Defence Research and Development Canada   Recherche et développement pour la défense Canada

DEFENCE R&D DÉFENSE

Validation of virtual environments incorporating virtual operators for procedural learning

Brad Cain, Lochlan Magee
DRDC Toronto

Courtney Kersten
CAE Professional Services

Defence R&D Canada
Technical Memorandum
DRDC Toronto TM 2011-132
September 2012

Canada

Validation of a virtual environment incorporating virtual operators for procedural learning

Brad Cain, Lochlan Magee
DRDC Toronto

Courtney Kersten
CAE Professional Services

Defence R&D Canada Toronto

Technical Memorandum
DRDC Toronto TM 2011-132
September 2012

Principal Author

Original signed by Brad Cain

Brad Cain

Defence Scientist

Approved by

Original approved by Linda Bossi

Linda Bossi

Head/Human System Integration Section

Approved for release by

Original released by Stergios Stergiopolous

Stergios Stergiopolous

Acting Chair, Knowledge and Information Management Committee

This work documented in this report was conducted under Thrust 14dn, Virtual Reality for Team Training.

In conducting the research described in this report, the investigators adhered to the policies and procedures set out in the Tri-Council Policy Statement: Ethical conduct for research involving humans, National Council on Ethics in Human Research, Ottawa, 1998 as issued jointly by the Canadian Institutes of Health Research, the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada.

Abstract

This paper presents the results of an experiment to assess the validity of a prototype simulation to train individuals to perform a task as part of a team. The application domain is Maritime Helicopter-Ship operations and the task selected is of a Landing Signals Officer (LSO) coordinating the approach and landing of a helicopter on board Canadian Forces frigates. The simulation includes physics based models of the helicopter, ship and the environment, as well as a human factors approach to representation of team mates by computer generated, behavioural agents. A reverse transfer of training experiment was conducted to assess how three groups, each initially differing in domain knowledge, acquired the necessary procedural knowledge, verbal communications and manual actions to complete the task without error. Thirty subjects participated: ten assigned to each of a Naïve, Aircrew and LSO group as determined by their initial domain knowledge. Learning rate results indicate significant differences among the groups and the effect sizes were sufficient to conclude that the approach is valid for training procedural tasks of the LSO occupation and, by extension, to other small team, procedural task trainers with similar user interface requirements. The simulation was not found to be adequate to train the fine, visual judgements involved in directing the helicopter over the deck, and improvements to the simulation have been proposed.

Résumé

Le présent document présente les résultats d’une expérience visant à évaluer la validité de la simulation d’un prototype d’entraînement de personnes à l’exécution d’une tâche au sein d’une équipe. Le domaine d’application est l’exploitation d’un hélicoptère maritime ainsi que d’un navire, et la tâche choisie est celle d’un officier de signalisation à l'appontage (LSO) coordonnant l’approche et l’appontage d’un hélicoptère se trouvant à bord de frégates des Forces canadiennes. La simulation comporte des modèles de l’hélicoptère, du navire et de l’environnement basés sur la physique, ainsi qu’une approche de la représentation des coéquipiers tenant compte des facteurs humains représentés par des entités reproduisant le comportement humain et générées par ordinateur. On a utilisé la méthode du transfert de formation inverse pour évaluer la façon dont trois groupes, selon leurs connaissances initiales du domaine, ont acquis les connaissances procédurales ainsi que les aptitudes à utiliser les commandes verbales et les actions concrètes nécessaires à l’exécution de la tâche sans erreur. Trente sujets ont participé; on les a répartis en trois groupes de dix novices, dix membres d’équipage et dix LSO, selon leurs connaissances initiales du domaine. La courbe d’apprentissage observée variait considérablement d’un groupe à l’autre et l’importance de l’effet était suffisante pour conclure que la technologie en question est utile à l’apprentissage du travail d’un LSO et, par extension, peut être utilisée par d’autres petites équipes dont l’entraînement exige une interface utilisateur similaire. La simulation s’est révélée inadéquate pour l’entraînement des jugements visuels excellents que nécessite la direction de l’hélicoptère au-dessus du pont, et on a proposé des améliorations à cette simulation.

This page intentionally left blank.

Executive summary

Validation of a virtual environment incorporating virtual operators for procedural learning. Brad Cain, Lochlan Magee, Courtney Kersten;

DRDC Toronto TM 2011-132; Defence Research and Development Canada (DRDC) Toronto.

Introduction: The apparent visual and behavioural fidelity of modern simulations are often thought to provide an effective learning experience when adapted for military training. Even though the fidelity of a Virtual Environment can be measured along some of its dimensions, it is difficult to assess all the relevant elements involved in determining whether a simulation is valid for training. In practice, validation is often a qualitative judgement based on a fitness for purpose in a given context rather than an objective assessment of training transfer. However, validation studies that focus on human performance and system effectiveness provide an approach that can determine whether a simulator is "fit for purpose", that is, valid for training. The study reported in this paper was conducted as part of a research project that is investigating a number of enabling technologies that have promise for affordable team training within virtual environments. The objective of the study was to demonstrate and validate the experimental approach using quantifiable evidence of its “fitness for purpose” while addressing an outstanding training need within the Canadian Forces (CF) Maritime Helicopter community.

Approach: The DRDC Toronto Helicopter Deck Landing Simulator and its counterpart at 12 Wing Shearwater (HelMET, Helicopter Marine Environmental Trainer) were leveraged by adding a Landing Signals Officer (LSO) workstation simulator. A Human Behavioural Representation (HBR) computer model of several members of the helicopter - ship team was developed in IPME (Integrated Performance Modelling Environment) to substitute for role players that would typically be required during team training. Thirty subjects (10 LSO, 10 Aircrew, and 10 Naïve) played the role of an LSO in a repeated measures experimental design, conducting 16 approaches and landings of a Maritime Helicopter onto a CF Halifax Class Frigate in a Reverse Transfer of Training experiment. Learning the LSO task was assessed by analyzing the proportion of correct verbal communications and manual actions made in each trial. The Reverse Transfer of Training hypothesis is that if the training technique is fit for purpose, expert subjects (LSOs) will adapt to the simulation quickly and demonstrate performance at criterion level; non expert subjects (Aircrew and Naïve groups) will initially perform poorly but then improve with training to approach criterion level. Failure of the expert group to perform well or the untrained group to improve at a reasonable rate is indicative of an environment that is not “fit for purpose.”

Results: The proportion of correct communications and actions was analyzed by several different approaches to address limitations in the data set. All of the analyses indicated that the percent correct metric was significantly different among the groups and that all groups improved with practice. The expert LSO group started at a high level of performance and quickly reached near perfect performance, consistent with high domain knowledge and ready adaptation to the simulator. The Aircrew group percent correct measure was initially of moderate performance, consistent with their familiarity with the environment but demonstrating a lack of specific LSO training; the Naïve group percent correct measure was initially low, reflecting their lack of exposure to the task. Nevertheless, the Aircrew and Naïve group percent correct measures both improved with repeated trials, eventually becoming indistinguishable from the expert LSO group, consistent with expectations for a simulator that is valid (“fit for purpose”).

Additional analysis indicated that the simulation was not adequate for training the conning of the helicopter over the flight desk, a time sensitive, tightly coupled manoeuvre. Further work is required to improve this feature of the simulation.

Significance: This study indicates that learning team tasks in a simulated environment with constructive Human Behaviour Representation operator models is feasible. It provides one method (Reverse Transfer of Training) of validating a training device using quantitative methods rather than relying on qualitative judgements. Validation of the simulator suggests that this approach could be used in similar applications, not only in the Maritime Helicopter domain but across the CF in many of the small team training situations. The results of this study have been used as the basis to provide advice to exploitation agencies within the Department of National Defence to apply the technologies advantageously in training applications.

Future plans: This study is one of several planned to study and demonstrate both techniques for validating training simulators and to assess emerging technologies such as virtual reality and human behaviour representation for use in military training. Subsequent studies will elaborate on the Reverse Transfer of Training approach and incorporate other approaches to improve our understanding and use of training simulator technologies as well as techniques to validate their use through evidence and performance based quantitative measures.

Sommaire

Validation d'un environnement virtuel intégrant des opérateurs virtuels pour l'apprentissage procédural. Brad Cain, Lochlan Magee, Courtney Kersten;

RDDC Toronto TM 2011-132; Recherche et Dévolopment pour la Défense Canada (RDDC) Toronto.

Introduction : L'apparente fidélité des simulateurs modernes sur les plans visuel et comportemental porte souvent à croire que ces derniers offrent un apprentissage de qualité en contexte militaire. Même si la fidélité d'un environnement virtuel peut se mesurer à partir de certaines de ses dimensions, il est difficile d'évaluer tous les éléments pertinents qui entrent en jeu dans la détermination de la validité d'une simulation pour l'entraînement. En pratique, on juge souvent sa validité du point de vue qualitatif en se basant sur son utilité en contexte militaire plutôt qu'en faisant une évaluation objective axée sur le transfert de la formation. Les études de validation portant sur la performance humaine et l'efficacité d'un système servent à déterminer si un simulateur donné est adéquat ou non, à savoir, dans le cas qui nous occupe, s'il est valide pour l'entraînement. L'étude dont il est question dans le présent document a été menée dans le cadre d'un projet de recherche et développement visant à examiner diverses technologies prometteuses en ce qui a trait à l'entraînement en équipe dans un environnement virtuel. Notre objectif était donc de démontrer et de valider la technologie en question en présentant des preuves quantifiables de son adéquation, tout en répondant à un besoin exceptionnel en matière d'entraînement au sein de la collectivité de l'hélicoptère maritime des Forces canadiennes.

Démarche : Pour cette étude, on a fait appel au simulateur d'appontage pour hélicoptères (SAH) de RDDC Toronto et à son analogue de la 12e Escadre Shearwater, le simulateur d'appontage en milieu marin (HelMET), auxquels on a ajouté un simulateur de poste de travail de LSO. Afin de remplacer les acteurs de soutien habituellement requis lors d'un entraînement d'équipe, on a employé un système informatique imitant le comportement humain des membres de l'équipage d'hélicoptère et de navire, un système conçu dans l'environnement intégré de modélisation des performances (EIMP). Les trente sujets (10 LSO, 10 membres d'équipage et 10 novices) ont joué le rôle de LSO dans le cadre d'une étude utilisant à plusieurs reprises des mesures expérimentales, en dirigeant 16 manœuvres d'approche et d'atterrissage d'un hélicoptère sur une frégate des FC de classe Halifax, dans le cadre d'une expérience de transfert de formation inverse. On a évalué l'apprentissage du rôle de LSO en analysant la proportion de commandes verbales et d'actions concrètes exécutées correctement lors de chaque essai. Le transfert de formation inverse part de la prémisse que si la technique de formation est adéquate, les sujets experts (les LSO) devraient maîtriser plus rapidement le simulateur de manière à satisfaire les critères de rendement que les sujets profanes (membres d'équipage et novices), qui devraient offrir un piètre rendement au début pour ensuite s'améliorer jusqu'à ce qu'à s'approcher des critères de rendement. Lorsque les sujets experts obtiennent de mauvais résultats ou que les sujets profanes ne s'améliorent pas à un rythme raisonnable, cela indique que la technologie en question est inadéquate.

Résultats : On a analysé la proportion de commandes verbales et d'actions concrètes exécutées correctement en utilisant différentes démarches pour pallier les lacunes de l'ensemble des données. Toutes les analyses ont indiqué que le pourcentage de mesures effectuées correctement présentait un écart considérable d'un groupe sujet à l'autre et que tous les groupes s'étaient améliorés en s'exerçant. Dès le début, les LSO ont offert un excellent rendement et ont rapidement atteint un niveau d'aptitude presque parfait, ce qui traduit de vastes connaissances du

domaine et une adaptation rapide au simulateur. Les membres d'équipage ont, quant à eux, d'abord présenté un rendement moyen correspondant à leur niveau de familiarité avec l'environnement en question, mais reflétant un manque de formation propre au rôle de LSO. Enfin, le rendement des novices était faible au début, comme il fallait s'y attendre vu leur manque d'expérience dans le domaine. Néanmoins, au fil des essais, tant les membres d'équipage que les novices se sont améliorés jusqu'à finalement réussir à présenter un rendement semblable aux LSO, ce à quoi l'on peut s'attendre d'un simulateur adéquat.

Une analyse supplémentaire a révélé que la simulation ne convenait pas à l'entraînement au contrôle de l'hélicoptère au-dessus du pont d'envol, manœuvre en configuration groupée au cours de laquelle le facteur temps est critique. Il faudra travailler pour améliorer cette fonction de la simulation.

Portée : La présente étude prouve qu'il est faisable de conduire un entraînement d'équipe dans un environnement simulé appuyé de modèles reproduisant le comportement humain. Grâce à la méthode du transfert de formation inverse, l'appareil d'entraînement a été évalué en fonction de critères quantitatifs plutôt que d'un point de vue qualitatif. La validation du simulateur donne à penser que l'on pourrait utiliser cette démarche dans des applications similaires, non seulement dans le domaine de l'hélicoptère maritime, mais également au sein de toutes les FC, dans de nombreuses situations d'entraînement de petites équipes. On a utilisé les résultats de cette étude comme base pour fournir des conseils aux organismes d'exploitation au sein du Ministère, afin d'appliquer avantageusement les technologies aux applications d'entraînement.

Perspectives : La présente étude s'inscrit dans le cadre d'une série d'expériences prévues dans le but d'examiner et de démontrer les différentes techniques de validation de simulateurs d'entraînement et pour évaluer l'utilité de nouvelles technologies, comme la réalité virtuelle et la reproduction du comportement humain dans les entraînements militaires. Dans les prochaines études, nous exposerons plus en détail la méthode du transfert de formation inverse et incorporerons d'autres démarches visant à améliorer notre compréhension et notre utilisation des technologies du simulateur d'entraînement, ainsi que des techniques de validation de leur utilisation, en s'appuyant sur des observations concrètes et des mesures quantitatives de la performance.

Acknowledgement

The authors would like to thank the men and women of the Canadian Forces at 12 Wing Shearwater, Nova Scotia who made this study possible. In particular, we would like to thank the Wing Commander, Col. S. Michaude, for providing access to personnel to validate our concepts despite busy schedules and the efforts of Capt. D. Philip in coordinating our efforts. Additionally, we thank Ms. B. Fraser for her support during the pilot study and the recommendations that shaped the final experiment. Finally, we would like to thank Capt. N. McCarthy for his unwavering support and enthusiasm to advance the use of modelling and simulation for training within the Maritime Helicopter community.

This page intentionally left blank.

Table of contents

Abstract ... i
Résumé i
Executive summary iii
Sommaire. v
Acknowledgement vii
Table of contents ix
List of Figures x
List of Tables xi
Abbreviations xii
Introduction 1
  Purpose and objectives 2
Experimental approach 3
  Approach 3
  Apparatus 4
  Role-playing virtual operators 6
  Procedure 7
  Subjects 9
Results 11
  Procedural communications and actions 11
    Curvefiting of Percent Correct data 13
  Visual judgements 16
    Landings 18
  Workload 19
    Overall Workload 19
    Mental Demand 22
    Physical Demand 22
    Temporal Demand 23
    Own Performance 23
    Effort 24
    Frustration 25
  Simulator sickness 26
Discussion 29
Conclusions 32
References 33
Appendix A. Experimental scenario 35
Appendix B. Experimental simulation hardware 41
Appendix C: NASA TLX subjective workload rating scale 43
Appendix D. Simulator Sickness Questionnaire (SSQ) 45

List of Figures

Figure 1. Sea King helicopter hovering over the flight deck of a CF Halifax Class frigate during deck operations. 1

Figure 2. Computer generated imagery showing the Sea King helicopter hovering over the trap on the flight deck. The left image shows the LSO's HOWDA in the lower left corner while the right image shows the LSO subject's view of the simulated environment through the HMD. 5

Figure 3. LSO subject wearing a Head Mounted Display in the HOWDAH simulator and the LSO console. 5

Figure 4. A segment of the virtual MH Pilot model as represented in IPME. 6

Figure 5. Conceptual layout of the links among the subject and the SimON virtual operators....... 7

Figure 6. Performance as measured by the combined correct verbal communications and manual actions expressed as a percentage for each group over the 16 trials. No imputed values are included for lost data in this figure. Data points are means and standard deviations. N=10 (nominally). 12

Figure 7. Regression values for the Exponential function (Plateau – *Y0*) coefficient representing the amount of information to be learned to achieve perfect performance. The Plateau was fixed at 100. Data are group means and standard deviations. 15

Figure 8. Regression values for the Exponential function learning time constant τ (*Tau*; group means and standard deviations.) 16

Figure 9. Regression coefficients for the simplified Power function of learning (no Experience coefficient). Values are group means and standard deviations. 16

Figure 10. 2-D histogram plot of the fraction of (a) correct and (b) incorrect verbal conning calls of the helicopter probe over the ship's trap from all subjects. Perspective is approximately that of the LSO's view of the trap from the HOWDAH. 17

Figure 11. Average (standard deviation) landing attempts by group. 18

Figure 12. Average (standard deviation) number of landings by trial. The dashed line, shown with the regression line, suggesting a slight improvement with practice, although the effect was not significant. 19

Figure 13. Unweighted sum of NASA TLX factors by Group and by Block. Values are average (standard deviations) without imputation. 20

Figure 14. Average (standard deviation) values of the sum of weighted NASA TLX factors...... 21

Figure 15. Average (standard deviation) values of the Mental Demand ratings by Block and Group. 22

Figure 16. Mean (standard deviation) values of the perceived temporal demand showing the interaction between Group and Block factors (dashed lines). 23

Figure 17. Average (standard deviation) values of the Own Performance ratings by Block. 24

Figure 18. Average (standard deviation) ratings of the NASA TLX Effort factor by Block and Group. 25

Figure 19. Average (standard deviation) ratings of the NASA TLX Frustration factor by Group. 26

Figure 20. Average (standard deviation) of the Total SSQ scores at the end of each of the two Sessions decomposed by Group. 27

List of Tables

Table 1. Aggregate subject characteristics by group. 9

Table 2. Descriptive statistics on subject experience categorized by LSO qualifications 10

Table 3. Summary of the preferred regression model based on the Akaike Information Criterion (AIC) for all subjects combined and broken out by group. 14

Table 4. Evaluation of the use of weighted factors in explaining variance in the NASA TLX ratings. SS is the sum of squares derived from the ANOVA tables. 21

Abbreviations

AGARD	Advisory Group for Aerospace Research and Development (NATO)
AIC	Akaike Information Criteria
ANOVA	Analysis of Variance
ASL	Above Sea Level
CF	Canadian Forces
CPF	Canadian Patrol Frigate
DOF	Degrees of Freedom
DRDC	Defence Research and Development Canada
ELVA	Emergency Low Visibility Approach
FLYCO	Flying Coordinator
FOD	Foreign Object Damage
HBR	Human Behaviour Representation
HDLS	Helicopter Deck Landing Simulator
HelMET	Helicopter Maritime Environmental Trainer
HMCS	Her Majesty's Canadian Ship
HMD	Head Mounted Display
HTA	Hierarchical Task Analysis
ICS	Internal Communication System
IMC	Instrument Meteorological Conditions
IOS	Instructor Operator Station
IPD	Interpupilary distance
IPME	Integrated Performance Modelling Environment
KIAS	Knots Indicated Air Speed
L/G	Landing Gear
LCD	Liquid Crystal Display
LSO	Landing Signals Officer
MH	Maritime Helicopter
NASA TLX	National Aeronautics and Space Administration Task Load Index
PC	Percent Correct
RADHAZ	Radiation Hazard
RAM	Random Access Memory

RCA	Radio Corporation of America
RSD	Rapid Securing Device
SA	Scientific Authority
SAC	Shipborne Air Controller
SHOPS	Shipborne Helicopter Operating Procedures
SimON	Simulated Operator for Networks
SOP	Standard Operating Procedure
SS	Sum of Squares
SSQ	Simulator Sickness Questionnaire
TACCO	Tactical Coordination Officer
VE	Virtual Environment

This page intentionally left blank.

Introduction

Defence Research and Development Canada (DRDC) Toronto has applied human factors to the design, development and evaluation of low cost simulators for affordable training within the Canadian Forces (CF). A recent example is an experimental development simulator, the Helicopter Deck Landing Simulator (HDLS), for training Maritime Helicopter (MH) Pilots the procedural aspects of landing a helicopter on a Canadian Patrol Frigate (CPF) under way. The HDLS is being used by the MH training community at 12 Wing Shearwater, renamed HelMET (Helicopter Maritime Environmental Trainer), for Advanced Force Generation training to provide pilots with a virtual experience of the approach and landing on a CF Halifax Class frigate under various environmental conditions.

The MH team, however, consists of more than the pilots; there is aft cabin flight crew as well as several members aboard the ship who contribute to the overall performance and safety of helicopter-deck operations. Currently, the principal training method for the MH team is to use operational equipment while at sea, an expensive (~ $40,000 per hour) and risky approach to training that may not be an optimal learning environment, but one driven by necessity due to a lack of a suitable alternative.

The HDLS has been extended by DRDC Toronto to include a simulator for the Landing Signals Officer (LSO) of a Halifax Class ship along with a Human Behaviour Representation (HBR) or computer model of the Sea King Pilot to demonstrate and validate emerging technologies that may lead to a more suitable learning environment that is more inclusive of the rest of the operational team (see **Figure 1** for task examples). The LSO is a member of the MH team aboard ship who is responsible for flight operations, assisting the pilots during launch and recovery of the helicopter from the ship's flight deck as well as directing other close-quarters activities.

Figure 1. Sea King helicopter hovering over the flight deck of a CF Halifax Class frigate during deck operations. Other helicopter-ship tasks are also performed requiring team coordination.

While the environment selected for the current demonstration and experiment is a MH-CPF scenario, the technologies under study are thought to be applicable to many CF team training domains, as are the methods used to validate the technologies. If the HDLS/HBR system proves to be an effective method for instructing LSOs on deck landings, then it seems plausible that this approach would provide an effective training tool for many team training tasks that the CF undertake that place similar learning and judgement demands on personnel.

Purpose and objectives

The objective of this study is to demonstrate and validate the collective use of immersive visual displays, affordable simulators and behavioural models of team mates in procedural training devices.

Although simulation and simulators have been used for quite some time, many have not been subjected to rigorous validation. Cost and technical difficulty are cited as reasons for not conducting validation studies and the compelling nature of modern computer generated imagery may be leading to over-reliance on face validation or opinion about the effectiveness of a technology in a training application. Nevertheless, there is increasing interest in evidence based learning, relying on quantitative metrics to assess and evaluate tools, methods and techniques.

Validation, in the current context, is taken to mean “fit for the purpose in which it is intended”, not that the simulator is “indistinguishable from the operational environment.” Modelling in general is about abstracting the important elements from “the real thing” and representing those elements such that they achieve an intended goal. In the case of training, the goal is the transfer of skills learned during purposeful practice into applied operations. The current study will use a reverse transfer of training paradigm (AGARD, 1980) in the validation of a specific instance of the approach and technologies. Reverse transfer of training is used to mitigate the risks associated with negative transfer that may occur in forward transfer of training paradigms for operational settings.

Much of the current responsibility for collective training rests with the operational units that are already struggling with personnel shortages in many skilled occupations. The conventional method of training new personnel relies on the availability of qualified personnel to act as role players. This places demands on the organization by taking qualified personnel away from other duties while they derive little if any training benefit from the training exercise. Team training also presents a coordination challenge in terms of scheduling training events to coincide with the participants’ availability, particularly as the team grows. In many instances, only part of the team needs training, yet all members of the team are required to participate for an effective experience.

There is an increasing interest in the use of computer generated actors to take on the role playing tasks, referred to as Human Behaviour Representations (HBR). DRDC Toronto began the Simulated Operator for Networks (SimON) project to explore the development of computer models of operators performing tasks as substitutes for operators in simulations, particularly simulations where plausible human behaviour is required. SimON operator models strive to get plausible performance, preferably relying on human performance models, rather than striving for optimal performance, to create a richer environment demonstrating plausible variability and errors that arise for reasons similar to humans performing the role (Pew & Mavor, 1998, pp.19-20). HBR models of human characteristics are thought to be particularly important in applications where these models go beyond being simple stimuli, to being agents with pertinent human-like characteristics, where interactions among agents and human participants are unscripted, requiring plausible decisions and behaviours.

Experimental approach

The experiment comprised a pilot study and a formal study, both of which are reported in this document. Both the pilot and formal studies followed the same protocol, L-697 *Validation of Simulator Based Training of Tightly Coupled Operations: Training for Helicopter Deck Landing Procedures*, as approved by the DRDC Human Research Ethics Committee. The experimental scenario (**Appendix A**) was adapted to meet the experimental objectives from a training scenario developed by 12 Wing Shearwater.

Approach

A number of experimental paradigms are available for assessing the validity of a training simulation based on human performance (AGARD, 1980). The most direct approach is a Forward Transfer of Training experiment in which the benefits of simulator training are assessed in the real world. Unfortunately, such studies are rare: there are methodological constraints (counterbalancing and small samples); they are prone to noise (subject dropout, changing administration priorities, changing experimental conditions); they can expose subjects to inadvertent, negative training transfer.

The experimental approach selected for this assessment was a Reverse Transfer of Training paradigm (AGARD, 1980) that evaluates a training device using at least two groups of subjects: one group that is qualified in the real world at the task to be performed; another group that is not qualified at the task. Reverse Transfer of Training is an alternative, less direct method that can avoid many of the risks and challenges of Forward Transfer studies.

The Reverse Transfer of Training paradigm hypothesizes the following for a valid simulation:

1. Experts who know the task will start at a high level of performance and asymptote quickly to a criterion performance level.

2. Non experts will start at a low level of performance and improve over time, eventually reaching the same asymptotic performance as the experts.

If initial expert performance is low or expert performance improvement is slow within the Virtual Environment (VE), the assumption is that the experts are accommodating to the VE and the implication is that the simulator is an inadequate representation of the real world. If non experts start at a level of performance similar to the experts or if they fail to approach the expert performance levels over a reasonable period (say, compared with field training) then it can be concluded that the simulator is failing to provide an appropriate learning environment. If the participants possess partial knowledge of the task, then intermediate levels of initial performance and amounts of practice required to achieve asymptote are predicted.

Subjective assessments of workload and simulator induced sickness were also considered indicators of the "fitness for purpose" of the simulator and these indices were evaluated using common measurement scales. The NASA TLX Workload measurement scale (**Appendix C**) and the SSQ (Simulator Sickness Questionnaire, **Appendix D**) were completed at various points in the study by the subjects as described in the Experimental approach. If the synthetic environment is a valid representation of the operational environment, we would expect that workload would be manageable by qualified LSOs while workload would be initially high for untrained personnel but approaching the qualified LSO level with exposure. Conversely, if workload is unmanageable by trained personnel, or workload does not reduce with time for personnel learning the task, then the synthetic environment may not be providing an adequate simulation of the operational environment to train the task effectively.

Further, we would also expect the incidence of simulator induced sickness to be low for all subjects exposed to the synthetic environment. If extreme levels of simulator sickness are observed, then it is reasonable to assume that some aspect of the simulator is inappropriate or inconsistent with effective use.

Hypotheses

The principal hypothesis of this experiment is that the combined HDLS/HelMET synthetic environment and the SimON HBR are a sufficiently valid representation of the Canadian Forces Maritime Helicopter (MH) deck landing environment that it provides an effective method for training the LSO in the procedural aspects of MH free-deck landing evolutions.

A secondary hypothesis is that this system will aid training LSOs to make accurate visual judgments of relative position of the helicopter over the trap, allowing them to learn how to conn (direct) the helicopter pilot to a successful landing.

Apparatus

The experimental simulation comprises an LSO simulator (**Appendix B**), a Sea King helicopter simulation, a Sea King Pilot simulation and an Instructor-Operator Station (IOS). This application entails a time sensitive, tight coupling of interactions among the simulations and the participating personnel. The HDLS is a real time simulation incorporating three dimensional models of the synthetic environment, the helicopter and the ship as well as moderate fidelity dynamic models of the aircraft aerodynamics, ship motion and the air wake over the flight deck of the ship. As a simulation of the Sea King Pilot was used in this experiment, the full HDLS/HelMET Sea King simulator was not required, only the visual representation from the underlying, physics based models. Nevertheless, the pilot model provided primary flight control displacement signals to the helicopter simulation to control the helicopter's flight path to demonstrate a flexible, modular team training concept where team positions could be staffed by students, computer agents or role players as desired.

The LSO's actual workstation, called the Howdah, is located at the foreward-starboard side of the flight deck looking aft; a simulated view as seen by the LSO subjects is shown in **Figure 2**. The LSO simulator is networked with the HDLS/HelMET simulation and presents visual imagery to the subjects using a fully occluded, stereo, colour head mounted display (HMD; for hardware details, see **Appendix B**). The instantaneous point of view was determined by a magnetic, head-tracking system, allowing for an unrestricted *field of regard*[1].

---

[1] "…the field of regard refers to the area within which the operator can move his or her head to see visual information…" retrieved from http://www.trainingsystems.org/TTCP/html/anatomy_of_simulations/concepts_and_terms.html#field%20of%20regard

Figure 2. Computer generated imagery showing the Sea King helicopter hovering over the trap on the flight deck. The left image shows the LSO's Howdah in the lower left corner while the right image shows the LSO subject's view of the simulated environment through the HMD.

The LSO subjects stood in front of a physical mock-up of the LSO console that provided the necessary switches and buttons for the training scenario (**Figure 3**). Two rotary switches (Bridge Clearance Request, Trafficator lights) and one toggle switch (Rapid Securing Device, RSD or the "trap", control) were used in the study. The subjects could not see their hands when they had to adjust the switches on the LSO console because of the occluded HMD; hands were not tracked and computer generated in the visual display. The virtual switches on the LSO console could be seen by the subjects and these virtual switches changed to reflect any changes subjects made to the physical switches. Subjects did adapt quickly and were able to locate the switches by touch after a few trials; LSOs often use touch to locate buttons on the actual console as their attention is usually directed out of the Howdah, viewing the helicopter and ship.

Figure 3. LSO subject wearing a HMD in the Howdah simulator and the LSO console.

Role-playing virtual operators

Three virtual operators were included in the scenario: the Shipborne Air Controller (SAC), the MH Tactical Coordination Officer (TACCO) and the MH Pilot. These virtual operators were created in the Integrated Performance Modelling Environment (IPME) using a Hierarchical Task Analysis (HTA) framework in a procedural, but unscripted representation of the tasks required during the MH's approach and landing on the ship. The representations of the TACCO and SAC roles were minimal, limited to providing and reacting to verbal stimuli early in the scenario. The Pilot model, which was the focus of the HBR modelling, was more detailed, monitoring goals, interacting with the subjects according to Standard Operating Procedures (SOPs) and providing corrective inputs to the helicopter simulator to fly the approach and landing.

A segment of the Pilot model landing phase procedure is shown in **Figure 4**; other portions of the model provided representations of monitoring communications, aircraft status, primary flight control inputs, etc. The Pilot model is organized according to Hierarchical Task Analysis principles (Annett, 2003; Annett & Duncan, 1967; Annett, Duncan, Stammers & Gray, 1971; Annett & Stanton, 2000; Shepherd, 2000), allowing for subsequent elaboration of procedures and tasks as required.

Representation of the model within IPME permits inclusion of stressor and performance moderator functions (such as workload) as well as variation in operator traits and states to provide a less predictable yet controllable interaction among the subjects and the HBR computer agents. IPME was networked with the HDLS simulation, receiving updates of a number of variables at 60 Hz. These variables were subsequently sampled by the Pilot model as dictated by the active tasks (approximately 3 to 4 Hz) to assess current task goal status.

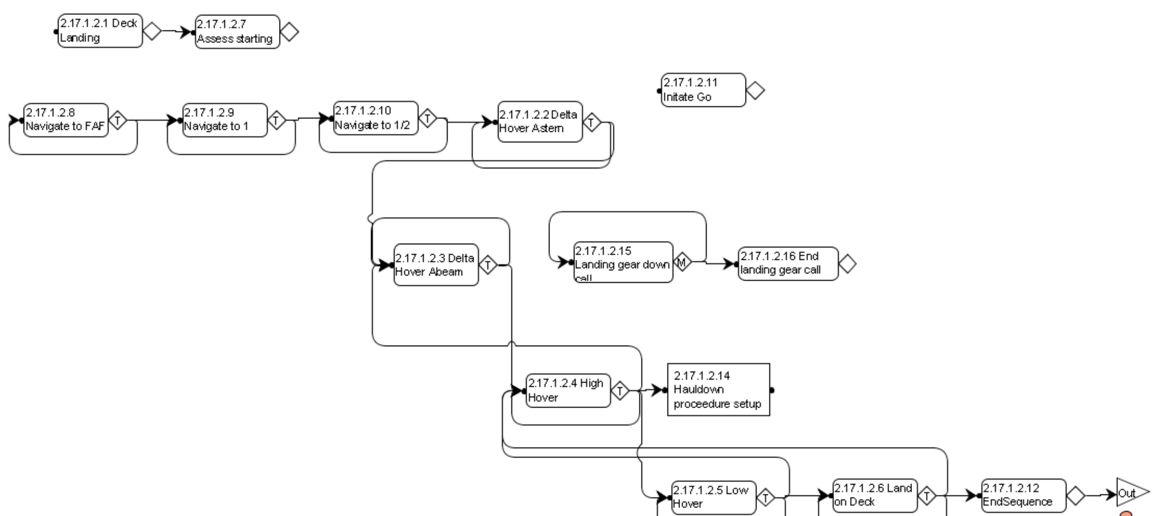


Figure 4. A segment of the virtual MH Pilot model as represented in IPME.

Verbal communication between the LSO subjects and the virtual operators was through a microphone (connected to speech recognition software) and speakers (through commercial speech production software, AT&T Naturally Speaking.) Speech recognition and production were handled though software clients networked with IPME. A conceptual layout of the audio and video configuration is shown in **Figure 5**.

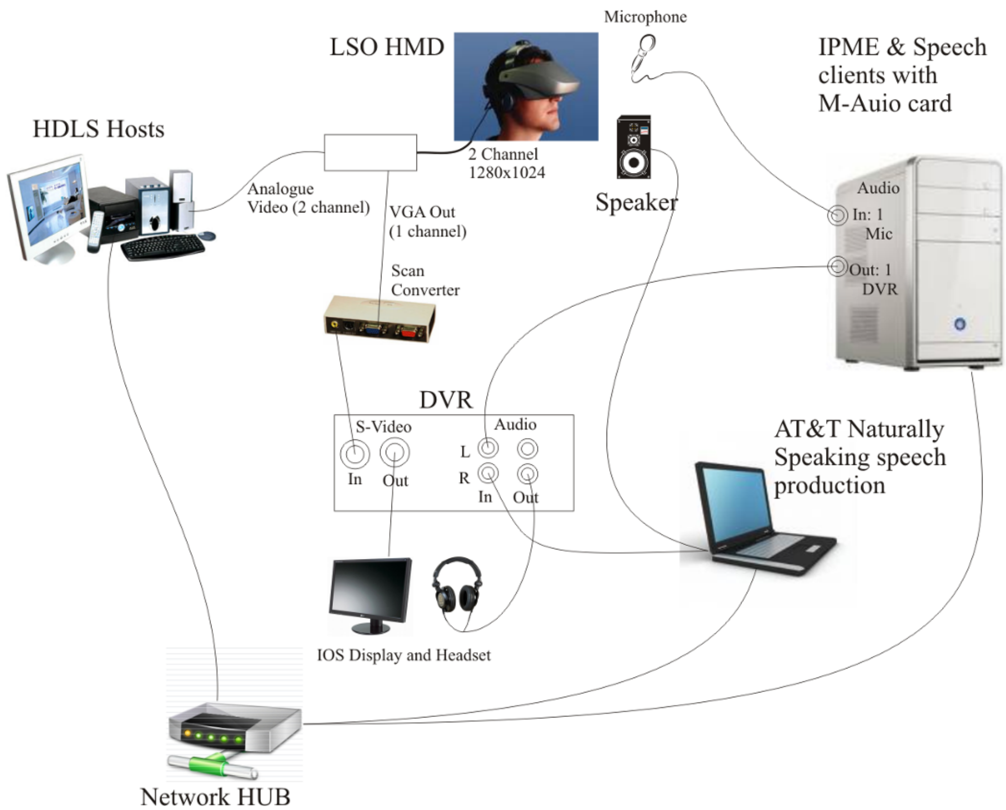


Figure 5. Conceptual layout of the links among the subject and the SimON virtual operators.

Procedure

At the beginning of each experiment, subjects were briefed on the experimental objectives and time commitment, benefits and risks associated with their participation. Subjects were informed that their participation was purely voluntary and that they had the right to withdraw from the study at any point in time without prejudice and at their own discretion. Prior to their participation, all subjects read and signed the subject consent form, giving their written consent to voluntarily participate in the study. Subjects also completed stress remuneration forms as part of the monetary compensation given for their participation in this study.

Each subject's stereo-acuity was then assessed using the Titmus Graded Circles Stereo-acuity Test and their interpupillary distance was measured. Subjects provided their age and an estimate of their height. Additionally, subjects in the formal study provided the number of flight hours they currently had in a Sea King helicopter and a rough estimate of the number of deck landings they had experienced as either a qualified LSO or as a member of the MH flight crew, as appropriate.

If the subject was assigned to the baseline pre-exposure SSQ group, they received an SSQ prior to participation in the first experimental session (**Appendix D**). If a subject was not assigned to the baseline pre-exposure SSQ, the research assistant sought verbal confirmation from the subject that they were in a general healthy state before proceeding.

All subjects were then given a sample scenario story-line to review that outlined the scenario and explained the subject's roles and responsibilities associated with playing the LSO role (**Appendix**

A). During this time, the HMD was disinfected, ensuring all skin contact areas were cleaned with an alcohol wipe. Subjects were provided with answers if they had any questions about the task.

The subjects were then briefed on the NASA TLX workload questionnaire (**Appendix C**), explaining how and when it would be required to be completed throughout the study. The subjects were then asked to read over the definitions associated with each workload demand, so that they became familiar with the definitions and understood how each workload demand was defined.

The subjects were briefed on the role of the research assistant: to provide corrective feedback for the LSO's verbal communications and manual actions, both throughout and at the completion of each trial, similar to on-the-job training received during current LSO training at sea. Subjects were then introduced to the 3 LSO switches (Trafficator, RSD and Bridge Clearance Request) situated on the LSO simulator console that subjects would be required to interact with during each trial. The starting position in which each switch must be placed at the beginning of each trial (Trafficator in RED, RSD in OFF and the Bridge Clearance Request switch in AIRBORNE) was identified. Additionally, the subjects were informed that in order to initiate any trial, they would be required to turn the Bridge Clearance Request switch to the RECOVER mode.

Subjects were then asked to place a microphone over their right ear and instructed to speak clearly with a normal cadence during each trial. Subjects then received instruction on how to adjust the HMD to ensure that it fit appropriately.

Once subjects had adjusted the HMD, the HDLS simulation was started. Subjects were encouraged at this point to explore the virtual environment. Once a subject felt comfortable with the simulation, the HDLS and SimON software began executing the scenario and the subject was instructed to place the Bridge Clearance Request switch into the "recover" mode to mark the beginning of each trial. Each trial involved the same scenario and lasted approximately 4 to 5 minutes.

During each trial, subjects listened to the virtual operators, provided verbal instructions to them and operated the LSO console switches. If time permitted, the research assistant provided corrective feedback immediately after an error was committed; if there was insufficient time, corrective feedback was provided at the end of the trial.

After the completion of the first block of 4 trials, subjects were given a break, removing the HMD and microphone. Subjects were allowed to sit and were provided with water. During the break, subjects completed the first NASA TLX workload questionnaire to assess their perceived workload demands. The second block of 4 trials began at the discretion of the subject (approximately 15 minutes later).

At the end of the second block of trials, subjects completed a second NASA TLX questionnaire as well as the SSQ. All subjects SSQ ratings were immediately reviewed by the research assistant and any simulator sickness symptoms indicated on the questionnaire were brought to the attention of the Scientific Authority overseeing the experiment. Before leaving the first experimental session, subjects were cautioned about potential issues surrounding simulation sickness.

This concluded the morning session; an interval of approximately 3 hours was observed before subjects returned for the second, afternoon session. Pre and post trial procedures were similar for the second experimental session, replicating the first experimental session of two blocks of 4 trials with the NASA TLX workload questionnaire completed after the 4th and 8th trials. However, at the end of the second experimental session, subjects additionally were asked to compare each workload demand rating, identifying the workload demand they thought to be the larger or more important contributor to their overall workload.

Subjects

Subjects were recruited by poster, both in the pilot study and in the formal study. All of the subjects were between the ages of 18 and 60 years.

In the pilot study, eleven volunteers (6 men and 5 women) were assigned random subject numbers on a first-come-first-served basis; one extra subject was included as a portion of another subject's data was lost during the testing due to equipment malfunction and this partial data set was eliminated from the data analysis. None of the Naïve subjects in the pilot study had any prior experience with MH operations.

In the formal study, CF flight crew volunteers from the MH community were placed in a pool by the 12 Wing Duty Officer. On each of the ten days of testing, the Duty Officer selected a pair of volunteers (1 LSO and 1 Aircrew) from the subject pool, who were then assigned random subject numbers by the SA. Selection from the pool was based on availability for the current day, which was constrained by operational commitments as subjects were prohibited from flying for 12 hours after exposure to the simulation. A total of 20 military personnel (19 male and 1 female) from within the MH flight crew community participated in the formal study at the HelMET simulator facilities at 12 Wing Shearwater. As implied above, 10 of the subjects were qualified or previously qualified LSOs and 10 subjects were Aircrew who had no formal LSO training. In order for a subject to be eligible as a qualified LSO in this study they were required to have obtained full LSO qualifications, thereby having the ability to fulfill the LSO role when at sea, but they did not have to be currently qualified (the number of qualified, current LSOs is small and operational duties precluded accepting only current LSOs.) Aircrew subjects in this study were required to be members of the MH flight crew, familiar with high level helicopter-deck landing procedures but not having received any formal LSO training. Ninety-five percent of the subjects in the formal study scored 100% on the Titmus Graded Circles Stereo-acuity Test. Descriptive statistics (mean, standard deviations (s.d.) and ranges) for age, height and interpupillary distance (IPD) are presented in **Table 1**.

Table 1. Aggregate subject characteristics by group.

	Height (m)				Interpupillary Distance (mm)			
Group	Mean	s.d.	max	min	Mean	s.d.	max	min
Naive	1.71	0.08	1.83	1.57	61.5	2.7	68.5	58.0
Aircrew	1.76	0.06	1.84	1.68	60.9	3.3	66.5	56.0
LSO	1.78	0.08	1.88	1.63	63.1	1.7	66.0	61.0

Additionally, subjects were asked to provide: 1) the number of flight hours they currently had accumulated in the Sea King helicopter and 2) an estimate of the number of deck landings they had either performed as a qualified LSO or had experienced second hand as a member of the MH flight crew. Flight hours are continually tracked by flight crew and readily recalled; deck landings are not tracked, so the values provided are crude estimates at best. For those subjects who provided an estimated range for the number of deck landings (i.e., 500-1000) rather than a single number estimate (i.e. 500), the average of their estimated range was calculated and presented within the following descriptive statistics. **Table 2** below provides the descriptive statistics (mean and standard deviations) of Sea King helicopter flight hours and estimated deck landings for both the qualified LSO and Aircrew subjects.

Table 2. Descriptive statistics on subject experience categorized by LSO qualifications

Subject Experience	**LSO**		**Aircrew**	
	Mean	**s.d.**	**Mean**	**s.d.**
Sea King flight hours	1369.9	559.1	1389.9	1192.4
Estimated number of deck landings	612.5	421.5	397.9	335.3

Results

Several subjects were not required to complete all of the trials as they reached criterion performance (two successive perfect task completions) in fewer than the maximum number of trials allowed. This resulted in lost data in the repeated measures analysis (18% for LSO subjects and 3% for the Aircrew subjects; none for the Naïve group). The task performance data were thus analyzed several ways to determine whether or not a consistent set of conclusions would be reached.

First, the Percent Correct data were analyzed as a mixed model Analysis of Variance (ANOVA) (3 groups x 16 trials) with no estimates for missing data. A p-value of 0.05 was considered statistically significant for all analyses. This resulted in several subjects' data being automatically removed from the analysis due to missing values (case-wise deletion of subjects) and hence unequal numbers of subjects within each group (10 Naïve; 9 Aircrew; 5 LSO). The data were then reanalyzed using imputed values for missing data – the last recorded value for each subject was used as an estimate for the missing values. Because the variance was correlated with the means, the data (with imputed values for missing data) were subsequently adjusted by a sine-logarithm transformation and again analyzed in a 3 by 16 mixed model ANOVA. Finally, learning curve functions (both power-law and exponential curves) were then fitted to the data by nonlinear regression (with no estimates for missing data) and the resulting curve fit coefficients were analyzed.

Secondary measures (workload, SSQ, etc.) that were recorded at the end of each block were less susceptible to missing data (10 Naïve; 9 Aircrew; 8 LSO) and so they were only analyzed using imputed values for missing data (typically required for the 4^{th} block only). The imputed values for each subject were estimated by assuming the values recorded for that subject in the previously completed block, as if they had reached an asymptotic value.

Procedural communications and actions

The verbal communications and manual actions recorded were converted into a percent correct score (PC) for each trial. Different numbers of communications or actions were possible during the landing phase of the simulation, as several landing attempts or different conning styles (differing communication frequencies) were possible. The count of the verbal communications and manual actions within the landing phase were normalized by the number of landing attempts (until the subject felt the landing was successful or a limit of 5 attempts was reached); the number of directional conning commands used to direct the pilot when positioning the helicopter over the trap was normalized by the total number of directional commands issued within a trial. This normalization process was done in an attempt to reduce any bias that might arise due to an unequal number of landing attempts or differing conning frequencies.

Values of the combined verbal communications and manual actions are shown in **Figure 6**; the interpolation lines were included to illustrate that the trends are exponential fits to each group's data. These results show that, as might be expected, initial performance depends on experience, with the most experienced, LSO group starting at a higher performance level than the other two groups, while the Aircrew group (with at least domain experience and likely some indirect exposure to LSO procedures) falling intermediate to the LSO and Naïve groups. All groups improve with practice and appear to asymptote to perfect performance as the number of trials increase, also as expected. There is an apparent performance decrement after the 3 hour break between Blocks 2 and 3, although no decrement is evident after the shorter breaks between Blocks 1 and 2 or Blocks 3 and 4.

Analysis of the data with no imputed values for lost data indicated a significant interaction of the Group (between) factor and the Trial (within) factors ($F_{30,315} = 13.786$, $p < 0.001$, $MS_{error} = 0.0043$) with main effects of both Group ($F_{2,21} = 27.493$, $p < 0.001$, $MS_{error} = 0.0439$) and Trial ($F_{15,315} = 94.248$, $p < 0.001$, $MS_{error} = 0.0043$). A similar analysis with imputed values for missing data indicated a similar pattern of outcomes with a significant interaction between Group and Trial ($F_{30,315} = 22.441$, $MS_{error} = 0.0039$, $p < 0.001$) and significant main effects of both Group ($F_{2,21} = 51.379$, $MS_{error} = 0.0352$, $p < 0.001$) and Trial ($F_{15,315} = 147.737$, $MS_{error} = 0.0039$, $p < 0.001$). The degrees of freedom have been manually reduced to reflect the imputed values approximations.

Although both of these results show similar outcomes, analysis of the data indicates that they fail the homogeneity of variance constraint (Levene's test) and this is evident by inspection of **Figure** 6 where it can be seen that the standard deviation decreases while performance increases with repetition, a commonly observed phenomenon in learning (Ritter & Schooler, 2002). A sine-logarithm transformation of the data considerably improved the normality and homogeneity of variance, although the data remained somewhat skewed due to the performance ceiling effect. Analysis of the transformed data showed an identical pattern of results to the previous two analyses, with both main effects (Group: $F_{2,27} = 41.714$, $MS_{error} = 0.1083$, $p < 0.001$; Trial ($F_{15,405} = 92.794$, $MS_{error} = 0.0131$, $p < 0.001$) and the interaction ($F_{30,405} = 6.426$, $MS_{error} = 0.0131$, $p < 0.001$) being significant.

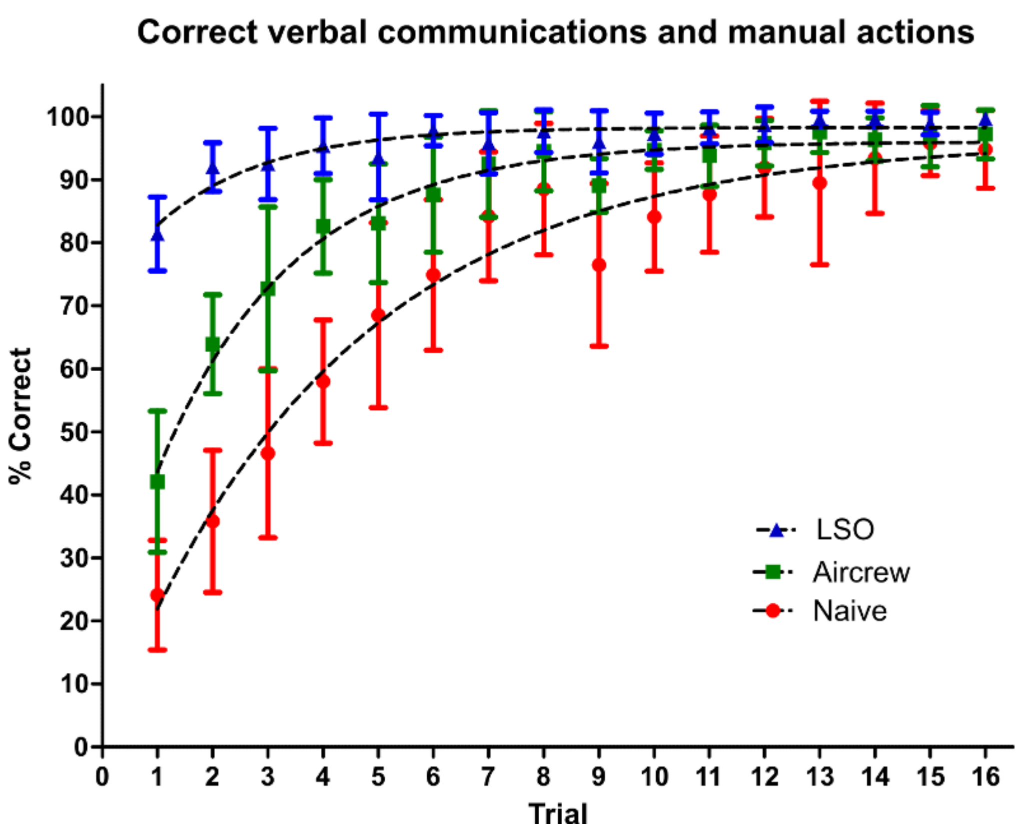


Figure 6. Performance as measured by the combined correct verbal communications and manual actions expressed as a percentage for each group over the 16 trials. No imputed values are

included for lost data in this figure. Data points are means and standard deviations. N=10 (nominally).

Curvefiting of Percent Correct data

The performance data were further analyzed by fitting a nonlinear curve (using GraphPad Prism 5, http://www.graphpad.com) and assessing various products of the curve-fitting procedure. Two sets of curves were initially considered based on inspection of the raw data and on forms commonly reported in the literature: a power series relationship (Anderson, 2001) and an exponential relationship (Heathcote, Brown & Mewhort, 2000). There has been some debate over the precise form that learning, forgetting and performance improvement curves should take (Anderson, 2001; Haider & Grensch, 2002; Newell, Mayer-Kress & Liu, 2006), although the arguments seem based on the results of regression rather than stemming from a theoretical basis.

The power series function generally reported in the literature is:

$$\textbf{Percent Correct} = A*(\textbf{Trial} + \mathit{Experience})^{-B} + \text{Plateau} \tag{1}$$

where A and B are coefficients that are optimized to fit the data. **PercentCorrect** is the fraction of correct manual actions and verbal communications in each trial corresponding to the **Trial** variable. The parameter Plateau reflects the expected level of performance once all learning has been completed; Plateau was constrained to be a constant value of 100 in expectation that performance would eventually show no error in an ideal model. The coefficient A reflects the initial level of performance coming into the task (**Trial** = 0). The coefficient B (often referred to as the learning rate) reflects the rate of change of performance with practice (i.e. learning the task).

The *Experience* coefficient is added to the **Trial** parameter to accommodate prior knowledge. Unfortunately, we do not have a reliable estimate of what the *Experience* value should be, only some crude estimates. The curvefitting procedure proved to be very sensitive to the *Experience* parameter and even small values (less than approximately 5) produced unrealistic regression coefficients; when left as a free parameter determined by the regression process, unrealistic and counter intuitive values resulted, a phenomenon noted by Boff and Lincoln (1988, Vol.II, Section 4.201). For subsequent analyses, the *Experience* confident was set to zero, resulting in the power law equation used in this analysis to:

$$\textbf{Percent Correct} = A*(\textbf{Trial})^{-B} + \text{Plateau} \tag{2}$$

The exponential function that was used to fit the data is:

$$\textbf{Percent Correct} = (\text{Plateau} - Y0) * \mathrm{e}^{-\textbf{Trial}/\tau} + \text{Plateau} \tag{3}$$

where the curve fit coefficients are $Y0$ and τ, while **PercentCorrect** and **Trial** remain the same. As in the power series, Plateau is a constant, constrained to 100, predicting perfect performance as practice increases. $Y0$ is similar to *Experience* in the power series relationship and initial estimates for $Y0$ were established similarly to the *Experience* initial estimates. In representations of the exponential function, some authors prefer to incorporate the Plateau-$Y0$ difference as a single coefficient, similar to A, representing the maximum performance improvement that can be achieved; the exponential function automatically accommodates pre-experiment knowledge in its representation. The coefficient τ represents the rate of performance improvement due to learning the task through practice, or the inverse of the learning rate λ (equation 4), that controls the nonlinearity of the exponential learning curve (similar to the coefficient B in the power function learning curve) which is assumed to be proportional to the amount left to be learned:

$$\partial \mathbf{PercentCorrect} / \partial \mathbf{Trial} = -\lambda \mathbf{PercentCorrect} \tag{4}$$

When each individual subject's data were fit with these curves, the exponential function was found to be superior to the power law in 21 of the 30 cases, based on the Akaike Information Criterion (AIC: Akaike, 1974, 1981). As mentioned, in most instances the power function would only converge when the *Experience* coefficient was set between 0 and 5, so the *Experience* coefficient was set to zero in these Power Function regression results. The range of fits for the exponential function was quite wide, with regression coefficient (R^2) values ranging from 0.12 to 0.98 with a median R^2 value of 0.88 (interquartile range: 0.15). The power function results with *Experience* set to 0 were similar, with R^2 varying from 0.21 to 0.96 and a median value of 0.81 (interquartile range: 0.16).

The AIC preference of the exponential function over the power function did depend on the group, with the strongest preference in the Naïve group followed by the Aircrew and LSO groups respectively, as shown in **Table 3**. This suggests that the exponential function is a better representation of the observed learning data than the power law when the changes are more extreme, such as when first learning a task, than in the latter, refinement stages where incremental learning is much smaller.

Table 3. Summary of the preferred regression model based on the Akaike Information Criterion (AIC) for all subjects combined and broken out by group.

	Preferred Model			
	Overall	Naïve	Aircrew	LSO
Power	9	0	3	6
Exponential	21	10	7	4
Average AIC	4.57	11.65	1.95	0.11
Standard deviation	4.57	11.65	1.95	0.11

A one-way analysis (n = 10/group) of the resulting exponential function regression coefficients indicated a main effect of Group ($F_{2,27} = 32.2$, $MS_{error} = 395.1$, $p < 0.001$) for the (Plateau - *Y0*) coefficient shown in **Figure 7**. As expected, the qualified LSO subjects have the least to learn while the Naïve subjects have the most to learn (effectively everything). Note that these data indicate that the results are skewed, at least for the Naïve subjects, suggesting that some subjects were able to remember some of the procedures from the initial exposure (reading the scenario and watching the video once) although this initial, single exposure does not appear to provide a substantial level of training.

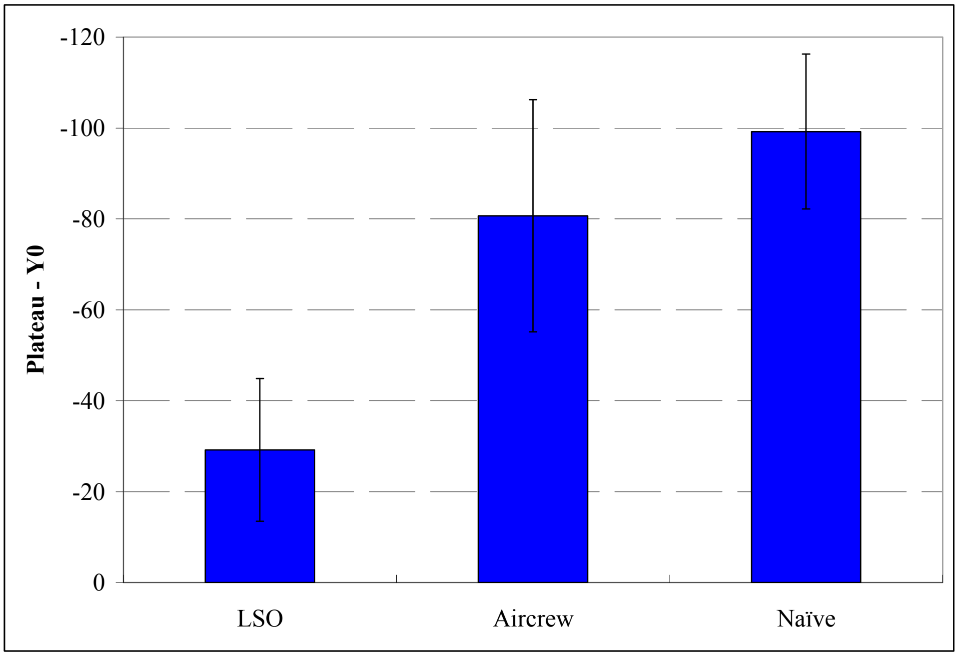


Figure 7. Regression values for the exponential function (Plateau – Y0*) coefficient representing the amount of information to be learned to achieve perfect performance. The Plateau was fixed at 100. Data are group means and standard deviations.*

The exponential function learning coefficient, τ was not statistically different between groups ($F_{2,27} = 1.976$, $MS_{error} = 5.0$, $p > 0.15$), suggesting that each group learned at approximately the same rate; that is, the amount of prior knowledge affected the time to reach criterion, but it did not appear to affect the rate at which performance improved.

The means for the exponential learning coefficient τ (equation 3) shown in **Figure 8** do suggest, however, that the LSO and Aircrew groups, which had similar values, did improve somewhat faster than the Naive group, possibly indicating an effect of familiarity with the environment or prior exposure to the task. A power calculation[2] indicated a low level of power to detect the observed effect sizes with only 10 subjects per group (power was approximately 0.15 to 0.35) and that approximately 30 subjects per group would be required to achieve a more desirable power level of 0.8.

An analysis of the power function regression coefficients provided similar conclusions (**Figure** 9): the initial rank ordering of the amount to be learned (coefficient *A*) increased from LSO to Aircrew to Naïve subjects ($F_{2,27} = 148.1$, $MS_{error} = 80.45$, $p < 0.001$); there was no difference in learning rates (coefficient *B*) among groups ($F_{2,27} = 1.967$, $MS_{error} = 0.097$, $p > 0.15$).

---

[2] Values calculated from http://euclid.psych.yorku.ca/cgi/power.pl . Not to be confused with the power function used in the regression analysis.

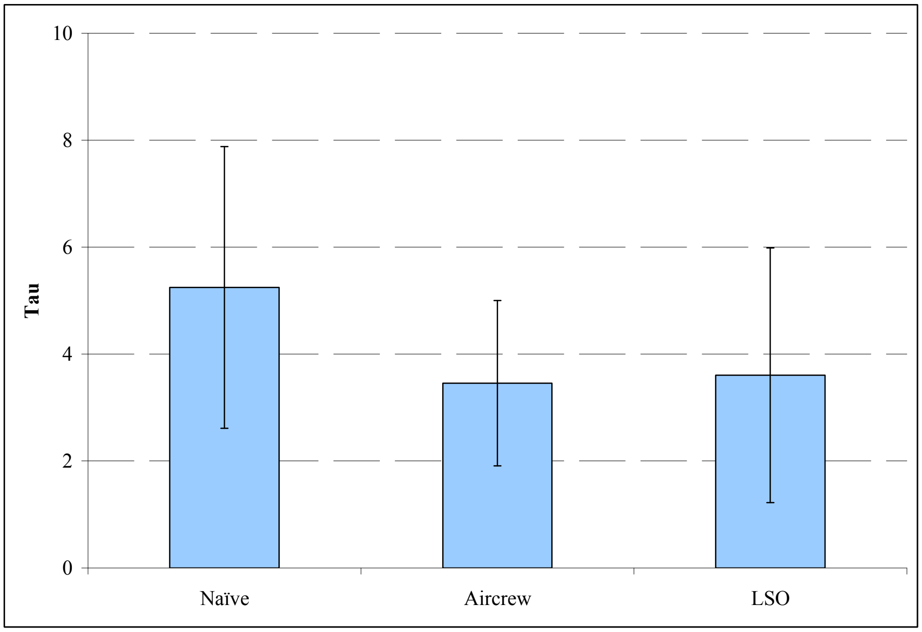


Figure 8. Regression values for the exponential function learning time constant τ *(*Tau*; group means and standard deviations.)*

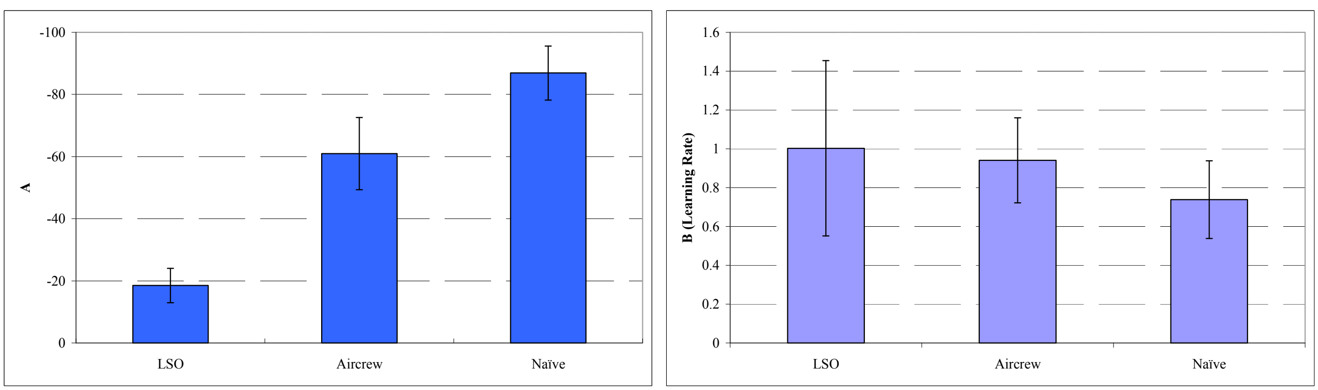
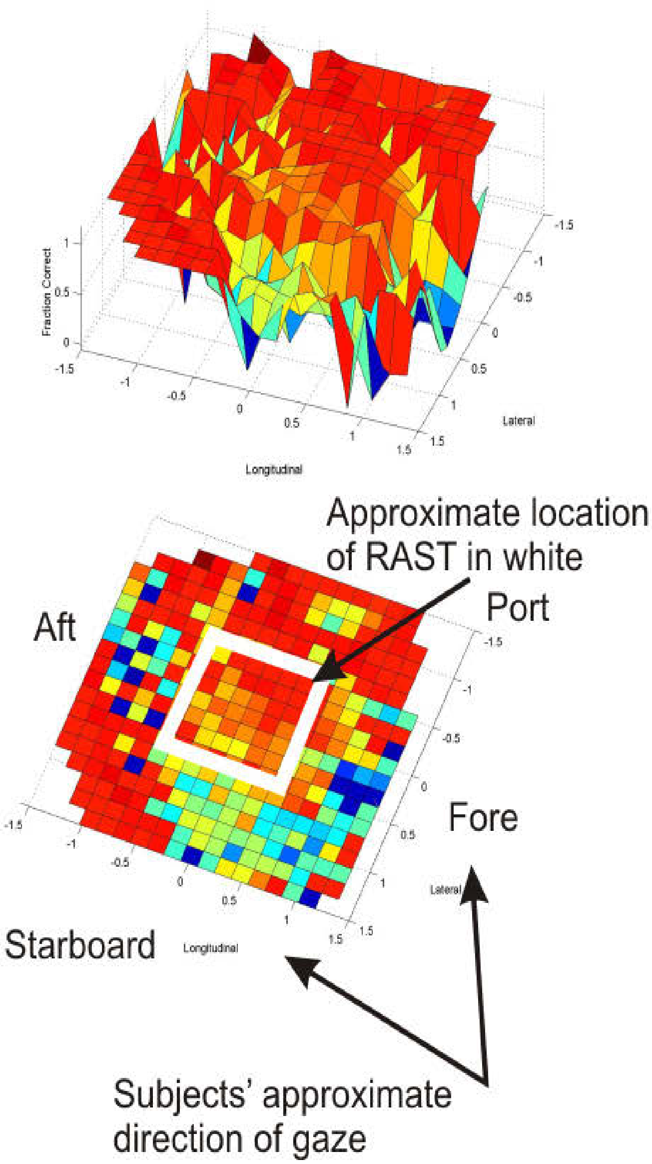


Figure 9. Regression coefficients for the simplified power function of learning (no Experience coefficient). Values are group means and standard deviations.

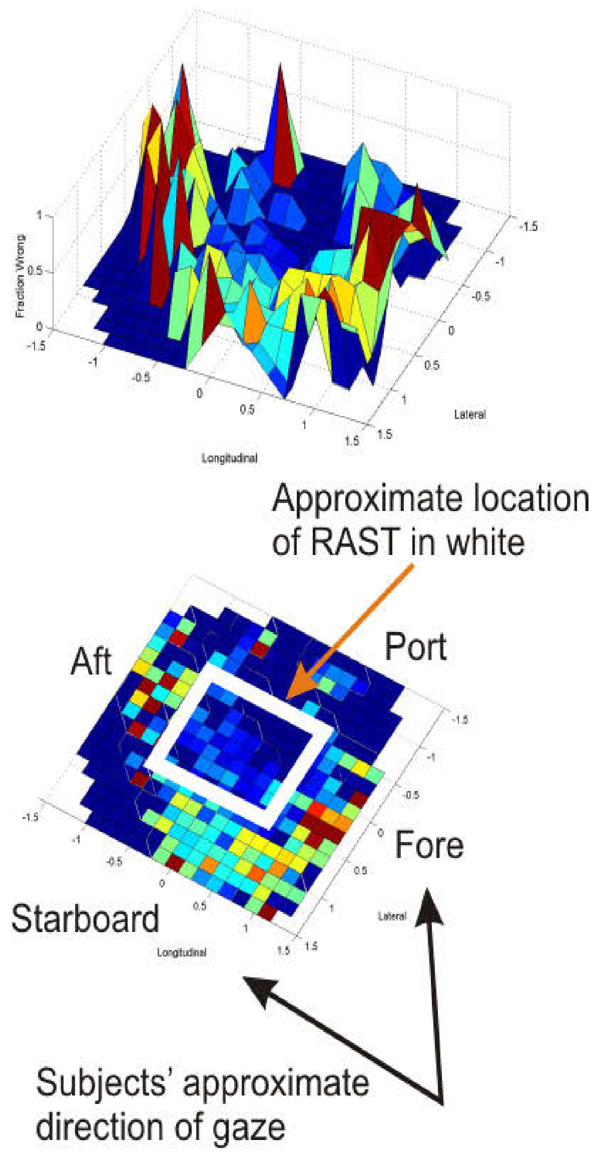
Visual judgements

The conning component of the LSO task required that a visual judgement be made of the helicopter probe relative to the ship's trap. The conning calls instructing the pilot to move appropriately when in the low hover were coded and expressed as fractions of percent correct and percent wrong. Calculations of movement-conning in the horizontal plane were made by comparing the probe position to the middle of the trap; "Landing now!", "Wave off!" and "In the trap." calls were evaluated by calculating whether the probe was within the boundaries of the trap.

The results were plotted in two dimensional histograms **(Figure 10a and 10b)** using Matlab. The trap area is shown in the plan-view in each instance. The values represent all subjects as the number of conning calls within any single group was insufficient to adequately map the distribution of events.



(a)



(b)

Figure 10. Two dimensional histogram plots of the fraction of (a) correct and (b) incorrect verbal conning calls of the helicopter probe over the ship's trap from all subjects. Perspective is approximately that of the LSO's view of the trap from the Howdah.

Landings

The subjects' conning instructions that resulted in successful landings on the first attempt and the number of landing attempts per trial were evaluated as another metric of the visual judgement validity of the simulation. There was no significant difference between groups or by trial: subjects were successful in landing the helicopter on the first attempt in approximately half of the trials, regardless of prior experience.

Analysis of the number of landings per trial with missing values (case-wise deletion) did not detect any significant effects, however, when imputed values were assumed for missing data, there was a significant main effect of Trial ($F_{15,315}$ = 1.72, MS_{error} = 0.84, $p < 0.05$) while the Group factor just failed to reach significance ($F_{2,27}$ = 3.297, MS_{error} = 2.167, $p = 0.052$); there was no significant interaction. Observed data without imputation are shown by group in **Figure 11** and by trial in **Figure 12**.

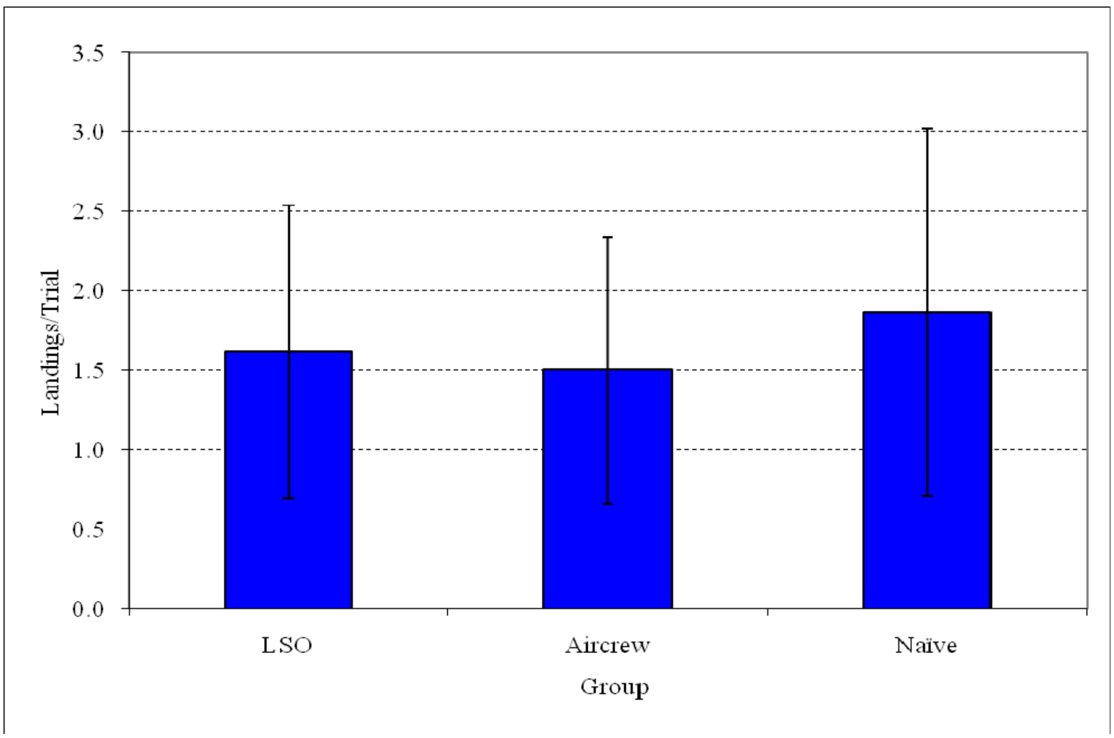


Figure 11. Average (standard deviation) landing attempts by group.

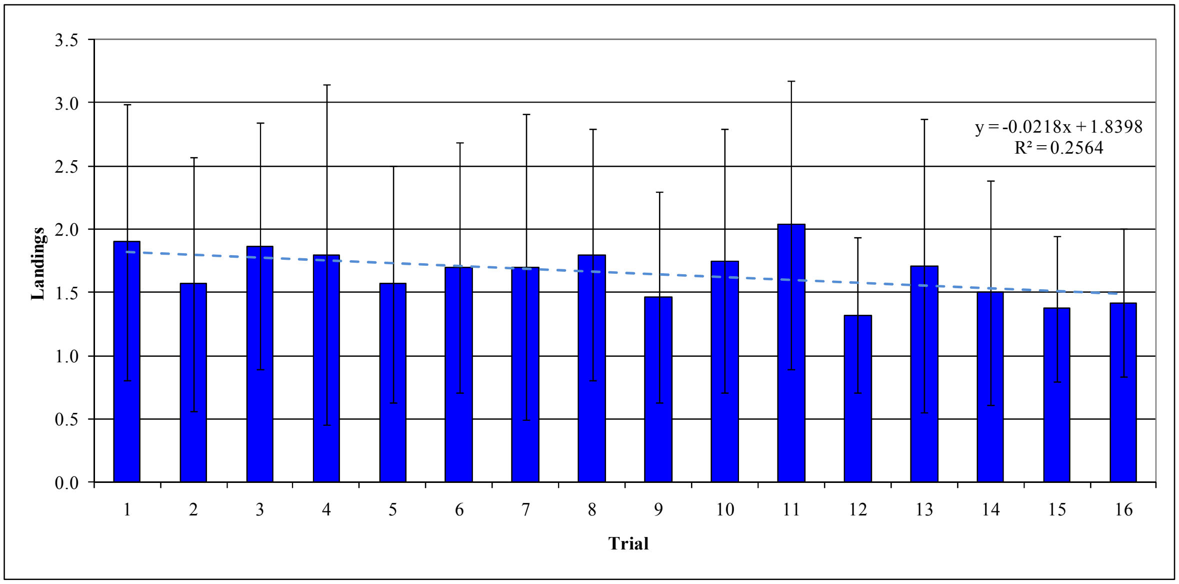


Figure 12. Average (standard deviation) number of landings by trial. The dashed line, shown with the regression line, suggesting a slight improvement with practice, although the effect was not significant.

Workload

NASA TLX subjective workload ratings were recorded at the end of each block of 4 trials; paired comparisons of the six NASA TLX factors (Mental Demand, Physical Demand, Temporal Demand, Own Performance, Frustration and Effort) were completed after the final block of trials. Overall Workload and the individual factors are analyzed in a 3 level between by 4 level within (Group by Block) repeated measures ANOVA.

There were some lost questionnaire data (NASA TLX and Simulator Sickness Questionnaire) due to subjects reaching criterion. All subjects completed the questionnaires for the first two blocks (n = 10/group). In the third block, eight LSOs completed the questionnaires (n_3 = 8 LSOs); all Aircrew and Naive completed the questionnaires (n_3 = 10/group). In the fourth block, 5 LSOs (n_4 = 5 LSOs), nine Aircrew (n_4 = 9 Aircrew) and all Naïve (n_4 = 10 Naïve) completed the questionnaires.

Overall Workload

The overall workload was assessed two ways. First, a simple sum of the unweighted ratings was analyzed based on an observation that weighting the ratings failed to improve the sensitivity of the NASA TLX technique beyond that achievable with an unweighted sum of the six factors (Hendy, Hamilton & Landry, 1993, p. 596). Then, an overall score derived from the weighted ratings was assessed as per the original authors' report (Hart & Staveland, 1988).

Analysis of the sum of the unweighted ratings (shown in

Figure 13) indicated a significant main effect for Block ($F_{3,63} = 6.72$, $MS_{error} = 71.7$, $p < 0.001$) with a significant interaction between Block and Group ($F_{6,63} = 2.257$, $MS_{error} = 71.7$, $p < 0.05$); the Group factor approached, but did not reach statistical significance ($p \sim 0.08$). From

Figure 13**,** it can be seen that the LSO group's perceived workload did not vary appreciably across blocks, while both the Aircrew and Naïve groups' perceived workload decreased. The Naïve group's perceived workload decreased the most and was largely undistinguishable from the other groups by Block 4.

Analysis of the sum of weighted factor ratings indicated a somewhat different outcome from the unweighted scores. A significant main effect remained for the Block factor ($F_{3,60} = 17.34$, $MS_{error} = 2.004$, $p < 0.001$), still moderated by a significant interaction between the Block and Group factors ($F_{6,60} = 8.157$, $MS_{error} = 2.004$, $p < 0.001$) as shown in **Figure 14,** but now there was also a main effect of Group ($F_{2,20} = 14.0$, $MS_{error} = 9.327$, $p < 0.001$). The LSO and Aircrew scaled workload ratings are more similar and vary somewhat less over the blocks. The Naïve group, however, appears distinctly different from the other two groups, decreasing significantly by block until it is (again) indistinguishable from the other two groups.

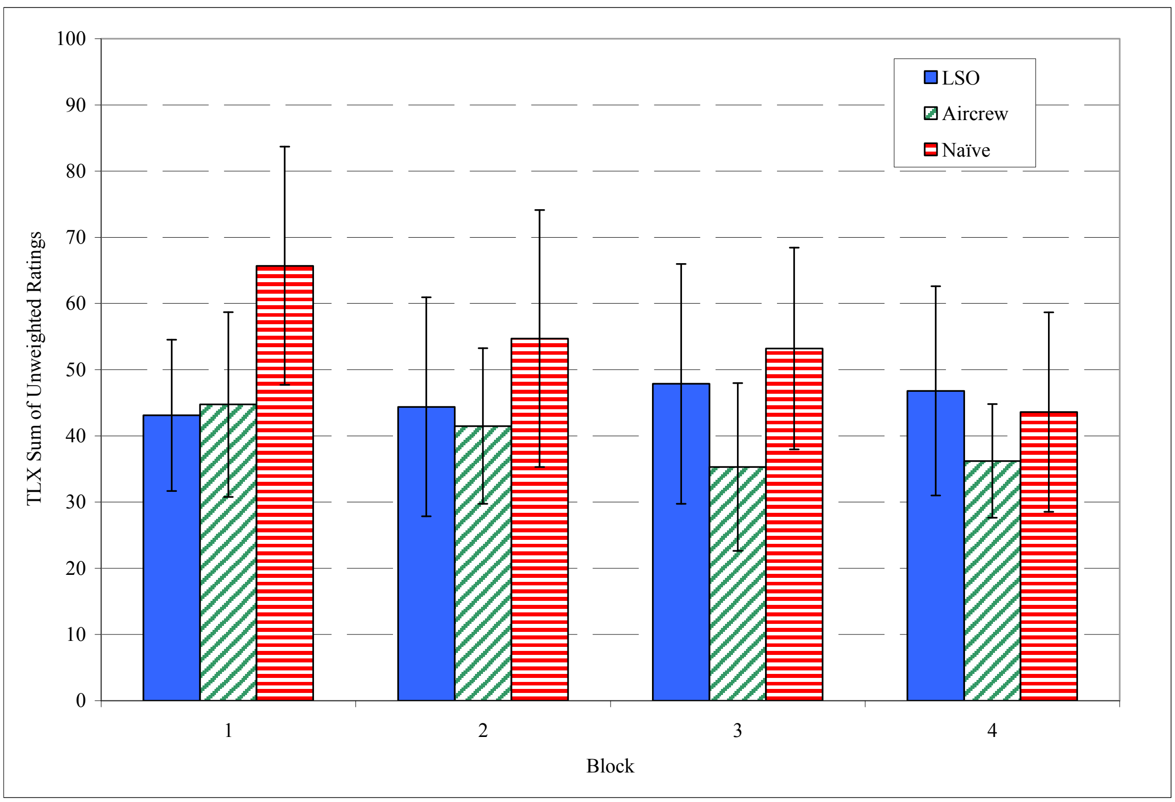


Figure 13. Unweighted sum of NASA TLX factors by Group and by Block. Values are average (standard deviations) without imputation.

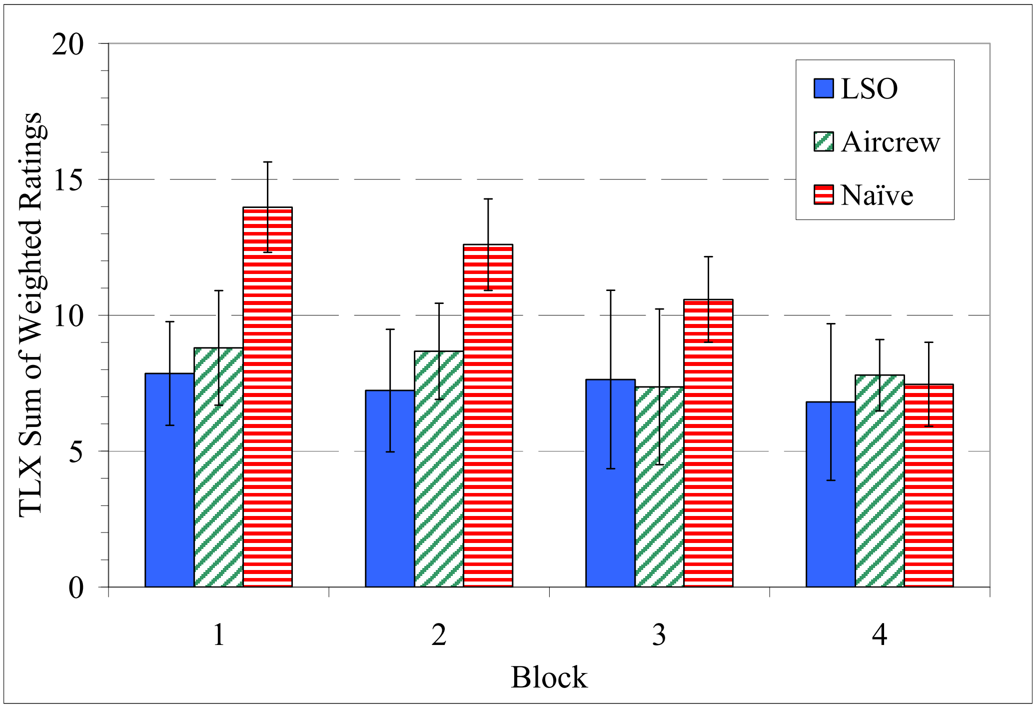


Figure 14. Average (standard deviation) values of the sum of weighted NASA TLX factors.

Further investigation of the ANOVA tables indicated that there was a substantial advantage to using the paired comparisons to weight the factor ratings rather than just using the raw scores in the overall workload calculation, more than doubling the amount of variance explained by the NASA TLX model as indicated in the regression coefficient, R^2 shown in **Table 4**. This result supports the observations and recommendations of Hart and Staveland (1988) when calculating the overall NASA TLX workload rating.

Table 4. Evaluation of the use of weighted factors in explaining variance in the NASA TLX ratings. SS in the ANOVA table is the sum of squares.

	Unweighted Sum	Weighted Sum
SS Effect	6694.5	463.5
SS Error	20062.2	306.8
SS Total	26756.7	770.3
R^2	0.25	0.60

Mental Demand

Analysis of the Mental Demand ratings indicated main effects of both Group ($F_{2,21} = 5.51$, $MS_{error} = 61.866$, $p = 0.012$) and Block ($F_{3,63} = 0.62$, $MS_{error} = 6.46$, $p < 0.001$). There was no significant interaction. As shown in **Figure 15**, Mental Demand was perceived to decrease with exposure, presumably due to increased familiarity both with the synthetic environment and with the task elements. Both the LSO and the Aircrew groups perceived the mental demands of the task to be lower than did the Naïve group, possibly due to their familiarity with the environment and the pattern of radio communications or simply their level of experience dealing with complex tasks on a daily basis. Generally, however, the rated Mental Demand for the task was low and it does not appear that the subjects considered the task overly challenging, despite the observed difficulties many had in correctly completing the verbal syntax and manual actions associated with the task's communications.

There was a moderately large, negative correlation between the average performance (as measured by the percent correct verbal commands and manual actions) by block and the Mental Demand rating for each group, although the magnitudes differed somewhat. The Naïve group had the greatest correlation between Mental Demand and percent correct (- 0.42) and the LSO group had the smallest correlation (- 0.26), with the Aircrew group intermediate (- 0.39), but more similar to the Naïve group than to the LSO group.

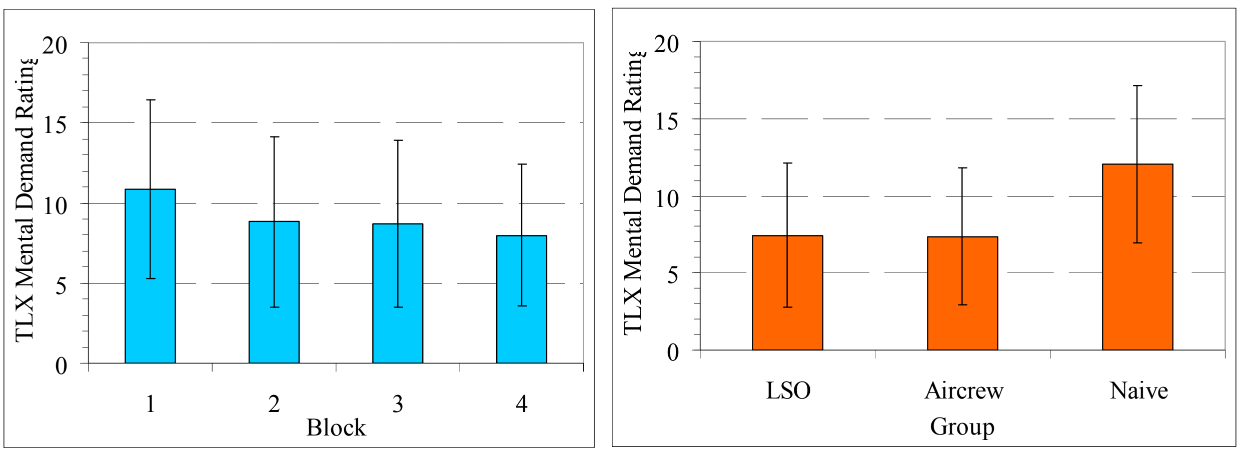


Figure 15. Average (standard deviation) values of the Mental Demand ratings by Block and Group.

Physical Demand

The Physical Demand ratings were not found to vary significantly across any of the experimental factors and were small in magnitude. The overall mean Physical Demand rating was 4.4, with a standard deviation of 3.6. There was a low, negative correlation between the Physical Demand rating and the Percent Correct scores for the Naïve (- 0.22) and the Aircrew (- 0.1) groups, but the LSO group had a negligible correlation (- 0.02), suggesting that the physical actions themselves had very little to do with the perceived workload or task demands, but what demand there was decreased with exposure.

Temporal Demand

The Temporal Demand ratings indicated a significant interaction between Block and Group ($F_{6,63}$ = 2.502, MS_{error} = 3.99, p = 0.03). Neither the Block nor the Group factors showed a significant main effect, although the Group factor approached significance (p = 0.07), presumably due to the initially high Temporal Demand ratings of the Naïve group relative to the other groups. The interaction shows that the Naïve group's perception of Temporal Demand decreased with exposure while the LSO group increased somewhat; the Aircrew's perception changed only slightly with exposure. Simple paired t-tests for the Naïve and LSO groups suggest that the change in the Naïve group's perception of Temporal Demand was significant, while the LSO group's perception of the change was not significant.

The groups seemed to be converging on a Temporal Demand rating of 3 to 5 (out of 20), which suggests that the task did not impose a substantial perception of time pressure. Even the initial Naïve group rating (approximately 8) was substantially below the maximum rating (20). The variation within the Naïve group, as indicated by the standard deviation in **Figure 16**, seemed to decrease with exposure and all groups had similar variability by the end of the trial.

The Naïve and Aircrew groups had low, negative correlations between the Temporal Demand ratings and the Percent Correct performance metric (- 0.33 and - 0.24 respectively), while the LSO group had a negligible, positive correlation (0.02), further suggesting that time pressure was not a substantial demand in the task, but the little time pressure that was perceived decreased with exposure.

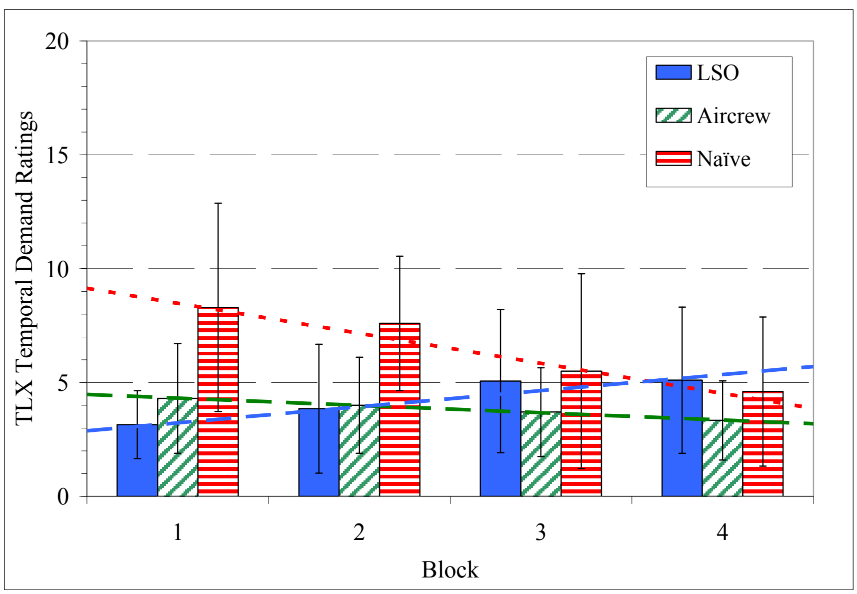


Figure 16. Mean (standard deviation) values of the perceived temporal demand showing the interaction between Group and Block factors (dashed lines).

Own Performance

The Own Performance ratings indicated a main effect of Block only ($F_{3,63}$ = 9.06, MS_{error} = 11.07, p < 0.001). **Figure 17** suggests that the subjects recognized a general trend of improved performance with exposure, but subjects also perceived that performance was poorer after the break between the two sessions (3 hour interval between Blocks 2 and 3).

Correlation with the Percent Correct scores by block indicated a moderate, positive correlation for each group (LSO: 0.31; Aircrew: 0.29; Naïve: 0.48) suggesting that the subjects were aware that their performance was improving with repetition.

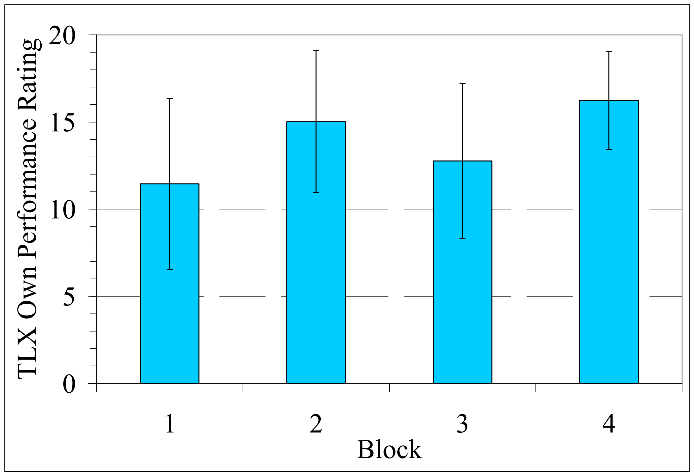


Figure 17. Average (standard deviation) values of the Own Performance ratings by Block.

Effort

The subjects indicated that the level of effort applied to the task decreased with practice ($F_{3, 63}$ = 5.34, MS_{error} = 6.1, p = 0.002), perhaps indicating an improving proficiency on the task as well as accommodation to the simulation. The Naïve group effort rating was approximately twice that of the Aircrew and LSO groups (**Figure 18**), probably reflecting their lack of familiarity with the domain as well as the simulation ($F_{2, 21}$ = 6.28, MS_{error} = 302.6, p = 0.007). The magnitudes of the ratings were low, in most cases less than 50 % of full scale.

The interaction was not significant (p > 0.09), however, it is suggestive of a differential effect; inspection of the trend lines for each group indicated that the rate of decrease of effort with block was greater for both the Aircrew and Naïve groups, while the LSO group effort ratings changed little over the blocks, likely reflecting the LSOs' familiarity with the task. This is interesting because, if the simulator had substantial differences from the real application (as far as training is concerned), one might expect the LSO-group effort to be initially high, then improve with exposure as the LSOs adapt to the simulator; one might also expect that the other two groups would show small changes in effort if those subjects had to both learn the task as well as struggle with a poor simulation. While this phenomenon did occur for the LSOs, its effect was weak; conversely, the other two groups showed marked reduction in effort with exposure, all suggesting that the simulation was suitable for learning the task, in support of the error data reported earlier. Moderate, negative correlations were observed between Effort ratings and Percent Correct scores for each group (LSO: - 0.30; Aircrew: - 0.34; Naïve: - 0.41).

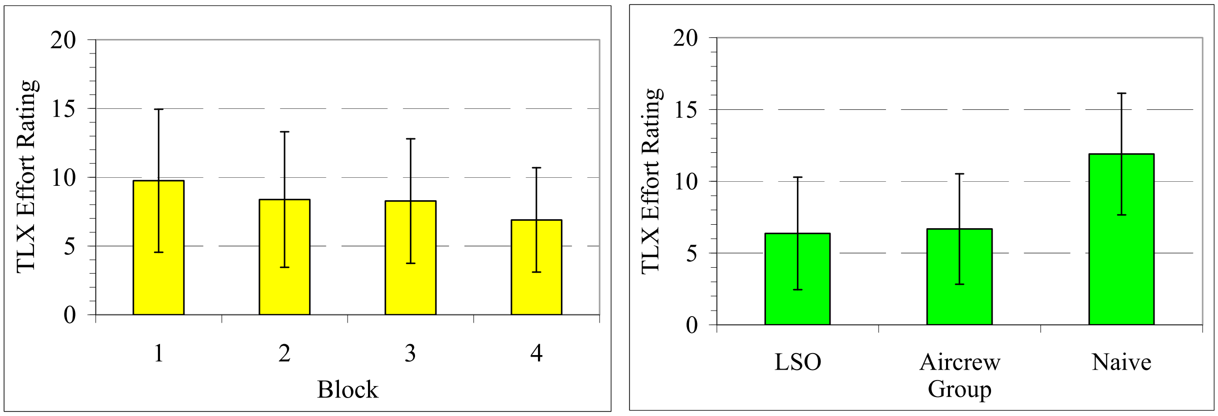


Figure 18. Average (standard deviation) ratings of the NASA TLX Effort factor by Block and Group.

Frustration

There was only a significant main effect of Group on the Frustration rating ($F_{2,21} = 6.48$, $MS_{error} = 46.6$, $p = 0.006$) that was due to the difference between the Naïve and Aircrew groups; neither the Aircrew and LSO groups nor the Naïve and LSO groups were statistically different. The difference is possibly due to the Naïve groups' lack of familiarity with the domain and the attention required to use specific syntax during communications. Nevertheless, the Frustration ratings are low as indicated in **Figure 19**.

Correlation between the Frustration ratings and the Percent Correct scores indicated a moderate, negative correlation for the Naïve (- 0.22) and Aircrew (- 0.40) groups, suggesting that these groups became more comfortable with the simulation and the task with repeated exposure. The LSO group had a negligible correlation between Frustration and Percent Correct (0.04); as LSO performance varied little by block, the lack of correlation with frustration can be attributed to random variation that is consistent with the observation that the LSOs were able to accommodate readily to the task in the simulator.

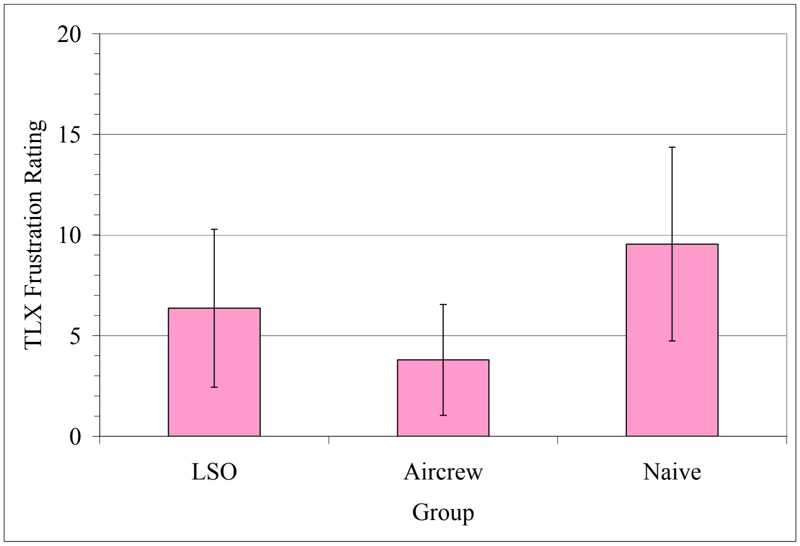


Figure 19. Average (standard deviation) ratings of the NASA TLX Frustration factor by Group.

Simulator sickness

Subjects completed the Simulator Sickness Questionnaire (SSQ: Kennedy, Lane, Berbaum & Lilienthal, 1993) at the end of each session. Half of the subjects in each group completed a pre-exposure questionnaire and all subjects completed the SSQ at the end of the first session (8 trials). Subjects then completed the SSQ at the beginning and end of the second session. These data were analyzed as a 2 (Conditioning: pre-exposure measurement) x 3 (Group) x 2 (Session: post Sessions 1 and 2) Repeated Measures ANOVA.

Two LSO subjects reached criterion in the first session, so they did not participate in the second session, resulting in 2 sets with lost data; three additional LSO subjects and one Aircrew subject reached criterion in the first block of the second session, so they only completed 4 trials in the second session before completing the SSQ. This meant that the amount of time spent in the simulator during the second session varied as subjects reached criterion, which likely affected the ratings for these subjects, possibly reducing the severity of any symptoms experienced in Session 2.

One other LSO subject recorded a noticeably higher SSQ score, but only at the end of the first session. This skewed the group results substantially as it was greater than 5 standard deviations from the mean (considering only the other group members; the score was 3 standard deviations greater when the subject's score was included in the total). The subject's score was comparable to the group's score at the second and third recordings, so this subject's SSQ score was treated as an outlier and the entire data record removed from the analysis.

Analysis of the composite, Total SSQ scores at the end of each session indicated main effects of Group ($F_{2,21} = 6.097$, $MS_{error} = 769.4$, $p = 0.008$) and Session ($F_{1,21} = 4.57$, $MS_{error} = 70.24$, $p = 0.04$); there was no significant interaction. As can be seen in **Figure 20**, the Naïve group score was significantly higher than both the Aircrew and LSO groups, and maintained a similar magnitude across sessions. The LSO group appears to have a lower Total SSQ score in Session 2 but there is considerable uncertainty in the Session 1 measurement and this difference is not significant. The Aircrew and LSO Total SSQ scores are not significantly different.

The Total SSQ score for Session 2 was smaller than that of Session 1, but this can be attributed to the fewer number of trials that subjects spent in the simulator. This is particularly true for the LSO group, as only half of the LSOs completed all 16 trials, the others having reached criterion by the end of the third block (12 trials). This explanation does not explain the suggested decrease for the Aircrew group, however, a paired t-test on the Aircrew post-session Total SSQ scores was not significant.

As there was evidence that the data were positively skewed, the Total SSQ scores from the end of the two sessions were reanalyzed after a square-root transformation without considering the Conditioning factor. The pattern of results was the same as the analysis for the untransformed data, showing significant main effects of Group ($F_{2,24} = 7.45$, $MS_{error} = 3.08$, $p = 0.003$) and Session ($F_{1,24} = 7.8$, $MS_{error} = 1.31$, $p = 0.01$), but no interaction.

The individual dimensions of the SSQ were subsequently analyzed as independent 3x2 (Group x Session) repeated measures ANOVA, considering only the subjective ratings at the end of each of the two sessions. The ratings within each dimension were transformed using a square-root transformation to reduce skewness and make the variance more uniform across groups.

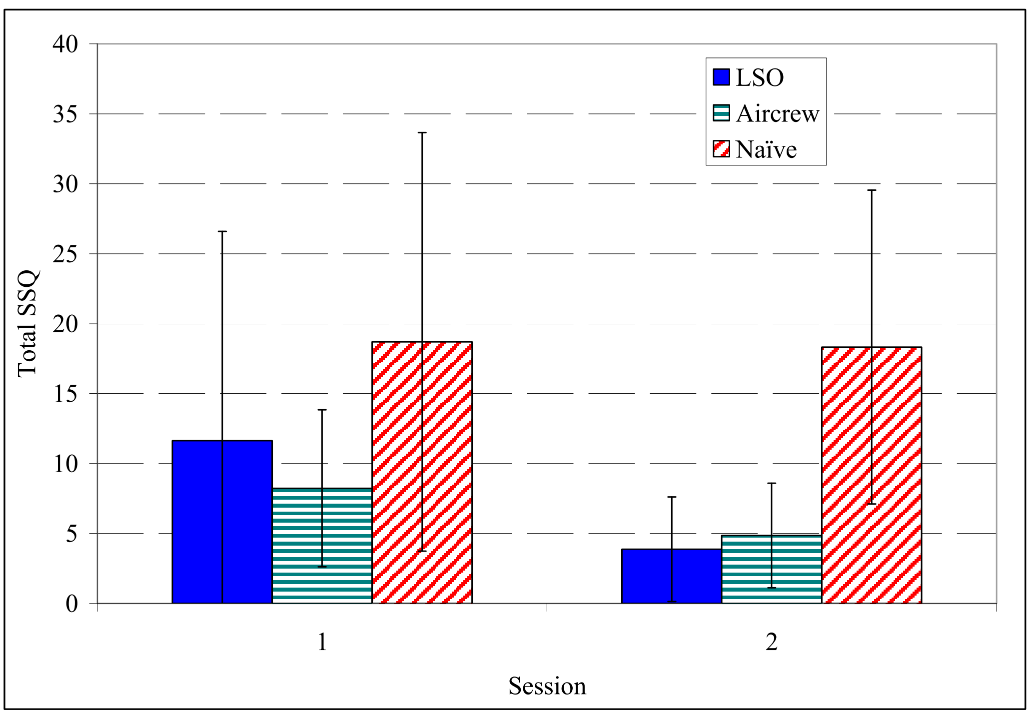


Figure 20. Average (standard deviation) of the Total SSQ scores at the end of each of the two Sessions decomposed by Group.

The Nausea dimension indicated that there was a main effect of Group ($F_{2,24}$ = 8.42, MS_{error} = 3.36, p = 0.002) where the Naïve group seemed more sensitive to the simulation (11.4 ± 9; actual mean and standard deviation) compared with the Aircrew (3.8 ± 6) and the LSO groups (1.8 ± 4), which were not significantly different. There was also a significant main effect of Session ($F_{1,24}$ = 4.49, MS_{error} = 1.26, p = 0.04), with the Nausea dimension in Session 2 having a somewhat lower rating (5.3 ± 8) than Session 1 (6.6 ± 7). While it is possible that there may have been some accommodation to the simulator, it is more likely that Session 2 had a lower score due to the lower number of trials completed (208 in Session 2 versus 280 in Session 1) resulting in lower ratings from some subjects than might be expected if they had to complete all 8 trials. This effect would be expected within each of the dimensions as well as the total SSQ score.

The square-root transformation of the SSQ Oculomotor dimension showed a similar pattern with significant main effects of Group ($F_{2,24}$ = 5.01, MS_{error} = 5.18, p = 0.002) and Session ($F_{1,24}$ = 4.68, MS_{error} = 1.75, p = 0.04). Tukey Honest Significant Difference (HSD) post hoc analysis of the Group effect indicated that the Naïve group Oculomotor rating (22.0 ± 17) was greater the Aircrew rating (8.0 ± 8) but not the LSO rating (12.8 ± 11). Analysis of the Disorientation dimension did not uncover any significant differences, although the Session factor approached significance (p ~ 0.06), consistent with the differences in the number of trials per session discussed above.

Discussion

Procedural learning effectiveness

All groups improved their performance (as measured by the Percentage of Correct verbal commands and manual actions) with repeated exposure. The rank ordering of the groups is consistent with the hypotheses, with the Naïve group initially having the most to improve and the LSO group having the least. No single analysis was perfect because of limitations in the data set, but all approaches indicate similar conclusions: the simulations of the environment and the virtual crew were sufficient to learn the procedural elements of the LSO's role in a helicopter-deck landing task, exchanging verbal information and making manual control actions in response to both visual stimuli and auditory communications from the simulations. While transformations to address violations of the ANOVA assumptions of normality and homogeneity of variance improved the data distribution, they did not change any of the conclusions reached from analyzing the original data.

The rapid adaptation of the LSO group indicated that the simulation did not overly constrain accommodation to the simulation or (re)learning of the task. Several LSO subjects, while having substantial experience as LSOs, were not current, having not been to sea for several months or years in some cases. Although this was not ideal from an experimental perspective, it was an operational constraint imposed on the availability of expert LSO subjects that emphasizes the operational need for alternative, shore-based training methods to develop and maintain capability. Some of the variance included in the LSO group can also be attributed to the measurement technique, where a stricter adherence to standard operating procedure syntax was enforced than is typically adopted in practice.

The high level, asymptotic Percent Correct performance realized by all groups is also consistent with the hypotheses for an effective training simulator. The substantial elimination of the initial differences indicates that the simulator presented no barrier to learning, at least within the scope of the experimental task, and that even Naïve subjects became proficient in the procedural aspects of the LSO's role during free-deck landings within a short time frame.

Evaluation of the nonlinear regression parameters for the learning curves suggested a modest advantage when using an exponential function rather than a power function for describing the rate of improvement with practice. This is at odds with the more common assumption of a power law relationship, although it is consistent with a growing segment of the learning literature. Analysis of the data through the regression parameters avoids a thorny issue of unequal number of observations in repeated measures ANOVA when using the raw data obtained with a performance criterion-based cut-off. The lack of a sound theoretical basis for choosing the form of the regression equation does, however, complicate the analysis somewhat. Perhaps more disquieting was the tendency of the power function to fail to converge to the individual subject data in some instances, although both relationships fit the group data adequately, consistent with other observations in the literature (Anderson, 2001; Haider & Grensch, 2002; Heathcote et al., 2000; Suzuki & Ohnishi, 2007). Nevertheless, the analysis of the regression parameters obtained from both the exponential function and the power function supported the observations obtained from the analysis of the raw data.

The overall NASA TLX workload metric indicated that the LSO group did not perceive a significant change in demand with exposure, supporting the hypothesis that an adequate simulation should not require substantial adaptation by experts in the task. Meanwhile, both the two non-expert groups showed a slight yet significant decrease in workload with exposure. Although we cannot attribute this solely to learning the task without any simulator adaptation, it is consistent with the hypothesis that the simulator is a valid training device for this task.

Correlations between each of the NASA TLX demand ratings and Percent Correct scores were also consistent with the hypothesis that the simulator was suitable for training the procedural aspects of the LSO deck landing task and that the subjects readily accommodated to the simulation. While this observation does not validate any single element of the simulation, it does provide a holistic assessment of the ensemble that indicates it is valid as a training tool, a conclusion that would be difficult to substantiate if any key element was inadequate.

While there was evidence of discomfort induced by the simulator, the level was generally low as measured by the SSQ, particularly for the LSO and Aircrew groups. It seems plausible that the experience in flight operations of these groups may make them more tolerant or less susceptible to simulator sickness, however, the literature does suggest that the two phenomena are distinct. The low level of simulator sickness reported by the subjects suggests that the simulation is not overly provocative and that useful training in the simulator may be readily achieved by managing exposure.

Landing and visual judgements

The analysis of the landings and the conning data suggest that the visual presentation of the helicopter over the flight deck in the simulation is insufficient to learn the visual discrimination aspects of the LSO's role in the landing task. Anecdotal evidence from the LSO group suggests that the visual discrimination of the relative position of the helicopter probe and the flight deck trap was more difficult in the simulator than at sea under comparable environmental conditions.

There was no significant difference between the groups and only a trend towards improvement with practice. If the simulation was valid for the visual discrimination in the LSO's conning activity, it would be expected that the LSO group would have had an advantage because of their experience, however, no advantage was evident in the conning data.

Most of the uncertainty in the visual judgements occurred along the viewing axis, particularly at the right-front and left-rear corners of the trap. Some subjects moved to gain a better perspective, however, none moved to the limits of the Howdah enclosure, suggesting that instruction about moving in the simulator may be required. Additionally, more visual detail in the simulated trap's texture may be necessary to better convey a sense of depth.

The response time of the simulated helicopter was also noticeably slow; several experienced LSOs commented on and attributed missed landings to the response delay. Within the simulation, several stages occur in series, which introduced latencies in the simulator response to verbal commands, including both the helicopter and pilot simulations. In some cases, computational demands drive the time required for a stage, indicating that improvements may only be made if more efficient computations are possible or if hardware speed improves; however, in other cases, some of the latency is due to operation timing, indicating that reductions in the latency may be achieved through optimization of the existing simulations.

The difficulties landing and conning the helicopter suggest that the simulator is not yet adequate to train the visual judgement of the relative positioning of the ship and helicopter in this tightly coupled, dynamic event, although the ship motion was anecdotally reported to be very realistic. The technical problems that have been identified (long latency and lack of visual texturing on the trap) should be investigated to determine if improvements will lead to improved performance.

Cost-benefit of the training simulation

Determining a cost of the simulator is complicated because of its development as a research project rather than as a commercial product acquisition; therefore, a true cost-benefit analysis considering the total amortized platform costs cannot be adequately performed on the current system. However, an estimate can be made of what the equivalent operating costs would be for training at sea as it is currently done.

An unofficial estimate of the operating costs for personnel and materiel can be obtained from the Department of National Defence Cost Factor Manual (DSFC, 2009). The total hourly costs to operate a CF Halifax Class frigate is approximately 9700 $CAD (DSFC, 2009, Table 4Tab41_e.xls) while the hourly cost of running a Sea King helicopter is approximately 29000 $CAD (DSFC, 2009, Table 3Tab31_e.xls). Although some training and activity is possible on the ship during flight operations, the range of activities is severely limited due to restrictions imposed while the helicopter is flying in close proximity to the ship, so much of the cost of running the ship should be attributed to the flight training exercise, for a total hourly operating cost of approximately 40000 $CAD.

The duration of each of the experimental trials was approximately 5 minutes, however, the scenario duration was contrived for experimental purposes to reduce the amount of time during the helicopter approach when there are few LSO tasks to be performed; similar manipulation of the scenario can be achieved in actual training simulations. In practice, training circuits of a take-off and departure followed by an approach from the "Final Approach Fix" and landing on board ship might be expected to take about 15 minutes each, for a cost of approximately 10000 $CAD/trial.

The Naïve group differed statistically from the Aircrew group until the sixth trial in the first session, but they did not differ for the 7th and 8th trials. This performance similarity was not, however, particularly robust, showing a marked decrement over the 3 hour break between Sessions 1 and 2. In Session 2, the Naïve group again differed from the Aircrew group until 11th trial (3rd trial, Session 2) after which there was no reliable difference. A similar pattern arose comparing the Naïve and LSO groups, with the Naïve group showing poorer performance until the 11th trial, after which performance was similar. Both the Naïve and Aircrew groups showed considerable variance due to differences in individual performance, while the variance displayed by the LSO group was substantially smaller presumably because of their prior knowledge, being experts in the domain already. The Aircrew differed reliably from the LSO group until approximately the 5th trial, after which their performances were not statistically different. There was a noticeable decrease in the Aircrew performance after the 3 hour break between sessions, however the difference was not statistically significant.

These calculations suggest that at a minimum, training the LSO deck landing role in the simulator would save 110000 $CAD for Naïve students and $50000 for Aircrew having some familiarity with the role. Note that this does not include estimates for the costs associated with overtraining (known to reduce skill-fade) nor does it include any of the other procedural tasks that a MH pilot has to become accomplished at before being qualified as an LSO.

Several LSOs in our study group had not been to sea in several months or years, and this was reflected in their initial scores; however, the performance of the LSOs who were out of practice improved quickly with exposure in the simulator and quickly became indistinguishable from the more current LSOs. Similar savings to those calculated above might be realized by practice in the simulator by LSOs who are requalifying after lengthy absences from shipborne helicopter duties.

Conclusions

The conclusions from this study and the implications for use of the simulation (both the physical simulation as well as the Human Behaviour Representation crew models) are twofold, although this is mediated by the study being limited to a Reverse Transfer of Training paradigm and not including a Forward Transfer of Training assessment.

First, the simulations were effective in providing an environment where subjects could learn the procedural aspects of the Landing Signals Officer's role during the approach and landing of a Maritime Helicopter on board a Canadian Forces Halifax Class frigate under way. The implications are that other LSO tasks that are procedural, containing verbal commands or manual actions, could be trained in the simulator if it was extended to incorporate the associated procedures. As the demands associated with the procedural learning of the experimental task are generic, it seems reasonable to assume that many other CF procedural team tasks could make use of the same technologies once adapted to the new application environment.

Second, the visual display or graphics presented to the LSO subject were not adequate for training the fine visual judgements required to determine when the helicopter was positioned over the trap. Additional study is required to determine exactly what the deficiencies are or how the display may be improved, but some potential improvement areas have been identified already. While tasks or applications that require relative visual distance discriminations may not be appropriate for training that particular aspect of the task in the simulator in the current state of development, the approach should be adequate for tasks where only straight-forward visual stimulation or feedback is required.

References

AGARD. (1980). *Fidelity of simulation for pilot training* (No. ISBN: 92-835-1 377-0). 7 Rue Ancelle - 92200 Neuilly-Sur-Seine, France: North Atlantic Treaty Organization (NATO), Advisory Group for Aerospace Research and Development (AGARD).

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control, AC-19*(6), 716-723.

Akaike, H. (1981). Likelihood of a model and information criteria. *Journal of Econometrics, 16*, 3-14.

Anderson, R. B. (2001). The power law as an emergent property. *Memory & Cognition, 29*(7), 1061-1068.

Annett, J. (2003). Hierarchical task analysis. In E. Hollnagel (Ed.), *Handbook of cognitive task design* (pp. 17-35): Lawrence Erlbaum.

Annett, J. & Duncan, K. D. (1967). Task analysis and training design. *Occupational Psychology, 41*.

Annett, J., Duncan, K. D., Stammers, R. B. & Gray, M. J. (1971). *Task Analysis*. London: Her Majesty's Stationery Office.

Annett, J. & Stanton, N. A. (2000). *Task analysis*. London: Taylor & Francis.

Boff, K. R. & Lincoln, J. E. (1988). *Engineering Data Compendium, Human Perception and Performance* (Vol. I, II, III). New York: John Wiley and Sons.

DSFC. (2009). *Cost Factor Manual*. Ottawa: Department of National Defence, Director Strategic Finance and Costing (DSFC 2).

Haider, H. & Grensch, P. A. (2002). Why aggregated learning follows the Power Law of Practice when individual learning does not: Comment on Rickard (1997, 1999), Delaney et al. (1998), and Palmeri (1999). *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*(2), 392-406.

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

Heathcote, A., Brown, S. & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review, 7*(2), 185-297.

Hendy, K. C., Hamilton, K. M. & Landry, L. N. (1993). Measuring subjective workload: When is one scale better than many. *Human Factors, 35*(4), 579-601.

Kennedy, R. S., Lane, N. E., Berbaum, K. S. & Lilienthal, M. G. (1993). Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology, 3*(3), 203-220.

Newell, K. M., Mayer-Kress, G. & Liu, Y.-T. (2006). Human learning: Power laws or multiple characteristic time scales? *Tutorials in Quantitative Methods for Psychology, 2*(2), 66-76.

Pew, R. W. & Mavor, A. S. (Eds.). (1998). *Modeling Human and Organizational Behavior. Applications to military simulations. Panel on Modeling Human Behavior and Command Decision Making: Representations for Military Simulations. Commission on Behavioral and Social Sciences and Education.* Washington, DC.: National Research Council. National Academy Press.

Ritter, F. E. & Schooler, L. J. (2002). The learning curve, *International encyclopedia of the social and behavioral sciences* (pp. 8602-8605). Amsterdam: Pergamon. Retrieved from *acs.ist.psu.edu/papers/**ritter**S01.pdf*

Shepherd, A. (2000). HTA as a framework for task analysis. In J. Annett & N. A. Stanton (Eds.), *Task analysis* (pp. 7-24). London: Taylor & Francis.

Suzuki, H. & Ohnishi, H. (2007). *Non-random deviations from the Power Law of Practice.* Paper presented at the International of Skill Sciences, ISSS'07, Tokyo, Japan. 18-20 September. 161-166.

Appendix A. Experimental scenario

Context

The objective of the experiment is to determine the effectiveness of the Helicopter Deck Landing Simulator and the SimON human model of a Sea King helicopter pilot for training small teams. The subject will play the role of the Landing Signals Officer (LSO) assisting the helicopter pilot simulation to land the helicopter by communicating and manipulating the LSO console according to formal SHOPS (Ship Helicopter Operating Procedures) procedures.

Background

General

While on patrol, meteorological conditions (Wx) in the patrol area have degraded abruptly and the Captain of the CF Halifax Class Frigate (call sign 'Warship') has ordered recovery of the CF Sea King helicopter (call sign 'HelMET 01') before the visibility degrades further to a point where an Emergency Low Visibility Approach (ELVA) would be required. An extensive fog bank surrounding HMCS Warship is obscuring visibility beyond 0.5 to 0.75nm. HelMET 01's Crew Commander is concerned that the weather will continue to deteriorate and the Pilot Flying is an exchange officer newly posted to the Squadron, so there is some urgency to recover to the ship.

Ship Status

1) The ship is now on the flying course. Deck motion is evident but the motion is currently within free-deck limits.
2) The FLYCO has reported a problem with the FLYCO trafficator switches that prevents setting them from the FLYCO console, requiring the LSO to change the trafficator lights from the LSO console. The Trafficator lights are an important component of communications between the ship and the helicopter.
3) Ship Configuration:
    a. FOD Rounds of the flight deck and boat decks are complete
    b. All flight deck equipment has been deemed serviceable by the Flight Deck Stokers and was tested during the post launch 'first of the day checks'.
    c. Ship is closed up at Flying Stations.
    d. All Fly Ops personnel are closed up in the respective positions
    e. The Deck Crew has just left the flight deck and entered the hangar; ready for the recovery.
    f. Lighting:
        i. Horizon bars are functional (i.e. following the earth model horizon) and the green elements of the horizon bars are illuminated.
        ii. Trafficators are RED.
    g. The Ship is RADHAZ SAFE.
    h. All communications checks are complete (i.e. all comms between ship and helo have been tested and are functioning correctly, other than the FLYCO control of the trafficator lights).
    i. Trap is in position for a free-deck recovery. RSD safety bar has been removed from the trap

    NOTE: In the simulation, the trap may not be in the normally open state after removal of the safety bar and the LSO must confirm that the trap is still open prior to clearing the helicopter for landing.
4) The FLYCO has completed the Flying Stations checklist and has just reported "AIR DEPARTMENT CLOSED UP AT FLYING STATIONS."

Helo Status

5) Helo is flying a RADAR Controlled Approach (RCA) in IMC. The Shipborne Air Controller (SAC) has already passed the numbers to the helo.

6) CH124 Configuration

   a. Helo heading: inbound on a RED 150° radial
   b. Helo range: 0.8 to 0.9 nm
   c. Helo Indicated Air Speed in Knots (KIAS) = 54 kts (relative wind plus 30)
   d. Helo Altitude: 100 ft (ASL)
   e. Landing Gear (L/G) Up
   f. Main Probe Down
   g. Tail Probe – UP

Environment

7) Sea: Light swell, causing discernable deck motion within free-deck limits

8) Atmosphere: True Winds from (direction True North/speed kts): 330 at 10kts

9) Wind relative to ship's bow (direction/speed kts): Red 13 at 20kts

10) Visibility: Limited to 0.5 to 0.75nm due to fog.

Starting the Experimental Learning Plan Simulation

Scenario Events

1) The Instructor/Operator will tell the subject when the simulation is ready to start the scenario; there is some delay while each of the simulations connects. After the Instructor indicates that the simulation is ready, the LSO starts the scenario by setting the Clearance Request switch on the LSO console to RECOVER. The BRIDGE will initially respond NO on the Clearance Request. When permission is received from the Captain, the BRIDGE will make the light YES.

   Note 1: This action has been adopted for experimental purposes but it is consistent with procedures.

   Note 2: The LSO verifies that the trap is open during a functional check early in the landing evolution preparation, but this action has been moved to the beginning of the simulation for experimental purposes.

   Note 3: Control of the trafficator lights is normally the responsibility of the FLYCO while the LSO monitors their state, but due to technical difficulties in this scenario, the LSO must both control and monitor the trafficators.

2) The first radio call is from the SAC when the helo reaches 1nm:

   "1 MILE, CALL VISUAL".

   NOTE: At this point, the helicopter-ship communications are generally abbreviated to exclude the formality of "C/S, C/S, <message>" but either format is acceptable for the scenario.

4) TACCO informs the ship that the helo can see the ship:

   "HELMET 01, VISUAL"

5) LSO:

   a. The LSO will advise the SAC when the LSO clearly sees the helo and is ready to assume responsibility for it by calling over the SHINCOM:

      "I HAVE HELMET 01 VISUAL"
      "READY TO TAKE CONTROL"

   b. Switch TRAFFICATORS to AMBER if they are still RED.

6) SAC:

   a. Acknowledges LSO's call via the Ship's SHINCOM:

      "ROGER"

   b. Informs the helo over the radios:

"HELMET ZERO ONE, WARSHIP
PADDLES HAS YOU VISUAL
CALL PADDLES FOR CONTROL"

7) TACCO:
   a. The helo confirms the SAC's instruction and contacts the LSO over the radios:

      "HELMET ZERO ONE, ROGER."
      "BREAK, BREAK"
      "PADDLES, HELMET ZERO ONE FOR CONTROL."

8) LSO:
   a. In this scenario, there are no complications, the ship motion is within free-deck limits and there is some urgency to get the helicopter on-board due to deteriorating weather so the LSO should signal a Free-Deck landing by calling over the radios:

      "HELMET ZERO ONE, PADDLES,
      SIGNAL CHARLIE FREE DECK"

   b. When the clearance call "SIGNAL CHARLIE FREE DECK" is made, the Trafficators are switched to GREEN.

9) DELTA HOVER ABEAM (PORT SIDE)
   a. CH124 Position:
      i. Helo pulls alongside ship into a 40' hover, rotors clear of nets.
      ii. Established in the hover port side at 40 feet ASL for approximately 5 to 10 seconds.
      iii. The flying pilot (typically the right seat pilot when the helo is on the ship's port side), just before commencing transition towards the flight deck calls, over the helo's ICS, for the landing gear to be lowered. The call is "GEAR".
      iv. Non-Flying Pilot (NFP) Lowers Gear, and acknowledges request over the helo's ICS by stating "IN TRANSIT". The NFP confirms both wheels are down by checking the cockpit gear indicators and the illumination of the bug light on the landing gear on NFP side by looking out the window or in the rear view mirror. When confirmed, the NFP calls over the ICS "GEAR DOWN AND LOCKED". The landing gear should be fully down and locked by the time the helicopter reaches the high hover. The LSO, the FLYCO and both pilots are all proactively verifying that the gear is down and locked.
         1. <u>If</u> the gear is observed in the down and locked position, no broadcasts over the radio are required.
         2. <u>If</u> the landing gear is not down and locked, then the LSO is expected to prompt the pilots to recheck the landing gear by calling over the radios "CHECK GEAR".
      v. The helo transitions to high hover over deck.
      vi. Flying Pilot should be looking slightly down onto the hangar top.

10) LSO
   a. As the helicopter begins to transition over the nets along the side of the ship or stops in the HIGH HOVER, the Trafficators are switched to AMBER.
   b. The LSO ensures that the tail probe is up, main probe is down, and landing gear begins to lower as the helicopter begins to transition laterally from hovering abeam the flight deck.

11) HIGH HOVER
   a. High hover is about 23 feet above the flight deck on RADALT.

b. Flying pilot waits in high hover directly overhead the trap until confident that with the pattern of the ship's motion.
   i. Read the deck motion such that the flying pilot can manoeuvre to the low hover by the time the ship reaches its steady state period. The high hover position is normally identified by placing the helo so the top of the pitch bar is visible and mid-way between both fore and aft horizon bars.

c. The pilot begins the descent to the low hover once confident about the motion of the ship and the helicopter is stable in the high hover.

12) LOW HOVER
   a. The low hover is approximately 5-7 feet above the flight deck, as indicated to the flight crew on the RADALT. Pilot should reference this visually by observing the hangar face based on previous experiences. Tail wheel bounce is indicative of an excessively low hover.
      i. The helo will descend toward the low hover. When the pilot is confident with the relative position of the helicopter and the trap, and the helicopter is relatively stable, the pilot will make a radio call:

         "READY TO LAND"

13) LSO:
   a. The LSO should be assessing the relative position of the main probe over the trap.
      i. If the LSO does not think the main probe is over the trap, commence conning to assist the pilot to improve the relative position of the Main probe by broadcasting, over the radios, the appropriate direction the pilot should move the helo using the following words only:
         1. LEFT
         2. RIGHT
         3. AHEAD
         4. BACK
         5. STEADY
            a. to remain in the current location if position is good but the deck is not suitable for landing or to stop moving in one direction in anticipation of moving in the opposite direction
         6. UP
         7. DOWN
      ii. When both the aircraft and the ship are in a good relative attitude, and the motion between the two bodies is relatively calm (i.e. quiescent"), and the main probe is in a good position relative to the trap such that the probe will enter the central area and a successful trap will result, the LSO will call over the radio

         "LAND NOW, DOWN, DOWN, DOWN:

      NOTE: If the pilot misses the call, the LSO will repeat the call when conditions are again suitable. More DOWN calls may be required if the helicopter hesitates too long in the low hover; they should continue until the helicopter is on the deck, or a wave off has been executed.
   b. When the LSO signals "LAND NOW", the trafficators are switched to GREEN.

14) WEIGHT ON WHEELS
   a. Until the LSO signals "IN THE TRAP, TRAPPED", the flying pilot should be prepared for a WAVE OFF.

15) LSO:

a. If the aircraft is in a safe position in the trap,

i. The LSO will call over the radio:

"IN THE TRAP".

ii. The LSO will close **the trap with the console switch** and when the trap has finished closing, the LSO will inform the helo that the aircraft is secure by calling:

"TRAPPED"

iii. Switch Trafficators to AMBER.

iv. Make the RSD switch OFF.

b. If the helo lands with the probe outside of the trap, the LSO instructs the pilot to abort the landing and to return to the high hover by calling over the radios:

"WAVE OFF. WAVE OFF. WAVE OFF."

c. On a "WAVE OFF", switch Trafficators to RED.

i. In the event of a wave-off, the pilot responds by repeating "WAVE OFF, WAVE OFF, WAVE OFF" and the helo returns to the High Hover.

ii. When the deck is secure, the helicopter is steady in the high hover; the LSO indicates that it is safe to resume the landing procedure by calling

"ALL CLEAR"

iii. and the Trafficators are switched to AMBER.

iv. When the pilot is ready, the helicopter drops to the low hover and procedure repeats as required.

16) LSO

a. When the helicopter is properly trapped on the deck, the LSO advises the pilot to lower the tail probe to prevent the helicopter from pivoting by calling over the radio:

"DOWN TAIL PROBE"

17) Pilot:

a. Lowers the tail probe in response to the LSO's direction.

18) LSO:

a. Over the radio, advises the pilot when the tail probe is lowered and secure in the rails on the deck of the ship:

"IN THE RAILS"

b. When the tail probe is secure, the LSO confirms trafficators are AMBER

19) LSO:

a. Indicate to the FLYCO that it is safe for the Deck Crew to come on deck, refuel the helicopter with the engines running, and then Shutdown the helicopter:

"FLYCO, LSO, DECK CREW ON DECK
HOT FUEL
SHUT DOWN"

b. Indicate to the Bridge that the helicopter is secure and that it is safe for the ship to manoeuvre with caution:

"BRIDGE, LSO, HELO TRAPPED ON DECK.
FREE TO MANOEUVRE WITH CAUTION"

c. Return the Clearance Request switch to OFF.

This ends the current scenario.

This page intentionally left blank.

Appendix B. Experimental simulation hardware

The LSO simulator comprises a physical mockup of the LSO's console, with active switches to control the bridge clearance request, the trap closure and the trafficator lights. The switch positions and the associated indicator displays correspond to their visual presentation in the LSO subject's occluded Head Mounted Display. Other LSO simulator equipment includes:
Dell Precision 670 computer
Dual Xeon CPUs, 3.6 GHz
Nvidia Quadro FX 4500 Video Card
4GB RAM
Linux Operating System (Fedora Core 4)

Polhemus, Liberty, Head tracker

NVis, Nvision SX 60, Head Mounted Display

Colour, stereo, LCD displays
1280x1024 pixels/eye
60° diagonal field of view, fully overlapped
120 Hz refresh rate

The Instructor Operating Station (IOS) comprises:
Dell Precision 530 computer
Xeon Processor, 2.0 GHz
2GB RAM
Nvidia Quadro FX 1300
Linux Operating System (Redhat 8)

The HDLS/HelMET simulation comprises
Concurrent Computer Corporation Imagen computer
Four dual-core AMD processors.
16 GB RAM
Two NVidia Quadro FX 5500 graphics cards
Three 250GB 7.2K SATA hard drives
Linux Operating System (Redhawk Linux)

The HDLS/HelMET simulator also incorporates two head mounted displays with optical head tracking, an electric 6 DOF motion base and primary flight controls. The physical simulators were not used in this study; however, all of the underlying physics based models remained the same. The pilot model provided primary flight control signals to the HDLS/HelMET simulation, bypassing the physical primary flight controls.

The SimON Human Behaviour Representation of the Pilot (TACCO and SAC) comprises:

Dell Precision T7400 computer (running IPME and Sphinx software)
Intel Dual-Quad Core Xenon E5430 CPUs, 2.66 GHz
4 GB RAM

M-Audio 1010 PCI audio interface
Linux Operating System (OpenSuSE 10.2)

Toshiba Tecra S2 computer (running AT&T software)
Intel M750 CPU, 1.8 GHz
2 GB RAM
Linux Operating System (Redhat)

Software
IPME 4.3.3 (Integrated Performance Modelling Environment, Alion Science Ltd., MA&D Operation)
AT&T Naturally Speaking, Rev. 1.4 (text to speech production software)
Sphinx v4-1.0 open source speech recognition software (Carnegie Melon University)
Countryman e6i microphone

Appendix C: NASA TLX subjective workload rating scale

NASA TLX Subjective Workload Questionnaire

Please place an “X” along each scale at the point that best indicates your experience with the display configuration.

<u>Mental Demand</u>: How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc)? Was the mission easy or demanding, simple or complex, exacting or forgiving?

Low |___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___| High

<u>Physical Demand</u>: How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the mission easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Low |___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___| High

<u>Temporal Demand</u>: How much time pressure did you feel due to the rate or pace at which the mission occurred? Was the pace slow and leisurely or rapid and frantic?

Low |___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___| High

<u>Performance</u>: How successful do you think you were in accomplishing the goals of the mission? How satisfied were you with your performance in accomplishing these goals?

Low |___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___| High

<u>Effort</u>: How hard did you have to work (mentally and physically) to accomplish your level of performance?

Low |___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___| High

<u>Frustration</u>: How discouraged, stressed, irritated, and annoyed versus gratified, relaxed, content, and complacent did you feel during your mission?

Low |___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___|___| High

NASA-TLX Mental Workload Factor Paired Comparisons

For each of the pairs of factors listed below, circle the factor that represents the more important contributor to overall workload in that pair.

Mental Demand	or	Physical Demand
Mental Demand	or	Temporal Demand
Mental Demand	or	Performance
Mental Demand	or	Effort
Mental Demand	or	Frustration
Physical Demand	or	Temporal Demand
Physical Demand	or	Performance
Physical Demand	or	Effort
Physical Demand	or	Frustration
Temporal Demand	or	Performance
Temporal Demand	or	Frustration
Temporal Demand	or	Effort
Performance	or	Frustration
Performance	or	Effort
Frustration	or	Effort

Appendix D. Simulator Sickness Questionnaire (SSQ)

Interpretation of the ratings were made in accordance with Kennedy et al. (1993)

Simulator Sickness Questionnaire

Symptom Checklist

Instructions: Please indicate the severity of symptoms that apply to you right now.

1	General Discomfort	None	Slight	Moderate	Severe
2	Fatigue	None	Slight	Moderate	Severe
3	Headache	None	Slight	Moderate	Severe
4	Eye Strain	None	Slight	Moderate	Severe
5	Difficulty Focusing	None	Slight	Moderate	Severe
6	Increased Salivation	None	Slight	Moderate	Severe
7	Sweating	None	Slight	Moderate	Severe
8	Nausea	None	Slight	Moderate	Severe
9	Difficulty Concentrating	None	Slight	Moderate	Severe
10	Fullness of head	None	Slight	Moderate	Severe
11	Blurred Vision	None	Slight	Moderate	Severe
12	Dizzy (Eyes open)	None	Slight	Moderate	Severe
13	Dizzy (Eyes closed)	None	Slight	Moderate	Severe
14	Vertigo*	None	Slight	Moderate	Severe
15	Stomach awareness	None	Slight	Moderate	Severe
16	Burping	None	Slight	Moderate	Severe
17	Boredom	None	Slight	Moderate	Severe
18	Drowsiness	None	Slight	Moderate	Severe
19	Decreased Salivation	None	Slight	Moderate	Severe
20	Mental Depression	None	Slight	Moderate	Severe
21	Visual Flahsbacks	None	Slight	Moderate	Severe
22	Faintness	None	Slight	Moderate	Severe
23	Aware of Breathing	None	Slight	Moderate	Severe
24	Loss of Appetited	None	Slight	Moderate	Severe
25	Increased Appetite	None	Slight	Moderate	Severe
26	Desire to move bowels	None	Slight	Moderate	Severe
27	Confusion	None	Slight	Moderate	Severe
28	Vomiting	None	Slight	Moderate	Severe

*Vertigo is a disordered state in which the person or his surroundings seem to whirl dizzily: giddiness.

This page intentionally left blank.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.) Defence R&D Canada – Toronto 1133 Sheppard Avenue West P.O. Box 2000 Toronto, Ontario M3M 3B9	2. SECURITY CLASSIFICATION (Overall security classification of the document including special warning terms if applicable.) UNCLASSIFIED (NON-CONTROLLED GOODS) DMC A REVIEW: GCEC JUNE 2010

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

Validation of a virtual environment incorporating virtual operators for procedural learning

4. AUTHORS (last name, followed by initials – ranks, titles, etc. not to be used)

Brad Cain; Lochlan Magee; Courtney Kersten

5. DATE OF PUBLICATION (Month and year of publication of document.) September 2012	6a. NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.) 64	6b. NO. OF REFS (Total cited in document.) 20

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Technical Memorandum

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

Defence R&D Canada – Toronto
1133 Sheppard Avenue West
P.O. Box 2000
Toronto, Ontario M3M 3B9

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)
10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.) DRDC Toronto TM 2011-132	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)
Unlimited distribution.

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))
Unlimited announcement.

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

This paper presents the results of an experiment to assess the validity of a prototype simulation to train individuals to perform a task as part of a team. The application domain is Maritime Helicopter-Ship operations and the task selected is of a Landing Signals Officer (LSO) coordinating the approach and landing of a helicopter on board Canadian Forces frigates. The simulation includes physics based models of the helicopter, ship and the environment, as well as a human factors approach to representation of team mates by computer generated, behavioural agents. A reverse transfer of training experiment was conducted to assess how three groups, each initially differing in domain knowledge, acquired the necessary procedural knowledge, verbal communications and manual actions to complete the task without error. Thirty subjects participated: ten assigned to each of a Naïve, Aircrew and LSO group as determined by their initial domain knowledge. Learning rate results indicate significant differences among the groups and the effect sizes were sufficient to conclude that the approach is valid for training procedural tasks of the LSO occupation and, by extension, to other small team, procedural task trainers with similar user interface requirements. The simulation was not found to be adequate to train the fine, visual judgements involved in directing the helicopter over the deck, and improvements to the simulation have been proposed.

Le présent document présente les résultats d'une expérience visant à évaluer la validité de la simulation d'un prototype d'entraînement de personnes à l'exécution d'une tâche au sein d'une équipe. Le domaine d'application est l'exploitation d'un hélicoptère maritime ainsi que d'un navire, et la tâche choisie est celle d'un officier de signalisation à l'appontage (LSO) coordonnant l'approche et l'appontage d'un hélicoptère se trouvant à bord de frégates des Forces canadiennes. La simulation comporte des modèles de l'hélicoptère, du navire et de l'environnement basés sur la physique, ainsi qu'une approche de la représentation des coéquipiers tenant compte des facteurs humains représentés par des entités reproduisant le comportement humain et générées par ordinateur. On a utilisé la méthode du transfert de formation inverse pour évaluer la façon dont trois groupes, selon leurs connaissances initiales du domaine, ont acquis les connaissances procédurales ainsi que les aptitudes à utiliser les commandes verbales et les actions concrètes nécessaires à l'exécution de la tâche sans erreur. Trente sujets ont participé; on les a répartis en trois groupes de dix novices, dix membres d'équipage et dix LSO, selon leurs connaissances initiales du domaine. La courbe d'apprentissage observée variait considérablement d'un groupe à l'autre et l'importance de l'effet était suffisante pour conclure que la technologie en question est utile à l'apprentissage du travail d'un LSO et, par extension, peut être utilisée par d'autres petites équipes dont l'entraînement exige une interface utilisateur similaire. La simulation s'est révélée inadéquate pour l'entraînement des jugements visuels excellents que nécessite la direction de l'hélicoptère au-dessus du pont, et on a proposé des améliorations à cette simulation.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

human modelling; validation; training; virtual reality; Landing Signals Officer; pilot model

Defence R&D Canada

Canada's Leader in Defence
and National Security
Science and Technology

R & D pour la défense Canada

Chef de file au Canada en matière
de science et de technologie pour
la défense et la sécurité nationale

DEFENCE DÉFENSE

www.drdc-rddc.gc.ca