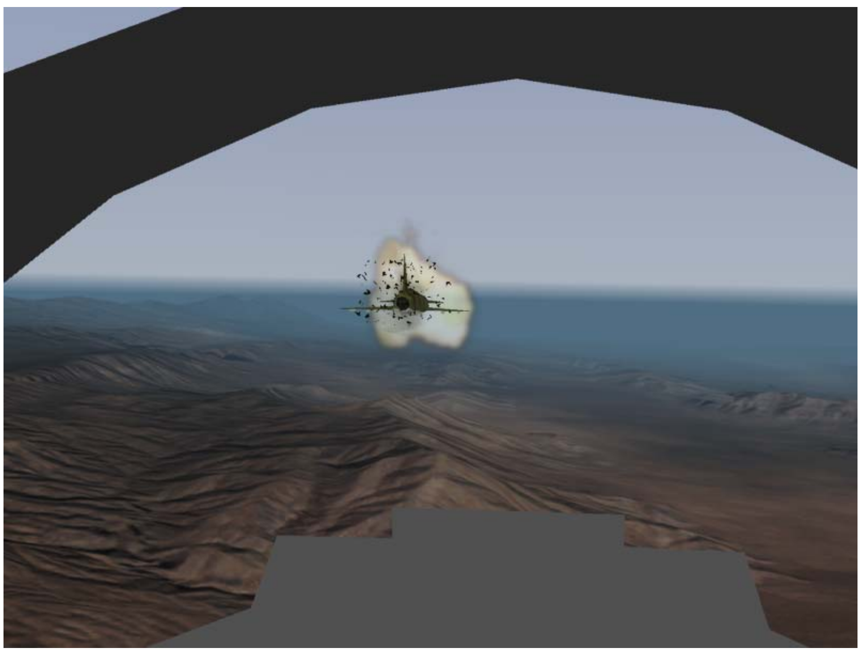
SWEDISH DEFENCE
RESEARCH AGENCY

FOI-R--0710--SE
December 2002
ISSN 1650-1942

Scientific report

Jonathan Borgvall, Staffan Nählinder, Jan Andersson

WVR-Illustrator Evaluation: Using Pilot Expertise for Future Development



Command and Control Systems
SE-581 11 Linköping

SWEDISH DEFENCE RESEARCH AGENCY

Command and Control Systems
P.O. Box 1165
SE-581 11 Linköping

FOI-R--0710--SE

December 2002

ISSN 1650-1942

Scientific report

Jonathan Borgvall, Staffan Nählinder, Jan Andersson

WVR-Illustrator Evaluation: Using Pilot Expertise for Future Development

TABLE OF CONTENTS

FIGURES

TABLES

1 WVR-ILLUSTRATOR EVALUATION – USING PILOT EXPERTISE FOR FUTURE DEVELOPMENT

This report is based on a study that was conducted during the spring of 2002 by the Swedish Defence Research Establishment[1] (FOI), department of Man-System-Interaction (MSI) in Linköping, Sweden. During the autumn of 2001, the development of a flight simulator for WVR-combat started of in cooperation between the Swedish Armed Forces Headquarters (HKV) and FOI. WVR is an acronym for Within Visual Range, which refers to all air-to-air combat situations where the participating aircrafts are within visual range of each other (normally within 15 kilometres). HKV[2] and FOI were responsible for the ordering and specification of the system. The main purpose of the project was to develop a prototype of a training environment for WVR-combat, in a short time and to a low cost. According to HKV and FOI guidelines, ISD Data AB[3] answered for the software development. In March 2002 the system was delivered to FOI-MSI for continuation of the project with research issues. The extensive goal of the project, depending on research findings, is to develop a simulator with comprehensive pedagogical tools for WVR-combat fighter pilot training (Danielsson, Svensson & Jenvald, 2002; Svensson, Nählinder, Danielsson & Jenvald, 2002).

The version of the system used in this study is called the Illustrator. The choice of name is done with regard to the system as not being an approved simulator. It is an illustrator of the possibility and potential of this kind of VR-systems. The Illustrator is based on commercial-off-the-shelf products (COTS) and developed for real-time simulation of WVR-combat. The system integrates aircraft controls, a motion tracking system and a head mounted image generation solution.

The main purpose of the work described in this report was to evaluate the Illustrator together with the end users, the pilots in the Swedish Air Force. The focus of the study was to collect expert opinions regarding psychological and technical aspects. The findings of the

---

[1] Please visit www.foi.se for closer information about FOI.

[2] Please visit www.hkv.mil.se for closer information about HKV.

[3] Please visit www.isd.se for closer information about ISD Data AB.

evaluation would then be the fundament for creating recommendations for future development of the system. There were two specific research issues of interest for this report:

- *Psychological evaluation* (psychological user acceptance)
  - Monitor opinions regarding three psychological aspects among the participants; attitude towards simulators in general together with commitment and challenge of the WVR-scenarios.
  - Explore the relation between the concepts of mental workload, situational awareness and performance.
  - Search for psychological predictors of performance in the Illustrator, that is to find factors that could explain why pilot A beats pilot B during a scenario.
- *Technical evaluation* (technical user acceptance)
  - Measure the participants' opinions of realism, limitation of performance and importance of realism for the training potential of a WVR-simulator, of different aspects or parts of the system.
  - Collect opinions regarding future improvements of the system.

1.1 PROJECT BACKGROUND

The Swedish Air Force has several simulators for training of fighter pilots (e.g. SUL37, FLSC). Concerning the visual presentation, the strength of those systems is to simulate beyond visual range-combat (BVR-combat). During this type of combat, the fighting aircrafts have no visual contact. The activities related to BVR-combat are therefore primarily based on instrument flying. This means that the demand on the visual presentation of the surrounding environment is restricted in comparison to within visual range-combat (WVR-combat). During a WVR-combat situation (called dogfight in daily terms), the fighting aircrafts have a maximum distance of approximately 15 kilometres, depending on weather conditions. During WVR-combat, the combatants have two main goals. First, they try to maintain visual contact. Secondly, by performing smarter manoeuvres than the enemy, they try to reach a position to fire at each other, or at least make sure that the other does not. In WVR-combat situations the instrumentation is primarily used only for controlling the status of the own aircraft, as speed, altitude and heading.

In the simulators of today, the visual presentation is static in the sense that the operators see a limited field of what is in front of the aircraft and the view does not change according to the operators' head movements. For BVR-combat simulators, this is believed to be satisfying but for WVR-combat it is not. In WVR-combat, the operator has to be able to follow the enemy visually in the simulated environment. The development of virtual reality (VR) offers this possibility. For a rather limited amount of money, VR-goggles including a head tracker with three degrees of freedom (DoF) can be bought. This technology opens up a potential for simulating situations with high demands on the visual presentation.

1.2 SIMULATORS

The aim of using simulators is to provide a synthetic presentation of a real situation. This presentation might be partial or full scale – it may contain only some features of the system or every detail. There are four broad categories for the use of simulators; training, research, evaluation and investigation.

Training simulators are used for individual or team training. It might for example simulate communicational procedures in a power plant or navigation in a naval vessel. *Research simulators* are used primarily for investigations of human performance. Since task decomposition almost always is necessary for research, low-fidelity simulators are often preferred for research into the domain of human performance (Stanton, 1996). *Evaluation simulators* are normally used to test future operational settings for the real equipment, for example instrument design. The last category, *Investigation simulators*, fills the purpose of task analysis of the operators' work situation.

Clymer (1981) has developed a classification of simulators based upon what kind of representation they have and the purpose of them. Any simulator may fall into one or more of the following categories[4]:

- *Replica* – An exact duplication of the real operational setting.

[4] Although the Illustrator is not an approved simulator, it could be interesting to classify the system according to Clymer's categories. The Illustrator would then be best described as a generic, basic-principles system (see method section for a closer description of the Illustrator).

- *Generic* – A representation of a class of systems, for example airliners. Generic systems are general models, not duplicates of any particular system.
- *Eclectic* – Include features, for example malfunctions, which could but do not normally happen under real settings. The aim is to broaden the experience of the operators.
- *Part-task* – Representations of parts of the operational system or parts of tasks performed in it.
- *Basic-principles* – A generic and/or part-task simulation that deliberately omits certain details of the operational setting to keep the cost low and the simplicity high.

There are many reasons for the use of simulators. First, a simulated environment is normally a lot safer than the real one. Some tasks might be too dangerous to practice in the real environment, not at least initially. Secondly, simulators provide an environment for human-performance measurement that is often more controllable and accessible than the real environment. Furthermore, simulators are extremely useful when there is a lack of availability or infrequency of use of the real environment. Another aspect that further justifies the use of simulators is the possibility of reducing task difference to a great extent, for example by altering temporal aspects. Finally, there are economical issues of importance. Simulators can save a lot of money for the user since they are often cheaper to use than the real environment. Research evidence indicates that the use of simulators may cost only 10% of the real operating cost (Stanton, 1996).

1.3 PERSONAL COMPUTER-BASED FLIGHT TRAINING DEVICES

In 1981, IBM released their first personal computer (PC). Aviation enthusiasts provided a market for the development of PC-based flight simulators. In those days, flight simulation programs were marketed as games and the users input possibilities were restricted to keyboard strokes. A few years later, there were flight sticks and control yokes available on the market. Rudder pedals, even with brakes in some cases, followed. As aviation pilots and aspirants experienced and started to show excitement in PC-based flight simulators, manufacturers in the simulation industry, developing for the armed forces and commercial air carriers, were pushing for greater technical complexity in the systems.

However, despite the primitive nature of PC-based systems, the general aviation public started to get a feeling of actually learning something of value by flying in those low-cost systems. In 1997, there was several PC-based flight simulators used for instrument training of already certified pilots. Today, the systems offer more and more parameter settings left for the user to elaborate with. There are a massive number of controls and visual systems available for the systems, for example flight sticks with motion feedback and VR-goggles (Koonce & Bramble Jr., 1998).

1.4 SIMULATOR FIDELITY

Simulator fidelity could be concluded as the degree of similarity between a simulator and the environment it simulates. There are several categories, or components that Rolfe (1985) calls them, of simulator fidelity namely the degree to which:

- The physical features and characteristics of the simulator match the real environment – *physical fidelity*.
- The simulator works as the real environment – *functional (dynamic) fidelity*.
- The simulator is built in the same way as the real operational equipment – *engineering fidelity*.
- The simulator, over a period of time, matches the operation of the real task situation – *operational fidelity*.
- The task domain of the simulator reflects that in the real operational situation – *task fidelity*.
- Transfer occurs despite the lack of other aspects of fidelity in the simulator – *psychological fidelity*.

According to Stanton (1996), the two main categories are physical and functional fidelity. The former could be described as to what degree the system “looks like” and the latter to what the degree the system “acts like”, the simulated environment. The main opinion in the literature is that a simulator should be as an exact duplicate of the real environment as possible (high-fidelity simulator) or that the maintenance of functional fidelity whilst physical fidelity is reduced (low-fidelity simulator) will not decrease the effect of training, rather vice versa (e.g. Boreham, 1985; Stammers, 1981). A report from a study by Welham (1986) supports the latter approach when expressing that 80% of the benefit often can be reached at 20% of the cost of a full fidelity simulation. In fact, there is evidence suggesting that planned departures from different aspects of fidelity actually can enhance learning. In 1987, Wightman and Sistrunk found better learning for part-task training than practice of the whole task.

Research has shown that the most crucial aspect for a simulator to be effective as a training device is that a particular user input should have the same effect as in the real environment. This does not mean that all features and components of the real system have to be provided in the simulator, but those that are implemented have to behave in a correct manner. This means that the simulator has to have a solid functional fidelity. Through the history of simulators, high physical fidelity has often been favoured with the argument that if it looks like the real thing it has to be better for training (often called face validity). However, in later years many results have shown a positive correlation between performance in a low-fidelity simulator and performance in the real environment. Hence, the criticism that low-fidelity simulators are inappropriate because they, more or less, lack physical fidelity has lost some force in recent years. Therefore, the use of low-fidelity systems has gained more and more attention. Further, for research into human performance, psychological and functional (dynamic) fidelity are considered to be of greater benefit than physical fidelity (Stanton, 1996).

As stated earlier, the Illustrator is a low-cost system based on COTS-products. It is also the first version of the system. Therefore, it is of great interest to explore how far this might reach in terms of WVR-combat simulator training. Thus, fidelity will be one of the major interests during the evaluation of the system.

1.5 SIMULATOR EVALUATION

In 1998, Bell and Waag proposed the following three categories of simulator evaluation; utility evaluation, in-simulator learning and transfer of training. From the categories for estimating the effectiveness of flight simulator training, Bell & Waag make a proposal for an evaluation model. The model consists of five stages; utility evaluation, performance improvement (in-simulator learning), transfer to alternative simulator environment, transfer to flight environment (transfer of training) and extrapolation to combat environment. This report will focus on the first stage of the evaluation model.

Utility evaluation is normally based on user opinion data. This type of data does not provide quantitative measures of neither performance improvement nor transfer of training. But according to Bell & Waag, it is a necessary first step in evaluating a simulator since user acceptance may support the decision about future development and more rigorous evaluations.

1.5.1 Utility evaluation

The purpose of the utility evaluation proposed in Bell and Waag's model (1998) is to a) evaluate the accuracy, or degree of fidelity, of the simulator environment, and b) gather user opinions regarding the potential value of the simulator as a future training environment. The primary aim of this approach is to identify the most central aspects of the simulator for the specific area of use and which of them that needs to be further improved. Those aspects or parts of the system are further in thesis called *fidelity levels*. Secondly, the user opinion data are of value in deciding whether the results are positive enough to continue with further development and more resource demanding evaluations. Not only are the results useful in deciding about further development, they also give hints or guidelines of what direction it should take according to the user.

As stated above, simulator fidelity could be concluded as the degree of similarity between a simulator and the environment it simulates. Thus, the concept of fidelity is closely related to realism. This implies that fidelity evaluations should include realism as an evaluation tool. However, Stanton (1996) is critical to simulator evaluation using the concept of realism. Different aspects, or fidelity levels, of the system requires different degrees of realism. For some task a very low degree of realism might be enough for some fidelity levels, but not for others. Stanton also argues that it is necessary and important to be careful when drawing conclusions about training effectiveness from data collected as described above. User opinion can be biased for many reasons. First, it assumes that the user is an appropriate judge of what constitutes an effective simulator. User acceptance might be a first evidence of the potential of a system but it does not guarantee training effectiveness. Secondly, user opinion of a system is biased of previous experience and hence opinions may vary because of that.

Stanton of course has a point in that user acceptance does not guarantee training effectiveness. But, the aim of conducting the evaluation proposed by Bell & Waag is not to draw conclusions about training effectiveness. Their idea is that user acceptance is a necessary but not satisfying condition for a system to have a potential as a future training environment. Monitoring user acceptance and expert opinions of the system and its parts, is according to them, a central aspect of modern simulator design. This is further supported by Salas, Bowers and Rhodenizer

(1998). They proposed three assumptions that characterized the current view of simulator training in 1998. One of them is the statement that *if the aviators like it, it is good*. In line with Stanton above, they mean that user acceptance does not provide an adequate measure of training success. However, in line with Bell and Waag, they conclude that subjective measures and expert opinions are important initially because they provide evidence for the user acceptance of the system.

1.6 Experience and Performance

Experience is strongly related to training since experience is created by training. Many studies imply a close relation between experience and performance. Most researchers in the area agree in that innate abilities affect performance, but they draw the conclusion that they play a relatively small role in explaining adult individuals' performance (e.g. Ericsson & Simon, 1993; Simon, 1990). Differences in innate abilities play a role, but superior performance is primarily a result of practice (Ericsson & Simon, 1993). Fisk and Lloyd (1988) have stated that "It is a truism that practice is needed to become skilled at an activity" (p. 36).

Ericsson and Simon (1993) have divided the development of superior performance into three phases. The primary assumption is that an individual's level of performance is increasing as a function of time spent on deliberate practice. Ericsson and Simon presume deliberate practice to be effortful and not intrinsically enjoyable.

a) The initial phase begins with the individual's introduction to the domain activities. This phase is characterized by relatively slow progress.
b) The first phase ends and the second phase begin when formalized instruction and deliberate practice are undertaken. This continues for an extended period of time, ending when the individual decides to make a full-time commitment toward improving performance in the domain.
c) This full-time commitment is followed by the third and final phase, where even more intensive practice is undertaken.

Ericsson and Simon (1993) have obtained support for this model in a study of violinists. They found that the level of skill was predictable from the violinists' estimated amount of deliberate practice. Similar results were found for pianists as well.

Research by Angelborg-Thanderz (1990) suggests that experience play a key role in explaining performance among fighter pilots. To explore differences in performance, Angelborg-

Thanderz divided the pilots in a study into several groups, depending on experience and flight status (i.e. if they were active or inactive as fighter pilots). When comparing the groups, the general result was that the more experienced pilot, meaning the older and active, performed best.

Since experience seems to be crucial for performance in general, and since earlier findings imply that experience is a central factor for pilot performance in specific, it will be explored whether this is the case during the WVR-combat scenarios in the Illustrator. Not only will different measures of flight experience be included. Different measures of personal computer experience will also be included to investigate whether this factor affect the outcome of performance. The purpose is to establish whether flight experience measures could predict performance in the Illustrator, and at the same time control personal computer experience measures. The result will then be interesting to compare with earlier research, as the findings by Angelborg-Thanderz above, since this indirectly could support conclusions about the technical status of the system and its need for future development.

1.7 MENTAL WORKLOAD, SITUATIONAL AWARENESS AND PERFORMANCE

Pilots in modern flight systems have to cope with a greater and greater amount of information, often during high physiological pressure. This often results in an increased risk of reaching mental overload (Svensson, Angelborg-Thanderz, Sjöberg & Olsson, 1997; Svensson & Wilson, 2002). The concept of mental workload (MW) has been studied since the late seventies (VINTHEC, 1997). According to Wilson (1997), mental workload refers to the capacity required by an operator when performing a certain task. When the situation becomes to demanding, the operator's resources are overstrained. In several studies, it has been shown that performance declines as a function of an increasing mental workload (Svensson, Angelborg-Thanderz, Olsson & Sjöberg, 1992; Svensson & Wilson, 2002).

Situational awareness (SA), or situation awareness, is another concept related to performance. There is no agreed, universal definition of situational awareness up to date, and it will probably stay that way since the definition is dependent on the context. Adam (1995) has presented a short definition as "Knowing what's going on so you can figure out what to do!" (p. 9/2). According to Endsley (1995a), situational awareness could be described as a person's mental model of the surrounding situation. Further, Endsley (1995b) has presented a general definition of SA as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (p. 36). SA is believed to be closely related to performance. Endsley (1995a) considers SA to be a precursor to the pilot's decision making, but at a stage separated from performance. Several studies have supported this idea (e.g. Svensson et al., 1997; Svensson et al., 1992; Svensson & Wilson, 2002).

As stated above, mental workload and situational awareness are both believed to be closely related to performance. Results not only suggest that the two concepts are related to performance, they seem to be related to each other as well. In fact, similar correlations and structural equation models (SEM) between the three concepts has been found at several occasions (e.g. Berggren, 2000; Magnusson, 2002; Svensson & Wilson, 2002). Figure 1 is an illustration of a causal model between the concepts, proposed by Svensson and Angelborg-Thanderz (1999, p. 17)[5], here modified for the purpose.

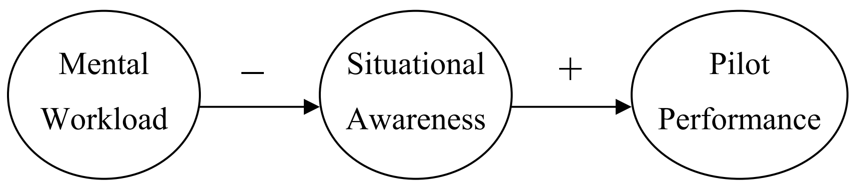


Figure 1. Svensson and Angelborg-Thanderz's proposal of a causal model illustrating the relationship between the concepts of mental workload, situational awareness and pilot performance.

The figure illustrates a negative relation between mental workload and situational awareness, and a positive relation between situational awareness and pilot performance. This means that the higher the mental workload is, the lower is the situational awareness and the lower is the performance. Research supporting this model has been found at several occasions, both for simulated and real flight (e.g. Magnusson, 2002; Svensson & Wilson, 2002). Interesting to note is that mental workload and pilot performance are not directly related, they are connected through the concept of situational awareness. So the relation between mental workload and performance, and situational awareness and performance should rather be understood as a relation between the three concepts together, with SA as the concept connecting mental workload and performance.

---

[5] The original picture includes "Mission Complexity" as the first factor. The idea is that mission complexity is positively related to mental workload. Mission complexity has been left out in this illustration since it is not of closer interest for this report.

The purpose of including measures of MW, SA and PP in the evaluation of the Illustrator is to investigate whether a similar pattern is obtained. Further, the three measures will be included in the analyses of performance predictors together with different experience measures, as described earlier.

2 EVALUATION METHODOLOGY AND CLARIFIED PURPOSE

The methodology used for the empirical study in the thesis is based on the first step of Bell & Waag's evaluation model – utility evaluation (1998). The purpose at this stage of the model is to gather user opinions regarding the potential value of the simulator as a future training environment together with evaluating the fidelity, of the environment. Bell & Waag's idea is that user acceptance is a necessary but not satisfying condition for a system to have a potential as a future training environment. However, they have not provided any evidence supporting their idea. Thus, whether user acceptance is a necessary but not satisfying condition for a system to have a potential as a future training environment or not, will be left unsaid. There is no purpose of deciding if the user acceptance is high enough for the future training potential here. Therefore, for the matter of this report, user acceptance should be understood as the overarching concept of the expert opinions of different aspects. It is the expert opinions of those specific aspects that will indicate the features of importance for improvement of the Illustrator, meaning that they will indicate which fidelity levels that should be prioritized in the further improvement according to expert opinions.

The evaluation described in this report consists of two parts; one psychological and one technical. In the psychological evaluation, three psychological aspects among the participants, attitude towards simulators in general together with commitment and challenge of the WVR-scenarios, will be measured. The aim is to observe any changes of the expert opinions regarding the three aspects from before to after the sessions. The specific purpose is to measure a) the psychological aspects initially, and b) how those aspects are affected by participating in the study (psychological user acceptance). Further, the concepts of mental workload, situational awareness and pilot performance will be measured by ratings performed by the participants at several occasions during the session. The aim is to explore the relation between the concepts in this study and compare it to earlier findings. Further, it will be investigated whether the factors included in the study could explain why pilot A beats pilot B.

The technical evaluation regards the measurement of the functional (dynamic) fidelity of the Illustrator. The aim is to collect expert opinions regarding the functional fidelity. To evaluate the functional fidelity of the Illustrator, seven fidelity levels of interest were identified; transition head-up/head-down, instrumentation, flight controls, graphics, visual feedback according to a manoeuvre, visual resolution and field of view. The seven fidelity levels were identified in advance of the study by a fighter pilot from the Swedish Air Force together with personnel from FOI-MSI. The fidelity levels could be put in relation to Stammers (1983; 1986) dimensions of simulation. Stammers has proposed nine different dimensions of simulation. Three of them are of closer interest here; stimulus/displays, responses/controls and display-control relationships. The dimensions could be seen as smaller instances of fidelity. *Stimulus/displays* concern the realism of the displays in the simulator, and whether they are complete compared to the real environment. In a flight simulator, for example, this would regard the instrumentation in the cockpit and the visual presentation of the surrounding environment. *Responses/controls* are a matter of to what extent the real control devices are represented in the simulator. *Display-control relationship* concerns the realism of the interaction between the controls and the displays of the simulator.

Transition head-up/head-down, graphics, visual resolution and field of view are treated as a matter of the stimulus/displays of the Illustrator. The flight controls concern responses/controls, and visual feedback according to a manoeuvre display-control relationships. The instrumentation is seen as related to both stimulus/displays and display-control relationships. The instruments provide information about the status of the aircraft (stimulus/displays), but they also provide a control for performed manoeuvres (display-control relationships). The fidelity levels were to be rated for three dimensions; realism, limitation of performance and importance of realism for the training potential of a WVR-simulator. The aim is to monitor expert opinions indicating a) the seven fidelity levels relation to each other regarding the three dimensions[6], and b) which of the seven fidelity levels that are of most interest for future development. The results are intended to be used for creating recommendations for future development.

As stated above, Stanton (1996) has criticized the evaluation of simulators regarding realism (fidelity), since different parts of the system are in different need of fidelity for different purposes

[6] Visual resolution and field of view were not rated regarding realism. This will be further discussed below.

of use. But, since the aim is to evaluate the Illustrator regarding activities related to WVR-combat specifically, this is not considered as a problem for the matter of this study.

Finally, it is the results of the technical evaluation that will indicate fidelity levels that should be further improved according to expert opinions. The results of the psychological evaluation will not leave a result pointing out any specific parts of the system, but an indication of how the system was experienced in general. To clarify the purpose of the study, the research questions and issues of interest in the psychological and technical evaluations respectively are presented below. The specific interest in the psychological evaluation was:

a) How are the participants' ratings of the three psychological aspects attitude towards simulators in general together with commitment and challenge of the WVR-scenarios distributed, and are the participants' opinions affected by participating in the study?
b) What is the relation between the concepts of mental workload, situational awareness and pilot performance, measured before and during the session?
c) Can any predictors of performance be found?

The aim of the technical evaluation could be closer described as:

a) What, or which, of the seven fidelity levels seem to be most crucial in the future development of the Illustrator, according to the participants' ratings of the fidelity levels realism, limitation of performance and importance of realism for the training potential of a WVR-simulator?

3 METHOD

3.1 PARTICIPANTS

Thirteen active military fighter pilots from the Swedish Air Force participated[7]. They were all male and belonged to the same Air Force Division. The fighter pilots' amount of different aspects of experience is crucial for the purpose of this report and is therefore presented in two tables below. The aspects of flight experience of interest were age in years (AGE), total amount of flight hours (TFH), flight hours in a fighter aircraft (FAFH), flight hours in fighter aircraft

[7] In advance of the study, a pilot study was completed at Blekinge Wing to evaluate the design of the experiment. Six male, active fighter pilots participated. The pilot study revealed that the prepared design worked appropriately, and only resulted in minor modifications of the questionnaires. The data from the pilot study has not been analyzed.

simulator (FASFH) and flight hours in a special fighter aircraft training facility centre in Sweden (FLSCFH) together with the participants rated experience and ability of WVR-combat (EXPWVR and ABWVR). The ratings were made on a scale running from one to seven, with one as the lowest degree of experience. The value for each pilot is presented in Table 1. The computer experience aspects of interest were experience of computers in general (CE), experience of computer games with joy stick (CGE), and experience of fighter aircraft computer games with joy stick (FACGE). Those aspects were rated by the participants (also from one to seven) and are presented in Table 2.

Table 1. The distribution and the means (*SD*) of the flight experience aspects among the participants.

PILOT	AGE	TFH	FAFH	FASFH	FLSCFH	EXPWVR	ABWVR
1	36	1500	450	150	40	5	3
2	32	1100	700	100	10	5	5
3	33	1400	300	100	15	5	5
4	29	1300	900	120	20	7	6
5	31	1250	860	105	6	7	7
6	34	1500	1100	130	5	6	6
7	26	380	150	100	0	2	4
8	32	1400	400	90	15	7	5
9	41	2700	1800	150	20	6	4
10	30	955	0	15	50	6	3
11	27	400	150	100	0	1	2
12	26	400	150	60	0	2	2
13	29	900	650	70	25	6	4
Mean	31.2	1168.1	585.4	99.2	15.8	5.0	4.3
(*SD*)	(*4,2*)	(*620,5*)	(*498,1*)	(*36,7*)	(*15,5*)	(*2,0*)	(*1,5*)

Table 2. The distribution and the means (*SD*) of the computer experience aspects among the participants.

PILOT	CE	CGE	FACGE
1	7	2	2
2	6	3	1
3	7	4	4
4	7	2	1
5	7	1	1
6	7	1	1
7	6	2	2
8	7	6	6
9	7	3	3
10	7	1	1
11	7	3	3
12	7	4	4
13	7	2	2
Mean	6,8	2,6	2,4
(*SD*)	(*0,4*)	(*1,4*)	(*1,6*)

All participants were concerned to be experienced and skilled in flying, and they all had experience of participating in studies. The experiment was conducted by three researchers from the Swedish Defence Research Agency (FOI). The main study was conducted at Blekinge Wing (F 17).

3.2 Materials

3.2.1 Questionnaires

Four different questionnaires were used in the study; QBackground, QBefore, QDuring and QAfter. Similar for all the questionnaires was that all ratings were made on a scale running from one to seven.

QBackground concerned the experience measures described initially, with one exception. They also rated their attitude towards simulator facilities in general (ATT). Regarding the computer experience measures, they were collected after the study had been performed since they were not included in the original design.

In QBefore, the participants performed ratings concerning their commitment (COMM) and challenge (CHALL) of the WVR-scenarios together with questions about mental workload (MW), situational awareness (SA) and pilot performance (PP). QBefore was completed before each session, but after QBackground.

In QDuring, the participants rated their experienced mental workload, situational awareness and pilot performance during the last scenario prior to the questionnaire. Mental workload was rated on the Bedford rating scale (BFRS) (Roscoe and Ellis, 1990), translated to Swedish by researchers at FOI. Situational awareness and pilot performance were rated on modified BFRS-scales. Except for the ratings of mental workload, situational awareness and pilot performance, the participants answered a multiple-response question concerning which fidelity levels, if any, that limited their performance (LIM) in the last scenario prior to the questionnaire. The fidelity levels rated regarding limitation in QDuring were transition head-up/head-down (HUHD), instrumentation (INSTR), flight controls (FC), graphics (GRAPH), visual feedback according to a manoeuvre (VFM), visual resolution (VRES) and field of view (FOV). Further, the flight model was included as an alternative here, to control whether the use of the F-16 model instead of the JA37 model affected the participants[8]. The fidelity levels and the flight model together with "Don't know", "Nothing" and "Other", were the alternatives. If the participants answered "Other" they were asked to specify what they aimed at. QDuring was performed after at least three of the scenarios in a session (the number of scenarios in a session varied between six and twelve[9]).

QAfter consisted of two parts. In the first part, three questions were asked again. Attitude towards simulator facilities in general (ATT) together with commitment (COMM) and challenge (CHALL) of the WVR-scenarios performed during the session. All of the fidelity levels were also rated. Three different types of questions were asked for each fidelity level (except for field of view and visual resolution); a) the degree of realism, b) the degree of limitation of performance and, c) the degree of importance of realism for the training potential of a future WVR-simulator. The first question (a) was not asked for field of view and visual resolution, due to ecological aspects concerning the Illustrator.

[8] Observe that the use of the term fidelity levels aims at the seven aspects described above, i.e. flight model is not included. The flight model was only included regarding limitation in QDuring and improvements in QAfter as a control question. Therefore, the flight model is not included regarding realism, limitation and importance in QAfter.

[9] Technical problems and/or personal time aspects were reasons that limited the number of scenarios in some sessions. The aim was a minimum of six scenarios for all sessions.

Furthermore, the participants answered a multiple-response question, similar to the one used in QDuring, but this time regarding future improvements (IMP) instead of limitation of performance. Flight model was an alternative here as well. Since there was a possibility that other fidelity levels than the ones included in the study were important, the second part of QAfter consisted of open-ended questions regarding fidelity aspects.

3.2.2 SCENARIO

One basic scenario was used for all of the sessions in the study. It was designed in cooperation between a fighter pilot from the Swedish Air Force and researchers from FOI. Two aircrafts started in the same direction, separated 2 km from each other at an altitude of 2000 meters. They were exactly parallel at the start of the scenario to guarantee initially equal fighting positions for the WVR-combat.

The goal for the participants was to shoot down each other. In the Illustrator, the enemy aircraft is shot down when the enemy is in front of you within certain angles and within a certain range[10]. The start of the scenario and all of the combat was performed during daytime over a flat landscape with good weather conditions.

3.2.3 APPARATUS

The Illustrator had five screens in total; two for each pilot station showing the instrumentation (see Figure 2) of the aircraft and the view a pilot had in the VR-goggles[11] respectively. The last screen showed a Gods-eye-view (showing the situation from above) of the WVR-combat and was used by the experiment leader.

[10] To have a shooting position, the attacking aircraft had to have the enemy aircraft within a cone of 45 degrees in front of him. Further, the attacker had to be within a cone of 45 degrees behind the enemy aircraft. The maximum range for a shooting position was 4 kilometres.

[11] This made it possible for the test leaders to observe what view the two participants had of the situation respectively.

Figure 2. The instrumentation in the Illustrator[12].

Five personal computers were used to run the system. The surrounding environment was presented to the pilot in VR-goggles (Sony Glastron) including a head tracker (see Figure 3) with three degrees of freedom; pitch, roll, and yaw. This means that the visual presentation followed the operators' head movements in three directions (see glossary for closer description). The visual resolution in the VR-goggles was 800 by 600 pixels. The goggles had a 30 degree field of view.

[12] The instrumentation presented above is a later version than the one used in during the study.

Figure 3. Sony Glastron VR-goggles including the head tracker system.

The aircrafts were manoeuvred with Saitek X35T Throttles and X36F Flight Control Sticks. Each pilot station had a sound system with two speakers and one subwoofer. A picture showing a pilot station is presented in Figure 4.

Figure 4. Pilot station with VR-goggles including head tracker, Saitek throttles and flight control sticks, sound system and a monitor presenting the instrumentation.

A generic F-16 model was used during the study, since that model is non-classified. The participants were informed of this fact in advance of the study.

3.3 SETUP

All five computers were placed under the table in front of the experiment leader. In the middle of the experiment leader's table, the screen presenting the "Gods-eye-view" was placed. To the left and right of that screen, the two screens presenting the current view of the two participants' VR-goggles were located respectively. The screens showing the instrumentation of each pilot station were placed with the backsides towards each other. The participants currently flying was sitting to the left and right of the experiment leader respectively, turned towards each other at about three meters distance.

3.4 Procedure

Each session started with the participants answering QBackground. Answers to all of the questionnaires, except part two of QAfter, were performed on laptops next to the pilot stations. After QBackground, QBefore was completed. The reason for not combining QBackground and QBefore was to make it possible for the participants to fill in the former in advance of the session. When finished, the participants were introduced to the Illustrator with advanced single flying. During this phase the two aircrafts were separated far apart so that the participants could concentrate on manoeuvring their own aircraft. The experiment leader suggested the participants some manoeuvres to practise. After they had practised all of the manoeuvres, they flew on their own for 20 minutes. Thereafter, the experiment leader stopped the first part of the training phase. The practise continued with advanced file flying. The experiment leader suggested some manoeuvres to practise, before the participants were aloud to fly on their own. After about 10 minutes of practise the participants were asked if they had any questions and if they wished to practise more, to make certain that they really felt comfortable with the system.

When both of the participants said that they felt familiar with the Illustrator they made themselves ready for the WVR-scenarios. The first two were training scenarios. After the training scenarios the experiment leader started the test scenarios. The time for each scenario varied from 30 seconds up to 4 minutes. After three scenarios in each session, the participants removed the VR-goggles and answered to QDuring. When they had finished QDuring, they made themselves ready for the next scenario. After all scenarios and the last QDuring had been completed, the participants answered part one of QAfter on a laptop, and part two with pen and paper. Each session took approximately one and a half hour. When a session was finished, the participants were debriefed and thanked for their participation.

3.5 Design

The study was performed as a within-participants design. In the psychological evaluation, the three psychological aspects will be used as dependent measures, monitored from before to after the session (3 x 2). Further, the psychological concepts of mental workload, situational awareness and pilot performance are dependent measures, a) in analyses of their relation to each other at a given occasion, meaning the pre measures and the three during measures respectively (i.e. MWpre, SApre and PPpre and similar for the three during measures), and b) in comparisons within each concept over time, meaning comparisons of the pre and during measures (i.e. MWpre, MWdur1-3 and similar for SA and PP).

In the technical evaluation, the dependent measures are the fidelity levels received ratings of realism, importance and limitation. Except for visual resolution and field of view, which were not rated for realism, every fidelity level was rated for those three dimensions (5 x 1, 7 x 2). Further, the fidelity levels are used as dependent measures in the analysis of the multi-choice questions regarding limitation in During (LIM) and future improvements (IMP) in QAfter.

Finally, regression analyses will be performed with "objective performance", that is a measure of wins and losses in the scenarios, as the dependent measure. Several of the psychological measures will be included as predictors of the performance criteria.

4 RESULTS

In this section, the results from the psychological and technical evaluations will be presented respectively. The first part will focus on the psychological measures and their relation to performance. Part two will deal with the fighter pilots fidelity ratings, that means, the results of the technical evaluation.

4.1 PSYCHOLOGICAL EVALUATION ANALYSES

4.1.1 ATTITUDE, COMMITMENT AND CHALLENGE

A two-way repeated measure analysis of variance (3 x 2) was performed for the three psychological aspects; attitude towards simulators in general (ATT), commitment of the scenarios (COMM) and challenge of the scenarios (CHALL) for two different occasions, before and after the session. The mean values (standard deviations) of the three aspects are presented in Table 3. The main effects and the interaction effect were not significant ($n = 13$, $ps > 0.05$). Thus, the WVR-session did not affect the fighter pilots attitude towards simulators in general. Further, there was no change in the participants rated commitment and challenge of the WVR-scenarios, from before to after the session.

Table 3. The mean values for attitude towards simulators in general, commitment of the WVR-scenarios and challenge of the WVR-scenarios, from before to after the session.

ASPECT	BEFORE	AFTER
ATT	5.5 (1.1)	5.5 (1.4)
COMM	5.5 (0.8)	4.8 (0.9)
CHALL	5.2 (0.9)	4.5 (1.6)
Overall mean	5.4	4.9

4.1.2 FLIGHT AND PC EXPERIENCE

Pearson correlation tests for the measured flight and personal computer experience aspects were performed (n = 13). The mean values for the flight experience aspects were presented in the method section (see Table 1). The correlations among the flight experience aspects are presented in Table 4. Age (AGE) correlated with total flight hours (TFH), flight hours in the fighter aircraft (FAFH) and flight hours in the fighter aircraft simulator (FASFH). TFH showed a correlation with FAFH, FASFH and rated experience of WVR-combat (EXPWVR). FAFH was correlated with FASFH. EXPWVR and rated ability of WVR-combat (ABWVR) were significantly correlated.

Table 4. Correlations between the flight experience aspects[13].

	AGE	TFH	FAFH	FASFH	FLSCFH	EXPWVR	ABWVR
AGE	1,000						
TFH	<u>0,951</u>*	1,000					
FAFH	0,724*	<u>0,823</u>*	1,000				
FASFH	0,611*	0,589*	0,658*	1,000			
FLSCFH	0,360	0,327	-0,045	-0,207	1,000		
EXPWVR	0,522	0,659*	0,522	0,128	0,480	1,000	
ABWVR	0,243	0,392	0,496	0,327	-0,158	0,712*	1,000

Thus, AGE, FASFH, TFH and FAFH correlated strongly with each other (THF correlated strongly with EXPWVR as well). EXPWVR and ABWVR were strongly correlated but did not correlate with the other measures, except for TFH and EXPWVR respectively. FLSCFH did not correlate with any of the other experience measures. In order to pin-point aspects that could explain performance in the WVR-illustrator, TFH and ABWVR were chosen to represent flight

[13] Further on in the result section, the correlations marked with * are significant at p < .05. The underlined correlations are still significant after Bonferroni correction for multiple comparisons.

experience in a regression analysis presented below. They were chosen since they capture different aspects of flight experience, that is, since they do not correlate strongly.

The mean values for experience of computer games were presented in the method section (see Table 2). The experience of computer games with joystick (CGE) correlated with experience of fighter aircraft computer games with joystick (FACGE). The correlations are presented in Table 5. FACGE was therefore chosen as a third experience measure in the regression analysis of performance.

Table 5. Correlations between the PC experience aspects.

	CE	CGE	FACGE
CE	1,000		
CGE	0,035	1,000	
FACGE	0,252	0,923*	1,000

4.1.3 Mental workload, Situational awareness and Pilot performance

MW, SA and PP were rated four times during a complete session. Pearson correlation tests were computed within each concept (n = 22[14]). One rating was completed prior to the session (pre) and three during the session (dur1-dur3). The rating means of mental workload did not reveal a variation over time (MWpre = 4.14, MWdur1 = 4.18, MWdur2 = 4.14, MWdur3 = 4.48). Thus, the fighter pilots experienced a moderate level of mental workload throughout the session. The correlation matrix reveals that MWpre correlated with MWdur2, and MWdur2 correlated with both MWdur1 and MWdur3 (see Table 6).

Table 6. Correlations between the pre and during measures of mental workload.

	MWpre	MWdur1	MWdur2	MWdur3
MWpre	1,000			
MWdur1	0,391	1,000		
MWdur2	0,444*	0,741*	1,000	
MWdur3	0,331	0,416	0,797*	1,000

The rating means of SA are presented in Figure 6. The figure reveal that pre ratings were lower than the during ratings, but that the during ratings stayed unaffected from SAdur1 to SAdur3.

[14] Data is missing from three of the participants for the third during measure of each aspect (n = 19).

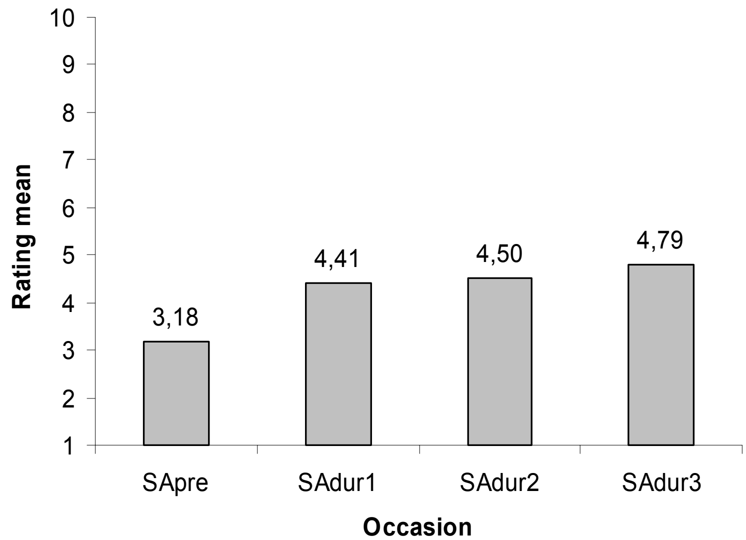


Figure 5. The mean values of situational awareness, before the session and at three occasions during the session.

SAdur1, SAdur2 and SAdur3 correlated with each other. SApre on the other hand, did not correlate with any of the during measures (see table 7). Thus, the pilots' ratings of situational awareness changed between the pre ratings and the ratings completed during the scenarios.

Table 7. Correlations between the pre and during measures of situational awareness.

	SAPRE	SADUR1	SADUR2	SADUR3
SApre	1,000			
SAdur1	0,009	1,000		
SAdur2	0,000	0,681*	1,000	
SAdur3	0,071	0,606*	0,722*	1,000

The mean values of the pilot performance ratings are presented in Figure 6. The figure reveal that pre ratings were higher than the during ratings, but that the during ratings stayed unaffected from PPdur1 to PPdur3.

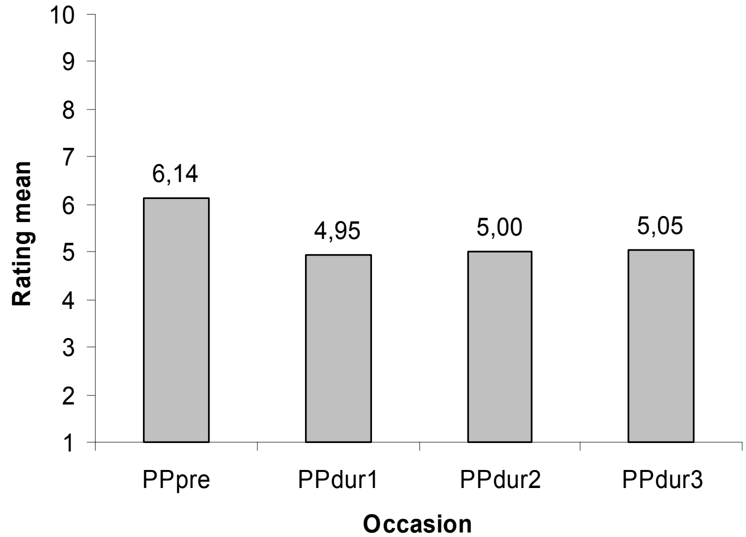


Figure 6. The mean values of pilot performance, before the session and at three occasions during the session.

The correlation pattern for PP was similar to the obtained pattern for SA, with one exception. PPpre correlated with PPdur2, even if the correlation was weaker than for the other significant correlations. The correlations are presented in Table 8.

Table 8. Correlations between the pre and during measures of pilot performance.

	PPPRE	PPDUR1	PPDUR2	PPDUR3
PPpre	1,000			
PPdur1	0,194	1,000		
PPdur2	0,470*	0,586*	1,000	
PPdur3	-0,035	0,857*	0,569*	1,000

Correlations were also computed between the pre measures (MWpre, SApre and PPpre) and similar for the three during measures. Regarding the pre measures, a correlation was found between MWpre and PPpre (see Table 9). For the three during measures, SAdur and PPdur showed correlations in all three cases (see Tables 10-12).

Table 9. Correlations between the pre measures of mental workload, situational awareness and pilot performance.

	MWPRE	SAPRE	PPPRE
MWpre	1,000		
SApre	-0,098	1,000	
PPpre	-0,441*	0,200	1,000

Table 10. Correlations between the first during measures of mental workload, situational awareness and pilot performance.

	MWDUR1	SADUR1	PPDUR1
MWdur1	1,000		
SAdur1	-0,310	1,000	
PPdur1	0,239	0,536*	1,000

Table 11. Correlations between the second during measures of mental workload, situational awareness and pilot performance.

	MWDUR2	SADUR2	PPDUR2
MWdur2	1,000		
SAdur2	-0,255	1,000	
PPdur2	-0,183	0,602*	1,000

Table 12. Correlations between the third during measures of mental workload, situational awareness and pilot performance.

	MWDUR3	SADUR3	PPDUR3
MWdur3	1,000		
SAdur3	-0,280	1,000	
PPdur3	-0,221	0,858*	1,000

The results suggest that higher rated mental workload was negatively correlated with rated pilot performance in the pre measurement. Situational awareness, on the other hand was found to be positively related to pilot performance in all three during ratings. The correlation even increased from 0,536 to 0,858 from the first to the third during measure. To further analyze the relation between mental workload, situational awareness and pilot performance, the values from each during measure for each of the three concepts were standardized. The correlations after standardization are presented in Table 13.

Table 13. Correlations between the standardized during measures of mental workload, situational awareness and pilot performance.

	MW	SA	PP
MW	1,00		
SA	-0.28*	1,00	
PP	-0.05	0.66*	1,00

MW and SA show a negative correlation, and SA and PP a positive correlation ($n = 63$, $p < .05$). To explore the causality between the three concepts, a structural equation model (SEM), ad modum LISREL[15] (Jöreskog & Sörbom, 1984), based on the correlations in Table 13, was calculated. This technique offers the possibility of testing the statistical goodness of fit in a population from which a sample has been drawn. The LISREL-analysis showed a model that further supports the relation between the three concepts (see Figure 7). Except the effects between MW and SA (-.28), and SA and PP (.66), the model also showed an indirect effect between MW and PP (-.18). All effects were significant ($n = 63$, $p < .05$). Adjusted Goodness of fit index was .86 and Root Means Square .055.

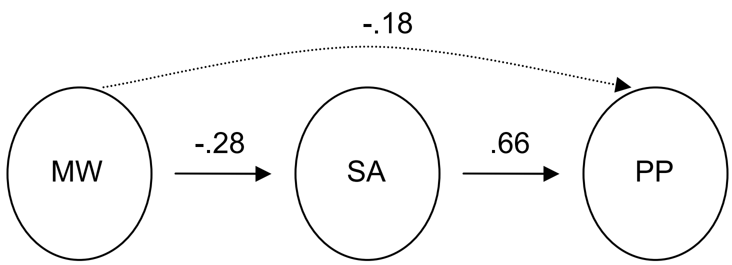


Figure 7. Illustration of the LISREL-model for the standardized during measures of mental workload, situational awareness and pilot performance.

Taken together, the general pattern reveals that different measures correlated strongly, that is a) different aspects of flight and PC experience) and b) the psychological concepts of mental workload, situational awareness and pilot performance over time (dur1-dur3). The pre measures did not correlate with the during measures to the same extent. It was further demonstrated how mental workload, situational awareness and pilot performance were related. The next step is to analyse how these measures relate to the "objective" performance measure that was obtained. This performance measure will therefore be described below.

4.1.4 PERFORMANCE

To analyze the performance of the participants during the sessions, three instances of the objective performance measure R will be used. For each scenario during the session the participants received zero for a loss and one for a victory. R1 is the result for each participant up to the point of the first during measure. Their obtained score is one if they had won one scenario and two if they had won 2 scenarios, and so on. R2 is the result from the scenario following the first during measure up to the point of the second during measure (and similar for R3). Thus, R1

[15] Analysis of LInear Structural RELationships.

to R3 are based on the performance for a separate number of scenarios. The value of R1 to R3 for each participant is presented in Table 14. Some pilots participated in several sessions and therefore have several occasions of R1 to R3.

Table 14. The distribution of the performance measures R1 to R3 for each of the participants and for each occasion of participation.

PILOT	OCCASION	R1	R2	R3
1	*1*	0	0	1
2	*1*	2	1	1
3	*1*	0	2	2
3	*2*	2	3	0
3	*3*	1	1	0
4	*1*	2	1	1
5	*1*	0	0	
6	*1*	0	1	0
7	*1*	0	2	2
7	*2*	2	1	2
7	*3*	3	3	
8	*1*	0	0	1
8	*2*	2	0	0
9	*1*	2	0	2
11	*1*	0	1	0
11	*2*	3	0	1
11	*3*	1	0	1
12	*1*	2	1	1
12	*2*	0	0	0
12	*3*	1	1	2
12	*4*	0	1	0
13	*1*	2	0	1

As the data in Table 14 might reveal, there were no significant correlations for the three performance measures R1 to R3 (ps > 0.05). This implies that there is no simple pattern for who win and who lose a specific scenario between two participants, even with the earlier performance result at hand. It is therefore problematic to treat these separate measures as an overall performance measure for the complete session. Thus, three different regression analyses were calculated.

Flight experience will be represented by total flight hours (TFH) and rated ability of WVR-combat (ABWVR), and computer experience by experience of fighter aircraft computer games with joystick (FACGE). Challenge of the WVR-scenarios (CHALL) will represent the three psychological aspects in the same way. TFH, ABWVR, FACGE and CHALLpre were included in all three regression analyses. The measures of mental workload, situational awareness and pilot performance, on the other hand, are more problematic. The pre measures and the first during measures of MW, SA and PP will be included when R1 is the criterion. Since the pre measures precede R1 closer than R2 and R3, they will only be used together with R1. When R2 is the criterion, the second during measures of MW, SA and PP will be included, and the third during measures will be included for R3. Thus, the regression analysis of R1 contained 10 independent factors. The analyses of R2 and R3 had seven factors.

The multiple correlations found in the three regression analyses were not significant. The values for R1, R2 and R3 were 0.78, 0.73 and 0.60 respectively. Thus, despite several measures that are believed to have a close relation to performance, no significant predictors of performance were found. However, the participants' ratings of pilot performance in the first and the third during measure showed to correlate with the corresponding performance measure respectively. The correlation between PPDUR1 and R1 was 0.465 ($n = 22$, $p < .05$) and between PPDUR3 and R3 it was 0.515 ($n = 19$, $p < .05$). PPDUR2 and R2 showed no significant correlation, although a tendency for a positive correlation was found. The two significant correlations between the pilot performance ratings and the performance measure were the only significant correlations found between R1 to R3 and any of the measures included in the three regression analyses. To continue, correlation tests were computed between the standardized measures of MW, SA and PP, used in the LISREL-model above, and the three performance measures R1 to R3. The test showed a significant correlation of 0.428 ($n = 63$, $p < .05$) between the standardized measures of PPDUR1 to PPDUR3 and R1 to R3 respectively. This indicates that the participants' rating of their performance is affected of whether they won or not before the rating.

4.2 Technical Evaluation Analyses

4.2.1 Realism, Limitation and Importance

The mean values (and standard deviations) of each fidelity level for each of the three rating dimensions are presented in Table 15.

Table 15. The mean values (standard deviations) for the seven fidelity levels concerning the three rating dimensions realism, limitation of performance and importance of realism for the training potential of a WVR-simulator.

FIDELITY LEVEL	REALISM	LIMITATION	IMPORTANCE
Transition head-up/head-down (HUHD)	2.1 (1.0)	6.5 (0.7)	6.2 (0.6)
Instrumentation (INSTR)	5.0 (1.5)	2.4 (1.3)	5.6 (1.2)
Flight Controls (FC)	3.4 (1.6)	2.7 (1.4)	4.5 (1.8)
Graphics (GRAPH)	4.8 (1.4)	2.9 (1.6)	4.4 (1.6)
Feedback (VFM)	4.9 (1.4)	3.7 (1.4)	6.2 (0.7)
Visual resolution (VRES)	-	4.3 (1.8)	6.2 (0.8)
Field of view (FOV)	-	6.5 (0.7)	6.6 (0.7)
Overall mean	4.0	4.1	5.7

The first analysis performed was to compare the ratings of realism with the real setting, meaning flying the real fighter aircraft (that is argued to be represented by the top value of seven on the scale). The analysis revealed that the rated realism (mean values for the five rated fidelity levels) was lower than the "true" value of realism, that is seven on the scale used[16] ($n = 13$, $p < .05$). The standard errors (SE) and t-values are presented in Table 16. Thus, it can be concluded that even if the realism values obtained were relatively high, the ratings were significantly lower than the "real thing".

Table 16. The pair wise differences obtained between the realism ratings of the five fidelity levels and the top value of realism.

	SE	t
HUHDreal < 7	0,288	-17,103
INSTRreal < 7	0,408	-4,900
FCreal < 7	0,432	-8,374
GRAPHreal < 7	0,378	-5,900
VFMreal < 7	0,390	-5,527

Multiple t-tests were performed for the seven fidelity levels values of each rating dimension ($n = 13$, $p < .05$). The general pattern obtained was that head-up/head-down (HUHD) and field of view (FOV) were the fidelity levels that diverged from the others (see Table 17). Both HUHD and FOV had lower ratings regarding limitation than several of the other fidelity levels. Further, HUHD also showed to diverge regarding realism. The third dimension, importance of realism for the training potential of a WVR-simulator, is more problematic to interpret. It could be the case

[16] To expect any other result might be a viewed as utopian.

that HUHD and FOV was receiving highest ratings of importance since they were the fidelity levels that received highest ratings of limitation *or* it could be the case that they actually are the most important aspects for a WVR-simulator to be experienced as a functional, effective practising tool for WVR combat. It is nevertheless beyond the scoop of the present work to pin-point that question.

Table 17. The pair wise differences obtained between the fidelity levels for each dimension.

	SE	t
Realism		
HUHD < INSTR	0,431	-6,789
HUHD < GRAPH	0,472	-5,703
HUHD < VFM	0,378	-7,323
FC < INSTR	0,368	-4,395
FC < VFM	0,386	-3,787
Limitation		
HUHD > INSTR	0,445	9,157
HUHD > FC	0,378	9,968
HUHD > GRAPH	0,460	7,852
HUHD > VFM	0,482	5,740
VRES > INSTR	0,459	4,186
FOV > INSTR	0,445	9,157
FOV > FC	0,426	8,848
FOV > GRAPH	0,446	8,100
FOV > VFM	0,482	5,740
FOV > VRES	0,465	4,635
Importance		
HUHD > FC	0,417	3,877
HUHD > GRAPH	0,323	5,472
FOV > FC	0,500	4,158
FOV > GRAPH	0,426	5,237

4.2.2 LIMITATION AND IMPROVEMENT

Two multi-choice questions were answered by the participants in QDuring. The data regarding which fidelity levels, if any, that limited the participant's performance during the last scenario prior to the questionnaire is presented in Figure 8. As could be seen in the chart, the fidelity level HUHD together with FOV had the highest number of hits, with 30 each.

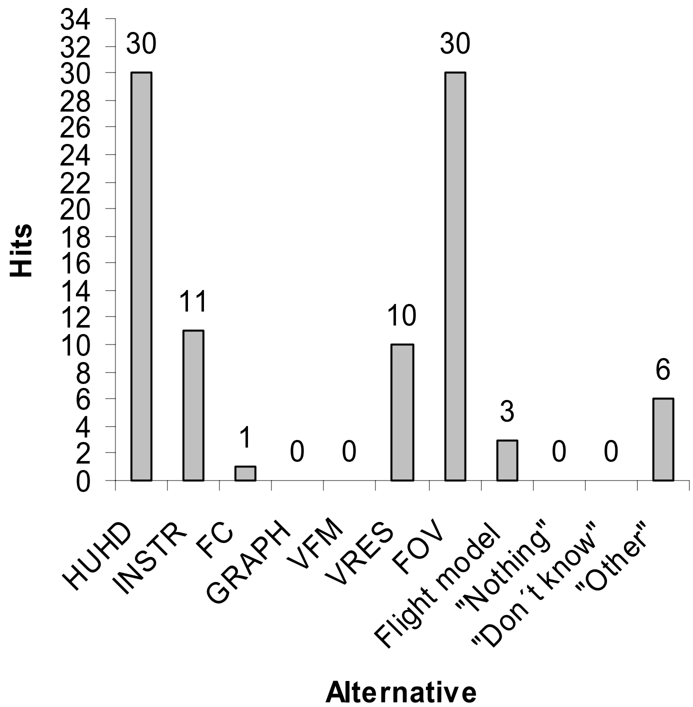


Figure 8. The number of hits for the alternatives in the multi-choice question in During, concerning limitation of performance.

The maximum value that could be obtained was 34 hits[17]. Instrumentation (INSTR) and visual resolution (VRES) had some more hits than the remaining fidelity levels, but rather few in comparison to HUHD and FOV. Regarding the alternative “Other”, one participant complained on a low “flight feeling” and one on graphical distortions. Three participants expressed that the knowledge about the aircrafts position in relation to the surrounding world together with the control of speed, altitude and height of the own aircraft was limited. It was interesting to obtain low frequencies on the alternative flight model since it was of interest to see if the participants felt limited by the F-16 model. As could be seen in the figure above, flight model only had three hits. Since the F-16 model was not of further interest for evaluation, it was not included in the ratings for three dimensions realism, limitation and importance in QAfter. In total there were 91 hits for the different alternatives.

---

[17] Ten of the participants performed QDuring at three occasions and two of them at two occasions. Data is missing from one participant.

In QAfter, the participants were asked what, if anything, that has to be improved in the Illustrator for future use as a training environment for WVR-combat. The alternatives were the same as for limitation of performance in QDuring that was presented above. The data is presented in Figure 9.

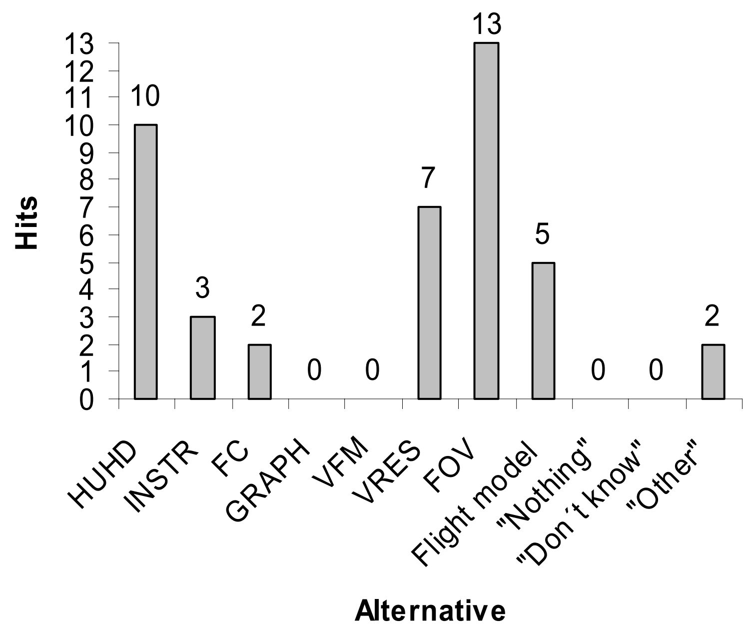


Figure 9. The number of hits for the alternatives in the multi-choice question in QAfter, concerning future improvements.

FOV and HUHD again had the highest number of hits, with 13 and 10 hits respectively. The maximum value that could be achieved was 13[18]. Further, VRES had a rather high value with seven hits. In total there were 42 hits for the different alternatives. Regarding the two hits for the alternative "Other", one participant meant that the sound picture should be improved to enhance the spatial awareness and another just that the sound picture should be improved. All together, the multi-choice questions revealed the same pattern as the questions about realism, limitation and importance presented above.

The results of the technical evaluation (technical user acceptance) revealed several interesting differences. Two fidelity levels, head-up/head-down and field of view, were rated significantly different compared to the other fidelity levels. Further, similar tendencies were

[18] The first occasion of QAfter for each of the thirteen participants was analyzed.

found regarding limitation in QDuring and improvement in QAfter (the two multi-choice questions). The results from the psychological and technical evaluations will be discussed further in the next section.

5 RESULT DISCUSSION

5.1 PSYCHOLOGICAL EVALUATION

No significant effects were found for the three variables attitude towards simulators in general, and commitment and challenge of the WVR-scenarios, from before to after the session. Attitude towards simulators in general showed the same value before as after the session. However, the values tended to drop from before to after for commitment and challenge. Commitment of the WVR-scenarios dropped from 5.5 to 4.8, and challenge of the WVR-scenarios from 5.2 to 4.5. This is interesting to relate to research performed for real flight. In several studies, it has been found that the activation among pilots is higher before flight than after (e.g. Svensson, Angelborg-Thanderz, Sjöberg & Gillberg, 1988; Svensson, Thanderz & Uneståhl, 1980). Before flight the pilots showed to be high in activity, but afterwards they were relaxed and tired, that means low in activity. However, to decide whether this was the case or not would have demanded comparisons with a control group, for example performing the same ratings before and after a real WVR-combat instead. Finally, to expect higher values initially would not be realistic at this stage of the development. To relate to the purpose of establishing the distribution of the participants' ratings of the three aspects initially and after the session, the values are believed to be rather high, both initially and after the session, and no significant changes were found for any of the three aspects from before to after the session.

Regarding the psychological concepts of mental workload, situational awareness and pilot performance, there are several interesting result. First, a clear tendency for a negative correlation between mental workload and situational awareness was found, and situational awareness and pilot performance showed a positive correlation at all three during measures. Further, the LISREL-model performed on the standardized data of MW, SA and PP is in line with earlier findings (e.g. Berggren, 2002; Magnusson, 2002; Svensson & Wilson, 2002). Thus, the purpose of exploring the relation between mental workload, situational awareness and pilot performance is then fulfilled with the interesting observation that a similar pattern between the concepts as in earlier findings was obtained here as well.

The performance analyses revealed that none of the included variables could explain variance in performance, since none of them could predict performance. As discussed in the introduction, earlier research has shown a strong relation between experience and performance in several areas (e.g. Ericsson & Simon, 1993), not at least among fighter pilots (e.g. Angelborg-Thanderz, 1990). Together with the finding that the three performance measures (R1 to R3) did not correlate, it seems like performance in the Illustrator is rather random. This is not in line with several earlier findings, since not a single measure could fill the purpose of predicting performance – not even the flight experience measures. The reason for this might be some major weaknesses of the Illustrator that were identified in the technical evaluation. Interesting to note however, is that a correlation was found between the three standardized during measures of pilot performance and the three performance measures (R1 to R3) respectively, indicating that whether pilot A did win or lose the scenarios affected his rating of his performance.

5.2 TECHNICAL EVALUATION

All of the five fidelity levels that were rated for realism had a mean significantly different from the "true" value of realism. This was rather expected since there are major differences between the real operational setting and the Illustrator.

To clarify the discussion about the fidelity levels, the rating scale from one to seven is divided into three interval sections. One to three are low values, three to five middle values and five to seven high values. The optimal values for a fidelity level would be a high value of realism, a low value of limitation of performance and a high value of importance of realism for the training potential of a future WVR-simulator[19]. During the discussion about the seven fidelity levels it should be remembered that the Illustrator is the first prototype of a low-cost system for future training of WVR-combat. This is primarily important when discussing the realism values of the fidelity levels. The top value of seven is utopian. Therefore, a value around five should be treated as very promising and a value around six as excellent.

Transition head-up/head-down is low in realism, high in limitation and high in importance, which means the worst-case scenario. Head-up/Head-down had a lower value of realism than

[19] A low value of importance would not be a problem other in the sense that the developmental resources could have been better distributed.

instrumentation, graphics and visual feedback according to a manoeuvre. Further, head-up/head-down showed significantly higher values regarding limitation of performance than all other fidelity levels except field of view and visual resolution. It also had the highest number of hits regarding limitation of performance in QDuring and the second highest of future improvements in QAfter. The alarming values for head-up/head-down were rather expected since the instrumentation in the Illustrator is presented on a screen in front of the user. In a WVR-scenario the operator has to turn his head in order to keep visual contact with another aircraft. Therefore it can be rather difficult to maintain control over the instruments. This problem is also affected by the rather limited field of view in the Illustrator, and the fact that the operator has to look under the lower edge of the VR-goggles to see the instrumentation. This will be further discussed below.

Instrumentation (INSTR) had a high value of realism, a low value of limitation and a high value of importance, meaning the optimal values as mentioned above. In QDuring however, instrumentation had eleven hits regarding limitation. That is a rather high value. According to the values of realism, limitation and importance in QAfter, a lower value would be expected. That the value of limitation in QDuring is a bit suspicious is also supported by the fact that only three participants chose instrumentation regarding improvements in QAfter. It might be the case that the ratings regarding limitation in QDuring were affected by experienced problems with transition head-up/head-down. In total, instrumentation is believed to show the most promising values of all the evaluated fidelity levels. This is supported by two participants who specifically mentioned that the instrumentation of the Illustrator were satisfying in part two of QAfter. Not a single participant mentioned the instrumentation as a weakness. However, for other activities than WVR-combat this might not be the case at all. WVR-combat is an almost exclusively visual activity with only short glimpse on the instrumentation. If the study had concerned some scenario with higher demand on instrumental flying, the results might have been different since the instrumentation would have been more frequently used. Perhaps that would have revealed problems that were not found in this study.

Flight controls (FC) received a middle value of realism, a low value of limitation and a middle value of importance. It further showed a very low number of hits regarding both limitations in QDuring and improvements in QAfter with one and two hits respectively. This is interesting since the controls in the Illustrator are low-cost products (speaking in economical terms of simulator development) that are available on the market to any enthusiast. It might be the case that controls of this type, with major differences compared to the real environment, are satisfying for the purpose of WVR-combat simulation. However, further evaluations will have to prove if that is the case or not.

Graphics (GRAPH) showed a middle value of realism, a low value of limitation and a middle value of importance. Regarding limitation in QDuring and improvement in QAFter, it received zero hits in both cases. The value of realism for graphics is rather high.

Visual feedback according to a manoeuvre (VFM) has a middle value of realism, a middle value of limitation and a high value of importance. However, in QDuring and QAfter VFM had zero hits regarding both limitation and improvement. Further, the value of realism for visual feedback according to a manoeuvre is rather high.

Visual resolution (VRES) had a middle value of limitation and high value of importance. Further, it had a high number of hits regarding limitation of performance in QDuring. For improvements in QAfter, visual resolution had the third highest number of hits, after field of view and head-up/head-down. However, the value of limitation of performance is not as high as for field of view and head-up/head-down. The main reason for those results is probably that it sometimes is hard to see the other aircraft against the sky in the Illustrator, since the aircrafts are grey. As mentioned earlier, the operator sometimes has to check the instrumentation during WVR-combat, in real flight as well as in simulated. This means a loss of visual contact with the enemy for a second or so. When trying to establish visual contact again the limited field of view seem to be a problem. During the study it was observed that the participants often swept by the enemy aircraft without noticing it against the blue-grey sky. One participant stated that the visual resolution is a weakness in the Illustrator since its hard to judge angles and the speed of an enemy aircraft. The visual resolution in the VR-goggles used in the study is 800 by 640 pixels. It might be the case that this is not enough.

Field of view (FOV) showed to be very high in both limitation and importance. The value of limitation of performance was significantly higher than for all the other fidelity levels except head-up/head-down. Further, field of view had most hits together with head-up/head-down regarding limitation of performance in QDuring. For future improvements in QAFter, all of the participants marked field of view. The field of view in the Illustrator is restricted to approximately 30 degrees. Compared to about 90-100 degrees for a person with normal sight, this is a very restricted view. During WVR-combat, the flight is almost exclusively based on visual references. Therefore the alarming results were rather expected. Further, field of view was mentioned as a weakness in part two of QAfter by all of the participants. Many of them stated that the limited field of view makes it hard to keep control of the position of the aircraft in relation to the surrounding world (further called spatial awareness) and to keep control over the instrumentation. This is closely related to the problems found with head-up/head-down. With a greater field of view, the possibility to maintain control over the instrumentation during flight would increase. As the Illustrator is today, a short switch from looking at another aircraft to the instrumentation and back again often results in a loss of visual contact[20]. Four of the participants mentioned this problem with head-up/head-down as a weakness in part two of QAfter. Further, several of the participants stated that the peripheral sight is the most important factor for spatial awareness. One participant wrote, "The field of view is too small. This results in abnormal difficulties with keeping up the spatial awareness, to control the instrumentation and to find a once lost enemy aircraft. The operator will therefore learn an inappropriate behaviour regarding instrument control which could lead to fatal consequences". Finally, it should be mentioned that finding the enemy aircraft after head-down in a real WVR-combat is not always an easy procedure either. However, the limited field of view in the Illustrator complicates this procedure even more. Thus, the need for improvements of field of view and head-up/head-down is obvious.

To conclude the discussion of the technical evaluation and to relate to the purpose of establishing eventual needs of future development, there seems to be two fidelity levels of major interest for the forthcoming development of the Illustrator, head-up/head-down and field of view. Visual resolution also seems to be a crucial fidelity level in the Illustrator, but to a lower extent and the obtained results might be a product of the limited field of view. However, remember that head-up/head-down, field of view and visual resolution all showed higher values regarding importance of realism than both flight controls and graphics. To finish the discussion of the results, the major problems in the Illustrator are within Stammers (1983; 1986) dimension of simulation called stimulus/displays.

---

[20] Meaning a constant loss and not a temporary loss caused of the transition head-up/head-down.

6 METHODOLOGY DISCUSSION

The results obtained in this study are only valid for this specific version of the Illustrator. However, general conclusions and recommendations could be put in relation to other systems. The problems with transition head-up/head-down and field of view could surely be generalized to other WVR-combat environments, similar to the Illustrator. Further, the systematic method of using expert opinions is applicable for the purpose of evaluations in other simulators and training environments.

6.1 QUESTIONNAIRES

The questions in the different questionnaires used in the study were designed by personnel at FOI. Many of the questions had been used earlier but some are new. The scale running from one to seven was chosen with regard to literature about psychological research methods and earlier studies conducted by FOI (e.g. Svensson et al., 1997; Svensson & Wilson, 2002).

As stated earlier, there was no question regarding realism for the two fidelity levels visual resolution and field of view. The reason for this was that both visual resolution and field of view are physical facts that we all face every day. Flight controls or instrumentation of an aircraft, for example, are designed artefacts. During the design of the questionnaires this was discussed and a decision was made to exclude the question regarding realism for visual resolution and field of view. However, if visual resolution and field of view had been rated for realism, at least the latter would probably have received a very low value. But, that might be wrong, and afterwards the opinion is that the question should have been asked. The graphics of the Illustrator was rated for realism and that surely is not an artefact. Further, all comparisons made in the data analyses could have been performed for all fidelity levels, which had simplified the statistical analyses.

6.2 SUBJECTIVE RATINGS

The study totally relies on subjective ratings made by the participants. The only data of a more objective nature collected was which one of the participants who won a scenario. The use of subjective ratings is sometimes criticised, especially when they alone constitute the collected data. Regarding the questions in the technical evaluation they all dealt with the Illustrator and its parts. For that reason, their reliability is not questioned. In the psychological evaluation however, ratings regarding attitude, commitment, challenge, mental workload, situational awareness and pilot performance were made. The ratings of those aspects are of a more introspective nature. Those ratings concerns internal states of the participants and it is often questioned how reliable results of this type are. However, the reliability of the data is supported by the fact that all participants were active fighter pilots, and they regularly participate in studies were they are to perform subjective ratings.

6.3 FLIGHT MODEL

For reasons discussed in the method section, an F-16 Fighting Falcon flight model was used during the study, instead of a JA 37 Viggen model. Since the participants were experts on JA 37 and not F-16, the flight model was not evaluated along the three dimensions realism, limitation and importance in QAfter. However, despite that the participants were not experts on F-16 they all had experience of WVR-combat. Their ability to express reliable opinions regarding the Illustrator and its parts is not questioned in any sense because of the F-16 model.

To control the participants' opinions of the F-16 model, flight model was included as an alternative regarding limitation of performance in QDuring and improvements in QAfter. In QDuring, flight model was chosen at three occasions. For improvements in QAfter, five participants chose flight model. This was rather expected since there are differences between an F-16 and a JA 37 Viggen[21]. In future evaluations it would be preferred to use a JA 37 model.

7 CONCLUSIONS AND RECOMMENDATIONS

Regarding the psychological evaluation, the monitoring of the three psychological aspects was positive. Although commitment and challenge dropped some from before to after, the values were initially high and they are still rather high after the drop. There seem to have been a positive attitude towards the Illustrator among the participants. During the further development, close contact with the pilots should be maintained to make sure that their expertise comes to use. As mentioned earlier, one of the most important aspects of modern simulator design is believed to be the use of operator expertise.

The analyses of mental workload, situational awareness and pilot performance revealed a pattern in line with several earlier findings in simulators as well as in real flight. To some extent that might be viewed as a positive result for the system, but primarily for the data collection

[21] Exactly what differences there are is beyond the scope of this report.

method. Keep in mind that none of the measures included in the regression analyses of performance showed the capacity of predicting performance. This suggests that performance in the environment is rather randomized. At least, no pattern explaining why pilot A beat pilot B was revealed. A major cause is believed to be technical weaknesses of the system. To some extent, the flight model also might have affected this result.

In the technical evaluation, both positive and negative results were found for the seven fidelity levels. Keep in mind that the results obtained regard the specific situations related to WVR-combat. Their validity for other areas of use of the Illustrator is unknown. Transition head-up/head-down and field of view turned out to be alarming problems. Regarding transition head-up/head-down, one participant suggested that it should be possible to see the instrumentation on a screen as today, but through the VR-goggles instead of under them. Further, the problems with transition head-up/head-down are closely related to the limited field of view. Therefore, improving the field of view would probably improve transition head-up/head-down as well. However, the aim should be a combination of greater field of view and improvements regarding the location of the instrumentation.

A greater field of view is of greatest priority in the future development of the Illustrator. Thirty degrees seem to be far too limited for the purpose of WVR-simulation. If the solution for transition head-up/head-down with semi-transparent VR-goggles is chosen, a greater field of view would probably improve transition head-up/head-down. As an initial and cheap solution, some of the participants mentioned that a symbol pointing towards the enemy aircraft presented in the VR-goggles would be of great help to compensate for the limited field of view. Also an indication of the distance between the aircrafts was suggested. In many ways, the limited field if view seems to the fundamental problem in the Illustrator.

The problem with the visual resolution could probably be solved by changing the colours of the aircrafts or in some other way make them more visible. This was addressed by one participant in part two of QAfter. Further, a greater field of view would probably affect the operator's experience of the visual resolution as well. As stated in the discussion, several participants in the study swept by the enemy aircraft without noticing it when trying to establish visual contact after head-down. To decide whether the visual resolution has to be improved, another study with a far greater field of view should be performed. It might be the case that the greater field of view and more visible aircrafts would be satisfying.

Regarding the visual feedback according to a manoeuvre, it might be the case that improved field of view would lead to improvement here as well. With a greater field of view and improved transition head-up/head-down, the operator's spatial awareness would probably increase in two ways. First, the operator would see more of the surrounding environment like clouds, the sun or the moon, and the ground. As discussed earlier, visual references are important for spatial awareness. Secondly, improved transition head-up/head-down would mean a more proper instrument control. This is also believed to improve the spatial awareness. Further, this would probably improve the visual feedback according to a manoeuvre, since the only way to establish that is to control the aircrafts behaviour with visual references and/or the instruments. If the operator's possibility of doing so is very limited, the visual feedback is not experienced properly. One participant also mentioned, in part two of QAfter, that the "sound image" changes according to for example performed manoeuvres, speed, rpm and alpha during real flight. To simulate those changes accurately is another idea for increasing the "flight feeling".

Flight controls and graphics are believed to have shown rather positive values in the evaluation. At this stage, the current flight controls are believed to be satisfying for the purpose of simulating WVR-combat. Future studies will reveal if that is a correct conclusion or not. For graphics, a rather high value of realism was shown together with no hits regarding future improvements. Thus, the problem with the visual resolution is recommended to be prioritized regarding the visual presentation.

The instrumentation showed the most promising values of all evaluated fidelity levels. According to the results, the instrumentation is believed to be satisfying for the purpose of WVR-combat at present. However, this might change if the field of view and transition head-up/head-down is improved, since this probably would mean a more frequent use of the instrumentation.

To continue, some recommendations of the participants (from part two of QAfter) regarding other areas of use of the Illustrator than WVR-combat, will be presented. Several participants mentioned training of file flying and other activities with high demand on visual references, for example start and landing procedures. Several mentioned that the Illustrator could be used in tactical as well as system development projects. One participant suggested that a system like the Illustrator should be integrated with FLSC. FLSC is an existing simulator environment, mainly for BVR-combat. This would, according to the participant, constitute a very potential training facility with great possibilities.

As seen above, a lot of ideas and suggestions were collected. Especially interesting is that there seem to be a belief that the system has a potential if further developed. As one of the participants wrote, "For the first time I believe that there might be a possibility of instructing a dogfight in a simulator". The road ahead might be long and winding but there seem to be a potential in this kind of PC-based systems. Not to forget, regularly performed empirical studies and the use of the pilots' expertise is of greatest importance.

The general conclusion of the study is that the Illustrator suffer of some technical weaknesses, mainly head-up/head-down and field of view, which is further supported by the fact that performance seem to be rather random in the sense that no predictors of performance were found and that the three performance measures did not correlate.

8 FUTURE RESEARCH

The results indicate that the Illustrator has a high potential if further developed. The first step after the Illustrator has been further improved according to the suggestions above should be another utility evaluation. In this evaluation, it is proposed that two other fidelity levels are included; the sound and the degrees of freedom in the head tracker. However, the obvious reason for performing this second utility evaluation is of course to establish whether believed technical improvements are perceived as improvements. Further, improvements of the major fidelity problems with head-up/head-down and field of view might shine light at other fidelity levels that were not perceived as a problem in the first evaluation.

The system will never be perfect regarding realism, but the most central aspect is not that the realism is total (since that will never happen). The aim should instead be to reach a level where the system can fill the purpose of use, simulator training of WVR-combat. To reach this level, the development should be based on regularly performed evaluations. First when the users' performances are not limited by the system to any greater extent, experiments measuring the training effects should be performed. To follow Bell and Waag's model (1998), the first step then is to conduct experiments measuring the in-simulator learning, with the aim of measuring performance improvement within the environment. The idea is that transfer to the real environment is unlikely if performance improvement within the system is not present. To establish in-simulator learning, test scenarios should be performed before and after training scenarios. The test scenarios should be similar but not identical to the training scenarios. Finally, relevant measures that reflect performance improvement have to be developed.

9 REFERENCES

Adam, E. C. (1995). Tactical Cockpits – Flat Panel Imperatives. In *Situation Awareness: Limitations and Enhancement in the Aviation Environment.* AGARD-CP-575. Pp. 9/1-7.

Angelborg-Thanderz, M. (1990). Military Flight Training at a Reasonable Price and Risk. *Swedish National Defense Research Establishment report C 50083-5.1.*

Bell, H. H. & Waag, W. I. (1998). Evaluating the Effectiveness of Flight Simulators for Training Combat Skills: A Review. *The International Journal of Aviation Psychology, 8*(3), 223-242.

Berggren, P. (2000). Situational Awareness, Mental Workload and Pilot Performance – relationships and conceptual aspects. *National Defense Research Establishment report FOA-R--00-01438-706--SE.*

Boreham, V. C. (1985). Transfer of training in the generation of diagnostic hypothesis: the effect of lowering fidelity of simulation. *British Journal of Educational Psychology, 55*(3), 213-223.

Clymer, A. B. (1981). Simulation for training and decision making in large scale control systems. In Cheremisinoff, P.N. and Perlis, H. J. (Eds.), *Automated process control*. London: Arbor.

Danielsson, B., Svensson, E. & Jenvald, J. (2002). Pedagogical Tools for Within Visual Range (WVR) Fighter Pilot Training. *Swedish Defense Research Agency report FOI-S--0453--SE.*

Endsley, M. R. (1995). Theoretical Underpinnings of Situational Awareness: A Critical Review. In Garland, D. & Endsley, M. R. (Eds.). *Experimental Analysis and Measurement of Situational Awareness.* Embry-Riddle Aeronautical University Press, Daytona Beach.

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors, 37*(1), 32-64.

Ericsson, K. A. & Simon, H. A. (1993). *Protocol Analysis* (Second Ed.). Cambridge: MIT Press.

Fisk, A. D. & Lloyd, S. J. (1988). The Role of Stimulus-to-Rule Consistency in Learning Rapid Application of Spatial Rules. *Human Factors, 30,* 35-49.

Koonce, J. M. & Bramble Jr., W. J. (1998). Personal Computer-based Flight Training Devices. *The International Journal of Aviation Psychology, 8*(3), 277-292.

Magnusson, S. (2002). Similarities and Differences in Psychophysiological Reactions Between Simulated and Real Air-to-Ground Missions. *The International Journal of Aviation Psychology, 12*(1), 49-61.

Rolfe, J. M. (1985). *Fidelity and its Impact on the Design, Precurement, Use and Evaluation of Flight Simulation*. Report as part of Defence Fellowship Study, Wolfson College, Cambridge.

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. *The International Journal of Aviation Psychology, 8*(3), 197-208.

Simon, H. A. (1990). Invariants of Human Behaviour. *Annual Review of Psychology, 41,* 1-19.

Stammers, R. B. (1981). Theory and practice in the design of simulators, *PLET*, *18* (2), 69-71.

Stammers, R. B. (1983). Simulators for training. In Kualseth, T. O. (Ed.), *Ergonomics of Workstation Design*. London: Butterworth.

Stammers, R. B. (1986). Psychological aspects of simulator design and use. In Lewins, J. and Becker, M. (Eds.). *Advances in Nuclear Science and Technology*. New York: Plenum Press.

Stanton, N. (1996). Simulators: a review of research and practice. In Stanton, N. (Ed.), *Human Factors in Nuclear Safety* (pp. 117-140).

Svensson, E. & Angelborg-Thanderz, M. (1999). Models of Pilot Performance for Systems and Mission Evaluation – Psychological and Pychophysiological Aspects. Report from USAF Research Laboratory.

Svensson, E., Angelborg-Thanderz, M., Sjöberg, L. & Gillberg, M. (1988). Military Flight Experience and Sympatho-adrenal activity. *Aviation, Space and Environmental Medicine, 59,* 411-416.

Svensson, E., Angelborg-Thanderz, M., Sjöberg, L. & Olsson, S. (1997). Information Complexity – Mental Workload and Performance in Combat Aircraft. *Ergonomics, 40*(3), 362-380.

Svensson, E., Angelborg-Thanderz, M., Olsson, S. & Sjöberg, L. (1992). Information Overflow – Mental Workload and Performance in Military Aviation. *Swedish National Defense Research Establishment report C 50097-5.1.*

Svensson, E., Angelborg-Thanderz, M. & Uneståhl, L.-E. (1980). Avslappningsträning, ett försök vid krigsflygskolan. *National Defense Research Establishment report C 590003-H9*.

Svensson, E., Nählinder, S., Danielsson, B. & Jenvald, J. (2002). Pedagogical Tools for Within Visual Range (WVR) Fighter Pilot Training – Another Step Forward. Proceedings of the First Swedish-American Workshop on Modeling and Simulation SAWMAS-2002.

Svensson, E. & Wilson, S. (2002). Psychological and Psychophysiological Models of Pilot Performance for Systems Development and Mission Evaluation. *The International Journal of Aviation Psychology, 12*(1), 95-110.

VINTHEC. (1997). *Dynamic Measures of Pilot Mental WorkLoad (PMWL), Pilot Performance (PP) and Situational Awareness (SA).* Technical Report: VINTHEC-WP3-TR 01.

Welham, D. J. (1986). Technology-based training and simulation. In Labinger, M. and Finch, P. J. (Eds.), *Technology-Based Training*. Maidenhead: Pergamon Infortech.

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

Wilson, G. F. (1997). Mental Workload. In Salvendy, G. (Ed.) (1997). *Handbook of Human Factors and Ergonomics.* New York: Wiley.

Issuing organization	Report number, ISRN	Report type
FOI – Swedish Defence Research Agency Command and Control Systems P.O. Box 1165 SE-581 11 Linköping	FOI-R--0710--SE	Scientific report
	Research area code 2. Operational Research, Modelling and Simulation	
	Month year December 2002	**Project no.** E 7048
	Customers code 5. Commissioned Research	
	Sub area code 21 Modelling and Simulation	
Author/s (editor/s) Jonathan Borgvall Staffan Nählinder Jan Andersson	**Project manager** Staffan Nählinder	
	Approved by Lennart Nyström	
	Sponsoring agency Swedish Armed Forces	
	Scientifically and technically responsible Staffan Nählinder	

Report title

WVR-Illustrator Evaluation: Using Pilot Expertise for Future Development

Abstract (not more than 200 words)

The Illustrator is a demonstrator of a future training simulator for within visual range (WVR) air-to-air combat, based on COTS-products. The focus of the study was to evaluate the Illustrator together with pilots from the Swedish Air Force, and to transform those expert opinions into guidelines for continued development. Two perspectives were undertaken, one psychological and one technical, with questionnaires and objective performance measures as foundation for the evaluation. The former concerned the measurement of several psychological aspects that were to be related to earlier research. In the technical evaluation, several aspects, fidelity levels, were evaluated regarding realism, limitation of performance, importance of realism and improvement need. Thirteen fighter pilots participated. The sessions consisted of several WVR-scenarios. Several results concerning the psychological evaluation were in line with earlier findings. However, not a single predictor of performance was found. Two fidelity levels, transition head-up/head-down and field of view showed to diverge from the other fidelity levels. They are therefore recommended to be prioritized during the further development. The general conclusion is that there seem to be a potential for the Illustrator to be a valuable tool for practicing WVR-combat if further development is undertaken.

Keywords

flight simulator, evaluation, fidelity, air-to-air combat, dogfight, fighter pilot, expertise, performance

Further bibliographic information	Language English
ISSN 1650-1942	**Pages** 54 p.
	Price acc. to pricelist

Utgivare	Rapportnummer, ISRN	Klassificering
Totalförsvarets Forskningsinstitut - FOI Ledningssystem Box 1165 581 11 Linköping	FOI-R--0710--SE	Vetenskaplig rapport
	Forskningsområde 2. Operationsanalys, modellering och simulering	
	Månad, år December 2002	**Projektnummer** E 7048
	Verksamhetsgren 5. Uppdragsfinansierad verksamhet	
	Delområde 21 Modellering och simulering	
Författare/redaktör Jonathan Borgvall Staffan Nählinder Jan Andersson	**Projektledare** Staffan Nählinder	
	Godkänd av Lennart Nyström	
	Uppdragsgivare/kundbeteckning Försvarsmakten	
	Tekniskt och/eller vetenskapligt ansvarig Staffan Nählinder	

Rapportens titel (i översättning)

Utvärdering av närstridsillustrator: Piloters expertis till stöd för framtida utveckling

Sammanfattning (högst 200 ord)

Illustratorn är en demonstrator av en framtida träningssimulator för närstrid (”dogfight”) mellan flygplan. Studien fokuserade på en expertutvärdering av Illustratorns överensstämmelse med en verklig närstridssituation. Målet var att omsätta experternas åsikter och kunskap till riktlinjer för framtida utveckling av Illustratorn. Utvärderingen genomfördes utifrån två perspektiv, ett psykologiskt och ett tekniskt, med enkäter och objektiva prestationsmått som underlag för utvärderingen. Den psykologiska utvärderingen omfattade mätning av flera psykologiska aspekter i syfte att relatera resultaten till tidigare forskning. I den tekniska delen utvärderades olika delar av Illustratorn utifrån realism, prestationsbegränsning, behov av realism och behov av förbättring. Tretton militära flygförare deltog i studien. En testomgång bestod av flera närstridskontakter. Den psykologiska utvärderingen gav flera resultat i linje med tidigare forskning. Däremot kunde inga prediktorer av prestation, huruvida pilot A vann eller förlorade en kontakt, påvisas. Två aspekter av systemet, växlingen ”head-up/head-down” och synfältet, avvek och utmärkte sig på flera sätt från de andra aspekterna. Rekommendationen för framtiden är framförallt en vidareutveckling och förbättring av dessa två aspekter. Den generella slutsatsen är att system av denna typ verkar inneha potential i riktning mot en framtida träningssimulator, förutsatt att föreslagna förbättringar och vidare utvärderingar genomförs.

Nyckelord

flygsimulator, närstrid, utvärdering, fidelitet, pilot, expertis, prestation

Övriga bibliografiska uppgifter	**Språk** Engelska
ISSN 1650-1942	**Antal sidor:** 54 s.
Distribution enligt missiv	**Pris: Enligt prislista**