actual system. In the LF simulator, the pace of the events, to a certain extent was determined by the operation of the trainees as no delays due to system response were programmed (i.e., there was no mathematical model driving the simulator). There was no possibility to test the effect of this condition but it should be interesting to evaluate the transfer from this to the actual system and assess whether the time scale affects the performance on the task and transfer to the actual system. Based on the transfer results from the LF to the HF simulator, our prediction is that it would not negatively affect performance accuracy. Most likely, the time used to perform the same tasks would be longer because the actual machinery aboard ship responds more slowly than the simulator. For example, the time for a high temperature alarm to develop in the simulator was almost immediate, in the actual system, this could take hours. The question here is: can we compress the time scale of training without compromising transfer, or is the time scale one of the informational invariants that should be simulated with HF?

In general, the information provided by the simulator contained many pieces of information which were not related to the tasks to be performed and did not have informative value, either for the trainees or for the trainers. We are referring to the names of software variables, which clearly reflect the software code but do not contribute in any way to the effectiveness of the simulator. On the contrary, they distract trainees in their search for meaningful cues. Of course, these can be very helpful for eventual debugging operations and system updates, as it allows straight communication between users and developers. Nevertheless, concerning the training purpose of this simulator system, these could be avoided at the user-interface level.

Trainees' prerequisites

In line with the informational value of the response-produced feedback provided by the simulator, an additional recommendation for future research should be made. The language used by the simulator (English) was not the mother tongue of the trainees. It is considered that knowledge of the language in which the training simulator presents its information to the trainees is a necessary condition for them to benefit from training. We did not find much evidence of impairment on trainees' performance due to this question. Nevertheless, in future research, effort should be made in order to ensure that the trainees understand fully the information provided by the simulator and errors due to this feature can be avoided (see, Frese and Altmann, 1989). This could be done by testing the trainees before the experimental sessions. Evaluating them after the language teaching session/s, and, if necessary, correct the possible remaining problems withfurther evaluation of their knowledge of the language after the experimental session/s. Another approach could be to select the experimental subjects according to their command of the language required. In this research, this could have resulted in the absence of a great number of subjects and thus it may have been impossible to carry out the experiments. Obviously, this did not restrict the execution of the research, but a lot of effort was required in order to overcome this contrariness.

Performance measures

Several measures of trainees' performance should be made available through the simulator. Overall scores like in the Space Fortress task (see, Mané and Donching, 1989) or the Kanfer-Ackerman Air Traffic Control (ATC) task (Ackerman and Kanfer, 1994) could be useful in providing assistance in refresher training for experienced operators. Performance measures provided in this fashion could greatly reduce instructor workload as well as objective assessment of training results. In detailed training programmes and for novice trainees, some other discrete measures should be incorporated. It is considered here, that overall performance measures would not be effectively diagnostic of the training needs not accomplished yet at a certain point in training. In our case, the use of different types of identifiable errors has proved valuable and discriminative of performance levels. The selection of performance measures for assessing training progress and overall results should take into account the type of tasks and the training goals at which simulator training aims.

In particular, I devised some measures, which were not directly available through the simulator. These were, firstly, the omission of necessary actions to perform the procedures or blank steps (BLK) as they have been termed in the experiments. Secondly, the number and type of parameter value checks or (CHK). The latter was conceived as a result of the implementation of the CPT or diagnostic task in Experiment 1. In this experiment, the trainees in the CPT group, on the first three runs, were required to mark on a sheet of paper representing the interactive control diagrams, which parameters they checked on the simulator screen in order to identify the malfunctioning device. It was obvious that this behaviour could not be registered by any means through the simulator. Even if we had had eye-movement tracking equipment available, the degree of inference in the judgement of whether the trainees where checking this or that parameter would have been extremely high. Another possibility could have been, as we have already argued, the use of video recording and concurrent think aloud protocols from the performance of the trainees. Thus, in Experiment 2, we implemented a relatively simple method to record and ensure that trainees checked the relevant parameters of the system in order to diagnose malfunctions. This was made by requesting the trainees to "click" on the parameter they were evaluating and registering this event as another step in their computer performance log or KP. By demanding this action from trainees in the AC group, it was assumed that the trainees searched for the parameter values they considered important, that they attended to them, and consequently, they extracted the information necessary for diagnosing and solving the malfunctioning piece of equipment.

Whether the trainees learned anything from these actions or not could be open to debate. However, at least, the minimum requirements for learning to take place were settled. In Experiment 2, we discussed the different malfunction solving strategies adopted by both groups AC and NAC. The AC group seemed to attempt a proper diagnostic behaviour before attempting to falsify their conclusions by trying to repair a certain device in the diesel generator. This information was not available from Experiment 1 or from the HF group in Experiment 3. Hence, information about the possible performance strategies adopted by the trainees in these experiments was not available. These questions should be taken into account in future simulator developments. Task analysis and cognitive task analysis, as well as information gathered through expert retrospective or concurrent verbal think aloud protocols could underpin the design of appropriate performance measures or training task specifications to obtain objective information about the strategic aspects of skill learning.

Despite the fact that many training programmes are evaluated by means of instructor or subject-matter experts' observations, additional objective measures should be provided by the simulators. For complex procedural tasks such as those investigated in this experiment, measures of accuracy or error which are meaningful and relevant to the tasks should be recorded. In our case, the number of checks performed, which is a measure of performance not available in the HF simulator used in Experiment 3, showed an important and meaningful value. This measure reflects the diagnostic behaviour displayed by the trainees. It showed which system parameters the trainees were monitoring in order to make a diagnosis of the device which is most likely malfunctioning. Performing the task in this way can help to avoid several undesirable errors. Some of these unsolicited errors are the attempts to repair devices which are actually not malfunctioning, unnecessary system shut downs, i.e., there are redundant pieces of equipment which can be repaired without shutting down the system, performing other actions which do not have a functional value or are unnecessary.

All these types of errors in trainees' performance not only detract efficacy from their task performance but also from the performance of the system, i.e., the system's goal might not be achieved due to those unnecessary actions on the part of the trainees which do not contribute to the normal functioning of the system. In real ship machinery operation, the efficiency of the system is of critical importance for economic and safety reasons. Other measures do not capture this aspect of trainees performance, i.e., monitoring the system state by visually inspecting the system parameters such as pressures, temperatures, levels, etc. Or otherwise, would need additional equipment in order to make them apparent, e.g., eye movement tracking and recording equipment. Therefore, I propose that measures which are relevant to the evaluation of task performance, though not very typical, are analysed carefully and incorporated to the simulator performance assessment functions. This issue should be taken into consideration by training developers, regardless of the degree of fidelity of the simulator and regardless of the technology used to provide training.

Memory representation

The task used by Benítez-Domínguez, in order to evaluate the degree of understanding of the system, was a valid measure of knowledge representation with the training methodology and for the purpose at which he aimed (Benítez-Domínguez, 1996). Nevertheless, this task was not considered valid for the purposes of Experiment 1. Hence, it is strongly suggested that pre-test/post-test tasks are devised in close relationship with the aims of the study. Too much abstraction of trainees' mnemonic representation of the tasks can be an oversimplification of the effects of training, provided that this representational issue was not directly addressed by the research hypotheses in any of the experiments. Additionally, the training provided by the simulator did not directly aim at providing the development of a particular memory representation or mental model of the task. The trainees throughout their training could have apprehended the functional dependencies between the different sub-systems. Nevertheless, it is understandable that trainees did not form a functional representation of the system in view of the task demands and the time constraints on performance.

Skill learning through practice on the tasks to which these skills apply has been demonstrated to follow a power function. The negatively accelerated reduction of performance time as practice progresses is proposed to result from chunking mechanisms (Newell and Rosenbloom, 1981). Some authors conceptualise these mechanisms as the progressive formation of procedures at higher hierarchical levels (e.g., Anderson, 1980; Singley and Anderson, 1989). Others consider these as the successive incorporation of units into a hierarchically organised schema (e.g., Adams, 1989; Gopher et al., 1989; Schmidt, 1975) or as mental models (e.g., Cannon-Bowers, Tannenbaum, Salas, and Converse, 1991; Glaser, 1990, Kieras, 1988; Kieras and Bovair 1984; Salas and Cannon-Bowers, 2001). Still, other authors view these mechanisms as the formation of retrieval structures composed of chunks (Ericsson, and Kintsch, 1995); or simply, as chunk structures in which the existing chunks progressively grow in size (Gobet, 2001; Miller, 1956; Saariluoma and Laine, 2001).

The similarities between these trends are larger than the discrepancies. Nevertheless, the key issues of these proposals, such as what is the content of the chunks represented in memory and how these chunk structures are organised, are not easily accessible by traditional experimental research (Gobet, 2001; Saariluoma and Laine, 2001). Hence, cognitive architectures, computational models, neural networks, and computer simulations, have been developed in order to examine the theoretical proposals about the nature and content of the structure of chunks in memory and which learning mechanisms explain better the development of skilled behaviour (see e.g., Gobet, 2001; Saariluoma and laine, 2001). Along this trend in cognitive research, and its relation with our area of study, is the concept of skilled memory posited by Chase and Ericsson (1982). Experts and novices differ at least in their memory representation (structure and content) of the system and the tasks to be performed. Perhaps, the kind of encoding strategies and the nature of the retrieval structures experts use could be elucidated by means of think aloud protocols and further validated through computer simulations as suggested by Saariluoma and Laine (2001). After analysing the data obtained in this fashion, training simulator tasks could be designed so that novice trainees could acquire faster (less than 10 years of deliberate practice, Ericsson and Lehmann, 1996) the kind of memory representation experts use to efficiently perform their daily tasks at work.

Analysis of expert performance should be carefully treated because it is not free from possible misconceptions. A likely mistake could be to consider the expert performance model as the most optimal in every single case (Hinds, Patterson, and Pfeffer, 2001). As Kieras (1988) indicated, the method or strategy the expert uses to perform the task is not necessarily the best model for the training programme (Hinds et al., 2001; Kieras, 1988). Even though expert performance is the most available source of information, the analyst should also perform a rational analysis of the task (Kieras, 1988).

Needless to say, extensive research should be carried out to test these hypotheses before this could be implemented in simulator training practices. In the meantime, we would like to pose two inter-linked questions: Can the skill acquisition process be accelerated more than it actually was through the experimental manipulations? Could this be done by providing structural information about the procedures at a higher hierarchical level, such as a graphical depiction (chunked), of the functional relationships of the different sub-systems, i.e., generator engine, lubrication sub-system, cooling sub-system, fuel sub-system, and turbocharger?

Simulator training advantages

Some of the advantages of simulator training were briefly indicated in the introduction of this research. Now, these will be extended to provide a more detailed view of the advantages and risks involved in simulator training.

First, training can be provided in tasks and conditions which are similar to the reality. System operators trained in this fashion are more likely to match their performance to that demanded by the system. This advantage is very well founded and does not provoke much controversy in the training field (see e.g., Salas and Cannon-Bowers, 2001; Stedmon and Stone, 2001; Stone, 2001). One risk associated to this advantage is the frequent one to one relationship which is established between the training simulator and the system to which the learned skills will be applied by the trainees. Usually, the training provided by one simulator is only expected to be transferred to one system. Therefore, if the operators have to face a similar but new system or a job or functional shift, those skills acquired in the previous simulator might not transfer adequately to the new conditions.

Associated to the previous advantage, and sometimes indistinguishable from this, training simulators can provide training in conditions and tasks which the trainees might never experience at work. The conditions referred to are emergencies, system’s failures, accidents, and so on, which in reality could never be trained for safety reasons. Training by means of simulators guarantees close similarity to real situations and safeguards personnel, system, environment and social damage. These benefits do not only relate to skill learning, but they also purport to economic savings. One can think about the training in emergency procedures which police, fire fighters, paramedical,

medical, and the lay people sometimes undergo in comparison to training simulators designed for the same purposes. The savings are immense. Comparative training effectiveness might not be possible or ethical to evaluate, if it should be comparative to real situations. The research community can easily tolerate the risks of not obtaining these comparative effectiveness measures.

The number of trainees that can be trained by means of simulators compared to actual systems is another of the advantages of this training technology. Many trainees can follow hands on practice in one or several simulators while not many trainees can do the same in the actual system. There may not be so many system units, or hands on practice may not be provided for operative/economical, ethical, or safety reasons.

Another benefit of simulator training is that practice can be provided throughout many sessions or for extended periods of time without incurring the costs of using the real system. Thus, if certain skills need to be acquired at the level of automaticity, i.e., requiring thousands of trials of practice, simulators are cost-effective where the actual systems are not. One possible risk with automating skills is that the trainee is shifted or promoted to a new system. Those automated skills might not be compatible with the operative procedures of the new system, the interface is incompatible, the interactive methods may have changed, i.e., new input devices such as voice, touch screen, optical pens, etc. Hence, trainers should pay attention to which skills or component skills are trained to the level of automaticity. Or, whether skill automaticity is a desirable training goal or not.

The content of training can be standardised in a manner that similar training scenarios and tasks can be provided to different trainees. Hence, training remains constant for a large trainee sample. This might not be an advantage per se. Perhaps, some would advocate for more individualised training to match the individual trainee characteristics and requirements. This is one of the objectives of intelligent tutoring systems, which are different training or instructional concepts. The advantage resides in the possibility to evaluate the results of training under similar conditions for different trainees. The results of this evaluation can then be used to improve the training strategies so that the effectiveness of simulator training can be increased. Additionally, these standardised scenarios can be used to explore and evaluate new working procedures or techniques, to compare expert professionals and novice trainees, to design refresher training, and to provide input to newer developments and technologies. Examples of these newer training technologies are virtual reality and advanced embedded training. More than a decade later (Caro, 1988; Glaser, 1990; Thomson and Spears, 1990), these training applications still seem to be more technology driven than research driven (see e.g., Andrews, et al., 1995; Stedmon and Stone, 2001).

Divergent from the previous benefit, the content of training can be designed so that the same trainees undergo many different versions of the training tasks. For instance, training in diagnosis or decision making skills can be designed so that many or all system failures can be experienced. In particular, certain symptoms, which can be provoked by combinations of failing devices, could be designed to allow for the widest variety of this kind of training. The absence of this characteristic can be a disadvantage of simulators, both HF and LF ones. For instance, due to the functional realism, which the simulator should display, and the development resources dedicated to this aspect, some of the training features may have been discarded from the development phase. Selection of certain requirements in favour of others should not affect the possibilities to implement any given training strategy if it is critical for the acquisition of skills. This applies to all training technologies including virtual reality and advanced embedded training.

Another advantage of training simulators is, or should be nowadays, the possibility to record the actions performed by the trainees while practising their training tasks. Maintaining a record of trainees' performance can help in debriefings, discussions between trainees, discussions between trainers in evaluations or in designing new training activities. These recordings can be shared between training centres, and so forth. We explicitly state that this should be possible nowadays because in the simulator used in this research it was not possible. However, the current state of the art in digital storage systems (i.e., HD, CD ROM, DVD) should allow this capacity.

Future training simulators

In the future, the question of which language the simulator interface uses could be solved by multi-language interface options. That is, enabling the simulator users to decide which language they prefer. Another possibility could be to use on-line translators such as those used over the Internet, and thus allowing the user to translate specific simulator information whenever he or she requires such help. Possibly, or preferably, the prospective customer should state this in the technical specifications if this kind of utility is considered necessary. The resources that such development or upgrading would require should also be taken into account as the price of the final simulators would definitely be affected. Nevertheless, within a user centred design framework, the availability of the technical characteristics concerning the simulator software should be discussed between the users and the provider of the simulator.

Software models differed between the LF and HF simulators. The differences between accurate software models (HF) against dynamic models based on contingencies (LF) were pointed out in Experiment 2. From the present experiment, no evidence has been found which could favour a highly accurate software model of the behaviour of the system. Neither from the transfer of LF to the HF simulator, nor from the contrast of the NAC group in Experiment 2 against the HF group in Experiment 3. Therefore, it is suggested that training goals are thoroughly analysed before deciding on any of both system behaviour models. Thus, the implementation of accurate software models of the behaviour of the system should be carefully considered in relation to the goals of training and the trainee population. Adopting this strategy could save training time and money.

If similar skill transfer can be achieved with an LF or HF simulator, the LF simulator could reduce the costs of the training programme (e.g., Goettl and Shute, 1996; Wightman and Lintern, 1985; Wightman and Sistrunk, 1987). Training costs are not only considered here in economic terms, but also in terms of the number of trainees per unit of time that a simulator can train. In the latter case, benefits and effectiveness issues should be carefully evaluated. Individual training can be better than collective training if the task under training must be performed by an individual operator rather than a team (see, Driskell and Salas, 1992).

However, simulator training should incorporate as much realism and functional fidelity as the training goals demand. Technology will keep affecting the development of training simulators in these aspects (e.g., Salas and Cannon-Bowers, 2001). Irrespective of the fidelity factor, simulators and other training devices should provide training designers with the possibility to implement as many training scenarios, training tasks, training sequences, training conditions, and training measures as the available technology permits. Training evaluation measures should be available in different formats, discrete, continuous, performance summaries, step by step action logs, etc. The training performance measures should be made adaptable or customisable so that trainers can choose the extent and format which better suits the training goals. The technology factor would not be the main constraint in attaining the training goals, rather, analysis, specification, and development resources will determine the final result of simulator training. That is, the human factors related to the activities associated to training simulator development will limit the end results more than the technological capacities of the equipment. From this perspective, and, as we have demonstrated in our research, training strategies and features that promote higher training results do not need to depend on technology but should be more efficiently supported by it.

The future challenge in simulator training could be stated as follows: We should describe where we are and propose where we want to go. Taking these into account, the means or resources to make the journey should be provided. Continuous evaluation of how far we are from the destination point, and why, should give us the opportunity to rest and obtain provisions, to correct the trajectory, or to change the destination point. This multidisciplinary journey has started long ago and will continue, probably, towards dynamically changing destinations and with increasing success.

YHTEENVETO

Simulaattoriharjoittelun tehokkuuteen vaikuttavat tekijät

Tutkimuksen tavoitteena oli, ensinnäkin hahmottaa keinoja parantaa simulaattoriharjoittelun tehokkuutta, ja toiseksi arvioida empiirisen tutkimuksen kautta parannusten tehokkuutta. Kolmanneksi, tarkoituksena oli käyttää kognitiivisen teorian ja simulaattoriharjoittelun käsitteitä pyrittäessä parantamaan taitojen omaksumista simulaattoriharjoittelun avulla. Ja viimeiseksi, tavoitteena oli tarjota tutkimukseen perustuvia suuntaviivoja harjoitussimulaattoreiden kehittämiselle tulevaisuudessa. Tutkimuksessa käytettiin kolmea koeasetelmaa selvittämään tekijöitä, jotka mahdollisesti vaikuttavat taitojen omaksumiseen simulaattoriharjoittelussa. Ensimmäisessä koeasetelmassa, oppilaat, joilla ei ollut aiempaa kokemusta asiasta, opettelivat laivaston dieselgeneraattorin käyttöä konehuonesimulaattorilla käyttäen osatehtävän, kriittisen osatehtävän ja kokonaisen tehtävän harjoittelumalleja. Toisena tehtävänä harjoittelijoiden piti oppia monimutkaisempi tehtävä samalla simulaattorilla. Tulokset osoittavat, että osatehtävä -harjoittelumalli oli tehokkain harjoittelun oppimis- sekä siirtovaikutuksissa. Samassa tutkimustilanteessa tehty toinen koe osoitti vahvistetun vihjeistyksen strategian olevan vastaavalla tavalla toteutettua vahvistamatonta strategiaa parempi. Lisäksi, tehokkuus nousi erityisesti harjoiteltaessa karkeammalla simulaattorilla käyttäen edellä kuvatun kaltaista strategiaa. Kolmannessa koeasetelmassa selvitettiin simulaattorin tarkkuuden yhteyttä harjoittelun tehokkuuteen. Vaikka merkittäviä eroja ei löytynyt tarkemman ja karkeamman simulaattorin välillä, alhaiset kustannukset puolsivat karkeamman simulaattorin käyttöä. Tutkimuksen tuloksia voidaan hyödyntää suunniteltaessa kustannuksiltaan edullisempia ja tehokkaampia harjoitussimulaattoreita tulevaisuudessa .

ACRONYMS

AC	Augmented Cueing
ACT*	Adaptive Control of Thought model
ANOVA	ANalysis Of VAriance
ATC	Air Traffic Control
ATT	false ATTempts to correct malfunctions in each run
BLK	number of actions not performed or left BLanK within the available time
CHK	parameter value CHecKs performed in each run
CIC	Combat Information Centre
CPT	Critical Part-Task
CRT	Cathode Ray Tube
CTER	Cumulative Transfer Effectiveness Ratio
DO	Diesel Oil
ERR	ERRors in each training run
FT	Full-Task
FW	Fresh Water
HF	High-Fidelity
HMI	Human-Machine Interface
HMS	Human-Machine Systems
KP	Knowledge of Performance
KR	Knowledge of Results
LAN	Local Area Network
LF	Low-Fidelity
LO	Lubricant Oil
MANOVA	Multivariate Analysis Of VAriance
MCRS	Machinery Control Room Simulator
MISC	Miscellaneous or other devices
NAC	Non-Augmented Cueing
NDGSTRT	Naval Diesel Generator Simulator Training and Research Tool
PC	Personal Computer
PFT	PerFormance Time or time to complete the task in each run
PT	Part-Task training
RCR	Runs needed to reach the performance CRiterion (3 consecutive correct runs)
RPM	Revolutions Per Minute
RT	Reaction Time
SW	Sea Water
TBCH	TurBoCHarger or turbo compressor
ToT	Transfer of Training

REFERENCES

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, *117*, 288-318.

Ackerman, P. L., and Kanfer, R. (1994). *Kanfer-Ackerman air traffic controller task© CD-ROM database, data collection program, and playback program*: Office of Naval Research, Cognitive Science Program.

Ackerman, P. L., and Kyllonen, P. C. (1991). Trainee characteristics. In J. E. Morrison (Ed.), *Training for performance: Principles of applied human learning* (pp. 193-229). Chichester: John Wiley & Sons.

Ackerman, P. H, Wickens, C. D., and Schneider, W. (1984). Deciding the existence of a time-sharing ability: A combined methodological and theoretical approach. Human Factors, 26,71-82.

Adams, J. A. (1971). A closed-loop theory of motor learning. *Journal of Motor Behavior*, *3*, 111-150

Adams, J. A. (1989). *Human factors engineering*. New York: Macmillan Publishing Company.

Alessi, S. M. (1988). Fidelity in the design of instructional simulations. *Journal of Computer-Based Instruction*, *15*, 40-47.

Alluisi, E. A. (1991). The development of technology for collective training: SIMNET, a case history. *Human Factors*, *33*, 343-362.

Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Lawrence Erlbaum.

Anderson, J. R. (1980). *Cognitive psychology and its implications*. San Francisco, CA: W. H. Freeman and Company.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*, 369-406.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review*, *94I*, 192-210.

Anderson, J. R. (1992). Automaticity and the ACT* theory. *American Journal of Psychology*, *105*, 165-180.

Anderson, J. R., and Fincham, J. M. (1994). Acquisition of procedural skills from examples. *Journal of Experimental Psychology: Learning Memory and Cognition*, *20*, 1322-1340.

Anderson, J. R. (Ed.) (1981). *Cognitive skills and their acquisition*. Hillsdale, NJ: Lawrence Erlbaum.

Andrews, D. H., Carroll, L. A., and Bell, H. H. (1995). The future of selective fidelity in training devices. *Educational Technology*, *35*, 32-36.

Annett, J. (1991). Skill acquisition. In J. E. Morrison (Ed.), *Training for performance: Principles of applied human learning* (pp. 13-51). Chichester: John Wiley & Sons.

Atkinson, R. C., and Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. Spence and J. Spence (Eds.); The psychology of learning and motivation (Vol. 2, pp. 89-195). New York: Academic Press.

Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.

Baddeley, A. D. (1997). Human memory: Theory and practice. Hove: Psychology Press Ltd.

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General*, *108*, 296-308.

Bailey, R. W. (1982). *Human performance engineering: A guide for system designers* Englewood Cliffs, NJ: Prentice Hall.

Bainbridge, L. (1993). Planning the training of a complex skill. *Le Travail Humain*, *56*, 211-232.

Bandura, A. (1977). *Social learning theory*. Oxford: Prentice Hall.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, *37*, 122-147.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Barnett, S., and Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*, 612-637.

Benítez-Domínguez, R. (1996). *Metodologías de entrenamiento en simuladores de cámara de máquinas* (Training methodologies in machinery control room simulators). Cádiz: Servicio de Publicaciones de la Universidad de Cádiz.

Branson, R. K., Rayner, G. T., Cox, J. L., Furman, J. P., King, F. J., and Hannum, W. J. (1975). *Interservice procedures for instructional systems development* (TRADOC PAM 350-30). Fort Monroe, VA: U. S. Army Training and Doctrine Command.

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., and Converse, S. A. (1991). Toward an integration of training theory and technique. Special Issue: Training theory, methods, and technology. *Human Factors*, *33*, 281-292.

Carlson, R. A., Sullivan, M. A., and Schneider, W. (1989). Component fluency in a problem-solving context. *Human Factors*, *31*, 489-502.

Carlson, R. A., Lundy, D. H., and Schneider, W. (1992). Strategy Guidance and memory aiding in learning a problem-solving skill. Human Factors, 34, 129-145.

Caro, P. W. (1988). Flight training and simulation. In E. L. Wiener, and D. C. Nagel (Eds.), *Human factors in aviation* (pp. 229-261). New York: Academic Press.

Chase, W. G., and Ericsson, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 16, pp. 1-58). New York: Academic Press.

Colley, A. M., and Beech, J. R. (1989b). Acquiring and performing cognitive skills. In A. M. Colley, and J. R. Beech (Eds.), *Acquisition and performance of cognitive skills* (pp. 1-16). New York: John Wiley & Sons.

Colley, A. M., and Beech, J. R. (Eds.) (1989a). *Acquisition and performance of cognitive skills.* New York: John Wiley & Sons.

Compton, B. J., and Logan, G. D. (1991). The transition from algorithm to memory retrieval in memory based theories of automaticity. *Memory and Cognition*, *19*, 151-158.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Science*, *24*, 87-185.

Cowen, M. B. (1993). Designing an instructional simulation for a program entry panel. *Simulation & Gaming*, *24*, 500-506.

Cowen, M. B. (1994). *Interactive hypermedia for training electronic surveillance personnel.* Paper presented at the February 1994 ADCIS Conference, Nashville.

Diehl, A. E., and Ryan, L. E. (1977). *Current simulator substitution practices in flight training.* Orlando, FL: US Navy, Training Analysis and Evaluation Group.

Doane, S. M., Sohn, Y. W., and Schreiber, B. (1999). The role of processing strategies in the acquisition and transfer of a cognitive skill. *Journal of Experimental Psychology: Human Percepction and Performance*, *25*, 1390-1410.

Dougherty, D. J., Houston, R. C., and Nicklas, D. R. (1957). *Transfer of training in flight procedures from selected ground training devices to the aircraft* (Human Engineering NAVTRADEVCEN TR-71-16-16). Port Washington, NY: US Naval Training Device Center.

Driskell, J. E., and Salas, E. (1992). Collective behavior and team performance. *Human Factors*, *34*, 277-288

Driskell, J. E., Willis, R. P., and Copper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology*, *77*, 615-622.

Drury, C. G., Paramore, B., Van Cott, H. P., Grey, S. M., and Corlett, E. N. (1987). Task analysis. In G. Salvendy (Ed.), *Handbook of Human Factors* (pp. 370-401). New York: John Wiley & Sons.

Eberts, R. E, and Brock, J. F. (1987). Computer assisted and computer-managed instruction. In G. Salvendy (Ed.), *Handbook of human Factors* (pp. 976-1011). New York: John Wiley & Sons.

Ericsson, K. A., and Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211-245.

Ericsson, K. A., and Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, *47*, 273-305.

Ericsson, K. A., Krampe, R. Th., and Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*, 363-406.

Eyrolle, H., and Cellier, J.-M. (2000). The effects of interruptions in work activity: Field and laboratory results. *Applied Ergonomics*, *31*, 537-543.

Fisk, A. D., Ackerman, P. L., and Schneider, W. (1987). Automatic and controlled processing theory and its application to human factors problems. In P. A. Hancock (Ed.), *Human factors psychology* (pp. 159-197). New York: North Holland.

Fisk, A. D., and Schneider, W. (1984). Memory as a function of attention, levels of processing, and automatization. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 1818-197.

Fitts, P.M. (1962). Factors in complex skill training. In R. Glaser (Ed.), *Training research and education*. Pittsburgh, PA: Robert Glaser.

Fitts, P. M. (1990). Factors in complex skill training. In M. Venturino (Ed.), *Selected readings in human factors* (pp. 275-295). Santa Monica, CA: Human Factors Society. First published in R. Glaser (Ed.) (1962). Training research and education. Pittsburgh, PA: Robert Glaser.

Fitts, P. M., and Posner, M. I. (1967). *Human performance.* Belmont, CA: Brooks Coleman.

Flexman, R. E., and Stark, E. A. (1987). Training simulators. In G. Salvendy (Ed.), *Handbook of human Factors* (pp. 1012-1038). New York: John Wiley & Sons.

Frederiksen, J. R. and White, B. Y. (1989). An approach to training based on principled task decomposition. *Acta Psychologica, 71*, 89-146.

Frese, M., and Altmann, A. (1989). The treatment of errors in learning and training. In L. Bainbridge, and A. Ruiz Quintanilla (Eds.), *Developing skills with information technology* (pp. 65-86). Chichester: John Wiley & Sons.

Gagnè, R. M. (1967). Instruction and the conditions of learning. In L. Siegel (Ed.), *Instruction: Some contemporary viewpoints.* San Francisco, CA : Chandler.

Gagnè, R. M., and Briggs, L. J. (1974). *Principles of instructional design.* New York: Holt, Rinehart and Winston.

Gander, T. J. (Ed.) (1992). JANE'S Military Training Systems. London: Butler and Tanner.

Glaser, R. (1990). The reemergence of learning theory within instructional research. *American Psychologist, 45*, 29-39.

Gobet, F. (2001). Chunk hierarchies and retrieval structures: Comments on Saariluoma and Laine. *Scandinavian Journal of Psychology, 42*, 149-155.

Goettl, B. P., and Shute, V. J. (1996). Analysis of part-task training using the backward-transfer technique. *Journal of Experimental Psychology: Applied, 2*, 227-249.

Goldstein, I. L. (1980). Training in work organizations. *Annual Review of Psychology, 31*, 229-272.

Goldstein, I. L. (1986). *Training in organizations: Needs assessment, development, and evaluation.* Monterey, CA: Brooks Coleman.

Gonzalez, N. (1996a). Physical System Parameters and Skill Acquisition - Field Orientation. EUCLID RTP 11.1, MASTER Deliverable C1.2. Prepared for the Directorate of Materiel of the Royal Netherlands Army under contract no. DMKL/EUCLID/RTP 11.1/016-92-7211.11.

Gonzalez, N. (1996b). Training and Sequencing Methods - Field Orientation. EUCLID RTP 11.1, MASTER Deliverable A2.2. Prepared for the Directorate of Materiel of the Royal Netherlands Army under contract no. DMKL/EUCLID/RTP 11.1/016-92-7211.11.

Gonzalez, N. (1996c). Transfer of Training - Field Orientation. EUCLID RTP 11.1, MASTER Deliverable B2.2. Prepared for the Directorate of Materiel of the Royal Netherlands Army under contract no. DMKL/EUCLID/RTP 11.1/016-92-7211.11.

Gonzalez, N., and Sanz, C. (1998). Transfer of training experiment - Naval systems. EUCLID RTP 11.1, MASTER Deliverable B2.3. Prepared for the Directorate of Materiel of the Royal Netherlands Army under contract no. DMKL/EUCLID/RTP 11.1/016-92-7211.11.

Gonzalez, N., Carro, J., and Prieto, G. (1994). Training systems, skill acquisition processes, and simulator training. OTRI Research Report, University of Salamanca, Spain.

Gonzalez, N., and Gonzalez del Campo, J. A. (1993). *Efectos de la interferencia contextual en tareas de búsqueda visual simple* (Contextual Interference Effects on Simple Visual Search Tasks), presented at the V Reunion de la Sociedad Española de Psicologia Comparada, Barcelona, September 22nd-24th.

Gopher, D., Weil, M., and Siegel, D. (1989). Practice under changing priorities: An approach to the training of complex skills. *Acta Psychologica*, *71*, 147-177.

Hagman, J. D., and Rose, A. M. (1983). Retention of military tasks: A review. *Human Factors*, *25*, 199-213.

Hays, R. T. (1980). *Simulator fidelity: A concept paper* (TR-490). Alexandria, VA: U. 2S. Army Research Institute for the Behavioral and Social Sciences. (AD-A101-474).

Hays, R. T., Jacobs, J. W., Prince, C., and Salas, E. (1992). Flight simulator training effectiveness: A meta-analysis. *Military Psychology*, *4*, 63-72.

Hesketh, B. (1997). Dilemmas in training for transfer and retention. *Applied Psychology: An International Review*, *46*, 317-339.

Hinds, P. J., Patterson, M., Pfeffer, J. (2001). Bothered by abstraction: The effect of expertise on kowledge transfer and subsequent novice performance. *Journal of Applied Psychology*, *86*, 1232-1243.

Holding, D. H. (1987). Concepts of training. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 939-962). New York: John Wiley & Sons.

Hurlock, R. E., and Montague, W. E. (1982). Skill retention and its implications for Navy tasks: An analytical review. (NPRDC-SR-82-21; 13-80-13) San Diego, CA: Navy Personnel Research and Development Center. (AD-A114-211/6).

Ivancic, K., and Hesketh, B. (2000). Learning from errors in a driving simulation: effects on driving skill and self confidence. *Ergonomics*, *43*, 1966-1984.

Jaspers, B. W. (1980a). *Simulator training of ship officers* (Report No.: 5915132-80-1). Delft, Netherlands: Instituut TNO voor Werktuigkundige Constructies. (PB84-164581).

Jaspers, B. W. (1980b). *Simulator training of ship engine room officers* (Report No.: 5915132-80-2). Delft, Netherlands: Instituut TNO voor Werktuigkundige Constructies. (PB84-164656).

Johnson-Laird, P. (1983). *Mental models*. Cambridge, MA: Harvard University Press.

Kalakoski, V., and Saariluoma, P. (2001). Taxi drivers' exceptional memory of street names. *Memory & Cognition*, *29*, 634-638.

Kieras, D. E. (1988). What mental model should be taught: Choosing instructional content for complex engineering systems. In J. Psotka, L. D. Massey, and S. A. Mutter (Eds.), *Intelligent tutoring systems: Lessons learned* (pp. 85-111). Hillsdale, NJ: Lawrence Erlbaum.

Kieras, D. E., and Bovair, S. (1984). The role of a mental model in learning to operate a device. *Cognitive Science*, *8*, 255-273.

Lane, N. E. (1987). *Skill acquisition rates and patterns: Issues and training implications*. New York: Springer Verlag.

Langley, P., and Simon, H. (1981). The central role of learning in cognition. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 361-380). Hillsdale, NJ: Lawrence Erlbaum.

Lassaline, M. L., and Logan, G. D. (1993). Memory-based automaticity in the discrimination of visual numerosity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 561-581.

Lee, F. J., and Anderson, J. R. (2001). Does learning a complex task have to be complex?: A study in learning decomposition. *Cognitive Psychology*, *42*, 263-316.

Lee, T. D., and Magill, R. A. (1983). The locus of contextual interference effect in motor-skill acquisition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *9*, 730-746.

Leplat, J. (1989). Cognitive skills at work. In L. Bainbridge, and A. Ruiz Quintanilla (Eds.), *Developing skills with information technology* (pp. 35-63). Chichester: John Wiley & Sons.

Lewis, C. H. (1979). Speed and practice. Unpublished manuscript.

Lintern, G. (1980). Transfer of landing skill after training with supplementary visual cues. *Human Factors*, *22*, 81-88.

Lintern, G. (1991a). Instructional strategies. In J. E. Morrison (Ed.), *Training for performance: Principles of applied human learning* (pp. 167-191). Chichester: John Wiley & Sons.

Lintern, G. (1991b). An informational perspective on skill transfer in human-machine systems. *Human Factors*, *33*, 251-266.

Lintern, G., Thomley-Yates, K. E., Nelson, B. E., and Roscoe, S. N. (1987). Content, variety, and augmentation of simulated visual scenes for teaching air-to-ground attack. *Human Factors*, *29*, 45-59.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492-527.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning Memory and Cognition*, *18*, 883-914.

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, *109*, 376-400.

Logan, G. D., and Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review*, *108*, 393-434.

Logan, G. D., and Stadler, M. A. (1991). Mechanisms of performance improvement in consistent mapping memory search: Automaticity or strategy shift? *Journal of Experimental Psychology: Learning Memory and Cognition*, *17*, 478-496.

Magill, R. A., and Hall, K. G. (1990). A review of the contextual interference effect in motor skill acquisition. *Human Movement Science*, *9*, 241-289.

Mané, A., and Donchin E. (1989). The space fortress game. *Acta Psychologica*, *71*, 17-22.

Masson, M. E. J. (1990). Cognitive theories of skill acquisition. *Human Movement Science*, *9*, 221-239.

McLeod, P. D., Poulton, E. C., du Ross, H., and Lewis, W. N. (1980). *The influence of ship motion on manual control skills* (DRIC-BR-77015). London, England: Royal Naval Personnel Research Committee. (AD-A095-853/8).

Miller, D. P., and Swain, A. D. (1987). Human error and human reliability. In G. Salvendy (Ed.), *Handbook of Human Factors* (pp. 219-250). New York: John Wiley & Sons.

Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81-97.

Montague, W. E., and Knirk, F. G. (1993). Findings matrix: What works in adult training. *International Journal of Educational Research*, *19*, 345-430.

Morrison, J. E. (Ed.) (1991). *Training for performance: Principles of applied human learning.* Chichester: John Wiley & Sons.

Newell, A., and Rosenbloom, P. S. (1981). Mechanisms of skill acquisition an the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale, NJ: Lawrence Erlbaum.

Newell, K. M., Carlton, M. J., Fisher, A. T., and Rutter, B. G. (1989). Whole-part training strategies for learning the response dynamics of microprocessor driven simulators. *Acta Psychologica*, *71*, 197-216.

Neves, D. M., and Anderson, J. R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 57-84). Hillsdale, NJ: Lawrence Erlbaum.

Norros, L. (1989). Simulation in industrial work training. In L. Bainbridge, and A. Ruiz Quintanilla (Eds.), *Developing skills with information technology* (pp. 315-327). Chichester: John Wiley & Sons.

O'Hara, J. M. (1990). The retention of skills acquired through simulator-based training. *Ergonomics*, *33*, 1143-1153.

Orlansky, J., and String, J. (1977). *Cost-effectiveness of flight simulators for military training: 1. Use and effectiveness of flight simulators* (IDA Paper P-1275). Arlington, VA: Institute for Defense Analyses.

Osgood, C. E. (1949). The similarity paradox in human learning: A resolution. *Psychological Review*, *56*, 132-143.

Patrick, J. (1980). Job analysis, training and transferability: Some theoretical and practical issues. In K. D. Duncan, M. M. Gruneberg, and D. Wallis (Eds.), *Changes in working life.* Chichester: John Wiley & Sons.

Patrick, J. (1991). Types of analysis for training. In J. E. Morrison (Ed.), *Training for performance: Principles of applied human learning* (pp. 127-166). Chichester: John Wiley & Sons.

Perez, R. S., and Seidel, R. J. (1990). Using artificial intelligence in education: Computer-based tools for instructional development. *Educational Technology, March*, 51-58.

Prophet, W. W., and Boyd, H. A. (1970). *Device-task fidelity and transfer of training: Aircraft cockpit procedures training* (HumRRO TR-70-10). Alexandria, VA: Human Resources Research Organization.

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man and Cybernetics, 13*, 257-266.

Rasmussen, J. (1993). Analysis of tasks, activities and work in the field and in laboratories. *Travail Humain, 56*, 133-155.

Reason, J. (1990). *Human error*. New York: Cambridge University Press.

Reeves, L. M., and Weisberg, R. W. (1994). The role of content and abstract information in analogical transfer. *Psychological Bulletin, 115*, 381-400.

Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shift and the automatization of cognitive skills. *Journal of Experimental Psychology: General, 126*, 288-311.

Robins, A. (1996). Transfer in cognition. *Connection Science: Journal of Neural Computing, Artificial Intelligence, and Cognitive Research, 8*, 185-203.

Roessingh, J. J. M. (2002). *The acquisition of complex skills*. Doctoral thesis, University of Utrecht, The Netherlands.

Roscoe, S. N. (1971). Incremental transfer effectiveness. *Human Factors, 13*, 561-567.

Rosenbaum, D. A., Carlson, R. A., and Gilmore, R. O. (2001). Acquisition of intellectual and perceptual-motor skills. *Annual Review of Psychology, 52*, 453-470.

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

Saariluoma, P. (2001). Chess and content-oriented psychology of thinking. *Psicológica, 22*, 143-164.

Saariluoma, P., and Kalakoski, V. (1997). Skilled imagery and long-term working memory. *American Journal of Psychology, 110*, 177-201.

Saariluoma, P., and Laine, T. (2001). Novice construction of chess memory. *Scandinavian Journal of Psychology, 42*, 137-146.

Salas, E., and Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology, 52*, 471-499.

Salvendy, G. (Ed.) (1987). *Handbook of human factors.* New York: John Wiley & Sons.

Schimmel, B. J. (1988). Providing meaningful feedback in courseware. In, D. H. Jonassen (Ed.), *Instructional designs for microcomputer courseware* (pp. 183-195). Hillsdale, NJ: Lawrence Erlbaum.

Schmidt, R. A. (1975). A schema theory of of discrete motor learning. Psychological Review, 82, 225-260.

Schmidt, R. A. (1991). Frequent augmented feedback can degrade learning: Evidence and interpretations. In G. E. Stelmach, and J. Requin (Eds.), *Tutorials in motor neuroscience* (Vol. 62, pp. 59-75). New York: Kluwer Academic/Plenum Publishers

Schmidt, R. A., and Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, *3*, 207-217.

Schneider, W. (1985). Training high-performance skills: Fallacies and guidelines. *Human Factors*, *27*, 285-300.

Schneider, W., and Detweiler, M. (1987). A connectionist/control architecture for working memory. In Bower, G. H. (Ed.), *The Psychology of Learning and Motivation* (Vol. 21, pp. 53-119). New York: Academic Press.

Schneider, W., and Fisk, A. D. (1983). Attention theory and mechanisms for skilled performance. In R. A. Magill (Ed.), *Memory and control of action* (pp. 119-143). New York: North-Holland.

Schneider, W., and Shiffrin, R. M. (1977). Controlled and automatic information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1-66.

Semple, C. A., Cotton, J. C., and Sullivan, D. J. (1981). *Aircrew training devices: Instructional support features* (AFHRL-TR-80-58). Brooks Air Force Base, TX: Airforce Human Resources Laboratory.

Semple, C. A., Hennessy, R. T., Sanders, M. S., Cross, B. K., Beith, B. J., and McCauley, M. E. (1981). *Aircrew training devices: Fidelity features* (AFHRL-TR-80-36). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Shebilske, W. L., Goettl, B. P., Corrington, K., and Day, E. A. (1999). Interlesson spacing and task-related processing during complex skill acquisition. *Journal of Experimental Psychology: Applied*, *5*, 413-437.

Shebilske, W. L., Regian, J. W., Arthur, W. J., and Jordan, J. A. (1992). A dyadic protocol for training complex skills. *Human Factors*, *34*, 369-374.

Shendel, J. D., and Hagman, J. D. (1991). Long term retention of motor skills. In J. E. Morrison (Ed.), *Training for performance: Principles of applied human learning* (pp. 53-92). Chichester: John Wiley & Sons.

Shiffrin, R. M., and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127-190.

Shlechter, T. M., Meliza, L. L., Burnside, B. L., and Bessemer, D. W. (1992). Developing automated feedback materials for a training simulator: An interaction between users and researchers. Paper presented at the Annual Conference of the American Educational Research Association, San Francisco, CA (ERIC Document ED-349-952) (IR-015-688).

Singley, M. K., and Anderson, J. R. (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard University Press.

Snoddy, G. S. (1926). Learning and stability. *Journal of Applied Psychology*, *10*, 1-36.

Speelman, C. P., and Kirsner, K. (2001). Predicting transfer from training performance. *Acta Psychologica*, *108*, 247-281.

Sticha, P. J., Singer, M. J., Blacksten, H. R., Morrison, J. E., and Cross, K. D. (1990). *Research and methods for simulation design: State of the art* (HumRRO-FR-PRD-88-27). Alexandria, VA: Human Resources Research Organization. (AD-A230-076).

Stedmon, A. W., and Stone, R. J. (2001). Re-viewing reality: Human factors of synthetic training environments. *International Journal of Human-Computer Studies*, *55*, 675-698.

Stone, R. J. (2001). Virtual reality for interactive training: An industrial practitioner's viewpoint. *International Journal of Human-Computer Studies*, *55*, 699-711.

Thomson, D. R., and Spears, R. E. (1990). A review of embedded training: Status and emerging role in training. In *Proceedings of Selected Paper Presentations at the Convention of the Association for Educational Communications and Technology* (ERIC Document ED-323-952). (IR-014-575).

Thorndike, E. L. (1903). *Educational Psychology.* New York: Lemke and Buechner.

Thorndike, E. L. (1931). *Human learning.* New York: Century.

TRADOC Regulation 350-7 (1987). *Systems approach to training (SAT).* Fort Monroe, VA: U. S. Army Training and Doctrine Command.

Tsang, P. S., Velazquez, V. L., and Vidulich, M. A. (1996). Viability of resource theories in explaining time-sharing performance. *Acta Psychologica*, *91*, 175-206.

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology*, *47*, 513-539.

Van Matre, N. H., Pennypacker, H. S., Hartman, W. H., Brett, B. E., and Ward, L. Q. (1981). *Computer-managed instruction in the Navy: V. The effects of Chartered feedback on rate of progress through a CMI course* (NPRDC-TR-82-16). San Diego, CA: Navy Personnel Research and Development Center. (ERIC Document ED-213-388).

Van Matre, N., Ellis, J., Montague, W., and Wulfeck, W. (1993). Computer-managed instruction in Naval technical training. *Instructional Science*, *21*, 295-311.

Welford, A. T. (1976). *Skilled performance: Perceptual and motor skills*. Glenview, IL: Scott Foresman.

Wickens, C. D. (1992). *Engineering psychology and human performance, second edition.* New York: HarperCollins Publishers.

Wightman, D. C., and Lintern, G. (1985). Part-task training for tracking and manual control. *Human Factors*, *27*, 267-283.

Wightman, D. C., and Sistrunk, F. (1987). Part-task training strategies in simulated carrier landing final-approach training. Human Factors, 29, 245-254.

Wikman, A.-S., Nieminen, T., and Summala, H. (1998). Driving experience and time-sharing during in-car tasks on roads of different width. *Ergonomics, 41*, 358-372.

Williams, A. C., and Flexman, R. E. (1949). *Evaluation of the School Link as an aid in primary flight instruction.* Urbana, IL: Institute of Aviation, University of Illinois, Aeronautics Bulletin 5.

Wulf, G., McNevin, N., and Shea, C. H. (2001). The automaticity of complex motor skill learning as a function of attentional focus. *The Quarterly Journal of Experimental Psychology*, *54 A*, 1143-1154.