FOI, Swedish Defence Research Agency, is a mainly assignment-funded agency under the Ministry of Defence. The core activities are research, method and technology development, as well as studies conducted in the interests of Swedish defence and the safety and security of society. The organisation employs approximately 1000 personnel of whom about 800 are scientists. This makes FOI Sweden's largest research institute. FOI gives its customers access to leading-edge expertise in a large number of fields such as security policy studies, defence and security related analyses, the assessment of various types of threat, systems for control and management of crises, protection against and management of hazardous substances, IT security and the potential offered by new sensors.

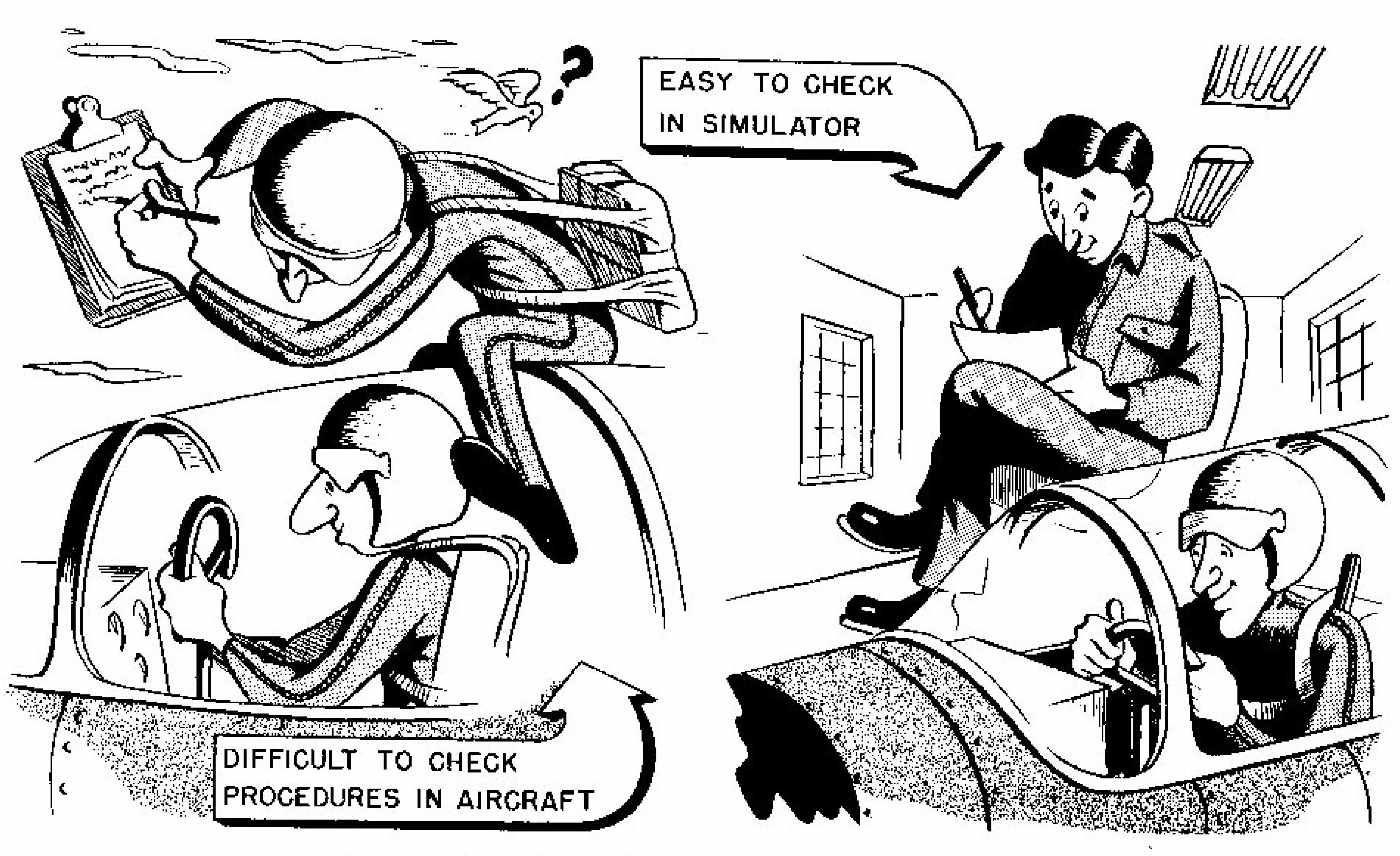
FOI
Defence Research Agency
Command and Control Systems
P.O. 1165
SE-581 11 Linköping

Phone: +46 13 37 80 00
Fax: +46 13 37 81 00

www.foi.se

Transfer of Training in Military Aviation

JONATHAN BORGVALL, MARTIN CASTOR, STAFFAN NÄHLINDER,
PER-ANDERS OSKARSSON, ERLAND SVENSSON



FOI-R--2378--SE Methodology report **Command and Control Systems**
ISSN 1650-1942 December 2007

Jonathan Borgvall, Martin Castor, Staffan Nählinder,
Per-Anders Oskarsson, Erland Svensson

Transfer of Training in Military Aviation

Titel	Överföring av träningseffekter i militär flygning
Title	Transfer of Training in Military Aviation
Rapportnr/Report no	FOI-R--2378--SE
Rapporttyp Report Type	Metodrapport Methodology report
Sidor/Pages	36 p
Månad/Month	December
Utgivningsår/Year	2007
ISSN	ISSN 1650-1942
Kund/Customer	FM HKV
Forskningsområde Programme area	7. Ledning med MSI 7. C4I
Delområde Subcategory	72 MSI 72 Human Factors
Projektnr/Project no	E7096
Godkänd av/Approved by	Jan Andersson

FOI, Totalförsvarets Forskningsinstitut	FOI , Swedish Defence Research Agency
Avdelningen för Ledningssystem	Command and Control Systems
581 11 Linköping	SE-581 11 Linköping

Sammanfattning

Föreliggande rapport redovisar resultat av en litteraturgenomgång samt egen forskning rörande träningseffekter och i synnerhet så kallade "transfer of training" studier. Mer specifikt handlar rapporten om överföring av kunskaper och färdigheter från träning i till operativ miljö. Särskild uppmärksamhet har riktats på studier som genomförts inom flyget. Syftet med rapporten är att tillhandahålla bakgrundskunskap för forskare som avser genomföra ”transfer of training" studier. Rapporten är därför utformad som en sammanställning av metodologiska erfarenheter och teoretiska frågeställningar kring ämnet ”transfer of training".

Nyckelord: träning, träningseffekt, överföring, flyg, simulering, fidelity, människa-system interaktion, MSI

Summary

The current report is the result of a review of the scientific literature as well as own experiences concerning transfer effectiveness and in particular transfer of training. More specifically the report relate to transfer of training from simulator to performance in the real task setting, with a special emphasis given to transfer of training studies performed within aviation. The report provides background knowledge for researchers conducting transfer of training studies. Thus the report is a compilation of methodological experiences and theoretical concerns concerning the subject of transfer of training.

Keywords: transfer of training, training effectiveness, aviation, simulation, fidelity, human factors, man-system interaction, MSI

Table of contents

1 INTRODUCTION

In today's world, simulation forms natural and useful parts of training and education in most domains. However, several factors limit for example flight training with real aircraft, for instance peacetime training rules, resource limitations, technical constraints, and security restrictions. To take one example, several factors limit live flight training, for instance peacetime training rules, resource limitations, technical constraints, and security restrictions. The use of flight simulators for training is therefore generally considered as a valuable complement to live training in the aircraft. However, with the increased use of simulation for training, the need to quantify the amount of training effectiveness and transfer of training increases.

Presented below are a number of question that, given the extensive use of simulator training, have been, or should have been, raised by decision makers at the Swedish Armed Forces Headquarters:'

- What is the effect on operative performance and/or flight safety when X hours of flight in the real aircraft is replaced by Y hours on a flight simulator?
- Can simulation make training even more effective – for example more efficient, quicker, and/or cheaper – than training with the real system?
- What is the optimal mix between live flight training and simulated flight training, and how should the structure of the total training regime be designed to provide maximum training value with high cost-effectiveness?
- What should be trained in the simulator and what must be trained in live flight? Can we identify critical phases or tasks that should receive special attention and focus in the simulator training?
- Can we identify some task components that are dangerous to train in the simulator due to negative transfer effects (i.e., live performance decreases as a function of simulator training)?
- How can we guarantee that the training received in a simulator is transferable to other environments and ultimately the combat environment?
- What are the benefits of mission rehearsal and spin-up exercises? Are they something conceptually different from more generic mission training in terms of what skills that can and should be developed? How can the effects of mission rehearsal and spin-up exercises be evaluated?
- The Swedish Armed Forces are currently introducing the Concept Development and Experimentation (CD&E) method in the development process. How is training at various levels (i.e., individual, team, collective, joint, coalition) considered and integrated in CD&E activities? How can training ideas and concepts be tested and developed iteratively under the CD&E paradigm?
- To what extent can low fidelity simulators provide valuable skill training that is transferable to task execution in a live environment? What is the relation between the development of complex skills and training complexity: can for example training with the Tetris computer game provide any value for a student pilot during basic flight training?
- What is more realistic and what provides the highest possible mission preparedness: live training with peace-time regulations or simulated training with combat regulations?

The answers to the questions above, but also the inability to sometimes answer them, provide the rationale for continued efforts on training effectiveness and transfer of training. Ever since the introduction of training simulators these questions have been identified as critical hot topics, but nevertheless, solid transfer of training studies are relatively rare. This is most probably due to the fact that the resource demands associated with these studies are rather extensive. However, when considering the amount of money invested in simulator training programmes worldwide, spending money on transfer of training studies to investigate the return seem to build a rather strong business case.

The need for transfer of training studies is further stressed by the fact that over the past few decades there has been a shift from traditional training of psycho-motor skills to higher levels of cognitive skills. The tactical operation of a modern fighter aircraft is for example more characterized by tactical decision making than by aircraft maneuvering. Furthermore, modern armed conflicts are characterized by ambiguous situations with a continuously increasing complexity, for example with regard to the identification of enemies and actions permitted by the Rules of Engagement (RoE) of an operation. In a meta-review of 53 articles, Arthur, Bennett, Stanush, & McNelly (1998) find evidence that performance on physical, natural, and speed-based tasks are less susceptible to skill decay than performance on cognitive, artificial, and accuracy-based tasks. The review also indicate that open-loop tasks such as continuous control are better retained, even for extended time periods (months or years), than closed-loop discrete tasks. Although the report concern skill decay it is highly probable that transfer of training effects also vary depending on the type of skill.

Although many of the references in this report refer to theories and studies of transfer of training in aviation the methods and the lessons learned are to a great extent directly applicable and transferable to other domains, military and civilian.

2 SCIENTIFIC AND EMPIRICAL BACKGROUND

2.1 Skill acquisition

Many theories and frameworks have been developed to explain the fundamentals of skill acquisition and learning. During the 1990-ties there have been tremendous theoretical developments in the field of training and training transfer research (Salas & Cannon-Bowers, 2001).

Ackermans (1992) theory of skill acquisition describe learning in early stages as characterised by knowledge as being declarative based largely on working memory and general intelligence. During a learning process components of the skill becomes automated and the demands on working memory decreases, while the importance of perceptual speed increases. Mollesworth and Wiggins (2006) describe three contemporary perspectives purport to explain how individuals transfer information from one context to another. According to the Structure-Mapping Model and the Pragmatic Schema Model the process of mapping occurs predominantly through the use of formal or abstract rules, based on concrete similarities between situations. Cased Based Reasoning on the other hand is based on the principle that the surface or superficial features of a problem play an integral role in transfer, sometimes at the expense of the abstract principles. The perspectives all assume that the information acquired during training must be retained and recalled to affect improvements in performance during testing.

The overarching goal of most military flight training regimes is to enhance the pilots' development of adaptive expertise. Adaptive expertise (Holyoak, 1991) entails a deep comprehension of the conceptual structure of the problem domain and is, with continuously increasing importance, a necessary requirement in order to successfully transfer training experiences to manage live tactical situations. Gopher, Weil, and Siegel (1989) describe complex tasks as "an organized set of response schemas", executed and coordinated by high-level schemas or strategies. In order to maximize the development of such schemas the intention of a training facility should be to provide and maintain a training environment which stimulates the motivation of deliberate practice and active learning among the trainees, improving their performance towards expertise levels (Ericsson, Krampe, & Tesch-Römer, 1993; Ericsson & Charness, 1994; Kozlowski, 1998).

Goettl and Shute (1996) discuss four challenges that a part-task training regime must overcome to be effective. These challenges are valid for any transfer study and highlight the need for a thorough analysis of training needs. The first challenge is that the training effectiveness of part-task training depends on the identification of valid critical component tasks. The second challenge is that the skills identified as most critical early in training may not be the skills most critical later in training. Thirdly, interactions among the component tasks play an important role in the whole task, and finally, individual differences in ability and style of learning play a large role in skill acquisition. Also worth remembering are Schmidt and Björk's (1992) point that the processes behind the acquisition, retention and transfer of skills are inseparable in order to understand the whole learning process.

2.2 Transfer of Training

The concept transfer of training refers to how previous learning influences behavior in a latter situation. Facteau, Dobbins, Russell, Ladd & Kudisch (1995) refer to transfer of training as the ability to apply what one has learned during training back to one's job. Transfer of training can be positive, nil, or negative. Positive transfer refers to improved real world performance of a given task following training in a training environment, nil to no effect, and negative transfer to degraded real world performance (Roscoe & Williges, 1980; Alexander, Brunyé, Sidman & Weil, 2005). The degree of positive or negative transfer can be calculated with somewhat different methods. The most commonly occurring formulas are percent of transfer, cumulative transfer, and incremental transfer (see Section 2.2.1 below). Percent transfer measures the ratio of time saved in simulator training relative to real-world training.

Alliger, Bennett and Tannenbaum (1995) has made an important distintion between what they call “different task, same environment” learning transfer paradigm and “same task, different environment” organizational training transfer paradigm. The former is interested in the generalization of learning from one task to another within the same environement (i.e., how performance on task B is facilitated by training on task A), while the latter focuses on the generalization of trained performance of a task from the training environment to the work environment (i.e., how performance on task A in the work environment is facilitated by training on task A in a training environment). This distinction is fundamental and seems rather obvious but confusion is often seen in definitions of transfer. Throughout this report, examples of both these paradigms occur.

2.2.1 Examples of measurements of transfer of training

The basic method of estimating transfer of training is the calculation of percent of transfer from the simulator to the real system or situation. For instance in a flight training setting, if percent of transfer is calculated for an experimental group having required 5 hours in the aircraft to reach task criterion, after having received 5 hours of simulator training, and a control group with no simulator training having required 10 hours in the aircraft to reach task criterion, the result is that the transfer of training from the simulator to the real aircraft is 50 percent (adapted from Roscoe & Williges, 1980):

$$\text{Percent of transfer} = \frac{Y_0 - Y_x}{Y_0} \times 100 \quad \Rightarrow \quad \frac{Y_0 - Y_5}{Y_0} = \frac{10 - 5}{10} \times 100 = 50\ \% \qquad (1)$$

Y_0 = time, trials or errors (in aircraft) required by the control group, with no simulator training, to reach performance criterion.

Y_x = time, trials or errors (in aircraft) required by the experimental group having received x units simulator training, to reach performance criterion

Two further formulas (not presented here) for calculation of percent of transfer have been described by Ellis (1965). A drawback to the calculation of percent of transfer is that the amount of practice in the simulator is not considered, thus no conclusions about the effectiveness of transfer of training is given (Roscoe & Williges, 1980). That is, as long as transfer of training is positive, the function of percent of training is increasing with time of simulator training. However, there is evidence that the transfer of the first hour in the simulator is higher than the second hour, and so on. Consequently, the effectiveness of transfer of training is decreasing over time in the simulator. When flight training curriculums are elaborated, estimation of the turning point when the cost of the simulator training exceeds the value of the transfer of training to the aircraft is of interest. Hence, formulas that calculate the effectiveness of training, by considering amount of time in the simulator can provide assistance.

The Cumulative Transfer Effectiveness Function (CTEF) gives the total time saving in the real aircraft, in relation to time spent in the simulator. For instance, if CTEF is calculated for an experimental group having required 5 hours in the aircraft to reach task criterion, after receiving 5 hours of simulator training, and a control group with no simulator training having required 10 hours in the aircraft to reach task criterion, the result is a time saving of one hour in the aircraft, for every hour spent in the simulator (adapted from Roscoe, 1971; Roscoe & Williges, 1980):

$$\text{CTEF} = \frac{Y_0 - Y_x}{X} \Rightarrow \frac{Y_0 - Y_5}{X} = \frac{10-5}{5} = 1 \text{ hour per hour} \qquad (2)$$

Y_0, Y_X: same as formula (1)

X: Amount of time, task iterations etc. received in the simulator.

The Incremental Transfer Effectiveness Function (ITEF) gives the incremental savings in the real aircraft, in relation to time spent in the simulator. For instance, if ITEF is calculated for an experimental group having required 52/3 hours in the aircraft to reach performance criterion, after receiving 4 hours of simulator training, and another experimental group having required 5 hours in the aircraft to reach performance criterion, after 5 hours simulator training, the result is that by increasing time of simulator training from 4 to 5 hours saves 2/3 hours in the real aircraft (adapted from Roscoe, 1971; Roscoe & Williges, 1980):

$$\text{ITEF} = \frac{Y_{x-\Delta x} - Y_x}{\Delta X} \Rightarrow \text{ITEF}_{5\text{-}4} = \frac{Y_{5-1} - Y_5}{5-4} = \frac{5\tfrac{2}{3} - 5}{5-4} = 2/3 \text{ hour per hour} \qquad (3)$$

Y_x *: Amount of time, task iterations etc. (in aircraft) of an experimental group, having received x units of simulator training, to reach the performance criterion (same as formula 1).*

$Y_{x-\Delta x}$ *: Amount of time, task iterations etc. (in aircraft) an experimental group, having received* $x - \Delta x$ *units of simulator training, to reach the performance criterion.*

ΔX *: Incremental amount of time, task iterations etc. in the simulator.*

Note, that when ITEF is computed it is not necessary to use the unitary amount of one hour, one iteration etc. as the incremental unit. Just as well, one half, two or three hours etc. can be used (Roscoe & Williges, 1980).

Also note that efficient training normally implies an increase in degree of difficulty for each training session, or in pace with the learner's rate of acquiring new skills. Therefore, proper use of the above presented formulas for calculation of transfer of training should be limited to training of well-defined skills in a limited period of time.

Even if the methods for estimation of transfer effects presented above are of practical value, they are simplistic with respect to explanatory power. In studies performed in the Swedish Air Force, we have adapted and used structural equation modeling techniques in estimations of transfer effects. By means of these techniques, we can estimate the (transfer) effects of an optimal linear combination of different, e.g., performance measures during simulated training on operational training in the air. Accordingly, the relative effects of different capabilities transferred can be estimated, as well as a combined measure of the transfer effect. The combined measure varies between -1.0 and 1.0. A transfer effect of for example .70 means that about 50 percent of the variance in operational performance can be explained by the pilots' performance level during simulated training. The measure can also be considered as a measure of to what extent the content of simulated training is adjusted to the operational situation. Examples of this technique will be presented below in section 2.4.9.

2.2.2 Negative transfer of training

When learning in one situation interferes with previous knowledge or skills from another situation, this is called negative transfer. Negative transfer most likely occurs when someone acts as if there are common features in two different settings. For instance, when learning a new language, one might be likely to (incorrectly) use experiences from a previously known language.

Positive transfer is when previous knowledge assists in learning of a new situation, for instance recognizing common concepts or features between different situations. Positive transfer is very effective in transferring previous knowledge or skills to a new situation. Negative transfer is also strong but counteracts the building of new knowledge and skills. In education positive transfer is often used to help students expand their knowledge, while trying to minimize negative transfer.

Negative transfer also occurs when someone incorrectly applies methods and techniques learnt in one environment in another environment. For instance, if one learns to drive a ship very close to the coastline in a simulator, and then does the same in real ship, one might be in real danger simply because the simulator did not correctly model the dynamics of the sea, the wind and the currents. This has been reported as a real danger in many settings and often used as an argument that the simulator must have very high fidelity to assure that no bad habits are learnt which are subsequently transferred to a real-life situation. Negative transfer can also occur when moving from a simulator to a real life setting, but also when transferring from one system to another, for instance moving from one type of aircraft to another.

Negative transfer is cited as a contribution in several accident reports at the NTSB (National Transportation Safety Board). For instance, after a general aviation accident in 2001, the NTSB concluded that the pilot transferred methods learnt in one aircraft (Katana) to another (Cessna 172) (National Transport and Safety Board, 2001):

> *The failure of the pilot-in-command to execute a proper landing flare, which resulted in an improper touchdown attitude and a subsequent loss of directional control. A factor in the accident was the difference between landing characteristics of the Katana and the Cessna 172. […] Since the pilot had received all of his primary training in a Katana, his tendency in any airplane would be to land it like a Katana.*

Even though the accident was non-fatal, the aircraft was substantially damaged.

2.2.3 Kirkpatrick's levels of training evaluation

Training evaluation can be defined as the "systematic collection of descriptive and judgmental information necessary to make efficient training decisions related to the selection, adoption, value, and modification of various instructional activities" (Goldstein, 1980). In the late 1950-ties Kirkpatrick presented a hierarchical model to be used in the evaluation of training programs (Kirkpatrick, 1994). Kirkpatrick's original model included analysis of (a) reaction, (b) learning, (c) behavior, and (d) results. The levels or types of evaluation of a training program or effort can be summarized as:

a. **Reaction:** How well the trainee liked the training program.

b. **Learning**: The knowledge acquired, skills improved, or attitudes changed as a result of training.

c. **Behavior:** Using those facts and skills learned on the job.

d. **Result:** Outcomes that appear on the job as a result of training.

The acceptance of Kirkpatrick's' model includes three key assumptions: 1) it is possible to arrange the hierarchical levels in increasing order of value (i.e., reaction, learning, behavior, results); 2) it is possible to causally link the levels, and 3) effects on the different levels correlate positively. Alliger, Tannenbaum, Bennett, Traver and Shotland (1997) expanded Kirkpatrick's reaction levels to include affective and utility reactions, and demonstrated a significant link between utility reactions and job performance. Another taxonomy similar to Kirkpatrick's levels of training evaluation is Bell and Waag's (1998) proposed simulator evaluation model in five stages: utility evaluation, performance improvement (i.e., in-simulator learning), transfer to alternative simulator environment (i.e., quasi-transfer), transfer to flight environment (i.e., transfer of training) and extrapolation to combat environment.

2.2.4 Supporting skill acquisition and retention

There are several aspects of a simulator training program that affects the training efficiency. First and foremost, the choice of pedagogical method is highly important. When introducing any training aid, be it literature, study groups or simulators, choices have to be made regarding pedagogical approach. A simulator cannot simply be added to a training environment without considering how it should be used.

A simulator (as well as any training tool) must fit into the overall pedagogic environment. However, different pedagogical environments require different aspects of a training facility. If one believes that learning takes place inside the student and that teachers are their coaches, a simulator may well be designed to encourage students to explore the possibilities and test the effects of certain actions. On the other hand, in a behavioristic pedagogical environment, the teacher is more important to rule out what is good and what is not good performance and behavior.

In several studies performed at FOI, the effects of embedded training tools have been analyzed (Berggren, Oskarsson, Nählinder & Borgvall, 2005). The studies were performed in a simulator environment called ACES (Air Combat Evaluation System) which is a dog-fight training simulator tailored for research purposes (Nählinder, 2004). The fidelity of the ACES system is far from perfect. The students normally fly a completely different aircraft than the one simulated in ACES, and one could easily expect that the students and instructors would be skeptical using the system. However, these studies performed, all conclude the appropriateness and usefulness of the embedded training tools in ACES, and both student pilots and instructors emphasized the effectiveness of training with them (Nählinder, Berggren & Persson, 2005).

The Swedish Air Force Air Combat Simulation Centre (FLSC) at FOI has had a strong focus on developing a pedagogical team training environment supporting the acquisition and retention of fast-jet pilot decision making, situation awareness and tactical execution skills. The skill acquisition model or philosophy at FLSC is influenced by Kolb's Experiental Learning Theory (Kolb, 1983), in the spirit that learning is the process whereby knowledge and skills are developed through the transformation of experiences. The practical implementation of this model is that the trainees train themselves using the tools provided in the facility under the support and guidance of instructors. One of the most critical parts of the training at FLSC is considered to be the after-action reviews and debriefings. Feedback and knowledge of results has been proven important for both motivation and performance improvement (Holding, 1987), and Freeman, Salter, & Hoch (2004) has declared that feedback in debriefings is particularly important to team training since teamwork itself does not necessarily produce immediate feedback from which the team members can learn during task execution.

The experience at FLSC is that effective debriefings can substantially enhance the effect of simulator-based team training. However, the complexity of scenarios in a multi-ship simulator including computer-generated forces and white force players stresses the need for pedagogical scenario replays, especially when considering that the pilots work through the debriefings themselves (although under the supervision of instructors). To support the debriefings, FLSC has developed a visualization tool that provides the opportunity for objective critique and reflections based on an audio-visual replay of the scenario. With the support of the tool and its embedded visualization aids, examples of good and bad performance can be discussed. Many studies have shown that allowing trainees to practice without specific feedback and guidance of good and bad performance can be detrimental (Wickens & Hollands, 2000) and may produce sub-optimal decision-making skills (Cannon-Bowers & Bell, 1997). So, in that sense, this tool enhances the skill acquisition process since it facilitates the pilots critical thinking about their own performance (Freeman, Salter, & Hoch, 2004), and supports their development of common ground (e.g., Clark, 1996).

2.2.5 Bell and Waags simulator evaluation model

In 1998, Bell and Waag proposed a simulator evaluation model in five stages: utility evaluation, performance improvement (i.e., in-simulator learning), transfer to alternative simulator environment (i.e., quasi-transfer), transfer to flight environment (i.e., transfer of training) and extrapolation to combat environment.

Based on Bell and Waag’s model, the simplest approach for estimating training effectiveness of flight simulation is utility evaluations. These are mainly based ratings by subject matter experts of the effectiveness of the simulation for training on a set of tasks or missions that they have performed in the simulator. The results do not provide quantifiable indices of performance improvement or training transfer. On the other hand, user acceptance of a simulator is of great importance; therefore initial utility evaluations should be performed. The next-step is in-simulator learning which reflects the belief that if performance in the simulator improves, then transfer to the aircraft is likely. Further evidence of training effectiveness is provided if skills are transferred from one simulator to another, which is the next stage of Bell and Waag’s model. This is often called quasi-transfer (e.g., Taylor, Lintern & Koonce, 1993; Brannick, Prince & Salas, 2005). The rationale for quasi-transfer is that specific situations best suited for evaluation, as combat and equipment failure, are not often encountered in practice or may not be suitable for performance measurement. For example, training on specific tasks in a PC based system can be evaluated, in a controlled setting, in a simulator with higher fidelity (Brannick, Prince & Salas, 2005).

The fourth step of Bell and Waag’s (1998) model is transfer of training which requires that improved performance can be shown in the aircraft (or in which ever platform that is studied). Many training researchers believe that transfer of training is the only adequate condition for establishing the effectiveness of simulator training. Finally, for military applications, the highest level of training effectiveness is if the transfer of skills to the real environment can be extrapolated to actual combat. In most cases ultimate training effectiveness criteria are probably met if skills are transferred from a simulator to live performance during peace time training and exercises. If the peacetime training is well-designed and conducted in the spirit “train as we fight”, a similar level of transfer could be expected also to the combat environment.

2.3 Fidelity

A number of factors are believed to affect transfer of training. Most of these factors concern different aspects of the fidelity of the simulation. Apart from the desire to safely train dangerous maneuvers on the ground, the major reason for building flight simulators is to provide training with reduced cost. Therefore, investigating which facets of the simulator that are contributory, or not, to transfer of training, is a crucial issue. High-fidelity simulations generally incur considerable expenses. According to Persing and Bellish (2005) the cost of development of the simulation models increases exponentially as fidelity increases. However, transfer of training does not seem to be a linear function of fidelity. Therefore, to avoid unnecessary expenses, the balance between fidelity and transfer of training should always be considered.

2.3.1 Definition of fidelity

Fidelity is the degree of similarity between the simulator and the equipment that is being simulated. Fidelity can be represented in two dimensions: the degree to which the simulator looks like (physical fidelity) and acts like (functional fidelity) the real operational equipment that is being simulated (Stanton, 1996).

Physical fidelity concerns the graphics of the simulation, such as screen resolution and type of display (e.g., CTR, LCD, plasma, projection, domes and VR displays). It also concerns the realism of instrumentation, flight stick, knobs, levers, and pedals, and the choice between motion-based and fixed or stationary simulation (which is further discussed in chapter 2.3.2 below).

Functional fidelity concerns presentation of proper scenarios, realistic environments, and relevant training tasks. It also concerns the realism and tactics of models (e.g., behavior of computer generated forces, radar models, weapons models, and electronic warfare models).

However, the separation of fidelity into two separate concepts is not unproblematic, since physical and functional fidelity are far from orthogonal. For instance, realistic radio communication in cockpit with air-traffic controllers (ATC) refers to functional fidelity, whereas equipment for sound generation, amplifiers, loudspeakers etc. refer to physical fidelity. However, if the sound system presents engine sound, the sound usually refers to physical fidelity.

2.3.2 Motion-based simulation

Stationary simulators are significantly less expensive than their motion based counterparts. Therefore, whether motion-based simulation contributes to transfer or training, or not, is a key question for optimized cost efficiency of simulator training.

In a series of joint Federal Aviation Administration (FAA)-Industry symposia, subject matter experts from industry, academia, and FAA expressed that they perceived that absence of motion cuing in fixed-base simulators are likely to have detrimental effect on pilot control performance. Especially in maneuvers entailing sudden motion-onset cueing with limited visual reference. However, they presented no scientific evidence that training in a fixed-base simulator would lead to degraded performance in the real aircraft (Bürki-Cohen, Go & Longridge, 2001).

In an extensive meta-analysis spanning two decades of research concerning the effects of motion simulation on training, Bürki-Cohen, Soja & Longbridge (1998), conclude that motion based simulators at that time had no more impact on learning than did motionless simulators. However, they do argue that better motion systems, better display systems and most importantly better synchronization between motion systems and visual displays will call for further studies of the impact of motion.

Bürki-Cohen, Go and Longridge (2001) report on a study performed to further investigate the question of motion based simulation. An experiment was performed in a wide field of view, motion based simulation of a 30 passengers, two engine, turboprop airplane; with experienced airline pilots as participants. For half of the participants, the control group, the motion system was shut off. The transfer of skills acquired by both groups during training was tested in the simulator with the motion system turned on (i.e., quasi transfer). Training and testing included engine failures on take-off with either rejected take-off, or continued take-off in low visibility.

This scenario included maneuvers described in the literature as diagnostic for the detection of motion requirement. No operationally significant support was given for the view that motion based simulation would affect the training progress, that is, transfer of training. Neither did motion affect the pilots', or the instructors/evaluators' subjective perceptions of the pilots' performance or workload, nor did motion affect the pilots' acceptability of the simulator. However, objective measurements of the motion characteristics of the FAA qualified simulator used in the experiment showed that lateral acceleration stimulation was minimal. Subsequent comparisons showed that attenuated lateral acceleration might be typical for the type of simulator that is used in airline training and evaluation (Bürki-Cohen et al., 2003).

In a more recent meta-analysis, Vaden and Hall (2005) found a small but positive effect of simulator platform motion on pilot training transfer. That is, analyzing eleven recent published studies, Vaden and Hall found small evidence that motion is better than no motion. However, they raise the point that the real question is not if simulator platform motion has a positive effect on training or not. Rather the real question is if the effect of motion is large enough to motivate the costs and other disadvantages of such systems. Also, they believe the effect of platform motion is largest at early stages of training, and that there might be other ways to cause the same (small) positive increase on training than by using platform motion.

The Swedish dynamic flight simulator (DFS) is a versatile high performance pilot training and research device. It has the capacity to expose subjects to high levels of G-forces in different directions, usually in the head-foot direction, corresponding to the most frequent occurring accelerations during real flying. In Levin (2006) the feasibility of one of DFS capabilities, the target chase mode, was studied. Although this study did not investigate transfer to real flight, the study shows that the participating pilots were exposed to the desired +Gz-levels when following an aerial target in the target chase mode of the DFS.

2.3.3 Does higher fidelity lead to better training?

What is the connection between fidelity and efficiency of training? It is commonly believed, and often argued, that high fidelity is a prerequisite for the possibility to achieve effective learning in a simulator training environment. Fidelity can be higher or lower in a variety of dimensions. The simulator industry drives the development of better graphics (higher resolution, wider presentation, higher update rate, larger terrain data bases, etc), better motion systems (electrical to replace hydraulics), better sound, better physical similarities (building simulators that look as similar as possible to the real thing) and so on. However, many of these different dimensions, even though having high face validity, are not scientifically proven to increase possibilities of transfer of training or training effectiveness.

On the other hand, several other dimensions do not have high fidelity at all. For instance, weather situations, metrological processes and radio communication are often not simulated realistically. Neither is radio communication which is believed by many civil pilots to draw much cognitive resources, and therefore one might expect that training could improve the operative performance.

Often, a simulators training value is assessed only through the degree of technical fidelity, for instance the size of visual field, latency times of the visual system, motion systems, etcetera. In some cases, a simulator with low technical fidelity can provide excellent training. The opposite is also true – even a very high fidelity top-of-the-line simulator might lack the possibility of getting the user involved to such a degree that meaningful training can be attained. However, in order to assess a training simulators' fidelity from a human-factors point-of-view, the user of the simulator must be considered (Bell & Waag, 1998; Salas, Bowers & Rhodenizer, 1998; Longridge, Bürki-Cohen, Go & Kendra, 2001).

It seems fidelity and its relation to training is a crucial but highly contextual issue. In order to avoid over-investment, simulator designers and developers has to determine appropriate levels of fidelity for achieving desired training effects. This philosophy is often referred to as targeted fidelity (the terms selected fidelity and tailored fidelity also occur). Even though most simulator developers and training researchers probably would agree to this approach there are at least two factors that limit its application: commercial interests (i.e., the simulator industry) and the challenges of establishing appropriate levels of fidelity prior to or early in the development process.

In order to determine the appropriate level of fidelity in all various aspects of a simulation environment, a carefully conducted training need analysis (TNA), or similar methodological approach, including proper elements of fidelity analysis currently seem to be the most fruitful. However, in many training simulator development projects the up-front analyses are often minimal and hence, a reliable and comprehensive TNA is rarely conducted. If it actually is, it most often lacks the inclusion of a solid fidelity analysis.

2.4 Examples of transfer of training studies

A review of transfer of training effectiveness in flight simulation between 1986 and 1997 by (Carretta & Dunlap, 1998) consistently showed improved training effectiveness for jet pilots who had trained in the simulator relative to training in the aircraft only. Improved training effectiveness was shown in studies of radial bombing accuracy, instrument and flight control, and landing skills. One study showed that beginning flight students who were given two practice lessons of landings skills in a flight simulator required 1.5 fewer pre-solo flying hours compared to a control group who had been given no simulator training.

2.4.1 Optimization of simulator training time

Time spent in the simulator is often determined by the availability of the simulator or by a pre planned class curriculum, rather than by systematic evaluation of the incremental learning that occurs across time. To optimize training time, a methodology has been developed to support continuous evaluation of performance across trials to identify a ‘plateau’ in the learning curve. Identification of the plateau is based on two parameters: 1) visual inspection that the slope has ‘leveled off’, and 2) mathematical deduction that the variability of the cumulative average between subsequent trials is less than x % for at least y consecutive trials, and that the values are in the last 1/3 of the lower y-axis. The goal of the methodology has been to obtain a time estimate for when trainees have achieved a substantial amount of training in a VE (Virtual Environment), and to avoid drawing conclusions about training amounts at points along a not yet stabilized learning curve. The methodology was developed for and tested in a virtual environment prototype trainer for Military Operation on Urbanized Terrain (MOUT). However, the parameters used may be generalized to other domains with some modification. One of the conclusions was that if performance is measured by many variables, the number of trials required to reach a plateau of all measures may not be the best option, given external resource constraints (Champney, Milham, Bell-Carroll, Stanney & Cohn, 2006).

Rantanen and Talleur (2005) performed a meta-analysis of 19 studies of ground-based flight training performed 1945 to 2005, on transfer of training. The results concerning percent of transfer of training seemed to indicate a clear benefit of using ground trainers.

2.4.2 Transfer between contexts

Training programs in applied environments often embody the expectation that information acquired in a training context will transfer to the operational environment with relatively little difficulty. However, empirical evidence suggests that unless an association is drawn between the training and test experiences, transfer will be less successful.

A study was performed to test training transfer when trainees were reminded of the relationship between the training and the test conditions. The study was performed with a PC-based flight simulator. The participants were divided in three groups (hint, no hint, control) with equal flying experience. Both the hint and the no hint group flew a low altitude mission. After flying they received feedback concerning their performance. The control group flew two left-hand circuits. The following week, all participants performed two low altitude missions in counterbalanced order. One mission was identical to the training mission, and the other was a low altitude mission in another context. In the "hint" group the hint was given that the training performed last week was relevant to this flight.

The results showed a higher mean altitude for the hint group compared to the control group when the identical type of mission was replicated, but no difference for the other context mission. The results indicate that a reminder only impact performance in the same context as where training is performed. The lack of transfer of training for flight in the other context confirms previous research concerning the difficulties of generalization. Of interest in this study is that in absence of a reminder, information from the training did not even transfer to the identical task. The difficulty seems to lie in the incapacity to recall information acquired during training. This suggests that training should be more memorable. The study highlights the need to systematically examine the potential utility of reminders as a base for facilitating the transfer of training (Molesworth & Wiggins, 2006).

2.4.3 Simulated visual cues

A study has been performed that investigated the effect of out-of-the window cues (visual hoops on the flight path) on training novice pilots on a flight simulator, with focus on development of a pedagogical model for training novice pilots. Compared to a control group there was no significant improvement during landing (alignment error, glide slope altitude error, time outside glide slope range). During learning to fly a 360º level turn hoops were given with both low density (8 sec between hoops) and high density (4 sec between hoops). The results showed a significant improvement for low density hoops compared to no hoops (altitude error, time spent outside allowable altitude range); but showed no difference for other measures (bank angle error, time spent outside allowable bank angle range). However, this test was about learning to fly in a simulator. If learning of a maneuver in the simulator is improved, it can be assumed to have positive transfer to real flying. The conclusion of the authors is that further research should address under which conditions out-of the-window visual cues are helpful for training of other basic flight maneuvers (Khan, Rossi, Heath, Ali & Ward, 2006).

2.4.4 CRM training study

A study of Crew Resource Management (CRM) for two person helicopter teams has been performed by Brannick, Prince and Salas. It was a study of quasi-transfer from a PC-based system, with flight simulator software (Artwick, 1989), to a high fidelity (full-motion TH-57) helicopter simulator. In both cases, the pilots communicated with air traffic control (ATC) via standard headsets with microphones. After the PC training an instructor pilot provided the pilot teams feedback on their performance. A control group, that were not given the PC based flight training, spent the same amount of time working together with problem-solving exercises, and a computer video game on the same platform. Participants were U.S. Navy pilots who had completed flight training, and were awaiting assignment to begin training as copilots in helicopter communities. The 96 pilots were randomly assigned to 48 teams (24 teams in each group).

The study showed successful transfer of training on coordination tasks trained on the PC flight simulation, as asking ATC for missing destination limit on clearance for takeoff, and checking legality/prudence of descending. On technical proficiency tasks not trained on the PC simulation, icing problem, boost pump problem, and electrical fire there were no differences compared to the control group.

The study supports the efficacy of CRM training, and that a PC-based flight simulator is useful for training of CRM behaviors. This underscores the utility of relatively inexpensive technology for training teamwork skills. This suggest that PC-based systems can be used to supplement, and perhaps in some cases replace, CRM training in high-fidelity simulators (Brannick et al., 2005).

2.4.5 Transfer of training study at the Swedish Air Force Flying Training School

In a recent study at the Swedish Defence Research Agency (FOI), the Department of Man System Interaction performed a transfer of training study together with the Swedish Air Force Flying Training School (FS). In this (yet to be published) study, transfer of basic maneuvering skills from simulator to real flying was studied. Before training of a specific maneuver in the real aircraft, half of the student pilots received instructor controlled simulator training for that particular maneuver. The other half of the students received no simulator training before flying, but received the same amount of training in the real aircraft. The simulator training was performed in ACES (Air Combat Evaluation System) which is a VR-based dog-fight training simulator tailored for research purposes, with a set of specially designed embedded training tools (Nählinder, 2004).

The study comprised simulator training of five different maneuvers. A sixth maneuver was only trained in the real aircraft. After each training session in the simulator, the student pilots were given an instructor lead debriefing, assisted by the embedded training tools. The study had a longitudinal between-groups design. After each training session of a maneuver in ACES, regular flying training of that maneuver was performed in a real aircraft. Then, the next maneuver was trained in ACES, followed by regular training of that maneuver in the real aircraft, and so on, for all five studied maneuvers. The control group received training in the real aircraft, of the same maneuvers, in the same order.

Before and after each training session, both in ACES and the real aircraft, the students answered questionnaires rating different aspects of difficulty and learning ability of the maneuver. The students who had received training in ACES were also asked some specific questions about the influence, or transfer, of this training. After each training session in the real aircraft, the flight instructors also answered a similar questionnaire, where they rated different aspects of the students' learning abilities and/or difficulties. After the last training session, the students also answered a summary questionnaire, with questions about the complete series of trained maneuvers.

Even though not yet fully analyzed, preliminary results indicate that training in ACES provides enhanced understanding of maneuvering, and particularly in relation to other aircraft. The student pilots, who received the extra simulator training, highly appreciated its value and believed that the simulator training helped them levitate their understanding of spatial understanding of maneuvering in the 3-dimensional space.

2.4.6 Red skies transfer of training study

The US Air Force Research Laboratory (AFRL) and the UK Defence Science and Technology Laboratory (Dstl) have conducted a number of collaborative research activities on Mission Training via Distributed Simulation (MTDS). In one recent effort, AFRL and Dstl studied transfer of training from the simulated spin-up exercise Red Skies to the live exercise Coalition Red Flag that took place at Nellis Air Force Base in March 2005 (Smith et al., 2007). This effort meant that research was extended from MTDS into subsequent live training to investigate training benefits (transfer of training) derived from being involved in the distributed spin-up exercise. During Red Skies, simulator facilities in the UK were connected to the AFRL Distributed Mission Operations (DMO) simulation facility in Mesa, AZ.

Since of the goals of Red Skies was to study transfer from a simulated exercise to an operational event it was important to also match the objectives and scenario characteristics between the two events. The missions flown during the spin-up were designed to be as similar to those anticipated at Red Flag as possible. The actual Air Tasking Order (ATO), Special Instructions (SPINS), and training rules from Red Flag were used during Red Skies, and the missions were flown in a geographical database covering all of the Nellis Air Force Base with target areas and airspace restrictions. Further on, a lot of the planning, briefing, and de-briefing during Red Skies was conducted via video-teleconference systems which often is the case during Red Flag. The AFRL and Dstl researchers, through Red Flag staff, also managed to schedule (task) the USAF and RAF crews to fly together at Red Flag in order to match the live performance with the simulator exercise.

Subjective and objective measures of performance were collected before, under, and after both Red Skies and Red Flag in order to study the transfer effects. In the conclusions of Smith et al. (2007), the matching of objectives and scenario characteristics for the simulated exercise and the live performance together with the matching of participants (i.e., that the same pilots fly in the same missions under the same conditions etc.) are pointed out as important prerequisites for a successful transfer of training study.

2.4.7 PC based simulations

A literature review by Koonce and Bramble (1998) presents a number of studies of successful transfer from PC based simulations to real flying. A study has shown that ten hours practice on a “Space Fortress” computer game improved performance in pilot training for initial flight cadets in the Israeli Air Force. The game did not look like an aircraft cockpit, and had no external view. Probably the principles the students learned had a beneficial effect on their attending to their surroundings in the aircraft (SA), their planning capabilities, and other cognitive components of in-flight performance (Gopher, 1994). One study showed that beginning flight students who were given two hours pre-training with a moderately detailed computer animated landing display, showed savings of 1.5 hours in flying time in the real aircraft (Lintern, Roscoe, Koonce & Segal (1990). In one study training on the Elite PC based flight simulator was compared with training in ILLIMAC, an enclosed cab simulator with full-sized electromechanical instrumentation. After ten hours of instruction on the PC simulator, the participants had significantly better pass/fail rates on the instrument stage check ride than a control group who had received the same amount of training in ILLIMAC (Phillips, Hulin & Landmayer, 1993). Another study compared a computer based-based training device with a government-approved flight and navigation procedures trainer, in the training of private pilots toward their instrument rating. On a check ride, no differences between those who had been trained in the two devices were found. Thus, by using the PC-based device significant cost savings can be made (Oritz, Kopp & Willenbucher, 1995)

A study by Olson and Austin (2005) investigated the effects of PC based flight simulation for training of novice flight students. The PC based simulator included a yoke, throttle quadrant, avionics panel, and rudder panels. Compared to a control group (matched pairs design), there was no significant difference for flying in the real aircraft. However, since some scientist and practitioners may be concerned about negative transfer, the authors conclude that it is an advantage that no evidence of this was found.

2.4.8 Psychophysiological comparisons of live vs simulated flight

Psychophysiological measures can be used to study similarities and differences between simulated and real flight. The idea is that if the psychophysiological responses in a simulator are similar to those in real flight, the simulator produces an environment that triggers the pilots to put as much effort and commitment into their job as they do flying the real thing. In a large study performed by the Swedish Defence Research Agency (FOI) at the Swedish Air force F17 Wing (Magnusson, 2002; Magnusson & Berggren, 2002) this was done. In a later study by FOI, Dahlström & Nählinder (2007) replicated the idea, gaining similar experiences as the Magnusson 2002 study.

One way of analyzing the similarities and differences between real flight and simulated flight is to study the person's reactions in both settings. If he or she reacts in a similar way in the simulator as in real flight, chances are the simulator produces the same amount of mental effort as the real deal, thus indicating a degree of correspondence. If the reactions were completely different, less transfer is to be expected. In this case it is concluded that the two settings were perceived differently by the person and there fore he/she will not put equal amount of effort into the settings.

2.4.8.1 Study at the F17 wing

In a study by FOI at the F17 wing, five fighter pilots from the Swedish Air Force flew the exact same type of mission in a simulator as in real flight. The mission was flown three times by each pilot in the simulator and later three times in real flight. The pilots' heart rate, heart rate variability and eye movements were continuously measured. Analyses of these data indicate that the pilot's psychophysiological reactions are very analogous in the simulator and in real flight, indicating that the pilots invest mental effort into performance in the simulator as they do in real flight.

This validates the use of the simulator, and therefore is a good indication that transfer of training can occur in this simulator. Besides psychophysiological reaction, subjective ratings where gathered. The results from the subjective data clearly support the conclusions from the psychophysiological data (Magnusson & Berggren, 2002). That is, the pilots' subjective ratings on mental workload, situation awareness and performance where indeed similar for each flight segment in real flight as it was in simulated flight.

2.4.8.2 Study at the Lund University School of Aviation

Heart activity, horizontal eye movements, and vertical eye movements were measured (further description on using psycho-physiological methods can be found in Stanton, Salmon, Walker, Baber, & Jenkins, 2005) while the participants flew a one-and-a-half hour flight in the simulator. Each participant answered one questionnaire before the simulated flight, and one after simulated flight. During the simulated flight, an instructor or qualified passenger rated the participants' mental workload during several different phases of the flight. One or two days after the simulated flight, the participants performed a similar flight in a real aircraft, answering questionnaires before and after flight. The same electrode configuration and equipment were used as in the simulated flight.

The heart rates were calculated as average values on two-minute segments centered on each flight phase. An observer either sitting just outside the simulator (in the simulated flights) or on a seat behind the student pilot (in the real flights) recorded the exact time for each flight phase. These times were used as center points for the calculation of average heart rate. These calculations were performed using a software tool developed by FOI (Nählinder, 2006).

In this study, results similar to Magnusson (2002) were found. That is, both psychophysiological data and subjective ratings from the student pilots themselves and instructors and observers indicate that most of the flight phases where perceived similar in real flight as they were in simulated flight. Therefore, it seems in these flight phases, the student pilots were as engaged in the simulator as in real flight.

However, in one phase (engine failure after take-off), reactions and ratings differed greatly, indicating that the students experienced more mental workload and higher stress level in the simulator than in real flight. After examination of the syllabus, the researchers found out that engine failure in real flight is simulated by the instructor putting one of the two engines into idle, whereas engine failure in simulated flight is caused by the instructor actually producing a real (simulated) engine failure. Also, in real flight this practice takes place at safe altitude whereas in the simulator the engine failed shortly after take-off at low altitude and while the aircraft was in climb (Dahlström & Nählinder 2007). There fore, the training of engine failure in the simulator was not similar to the training of engine failure in real flight. However, training engine failure in the simulator might have positive impact on (real) engine failure in real flight.

2.4.9 Re-acquisition study in the Swedish Air Force

During the 1980-ties the Swedish Air Force ordered a study from the Swedish Defence Research Establishment (FOA, the predecessor to the Swedish Defence Research Agency, FOI) on the possibilities of re-acquisition for pilots who had left the Air Force. Six of 16 pilots from a re-acquisition training study in a simulator were selected and provided live operational training (i.e., training in the air). Nineteen intercepts were chosen to constitute the objective of a fully trained J35 Draken pilot. The researchers at FOA developed corresponding simulator training intercepts, analyzed those down to their components, and created checklists of the decisive ones. Sixteen pilots who had not flown the J35 Draken fighter for a time varying from six months to 12 years were chosen for the initial study. Their experience on the specific system (J35F) ranged from 380 to 2400 hours. The pilots were tested on 19 intercepts in the simulator. On the basis of their results an intensive and individualized training program for each pilot, in the simulator, was provided. Six of them were later provided live operational training. (Angelborg-Thanderz, 1989, 1990).

In this second step of the study we performed 154 flight missions and 78 variables were measured. Performance was a crucial measure in the study. From the simulation study we knew that a simulator could very well be used for determining the pilot's skill and knowledge about the system and the direct handling of the tactical equipment. To estimate the pilots' performance and skill in the air was trickier.

Mental workload was considered of specific importance in a study of live flight. It has often been said that there is a genuine difference between simulated and live flight with respect to psychological stress and mental workload. Could the pilot's mental workload, due to lack of training, be so high that he would not be able to use his knowledge, and utilize the training efficiently in the air? That is the main reason for measuring not only aspects of the pilots' performance but also their corresponding mental workload.

In most studies of man-machine systems, one has emphasized cognitive and perceptual functions and neglected emotional and motivational ones. However, there is evidence that those latter aspects are important as well, especially when subjects must work under time pressure and high mental stress during live flights. Under some circumstances they can be the most important aspects. Therefore emotional and motivational factors were considered. The pilots rated their mental effort, motivation, and their moods, in terms of perceived activation, psychological stress, and extraversion. Furthermore, the pilots also rated the difficulty and the risk of the missions. All the judgments and evaluations were made both before and after the flights.

How did the performance in the simulator relate to and/or affect the performance in the air, i.e. to what extent was the simulation training effects transferred to real flight? By means of modeling ad modum LISREL, we estimated the optimal transfer effects. Figure 1 presents a model in which we used the performance ratings of the instructors, the performance ratings of the pilots as well as the pilots' ratings of their mental workload.

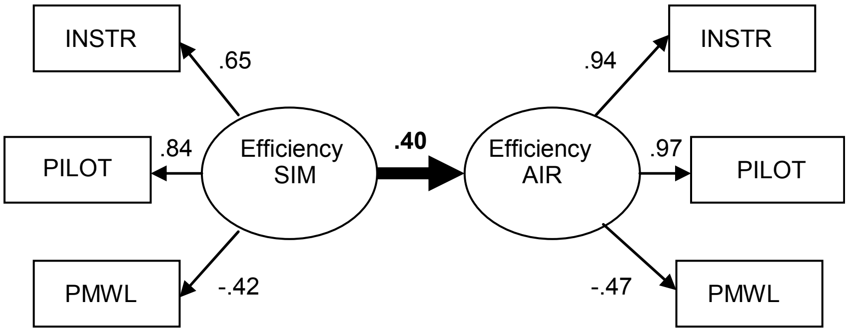


Figure 1. A model of the causal relationships between the factors efficiency in simulation and efficiency in the air (INSTR= Instructor ratings of performance, PILOT=pilot ratings of performance, PMWL = Pilot Mental Workload). $\chi 2 = 13.71$, df = 8, p = 0.09. Root Mean Square Residual (RMR) = 0.09. Goodness of Fit Index (GFI) = 0.94. Adjusted Goodness of Fit Index (AGFI) = 0.83.

As can be seen from the figure, the three markers form a factor called efficiency. The factor loadings for the performance measures are positive and the loadings for mental workload are negative. Accordingly, we have an efficiency measure ranging from high performance during low mental effort to bad performance during high mental effort. From the model we can see that there is a significant transfer effect (.40) from simulation to live flight. This means that 16 percent of the variance of operational performance in the air could be explained by the variance of performance in the simulator.

In a second model of transfer effects, we included emotional and motivational aspect. Also a factor called ‘fighting spirit’ with markers of the moods of activation and extraversion was formed.

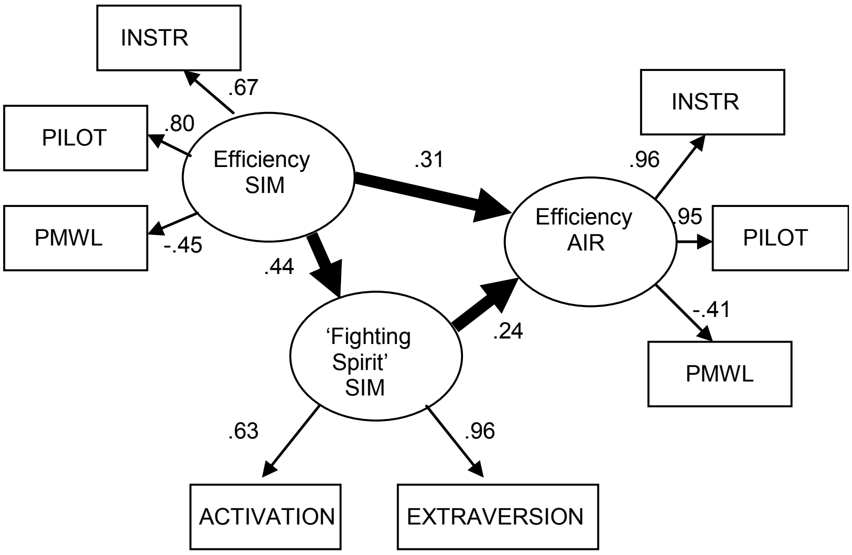


Figure 2. A model of the causal relationships between the factors efficiency in simulation, ‘fighting spirit’ in simulation, and efficiency in the air. $\chi 2$ = 34.27, df = 17, p = 0.01. Root Mean Square Residual (RMR) = 0.09. Goodness of Fit Index (GFI) = 0.88. Adjusted Goodness of Fit Index (AGFI) = 0.74.

From the model in figure 2, we found that efficiency in the air was affected not only directly from the efficiency in the simulator, but also indirectly by the ‘fighting spirit’ factor. The combined effects of efficiency in the simulator and ‘fighting spirit’ on efficiency in the air are .47. This means that 22 percent of the variance in efficiency in the air was explained by the variances in efficiency in the simulator and ‘fighting spirit’. Accordingly, good performance in the simulator resulting in a high ‘fighting spirit’ has a higher predictive power than the exclusive efficiency factor.

Our early studies in the Swedish Air Force indicated problems with respect to the applicability of classical experimental designs and optimal utilization of subjects, and over the years we have adapted designs and analyses techniques to the specific requirements of the situation. Classical experimental control groups designs have been and are the standards recommended for studies in applied settings. However, classical designs are basically developed for experiments i.e. situations where one or a few aspects vary while all others are controlled for. Using strict classical designs in applied settings with high complexity will often influence the realism and the authenticity negatively. Accordingly, designs of studies in applied settings have to compromise in order to balance scientific stringency and situational authenticity. These inconsistencies were apparent in our applied research performed in the Swedish Air Force. A specific complication concerns the experimental cases i.e. the subjects used. In applied situations, as military flight operations, the number of subjects available is comparatively low, and generally, the number of cases does not satisfy statistical demands. Repeated measurements designs represent here a practicable solution, and by means of these designs the number of measurement points will fulfill scientific requirements. There is a fundamental difference between classical and repeated measurement designs with respect to the origin of variance: The first is based on inter-individual variance and the second on both inter- and intra-individual variance. The difference may influence the possibilities to generalize from sample to population. An obvious advantage of using repeated measures is that the intra-individual variance represents changes over time, i.e. describes dynamic changes or processes. Military flight operations are dynamic processes indeed, and accordingly repeated measures designs are to be preferred. In classical designs, groups are compared with respect to different characteristics at a specific point in time or time period. In repeated measures designs, on the other hand, the co-variations between a large numbers of different dynamic variables can be compared. From these co-variations the number and content of latent variables or factors can be extracted by means of factor analyses. These factors then form base of causal models ad modum LISREL. By means of these models the effects (in terms of explained variance) of different aspects on dependent factors can be estimated.

2.4.10 Motion in flight control studies

In a two-part Swedish study of simulated landings in turbulence conducted by the Swedish Defence Research Establishment (FOA, FOIs predecessor) motion was compared to no-motion (Svensson & Angelborg-Thanderz, 2000). The objective of the study was to analyze whether a moving base dome simulator could produce more realistic cues than a fixed base simulator. In the first part of the study half of 60 landings were performed with the motion system disengaged. The pilots rated risk, difficulty, workload, performance, handling qualities, and pilot induced oscillations. A repeated measurement design was used, and two experienced test pilots performed 30 simulated offset landings, each under five different levels of turbulence.

The conclusions were that stick activity and difficulty were rated higher and performance rated lower under the motion condition. The 'handling qualities' tended to be lower, and the perceived 'risk of an accident' tended to be higher during the same condition. Thus, in the flight task analyzed, the pilot's stick inputs were reduced, when the motion system was disengaged. Accordingly, from the reduced stick inputs under these conditions, one runs the risk of overestimating the handling qualities of the real aircraft. All correlations between 'performance' and the other variables were significant under the 'motion' condition, but none were significant under the 'no motion' condition. These findings are of special importance with respect to the questions addressed. The pilot's acquisition of skill is based on his performance feedback, i.e., his ability to correctly perceive his performance. Lack of relevant cues diminishes his possibility to get performance feedback. Furthermore, this lack of relevant non-visual cues makes a positive transfer of training from simulations to real landings less likely.

In the second part of the study four experienced test pilots performed 48 simulated landings, each under five different levels of turbulence. The correlation matrix from the 'motion' condition of this study was compared with the corresponding matrices ['no motion' respectively 'motion'] in the first part of the study. Directly after each landing the pilots rated the same aspects as previously, as well as their 'fatigue' and 'psychological stress'. Furthermore, a combination of aircraft motions (in pitch, yaw, and roll) at touch down was added as an objective measure of performance. In this part of the study the importance of motion in turbulent landings was verified and analyses of the pilots' control responses showed that there were inter-individual differences.

When comparing the matrices a close correspondence between the 'motion' conditions of the two parts of the study was found. The relations between 'performance' and the other variables were almost identical under the two 'motion' conditions. These relations are important, because the pilots’ acquisition of skill is based on their performance feedback, i.e., their ability to correctly perceive their performance. The conclusion was that motion produces more realistic and useful cues (i.e., adequate sensory feedback) to the pilot than 'no motion'.

In the first part of the study significant differences between the pilots with respect to 'stick activity' were found. This inter-individual difference in 'stick activity' was verified in the second part. Furthermore, there were differences in the pilots’ increase in gain as a function of 'turbulence'. Two of the pilots presented a linear increase in gain, while two presented a decelerated increase, when the 'turbulence' was high. By means of partial correlation analyses a genuine increase in 'fatigue' as a function of the 'landing sequence' (when the effects of 'turbulence' were nullified) was found. In the same way a genuine increase in 'performance' and a decrease in 'stick activity' as a function of the 'landing sequence' (figure 3) was found. Our conclusion is that these changes reflect learning processes and that the pilots change their techniques to cope with the landings. It is interesting to note that 'stick activity' changes over the landings, as it is often considered as a stable characteristic.

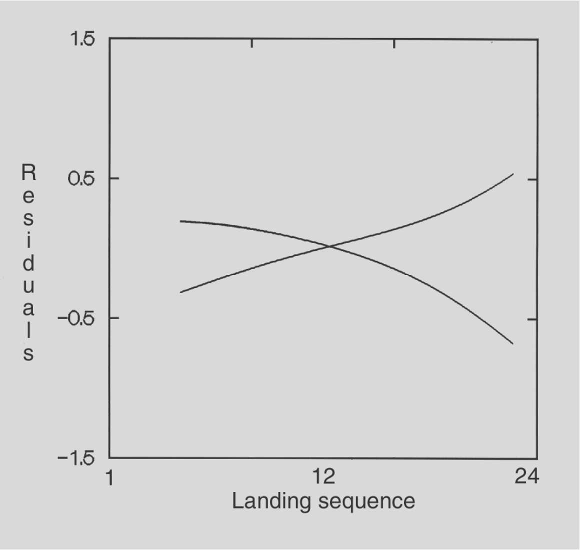


Figure 3. Changes (z-scores) in the residuals for perceived performance (increasing curve) and stick activity (decreasing curve) as a function of landing sequence, when the effects of turbulence are held constant. The curves have been smoothed by means of distant weighted least squares regression.

The pilots of the studies were experienced and their learning curves should have leveled out. However, the changes in the residuals for 'performance' and 'stick activity' as a function of 'landing sequence' seem to contradict this assumption, and the curves reflect an accelerating learning process?

Model analyses showed that turbulence affects workload, and that workload, in its turn, influences performance. It was found that handling qualities and induced oscillations could be predicted from the other variables. The variables turbulence and motion of the aircraft explains 65 percent of the variance in handling qualities ratings and 36 percent of the variance in ratings of pilot induced oscillations. Accordingly, handling qualities and pilot induced oscillations can be estimated and predicted in situations ‘without man in the loop’. The estimates are of practical interest because they can be used as guides by systems developer.

Factor analyses indicated that eight of the manifest or measured variables could be reduced to two latent factors: 'workload' and 'efficiency'. The reliability of the factors was established and both are combinations of subjective ratings and objective measures, which support their validity. The causal relationships between the independent variable ''turbulence', and the two factors 'workload' and 'efficiency' were tested statistically by means of Structural Equation Modelling, see figure 4.

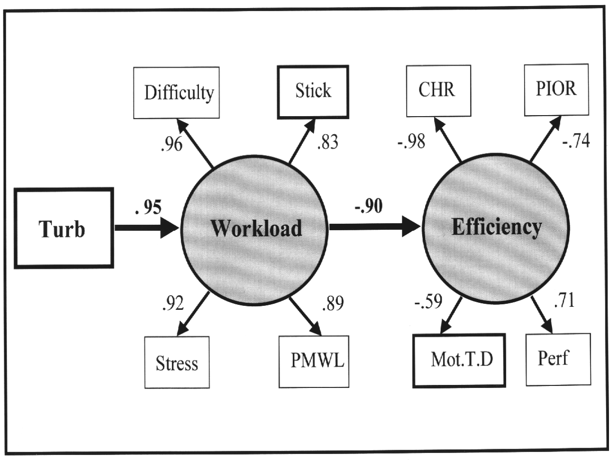


Figure 4. Structural equation model of the relationships between the nine manifest variables. Adjusted Goodness of Fit Index (AGFI) = .81. Rectangles indicate manifest variables and ellipses denote factors. Thick rectangles indicate objective measures. Thin arrows indicate factor loadings and thick arrows causal effects. TURB=turbulence, PMWL=pilot mental workload, CHR=Cooper-Harper scale, Mot. T.D.= motion at touch down. All effects and factor loadings are significant (p< .001).

As can be seen from the model, increases in 'turbulence' are followed by increases in 'workload'. Increases in 'workload' are, in their turns, followed by decreases in 'efficiency'. Thus, 'workload' mediates the effects of 'turbulence' on 'efficiency'. The plots of empirical data indicated that low 'workload' predicts high 'efficiency', but also that high 'workload', not by necessity, predicts low 'efficiency.' The variance in 'efficiency' increases as a function of 'workload'. The increased variance shows that the precision in the predictions of the pilots' 'performance' decreases, when the 'workload' increases.

2.4.11 Utility evaluation study

Bell and Waag (1998) has suggested a simulator evaluation model in five separate stages: utility evaluation, performance improvement (i.e., in-simulator learning), transfer to alternative simulator environment (i.e., quasi-transfer), transfer to flight environment (i.e., transfer of training), and extrapolation to combat environment. Focusing on the first stage of the model, utility evaluation, the purpose is to a) evaluate the accuracy, or degree of fidelity, of the simulator environment, and b) gather user opinions regarding the user acceptance and potential value of the simulator as a future training environment. The idea is that user acceptance is a necessary but not satisfying condition for a system to hold a potential as an effective training media. Monitoring user acceptance and expert opinions of a system and its parts, is according to Bell and Waag, a central aspect of modern simulator evaluation and design. Hence, even though utility evaluation based on user opinion data does not provide a quantitative measure of neither performance improvement or transfer of training, it provides valuable support about potential training value, future development needs and more rigorous evaluations.

Influenced by Bell and Waag's approach, (Borgvall, in press) conducted a simulator prototype evaluation study based on pilot assessment of utility and functional fidelity at the Swedish Defence Research Agency (FOI). The study was incorporated in a development project of a simulator prototype for within-visual-range air-to-air combat training. It was conducted to optimize continued development in relation to designated effects. Following scenarios in the simulator, participants completed a questionnaire evaluating different aspects of the simulation such as visual feedback, instrumentation, flight controls, graphics, and field-of-view. This gave valuable support for which aspects that should be prioritized during the next design cycle. The results generated recommendations for continued simulator design together with directives and criteria for continued evaluation.

2.4.12 Mission training via distributed simulation

One Swedish simulator facility where several demonstrations of distributed simulations have been performed is the Swedish Air Force Air Combat Simulation Centre (FLSC). The facility provides multi-ship flight training where visiting teams of pilots and fighter controllers train decision-making, situation awareness, and tactical execution skills. The visiting teams gain experiences both in traditional BVR situations and in Coalition Peace Support Operations (PSO) on an international arena, with a current emphasis on the latter type of scenario. The PSO training scenarios at FLSC were carefully designed and evaluated with the support of senior pilots possessing extensive experience from operational deployments, before implemented in the training program. To further develop the training, the Mission Essential Competencies (MEC) (Colegrove & Alliger, 2003) required for Swedish PSO Air to Air missions (Bennett et al., 2006, Borgvall & Castor, 2006; Borgvall, Castor & Lavén, 2006; Borgvall, Castor & Lavén, 2007) have recently been mapped with the support of Air Force Research Laboratory (AFRL/HEA, Mesa, AZ) Together, the Swedish Air Force, FOI and AFRL mapped knowledge and skills essential for SwAF PSO and their developmental experiences across environments were identified and evaluated them with regard to current training.

A number of distributed events between FLSC and AFRLs simulator facility in Mesa have been planned as focal points of the activities in the International Mission Training Research Project Agreement (IMTR PA). During the IMTR PA, methods, tools, and strategies for effective and efficient Mission Training via Distributed Simulation (MTDS) will be tested and evaluated.

3 METHODOLOGICAL ADVICE

Use the lessons learned in human performance measurement research

Given that human behavior is complex and adaptive and often hard to measure, a battery of different measures is usually needed to assess human performance. This also applies to transfer of training studies. A survey of human performance measurement literature (e.g., Castor et al., 2003; Wilson et al., 2002; Alfredson, Oskarsson, Castor & Svensson, 2003) reveal a wide variety of measures that have been used in scientific human performance studies. A high level categorization of different approaches to human performance measurement consists of the following types of measures:

- Subjective measures (own ratings; instructor and/or observer ratings)
- Psychophysiological measures (heart activity; eye activity; brain activity)
- Logging measures (primary task measures; secondary task measures; behavior measures such as checklist execution)
- Analysis of verbal communication

Each of the measures above have shortcomings and methodological problems, however in combination they can provide a robust and reliable picture of human performance. It is important to remember that it is not only the end-state performance that is significant, but rather the entire process leading up to that state that is important to study.

A recent review of FOI research activities in the measurement areas listed above is provided in Castor & Alfredson (2006). A transfer of training study will use some types of these different measures, as exemplified by the studies describes in section 2.4. A trade-off between the intrusiveness with regard to task performance and the possibilities to manage the data collection versus the need for data has to be made. Subjective ratings and questionnaires are useful tools in a transfer of training study, but they must be carefully designed in order to actually address transfer effects. To get the best possible understanding of training effects and transfer effects, a combination of objective, quantifiable and qualitative measures should be used.

Not everything can be trained in a simulator, and not everything is trained even though it could be trained in a simulator. It is however important to be aware of the simulator environments limitations regarding various aspects of training. What should be trained in a simulator and what is not trained?

Perform a thorough analysis of training needs

The fundament of any transfer of training study is a thorough analysis of training needs and the current competence status of the persons receiving the training. Examples of two suitable methods for this are Training Needs Analysis (TNA) (e.g., JSP 502, 2001) and Mission Essential Competencies (MEC) (Colegrove & Alliger, 2002). The analysis of the tasks to be trained should entail the basis for the dimensions along which the transfer effects are assessed. In order to follow the transfer effects and the skill development, efforts should be invested in a decomposition of the relevant tasks so that transfer effects for different types of skills and parts of full skill sets can be studied. If, for example, the training goals of an air force squadron are described on a high level, it might be beneficial for research purposes to support the instructors decomposition of the training goals so that goal achievement and transfer effects for each detailed subgoal can be analyzed.

Manage the lack of a control group

In an ideal transfer of training study a classical experimental design, a control group not receiving any training would be used. When studying transfer outside laboratory settings, for example the transfer effects from a simulated spin-up exercise before a live exercise for fast-jets, all participants normally, and ultimately, receive the spin-up training (since the reason for a spin-up exercise is to prepare each individual for a live event, any support for manipulating the participation for research purposes is highly unlikely). Accordingly, the classical experimental-control group designs are normally not applicable under operational settings on the field when it comes to transfer of training. Assuming that all trainees, to various extents have trained in simulators, the following design seems to be preferred:

- Identify aspects of critical importance for operational performance.
- Identify measures of these aspects and use these measures during simulation as well as in the air.
- If there are true relationships between the performance status in the simulator and performance status in the air for the critical aspects, significant correlations will be found over the trainees' status levels during simulation and real flight.
- In subsequent training, relative changes based on the initial performance measures can be used, and, in the same way, correlations be extracted.
- By means of this procedure differences in transfer effects at different skill- or performance levels can be disclosed by curve-linear relations between performance in simulation and air.
- The models of transfer presented in section 2.4.9 in part represent this design.

Psychometrical considerations

Psychometrically adjust for what sometimes has been called beta effects of an intervention, i.e. when the intervention or in this case the training, presumably may change the whole frame of reference for the subjects. So instead of just one rating of performance concerning some aspect before training and then one rating after training, the rating afterwards should be split into two ratings. One should be a regular rating of performance performance, but when the subject now sees his or her performance before training in a new light, a rating of performance before the intervention or training should be made. The realization that no such thing as totally non-intrusive measurement exists must also be made. If subjective questions of learning and transfer effects are used, this will to some extent affect the learning process of the subjects under study.

Operationalize training effects on all Kirkpatricks levels

A point in Alliger et als (1997) paper is that very few studies that describe the relation between the training effects on Kirkpatrick's different levels have been presented in the scientific literature. The optimal study design should include thus include operationalization of all of Kirkpatrick's levels (see section 2.2.3), with Alliger et al. extensions, in order to contribute with empirical data of how the effects on the levels actually relate to one another.

4 DISCUSSION

During the survey of previous effort into training of transfer issues a number of theoretical considerations and questions have been discussed by the authors of the report.

- In earlier Swedish analyses of transfer effects, see for example section 2.4.9, it has been found that the pilots' efficiency in the simulator, to some extent, can explain the efficiency in the air. In these analyses, and with the restricted arsenal of measures available, about 20 percent of the variance could be explained. This might not seem impressive, but an explanatory power of 100 percent is not to be expected. What is the maximum variance explainable? Well, the maximum explained variance is restricted by the reliability of measures used. If, for example, the reliability indices of the performance and workload measures is 0.80, the maximum expected correlation between the measures in the simulator and the corresponding measures in the air is 0.64 (i.e., 0.802). Accordingly, a more appropriate estimate of the proportion transferred is 0.40/0.64, i.e., 0.63. Seen in this perspective 20 percent is considerable, and of practical importance.
- Is there any distinction between what historically has been studied in transfer of training studies and studies of the effects of mission rehearsal or "spin-up" in simulators? During a mission rehearsal the particulars of a specific missions are implemented in the simulator (e.g., a geographical database of the mission area and specific Air Tasking Orders, ATO), while transfer of training studies have the focus on the more general tasks and subtasks that are need in order to execute a mission? The answers to these questions affect how training effects can and should be measured.
- To what extent is it valid to discuss a conceptual construct called team transfer? Can a team or military unit be considered as a "knowledge container" that permits indirect transfer effects? If a number of pilots in a unit receive training on a certain task or the execution of a mission during a spin-up or mission rehearsal, they will probably share this knowledge with their colleagues. Thus, given that the unit and its performance is considered as the most appropriate level of analysis, questions of team transfer will emerge as important. For example, most often it is not a full squadron coming to FLSC for training but rather some portion of it. Is it reasonable to believe that skills that are transferred to live settings for the individuals that have taken some tailored training also might be mediated to individuals who have not yet taken the training but that act together with the more experienced individuals?
- The simulator industry is growing, as the technical developments and reduced costs make simulation more accessible for training purposes. However, it is interesting to note that the emphasis on development of graphical systems and motion systems continues, even though these aspects of the simulation are not always the most important ones, especially for many training purposes. As noted by Bürki-Cohen et al. (2001), the lack of realistic radio communication is a deficiency in current full flight simulators for airline pilots. The observation that pilots, after completed training in expensive simulators, still need to develop new skills for some procedures, components or systems (e.g., radio communication procedures) when they start their regular service is one example of this problem. Methods for assessing level of fidelity required for different aspects of a training simulator are a very important topic that should receive higher priority. This is often called targeted or selective fidelity in the literature. But deliberately allowing lower levels of selective fidelity is not often used in procurement processes. In this area, human-factors evaluation and research could be of great use.

- The authors of this report believe simulators should not be limited to perform training of certain specific maneuvers and events. Rather, it might be of great benefit to allow trainees to explore outcomes of actions not suitable for testing in real-life setting. That is, simulators should be used to let students test any wild and crazy ideas they might have in order to fully understand what the effects of certain behaviours are. Taken to its extreme, one might even consider letting the trainees use the simulator freely to play with, explore and test whatever they want to try out. Providing high accessibility and positive attitude towards this type of activity might improve simulator training efficiency further.

5 REFERENCES

Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. *Journal of Applied Psychology*, 77, 598-614.

Alexander, A.L., Brunyé, T., Sidman, J., & Weil, S.A. (2005). *From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in PC-based simulations and games*. http://www.darwars.net

Alfredson, J., Oskarsson, P-A., Castor, M., & Svensson, J. (2003). Development of a meta instrument for evaluation of man-system interaction in systems engineering. In *Proceedings of 35th Annual conference of the Nordic Ergonomics Society*. Reykjavik, Aug. 10-13, 2003.

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Jr., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology, 50,* 341-358.

Angelborg-Thanderz, M. (1989). Assessing pilot performance in training simulators: A structural analysis. *In proceedings of The Royal Aeronautical Society spring convention, Flight simulation*: Assessing the benefits and economics. London, United Kingdom: The Royal Aeronautical Society.

Angelborg-Thanderz, M. (1990). *Military flight training at a reasonable price and risk*, Economics Research Institute, Stockholm School of Economics and FOA report C 50083 - 5.1. (report in Swedish, summary in English).

Arthur, W. Jr., Bennett, W. Jr., Stanush, P. L., & McNelly, T., L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance*, 11, 57-101.

Artwick, B. (1989). *Microsoft flight simulator* (Version 4.0) [Computer software]. Redoman, WA: Microsoft.

Bell, H. H., & Waag, W. L. (1998). Evaluating the effectiveness of flight simulators for training combat skills: A review. *The International Journal of Aviation Psychology*, 8(3), 223-242.

Bennett, W., Jr., Borgvall, J., Lavén, P., Castor, M., Gehr, S. E., Schreiber, B., Alliger, G., & Beard, R. (2006). International mission training research (IMTR): Competency-based methods for interoperable training, rehearsal and evaluation. Proceedings of the European Interoperability Workshop (EuroSIW) 2006. Stockholm, Sweden: Simulation and Interoperability Standards Organization.

Berggren, P., Oskarsson, P-A., Nählinder, S., & Borgvall, J. (2005). E*xperimentell procedur vid utvärding av ACES.* FOI-R--1688--SE. FOI: Linköping, Sweden.

Borgvall, J. (in press). Creating recommendations for simulator design through pilot assessment of prototype utility. *Military Psychology*.

Borgvall, J., & Castor, M. (2006). Mission Essential Competencies: Kompetens – Operativa Krav. FOI-R--2106--SE [In Swedish]. Linköping: Institutionen för Människa-System Interaktion.

Borgvall, J., Castor, M., & Lavén, P. (2006). JAS39A PSO A-A MECs. FOI-RH--0607--SE [In Swedish]. Kista: Avdelningen för Stridssimulering (FLSC)/Institutionen för Människa-System Interaktion.

Borgvall, J., Castor, M., & Lavén, P. (2007). Träningsbehov JAS39A jaktroll PSO mission. FOI-RH--0679--SE [In Swedish]. Kista: Avdelningen för Stridssimulering (FLSC)/Institutionen för Människa-System Interaktion.

Brannick, M. T., Prince, C., & Salas, E. (2005). Can PC-based systems enhance teamwork in the cockpit? *The International Journal of Aviation Psychology*, 15(2), 173-178.

Bürki-Cohen, J., Soja, N. N., & Longbridge, T. (1998). Simulation platform motion – The need revisisted. *The International Journal of Aviation Psychology*, 8(3), 293-317.

Bürki-Cohen, J., Go, T. H., Chung, W. W., Schroeder, J., Jacobs, S., & Longridge, T. (2003). Simulator fidelity requirements for airline pilot training and evaluation continued: An update on motion requirements research. *Paper presented at the 12th International Symposium on Aviation Psychology*, Dayton, OH, USA.

Bürki-Cohen, J., Go, T. H., & Longridge, T. (2001). Flight simulator fidelity considerations for total air line pilot training and evaluation. *Paper presented at the AIAA Modeling and Simulation Technologies Conference and Exhibit* 6-9 August 2001, Montreal, Canada.

Cannon-Bowers, J. A., & Bell, H. H. (1997). Training decision makers for complex environments: Implications of the naturalistic decision making perspective. In C. E. Zsambok & G. Klein (Eds.), Naturalistic decision making (pp. 99-110). New York, NY: Lawrence Erlbaum Associates.

Castor, M. & Alfredson, J. (2006). *MSI Metodutveckling Slutrapport 2004-2006*. FOI-R--2213--SE. FOI Linköping. (Report in Swedish, summary in English).

Castor, M., Hanson, E., Svensson, E., Nählinder, S., LeBlaye, P., MacLeod, I., Wright, N., Alfredson, J., Ågren, L., Berggren, P., Juppet, V., Hilburn, B., & Ohlsson, K. (2003). *GARTEUR Handbook of Mental Workload Measurement*. GARTEUR TP 145.

Carretta, T. R., & Dunlap, R. D. (1998). *Transfer of training effectiveness in flight simulation: 1986 to 1997* (AFRL-HE-AZ-TR-1998-0078). Mesa AZ: United States Air Force Research Laboratory, Human Effectiveness Directorate.

Champney, R., Milham, L., Bell-Carrol, M., Stanney, K. M., & Cohn, J. (2006). A method to determine optimal simulator training time: Examining performance improvement across the learning curve, *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, San Francisco, CA (pp. 2654-2658). Santa Monica, Ca: Human Factors and Ergonomics Society.

Clark, H. H. (1996). Using language. Cambridge: Cambridge University Press.

Colegrove, C. M., & Alliger, G. M. (2003). Mission essential competencies: Defining combat readiness in a novel way. In proceedings of NATO Research and Technology Organization (RTO) System Analysis and Studies Panel's (SAS) Symposium, Air *Mission Training Through Distributed Simulation (MTDS) – Achieving and Maintaining Readiness (SAS-038).* Brussels, Belgium: NATO RTO.

Dahlström, N., & Nählinder, S. (2007). Mental Workload in Simulator and Aircraft during Basic Civil Aviation Training. SUBMITTED to *International Journal of Aviation Psychology.*

Ellis, H. (1965). *Transfer of Learning*. New York: MacMillan.

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, 49(8), 725-747.

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

Facteau, J. D., Dobbins, G. H., Russell, J. E. A., Ladd, R. T., & Kudisch, J. D. (1995). The influence of general perceptions of the training environment on pretraining motivation and perceived training transfer. *Journal of Management*, 21, 1–25.

Freeman, J., Salter, W. J., & Hoch, S. (2004). The users and functions of debriefing in distributed, simulation-based team training. Proceedings of the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC) 2004. Orlando, FL: National Security Industrial Association.

Goettl, B. P., & Shute, V. J. (1996). Analysis of Part-Task Training Using the Backward-Transfer Technique. *Journal of Experimental Psychology: Applied*. Vol. 2. No. 3,227-249.

Goldstein, I. (1980). Training In Work Organisations. *Annual Review of Psychology*, 31, 229-272.

Gopher, D. (1994). Transfer of skill from a computer game trainer to flight. *Human Factors*, 36, 387-405.

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

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

Holyoak, K. J. (1991). Symbolic connectionism: Toward third-generation theories of expertise. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise (pp. 301-336)*. Cambridge, United Kingdom: Cambridge University Press.

JSP 502. (2001). The Tri-service guide to training need analysis for acquisition projects.

Khan, M. J., Rossi, M., Heath, B., Ali, S. F., & Ward, M. (2006). An experimental study of the effect of out-of-the window cues on training novice pilots on a flight simulator, *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, San Francisco, Ca (pp. 2659-2663). Santa Monica, Ca: Human Factors and Ergonomics Society.

Kirkpatrick, D.L. (1994). *Evaluating Training Programs: The Four Levels.* San Francisco, CA: Berrett-Koehler.

Kolb, D. A. (1983). *Experiental learning: Experience as the source of learning and development*. Upper Saddle River, NJ: Prentice Hall Inc.

Kozlowski, S. W. J. (1998). Training and developing adaptive teams: Theory, principles, and research. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 115-153). Washington, DC: American Psychological Association.

Koonce, J. M., & Bramble, W. J., Jr. (1998). Personal computer-based flight training devices. *The International Journal of Aviation Psychology*, 8(3), 227-292.

Levin, B. (2006). Applications of pilot training in the Swedish dynamic flight simulator, *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, San Francisco, Ca (pp. 1982-1985). Santa Monica, Ca: Human Factors and Ergonomics Society.

Lintern, G., Roscoe, S. N., Koonce, J.M., & Segal, L. (1990). Transfer of landing skills in beginning flight training. *Human Factors*, 32, 319-327.

Longridge, T., Bürki-Cohen, J., Go, T. H., & Kendra, A. J. (2001). Simulator Fidelity Considerations for Training and Evaluation of Today's Airline Pilots. P*aper presented at the 11th International Symposium on Aviation Psychology*, Columbus, OH.

Magnusson, S. (2002). Similarities and Differences in Psychophysiological Reactions Between Simulated and Real Air-to-Ground Missions. *International Journal of Aviation Psychology*, Vol. 12, No. 1.

Magnusson, S., & Berggren, P. (2002). Dynamic assessment of Pilot Mental Status. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* Baltimore, MD. Santa Monica, CA: Human Factors and Ergonomics Society.

Molesworth, B. R. C., & Wiggins, M. W. (2006). Limitations of skill generalization beyond training*, Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, San Francisco, Ca (pp. 2620-2623). Santa Monica, Ca: Human Factors and Ergonomics Society.

National Transportation and Safety Board. (2001). NTSB-ID: LAX01LA150. *NTSB Factual Report Aviation,* NTSB, USA.

Nählinder, S. (2004). *ACES - Air Combat Evaluation System.* FOI-R--1368--SE. FOI Methodology Report. FOI: Linköping, Sweden.

Nählinder, S. & Berggren, P. & Persson, B. (2005). Increasing Training Efficiency Using Embedded Pedagogical Tools in a Combat Flight Simulator. *Proceedings of the Human Factors and Ergonomics Society. 49th Annual Meeting*. Orlando, FL: Human Factors and Ergonomics Society.

Nählinder, S. (2006). A *Human-Factors Perspective on Simulator Fidelity Assessement.* FOI-R--2047--SE. FOI Scientific report, FOI: Linköping, Sweden.

Olson, R., & Austin, J. (2005). A step toward early PC-based training that reduces risk: The effects of practicing an "instrument referenced" skill pattern on "visually referenced" performance of beginning flight students*. Journal of Aviation/Aerospace Education and Research*, 14(2), 47-51.

Oritz, G. A., Kopp, D., & Willenbucher, T. (1995). Instrument training using a computer-based simulation device at Lufthansa Airlines. In R. S. Jensen & L. A. Rakovan (Eds.), *Proceedings of the Eighth International Symposium on Aviation Psychology* (pp. 1047-1050). Columbus.: The Ohio State University.

Persing, T. R., & Bellish, C. A. (2005). An investigation of the impact of aerodynamic model fidelity on close-in combat effectiveness prediction in piloted simulation. *Paper presented at the AIAA Modeling and Simulation Technologies Conference and Exhibit*, 15-18 August 2005, San Francisco, California.

Phillips, S. I., Hulin, C. L., & Lamermayer P. J. (1993). Uses of part-task trainers in instrument flight training. In R. S. Jensen & D. Neumeister (Eds.), *Proceedings of the seventh International Symposium on Aviation Psychology* (pp. 743-746). Columbus: The Ohio State University.

Rantanen, E. M., & Talleur, D. A. (2005). Incremental transfer and cost effectiveness of ground-based flight trainers in university aviation programs, *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*, September 26-30, 2005, Orlando, Florida (pp. 764-768). Santa Monica, Ca: Human Factors and Ergonomics Society.

Roscoe, S. N. (1971). Incremental transfer effectiveness. *Human Factors*, 13(6). 561-567.

Roscoe, S. N. & Williges, B. H. (1980). Measurement of transfer of training. In S. N. Roscoe (Ed.), *Aviation Psychology* (pp. 182-193). Ames: Iowa State University Press.

Salas, E., & Cannon-Bowers, J.A. (2001). The Science of Training: A decade of progress. *Annual Review of Psvchology*, 52, 471-499.

Salas, E., Bowers, C. A., & Rhodenizer, L. (1998). It Is Not How Much You Have but How You Use It: Toward a Rational Use of Simulation to Support Aviation Training. *International Journal of Aviation Psychology*, 8(3), 197-208.

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

Smith, E., McIntyre, H., Gehr, S. E., Schurig, M., Symons, S., Schreiber, B., & Bennett, Jr., W. (2007). Evaluating the impacts of mission training via distributed simulation on live exercise performance: Results from the US/UK "Red Skies" study. AFRL-HE-AZ-TR-2006-0004. Mesa, AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Stanton, N. (1996). Simulators: a review of research and practice. In N. Stanton (Ed.), *Human factors in nuclear safety* (pp. 117-140). London: Taylor & Frances.

Stanton, N. A., Salmon, P. M., Walker, G. H., Baber, C., & Jenkins, D. P. (2005). *Human Factors Methods – A Practical Guide for Engineering and Design*. Ashgate, England.

Svensson, E., & Angelborg-Thanderz, M. (2000). Simulated landings in turbulence – motion, predictive modelling, and psychometric aspects. *American Institute of Aeronautics and Astronautics*, AIAA-2000-4076.

Taylor, H. L., Lintern, G., & Koonce, J. M. (1993). Quasi-transfer as a predictor of transfer from simulator to airplane. *Journal of General Psychology*, 120, 257-276.

Vaden, E. A., & Hall, S. (2005). The Effect of Simulator Platform Motion on Pilot Training Transfer: A Meta-Analysis. *The International Journal of Aviation Psychology*, 15(4), 375-393.

Wickens, C. D., & Hollands, J. G. (2000). Engineering psychology and human performance (3rd ed.). Upper Saddle River, NJ: Prentice Hall Inc.

Wilson, G., Balkin, T., Beaumont, M., Burov, A., Edgar, G., John, F. W., et al. (2003). *Operator functional state assessment* (NATO RTO HFM-056/TG-008).