Nationaal Lucht- en Ruimtevaartlaboratorium
National Aerospace Laboratory NLR

NLR-TP-2002-693

Skill-transfer of PC-based simulation to real flight - a comparison of in-flight measured data and instructor ratings

J.J.M. Roessingh

Nationaal Lucht- en Ruimtevaartlaboratorium
National Aerospace Laboratory NLR

NLR-TP-2002-693

Skill-transfer of PC-based simulation to real flight - a comparison of in-flight measured data and instructor ratings

J.J.M. Roessingh

This report is based on a presentation held at the 12th International Symposium on Aviation Psychology, at Dayton, Ohio USA on April 14-17, 2003.

Customer: National Aerospace Laboratory NLR
Working Plan number: V.1.C.2
Owner: National Aerospace Laboratory NLR
Division: Flight
Distribution: Unlimited
Classification title: Unclassified
December 2002

Summary

Three groups of novice pilots received training to fly a complex sequence of aerobatic maneuvers in a light aircraft. Trainees in the control group received in-flight instruction and were given the usual pre- and post-flight briefings by a flight instructor. Trainees in the two experimental groups received extra ground-training: each in-flight lesson was preceded by solo practice with a PC 'game' that simulated the aerobatic maneuvers to be flown in the aircraft. The difference between the two experimental groups concerned the training equipment. One group used a basic PC-configuration, the other used a more advanced PC-configuration in a cockpit-like environment and received more advanced instructional feedback.

Progress in individual skill level was measured by the accuracy with which trainees could fly the maneuvers in subsequent flights in the real aircraft. A total of 2053 aerobatic maneuvers were analyzed on the basis of flight-data recordings on board the aircraft. In addition, instructor ratings were collected for each maneuver. We hypothesized that complex manual flying skills, learned on the ground, transfer to the aircraft.

The experiment, however, provided no objective support for this hypothesis. There were no significant differences between the three groups as measured by the flight profiles, nor significant differences between the learning curves of the groups. The only difference between the control group and the experimental groups was that the latter groups flew significantly more maneuvers in the same amount of flight time in the aircraft, indicating that PC-based simulation results in a procedural advantage in the air. However, this did not result in more accurate flight profiles. We also found that instructor ratings deviated significantly from the in-flight measured data. In the discussion, we compare these findings with ten published transfer-experiments with PC-based simulation. In these ten studies it is hard to find any evidence for the transfer of manual flying skills from PC-based simulation to real flight.

Contents

(34 pages in total)

1 Introduction

1.1 Low-fidelity training in flight instruction

Traditionally, the design of flight simulators for pilot training has been based on the assumption that the more a simulator behaves, responds, feels and looks like a real aircraft, the better will be the training. This is what we call the "high-fidelity view of flight simulation". However, this view has changed somewhat over the last few decades.

Positive transfer to the in-flight environment has been demonstrated with low-fidelity PC-based simulation training in a number of studies that focus on the initial stage of flight training. Ten such positive transfer studies have been found in the public literature; these can be grouped roughly in three categories (1) pre-solo training of aircraft landing, (2) training of basic flight maneuvers, and (3) training of instrument maneuvers.

Three studies in the first category focus on aircraft landing, using out-of-the-window visual information. Four studies in the second category focus on basic aircraft maneuvers, such as basic patterns and turns. Three studies focus on instrument maneuvers.

BFITS, a PC-based Basic Flight Instruction Tutoring System (Koonce, Moore & Benton, 1995) falls into the 'landing' category. Novice trainees received ground-training with BFITS on how to fly an airplane in 31 lessons. Application resulted in a significant reduction in the number of landings that had to be made with an instructor, and generally in a reduction in flight-time. However, in addition to simulated flight, BFITS also comprises PC-based teaching modules on the principles of flight; because these modules use text, graphics and animation and give extra feedback to the trainee, it is not clear to what extent practice with the simulation contributes to the transfer-of-training.

Two other studies in the 'landing' category also demonstrated positive transfer from landing trials in a low-fidelity simulation to real landing. In the first study, Lintern, Roscoe, Koonce & Segal (1990) applied adaptive guidance and a moderately detailed out-of-the-window scene in the landing simulation. In the second study, Lintern, Taylor, Koonce, Kaiser & Morrison (1997) added an experimental group that was trained with a low-detailed out-of-the-window scene. In contrast with the high-fidelity view, the latter study demonstrated that training with the low-detail scene in combination with the adaptive guidance yielded higher transfer to aircraft landing than did the moderately detailed scene in combination with the adaptive guidance.

An even sharper contrast with the high-fidelity view is provided by the studies in the category 'basic maneuvers' by Gopher, Weil & Bareket (one study, reported in 1992 and in 1994) and Hart & Battiste (1992). In both studies, trainees in the experimental group received ten hours ground-training with the Space Fortress game, a PC-based computer game, developed for research into the development of complex skills. The elements and parameters of this game are physically remote from those of the flight situation. However, the authors report significant transfer effects from the ground to real flight, particularly for those performance measures that are related to the control of attention and coping with high workload. Two additional studies in the category 'basic maneuvers' of Ortiz (1994) and Dennis & Harris (1998) also report that groups that received training with PC-based simulation performed much better than a control group who did not receive training with PC-based simulation.

In the 'instrument' category, positive transfer effects of instrument flight skills from the ground to the air using a commercially available PC-based flight simulator have been reported by Phillips, Hulin & Lamermayer (1993) and Ortiz, Kopp & Willenbucher (1995). The two studies used the ELITE software package. The latter study revealed that this package provided the same transfer as a more advanced flight and navigation procedures trainer.

In a more recent study of PC-based instrument training by Taylor, Lintern, Hulin, Talleur, Emanuel & Phillips (1999), the ground-training package is more advanced than that in the previously discussed studies, in that it involves instructor supervision and additional hardware. Ground-training with 13 of the 21 investigated instrument tasks yielded significant positive transfer. However, the latter study revealed that positive transfer of instrument flight skills is largely confined to the early stage of training, that is, when the instrument tasks are newly introduced in the air. This suggests that the effectiveness of ground-training decreases with practice.

The ten cited studies demonstrated positive transfer for standard flying tasks, such as basic instrument tasks, visual landings, visual turns, etc. in the initial stage of flight training. Intentional deviations from fidelity may even contribute positively to transfer, as was demonstrated by the visual landing scene experiments by Lintern et al. (1997). This evidence is supported by other transfer experiments that fall outside the PC-based category. An example is transfer to landing skills under crosswind conditions, which was higher after training without crosswind than after training with crosswind (Lintern & Garrison, 1992). A number of experiments with 'augmenting' visual objects (such as gradient lines and poles) in the simulated landing scene also demonstrated an increase in transfer to the flight situation (e.g. Reisweber & Lintern, 1991, Lintern & Koonce, 1992). Lintern (1992) concluded: 'It is this type of result that establishes the need for a theoretical conception of skill transfer that does not rely on the notion of fidelity'.

1.2 Experimental approach

In this research, we investigate learning curves of trainees who practice aerobatic maneuvers in an aircraft under supervision of an instructor. The goal task was to fly five aerobatic maneuvers (the loop, the slow roll, inverted flight, the Immelmann and the split-S) in a fixed-order continuous sequence (figure 1). We measured the skill level of each trainee by the accuracy with which each maneuver was flown, during ten consecutive flight lessons of 30 minutes each. The learning curve is the plot of skill level (accuracy expressed as performance score) against the number of practice hours in the aircraft.

Trainees were assigned to groups with different training regimes, that is, ‘no ground-training’, ‘ground-training with a standard PC-based simulator’ and ‘ground-training with a PC-based simulator with extra features’. Assignment to groups was balanced with respect to capabilities of trainees.

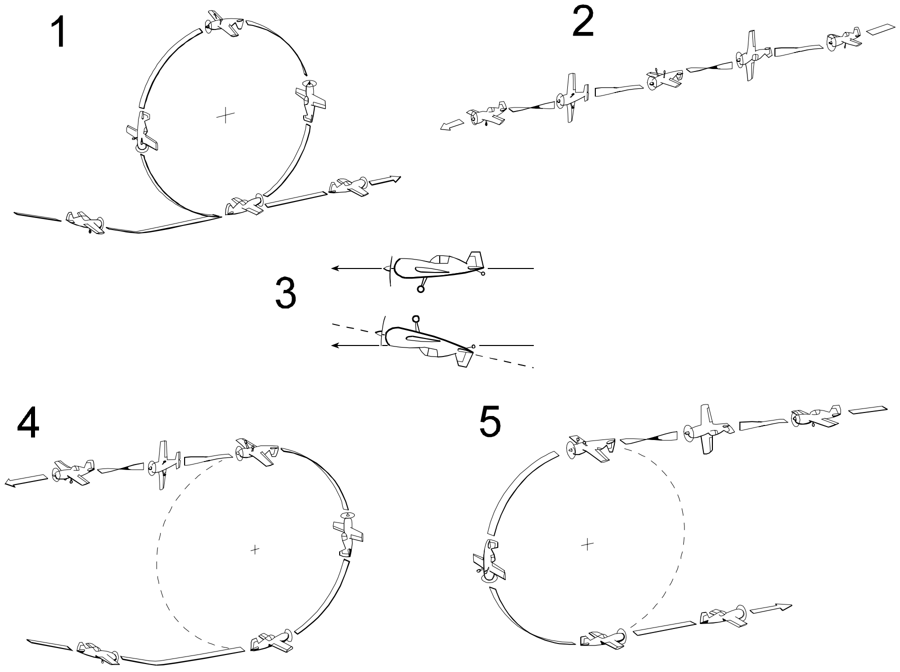

Figure 1: The five aerobatic maneuvers that consititute the experimental task: (**1**) loop, (**2**) left slow roll, (**3**) inverted flight (½ left roll, 10 seconds inverted, ½ left roll), (**4**) Immelmann (½ loop, ½ left roll), and (**5**) Split-S (½ left roll, ½ loop).

Aerobatic skills are needed for the execution of a series of complex aircraft control actions while unusual attitudes and forces are being experienced. The working hypothesis in training applications is that these skills only transfer from a situation that is identical or almost identical to the in-flight situation. This is a solid hypothesis in the absence of convincing counter-evidence. Moreover, with regard to these skills, one should not rule out the possibility of negative transfer from training environments that depart considerably from full fidelity.

The alternative hypothesis, which is investigated in this research, is that aerobatic skills learnt in a low-fidelity environment have a positive transfer to the aircraft. To this end, one ‘suitable’ low-fidelity environment may be sufficient to prove the alternative hypothesis. However, critical transfer elements in simulations are difficult to identify, since there are no precise theoretical principles (Lintern, 1991, 1992, 1995) such as how to configure such a simulation in order to induce measurable transfer. As Lintern (1992) put it: ‘Identification of transfer elements is a major challenge that will require a set of converging experimental techniques and that will test the ingenuity of the researcher.’ To date, few of these transfer elements have been investigated, and only for specific flight tasks, such as the horizon-aim-point angle (H-angle) and the relative gradient for the approach to landing (e.g. Wewerinke, 1980, Lintern & Liu, 1991, Galanis, Jennings & Beckett, 1998, Lintern, 2000).

Because there are no clear guidelines for transfer elements for manual flying skills in the simulation of aerobatic maneuvers, we dealt with this issue in a practical manner. Before the experiment, we asked a number of aerobatic instructors to fly the sequence of maneuvers with the simulation software package (that is, Flight Unlimited from Looking Glass technologies, 1995) and to give their assessment of the value of the package in training the experimental task.

There was agreement among these aerobatic instructors that, despite the limitations of PC-based simulation, the package provided specific features, including a relatively sophisticated aerodynamics software model, out-of-the-window view, instrument panel and engine sounds, all specific for the aircraft type on hand, which in the past could only be achieved on expensive and complex simulation systems.

In the development phase preceding the experiment we additionally undertook a pilot training with novices and experienced pilots. On this basis we speculated which optional features of the software could promote transfer. For example, for learning to fly a maneuver such as the loop, only one reference line on the ground is needed; thus terrain detail did not seem to be overly important. On these grounds we defined two different configurations of the low-fidelity environment in order to extend the range of possible outcomes of the experiment.

The first configuration consists of a software package, with which simulated aerobatic maneuvers can be practiced, installed on a standard PC with basic options. The second configuration has a number of extra features that are thought to improve the first configuration: a cockpit-like physical environment with a more realistic stick, rudder pedals and throttle. A more realistic layout of these controls improves the mapping of motor responses. Additionally, automatic instructional feedback is provided, which will be explained in the methods section. We additionally tested and fine-tuned the second configuration with input from aerobatics experts. There was a consensus that there was ‘a good chance’ that the thus configured PC-based simulations could improve the manual flying skills of aerobatics trainees. These expectations form the basis for the experiment. The real version of the aircraft on which the software package is based will be used for the in-flight evaluations.

Unlike experimental designs in which the goal is to evaluate a certain aspect of the simulation, such as the level of detail of the visual scene, the current experimental configurations intentionally differ from each other in multiple aspects. We thus increase the probability that differences in transfer are measurable if there is any (either positive or negative) transfer at all, relative to a control condition. Also, we thus disregard the option to determine systematically which elements of the second configuration will lead to differences in transfer.

In the following section we present the empirical method in detail. Thereafter, the results section starts with some summary statistics of the collected data. We analyze the differences in the learning of the three groups, using both the measured flight-data (with equipment installed in the aircraft) and the instructor ratings. We quantify to what extent ground-training with PC-based simulation contributes to final aerobatic performance. We compare this relative contribution with the contribution of other factors, such as the relative effect of total flight-time in the aircraft and general pilot ability, as measured before the experiment. In the discussion we compare the findings with those of the ten previously mentioned transfer studies and point to the benefits of PC-based simulation for the training of aerobatic maneuvers.

2 Method

2.1 Task

The task was to fly five aerobatic maneuvers in a fixed-order continuous sequence on a light aerobatic aircraft. The five aerobatic maneuvers are the loop, slow-roll, inverted flight, Immelmann and split-S, the trajectories of which are sketched in figure 1. Each of these maneuvers takes a skilled pilot approximately 20 seconds to complete.

Fully satisfactory completion of the task required the achievement of five binary (pass/fail) criteria per maneuver, resulting in 25 criteria in total, which are listed in appendix A. The criteria were chosen in consultation with aerobatics experts. The criteria were selected because of their importance for the maneuver and on the basis that the flight instructor should be able to decide, if necessary with the aid of cockpit instruments, whether the criterion has been satisfied or not. Moreover, a suitably low number of criteria were selected to allow the instructor to complete the score form, while seated behind the trainee in the aircraft

2.2 Trainees

Twenty-four trainees were selected from a larger group of 60 candidates, all students from a school for commercial pilots. Initial selection took place on the basis of body weight (maximum 80 kg, a limit dictated by aircraft performance), age (maximum 27 years), fixed wing flying experience (maximum 250 hours) and absence of aerobatic experience. This initial selection excluded 29 candidates. Three more subjects were excluded for physical reasons during an in-flight aerobatic resistance test. The remaining 28 candidates completed the Aiming Screening Task (Foss, Fabiani, Mane & Donchin, 1989), a task that is known to be a reasonable predictor for training success on complex tasks. Average score on this task was 800, with a standard deviation of 160 and scores ranging from 480 to 1100. The distribution of scores over the candidates was an approximately normal distribution.

On the basis of the score on this task, each trainee was assigned an initial ability level: Low (L), Low-Medium (LM), High-Medium (HM) or High (H). Three groups of eight trainees were formed, each group containing two Ls, two LMs, two HMs and two Hs. The remaining four candidates were discharged. The groups were randomly assigned to conditions: normal treatment (the Control group, which we will label as the “C-group”), ground-training with a Standard PC-configuration (the “S-group”) and ground-training with a PC-configuration with extra features (the “X-group”).

During the course of the experiment, three more trainees failed to meet the procedural standards described in this section. The flight school, at which the trainees were recruited, initially provided incorrect flight-medical information for one trainee from the C-group. On the basis of later information, this trainee could not be considered representative of the population under study. Another trainee, from the S-group, dropped out during the course of the experiment for previously unnoticed medical reasons. Finally, one trainee from the X-group had not reported properly regarding his experience (expressed in number of flight hours) during selection for the experiment. Later information revealed that this trainee had too many flight hours to be representative for the trainee-population under study. The performance scores of these trainees will be excluded from the subsequent analysis. The three trainees were not replaced, consequently the subsequent analysis is based on three groups of seven trainees.

2.3 Procedure

General All trainees received ten flight lessons of thirty minutes in a light aircraft (see equipment section), with an instructor in the back seat, who was responsible for in-flight instruction and rating. Each trainee received only one flight lesson per day.

The experiment started with the flights of the C-group. Subsequently, the flights of the S-group were flown and finally the flights of the X-group were flown. Since all trainees attended the same flight school daily, this schedule was chosen since it was likely to cause the least ‘cross-talk’ between trainees of the different groups. Hence, trainees of the C-group were prevented from becoming acquainted with the extra training and simulation-configurations of the S- and X-groups, and trainees of the S-group could not be exposed to the simulation-configuration of the X-group. However, it was practically impossible to keep the three flight-instructors unaware of the experimental setup.

Theory Well before the start of the experiment, all trainees received a paper manual, which included essential information about performing the aerobatic task. Just before the start of the training, all trainees had to pass a formal theoretical test with 23 questions on the principles of the task. In addition, each trainee received a set of five ‘cue cards’, one for each maneuver to be flown. Each of these cards depicted a maneuver and specified the criteria for acceptable performance.

Flight Instruction Three instructors relieved each other during subsequent flights and days on the basis of their availability. As a result, trainees saw two or three different instructors during the ten flight lessons. Due to the weather, flights could be postponed until a later date. Sometimes the weather was unclear or the cloud base was too low to allow visual flight.

The trainees received normal briefings and debriefings from the designated flight-instructor before and after each flight lesson. Each flight lesson included the transit flight from the airport to the nearby flying area where the maneuvers were flown and vice versa.

The first flight lesson consisted of a practical introduction to flying aerobatics and familiarization with the aircraft. The instructor demonstrated the sequence of five maneuvers, after which the trainee could have a first try at flying the sequence, while being talked-through by the instructor.

In the following lessons, each maneuver was practiced one-by-one in the order listed in the task description. The instructor rated each maneuver flown by the trainee on the basis of the binary criteria of appendix A. The pass or fail (0 or 1) relating to each criterion was scored on a pre-printed score form, which was held on a knee-pad during the flight.

Once a maneuver was mastered, according to criteria listed in appendix A, the next maneuver was practiced until the trainee could fly the whole sequence of five maneuvers in a continuous fashion. However, the instructor could decide to deviate from this schedule whenever this seemed advisable (for example, depending on wind, presence of clouds, physical state of the trainee, etc.). In the last two lessons the trainees had to fly the whole sequence twice per lesson, as accurately as possible.

Ground-training Trainees in the C-group received no specific ground-training but had to prepare themselves before each flight lesson using the training manual.

The S- and X-groups received simulation sessions preceding each flight lesson. These sessions were organized identically for the two groups. Each session took approximately 50 minutes. The first simulation session consisted of familiarization with the simulation package, a trial on each maneuver and a trial on the complete sequence. The following simulation sessions consisted of the repetition of problematic maneuvers as indicated by the flight instructor during the previous flight lesson and the preparation of new maneuvers.

When the trainee practiced a maneuver, instructional feedback was presented automatically in text on the computer screen. During the maneuver, the name of the maneuver (‘loop’, ‘slow-roll’, etc.) was displayed. In the center of the computer screen, an arrow indicated the direction in which the aircraft should be flown. After each maneuver, a performance rating was displayed. The nature of the deviations from the desired flight profile was presented in terms of altitude, heading, etc. and the most problematic part of the maneuver was indicated. There was no mediation by a flight instructor during the simulation sessions. However, the sessions were supervised by the experimenters and assistance was provided in case of problems with the equipment.

2.4 Equipment

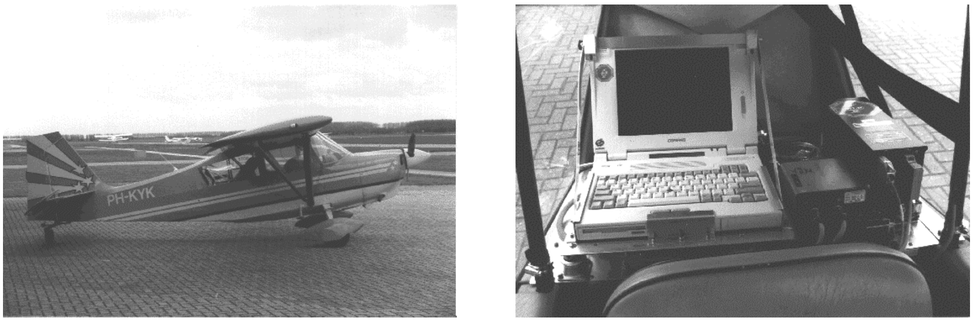

Figure 2: The Bellanca Super Decathlon used in the experiment, with the flight data measurement equipment in the back of the aircraft, used for 16 Hz sampling of 12 flight parameters.

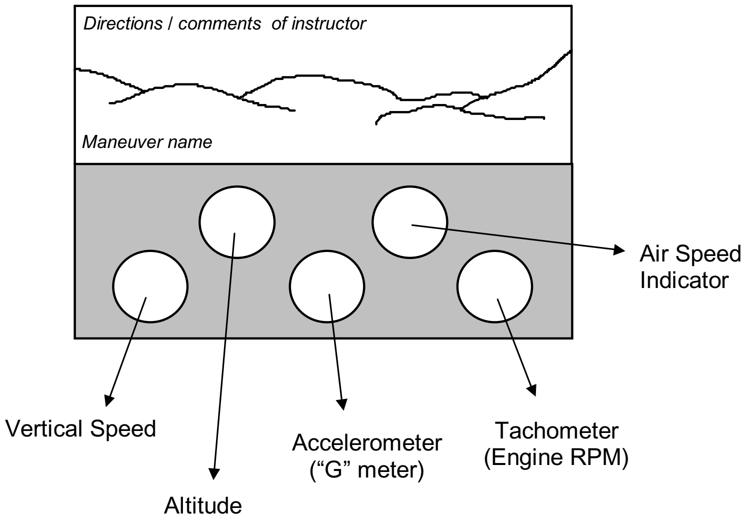

Figure 3: Structure of the cockpit view during the simulation lessons for S- and X-group.

Aircraft and on-board equipment A light propeller aircraft, the Bellanca Super Decathlon, suitable for aerobatic operations, with a single piston engine of 180 hp was used for the training (see figure 2). The following flight data were measured and logged with specially installed on-board equipment: altitude, indicated air speed, the three orientation angles, the three angular rates, the three linear accelerations and type of maneuver (the latter recorded manually via a switch operated by the instructor).

Ground Equipment and Software A commercially available software package (Looking Glass Technologies, 1995) allows for the practice of aerobatic maneuvers on a PC by simulated flight. The package has relatively accurate aircraft flight models, including that of the Bellanca Decathlon, which is used for the flight lessons. However, the specifications of the simulated Bellanca Decathlon do not fully match those of the real aircraft. A noticeable difference is that the entry speeds for the real aerobatic maneuvers are approximately 20% lower than the entry speeds that are required for acceptable performance in the simulation.
The package was installed on a Pentium PC, equipped with sound card and stereo loudspeakers. The lower half of the color PC screen (640 x 480 pixels) depicted the cockpit instruments; the upper half of the screen presented the forward out-of-the window view (see figure 3). The renderer was set up such that the terrain (3D photo-realistic) was represented with only a low level of detail. The sky was presented with some haze but without clouds. Engine noise and wind effects were clearly audible. The standard keyboard was replaced by a small keyboard with only the six keys that were needed to run the simulation sessions. Three of the keys, <glance up>, <glance left>, <glance right>, could be used with one hand to replace the forward out-of-the-window view by the view through the roof window, left window or right window, respectively.

The S-configuration, with which the S-group was trained, was equipped with inexpensive plastic spring-loaded game controls: Pro Throttle, Pro Pedals and a right hand stick Flightstick Pro (all by CH-products). Furthermore, this configuration consisted of a 17-inch monitor and standard furniture.

The X-configuration, with which the X-group was trained, was equipped with steel spring-loaded controls mounted on a fixed base such that their shape, position and displacement stroke resembled these characteristics of the controls in the aircraft. The seat resembled the seat in the aircraft and was adjustable as in the aircraft, such that the workspace closely mimicked that of the real cockpit with respect to position and stroke of stick, rudder pedals and throttle. A 21-inch color monitor could be tilted and adjusted in height and provided approximately 30 degrees horizontal and 22 degrees vertical field-of-view.
In this second configuration, two extra instructional options of the software package were installed. First, when the deviations from the desired flight profile exceeded particular limits during a trial on a maneuver, a so-called “stick arrow” and/or “pedal arrows” appeared on the screen. These arrows indicate in which way the stick or the pedals should be moved in order to correct the maneuver. All possible arrow symbols are shown in figure 4. Second, the instructions and automatic feedback that were presented in text and in graphics were also presented in audio via a set of loudspeakers. These spoken comments and directions from a “synthetic instructor” were more elaborate than the plain text and graphics displayed on the computer screen.

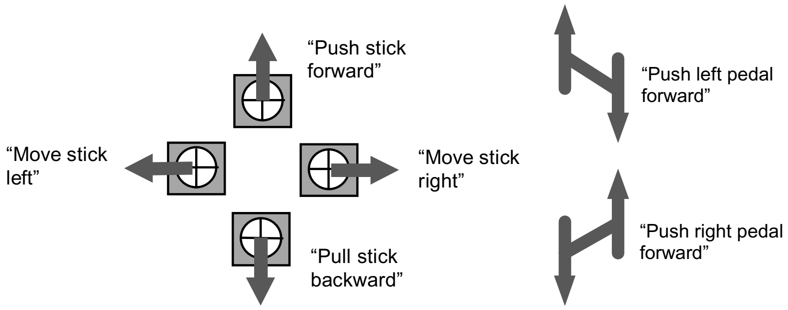

Figure 4: The four possible stick symbols and two possible pedal symbols that could appear on the computer screen. A stick symbol and/or a pedal symbol showed up automatically at deviations, an instructional feature for the X-group.

3 Results

3.1 Number of maneuvers flown

All flight lessons had a fixed duration of 30 minutes. The number of maneuvers to be flown in the first lesson was fixed. This lesson consisted of a demonstration of the sequence of 5 maneuvers, followed by an exercise in which the trainee had to fly the sequence while being talked through by the instructor. In lessons 2 to 8 the number of maneuvers to be flown during each 30 minutes of flight time was flexible. In lessons 9 and 10, a fixed number of two sequences (10 maneuvers) had to be flown.

The 21 trainees who finished the experiment flew a total of 2053 aerobatic maneuvers. The C-group flew 623 maneuvers, the S-group 680 maneuvers and the X-group 750 maneuvers in total. We analyze the differences in the number of maneuvers flown during the 'flexible' lessons 2-8 (at a 5 percent significance level).

Trainees in the S- and X-groups flew more maneuvers per lesson than trainees in C-group. Trainees in the C-group completed on average 9.0 maneuvers per lesson, trainees in the S- and X-groups completed on average 10.2 and 11.4 maneuvers per lesson, respectively. We tested the differences with a two-tailed t-test for differences between means; differences in number of maneuvers per lesson were matched on the basis of lesson number.

The differences between the S- and C-group and between the X- and C-group are significant ($t(48)=2.30$, $p=0.023$ and $t(48)=4.52$, $p=10^{-5}$, respectively). The difference between S-group and the X-group is not significant ($t(48)=1.97$, $p=0.052$).

All three groups demonstrated a significant increase in the number of maneuvers flown per lesson. The average start level for all three groups was 7.6 maneuvers per lesson. In each subsequent lesson, trainees in the C-group managed to fly an extra 0.3 maneuver, and trainees in the S- and X-groups managed to fly an extra 0.5 and 0.75 maneuvers respectively. Again, the differences between the S- and C-group and between the X- and C-group are significant ($t(41)=2.0$, $p=0.046$ and $t(41)=4.75$, $p=5\cdot10^{-6}$, respectively). The difference between the S- and X-group is also significant ($t(41)=2.48$, $p=0.015$).

However, it should be noted that the number of maneuvers flown per lesson of thirty minutes was not explicitly controlled by the experiment. Firstly, the time needed for transit from the airport to the flying area and vice versa is not necessarily constant. Each maneuver takes only about 20 seconds, and the time between maneuvers is taken up by non-specific flying activities, which can be a variety of activities such as physical and mental recovery of the trainee after a maneuver, verbal feedback and instruction, relocation to avoid clouds, etc.

3.2 Pre-flight briefing

A side effect observed in the experiment was that trainees in the S- and X-groups needed less pre-flight briefing time after each 50 minutes of simulation. Since trainees reviewed the maneuvers of the previous flight lesson and prepared the maneuvers for the next flight lesson with the aid of simulation, briefing times went down from approximately 15 minutes for the C-group trainees to approximately 5 minutes for the S-group trainees and to almost zero briefing time for the X-group trainees.

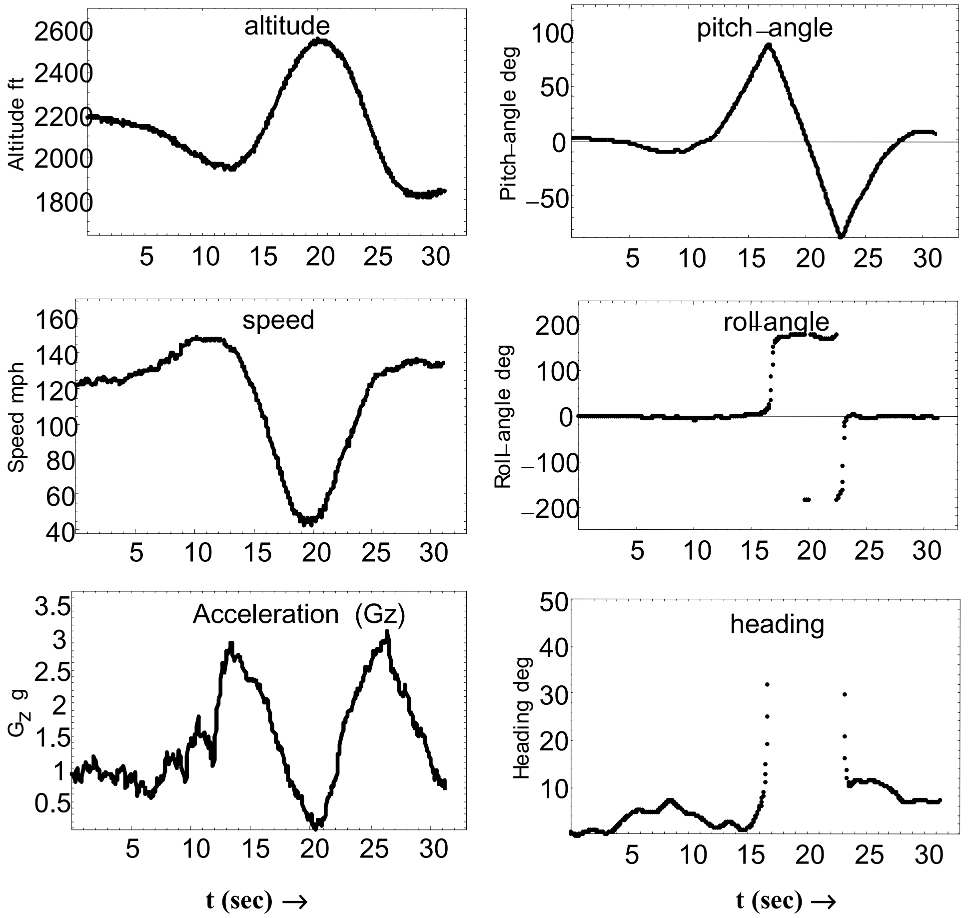

Fig. 5: Record of flight parameters for the loop – trainee S7 (S -group) – flight 10.

3.3 Accuracy of performance

All the maneuvers were recorded with the in-flight data recording equipment. The in-flight performance was analyzed after the experiment using predefined criteria as listed in appendix A. In addition, the instructors scored the performance of the trainee during the flight on the basis of observations of aircraft behavior and instrument readings, using criteria identical to those listed in appendix A.

As an example, we graphed six of the recorded flight parameters during one loop of one of the trainees (figure 5); from these we extract a performance score as follows.

From the graphs in figure 5 it can be seen that loop-entry starts at t=10 s. At this point, the aircraft has gained speed (in this case from 130 mph to just over 150 mph) and is straight and level. The aircraft loses speed as soon as the nose of the aircraft is pulled up. The trainee pulls the stick towards him to raise the nose of the aircraft and to gain altitude. After few seconds (at t=14 s) the aircraft will have maximum acceleration (G_z), which, as apparent from the acceleration-graph in figure 5, is only 2.9 g in this case.

At the top of the loop (at about t = 20 s), the aircraft is completely upside-down. Note that the pitch angle (i.e. the angle over which the nose of the aircraft is raised) passes the zero degrees level. At this point the trainee has to ensure that the aircraft has its wings level; this criterion is fulfilled in the case depicted in figure 5: absolute roll angle is approximately 180 degrees in the top of the loop.

After the top, the aircraft should regain speed and acceleration in a controlled fashion, towards the bottom of the loop. When the aircraft has leveled-off, at loop-exit (t=30 s), its altitude should be approximately equal to that at loop entry. However, the altitude graph in figure 5 shows that the trainee has lost some altitude. He started the loop well above 1900 ft and ended just above 1800 ft, a loss of well over 100 ft. Also, the compass heading at the exit of the loop should equal that at entry of the loop, which is indeed the case.

Thus, if we apply the criteria listed in appendix A to the loop illustrated in figure 5, the result is that the entry speed is more than 10 mph too high, the acceleration at pull-up is 1.1 g too low, and the altitude loss is more than 100 ft. Consequently, this trainee fulfilled only two out of five criteria, and hence obtains a performance score of only 40% for this maneuver. All 2053 maneuvers were analyzed in this way, using the criteria listed in appendix A.

3.4 Group learning curves

In figure 6(a), we have depicted the in-flight learning curve of the C-group (control group) for lessons 2 to 10. The learning curve was reconstructed from an analysis of the in-flight recorded data. The performance scores of the first lesson are not included since this lesson was a familiarization lesson, with an organization different from that of the subsequent lessons. The performance scores have been averaged over all maneuvers per lesson and over the seven trainees in the group. The error bars represent the standard deviation in the score of the control group, one standard deviation in upward direction and one in downward direction.

It is apparent that the C-group, as a whole, demonstrated substantial progress over the 10 flight lessons (roughly 20 percent performance improvement as measured by the predefined criteria). However, there are considerable differences between individuals, which is typical for all three groups. These will be further analyzed in the next section.

In figure 6(b), we have depicted the in-flight learning curves for all three groups: (1) the control group (C), (2) the group that received ground-training on a standard PC (S) and (3) the group that received ground-training on a simulation configuration with extra features (X). It is apparent that all three groups demonstrated considerable progress over flight lessons 2-10 and that the learning curves on the basis of in-flight recordings reveal few differences between the groups. The S- and X- groups do not seem to benefit from the additional ground-training with PC-based simulation. We will further analyze the effects in the next section.

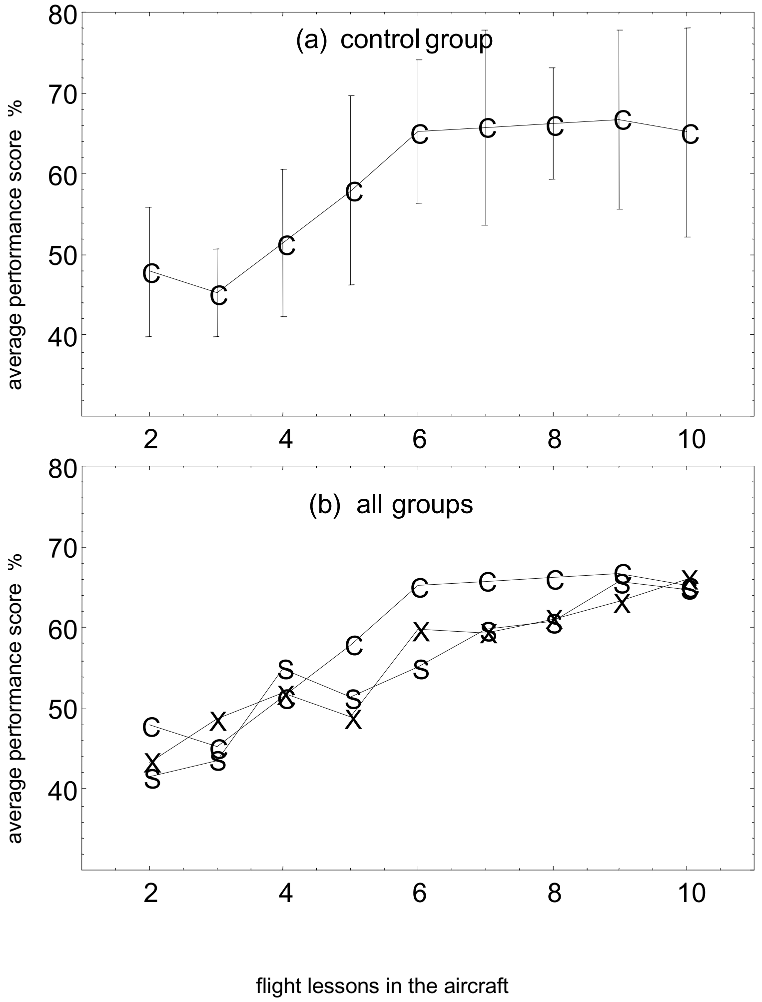

Figure 6: The top graph (a) depicts the in-flight learning curve for the control group, extracted from the in-flight data recordings and averaged over trainees and maneuvers. Error bars represent the standard deviation in the score per lesson for the control group. The bottom graph (b) depicts these in-flight learning curves for all three groups; the control group (C), the standard PC-group (S) and the extra PC-configuration –group (X).

3.5 Regression analysis of the in-flight data

In order to determine to what extent the additional ground-based simulation contributed to the in-flight performance of the S-group and the X-group, we carried out a standard linear regression analysis to identify the main effects. The dependent variable is the performance score (*P*, 0-100%), which is the average score of an individual trainee per flight-lesson, derived from the in-flight recorded data, as explained in the previous section. Thus, the score in each flight lesson is considered as an observation. The experimental treatment (C, S or X) is the qualitative independent variable. We used two binary dummy-variables s and x to encode the treatment quantitatively. That is, s is 1 for each flight of a trainee from the S-group, and otherwise 0. Likewise, x is 1 for each flight of a trainee from the X-group, and otherwise zero. Because of

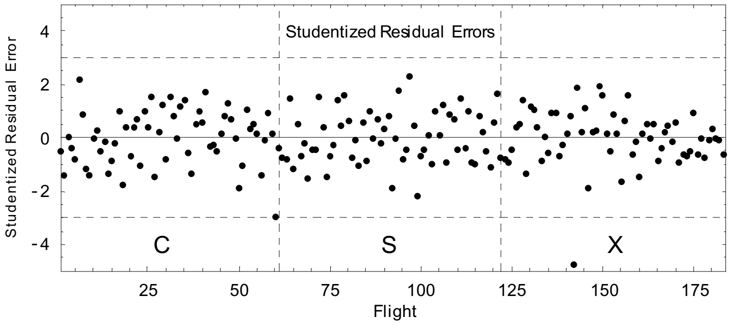

Figure 7: Plot of studentized residuals (residual errors scaled by their standard error), for the fit of equation (1) to in-flight data. The vertical dashed lines separate the flights of the three different groups. The horizontal dashed lines denote the criterion for outliers. One outlier is detected, an unusually low score for a trainee from the X-group.

the large variability in performance, as reflected by the learning curves of figure 6, we used three extra independent variables:

- First, because the learning effect as a result of the flight-lessons is evident, we used lesson-number (n, $2 \leq n \leq 10$) as a covariate.
- Second, we received, for each trainee, a data-file which was prepared before the experiment by the Aeromedical Institute, the central pilot selection agency in the Netherlands. These files reported the scores on twelve different tests, which are thought to measure different pilot abilities. The average score for an individual on these twelve tests is a notable figure used in pilot selection by this agency. After linear scaling, we used this score as a covariate, denoted by the symbol a. Average a was 86%, with a standard deviation of 7% and a range from 74%-100%.
- Third, since three different instructors (U, V, W) served as in-flight instructors, we encoded the presence of the instructor in a flight with two binary dummy-variables, u and v. For a flight with instructor U, u is 1. For a flight with instructor V, v is 1. For a flight with instructor W, both u and v are zero.

Hence, our model for performance score P on the basis of flight data recordings becomes:

$$P_i = \beta_0 + \beta_1 s_i + \beta_2 x_i + \beta_3 n_i + \beta_4 a_i + \beta_5 u_i + \beta_6 v_i + e_i \,, \qquad (1)$$

in which e_i is the residual error.

Studentized residuals (residual errors scaled by their standard error) for the best fit of this model to the data are plotted in figure 7, which shows that the distribution of residual errors is balanced. A studentized residual that is larger than three (in absolute value) is considered to be an outlier. Figure 7 shows that the data contain one such outlier; this concerns an unusually low score (a 20 percent score) for the fifth flight lesson of an X-group trainee. After removal of the outlier, new estimates of the coefficients $\beta_0..\beta_6$ have been calculated to minimize the sum of squared errors and the results are summarized in table 1.

Table 1
Multiple regression analysis results of the aerobatic performance scores in 9 flight lessons (2-10) on the basis of in-flight recordings. Significant effects (at the 5 percent level) are denoted by an asterix (*).

Model for Aerobatic Performance (equation (1))
$R^2 = 0.48$, $F(6, 181)=26.5$, $p<10^{-6}$

Variable	**Coefficient β**	**T-value**	**p-level**	
Constant (intercept)	3.49	0.41	0.69	
Standard PC-configuration *s*	-2.65	-1.52	0.13	
Extra PC-configuration *x*	-1.79	-1.07	0.29	
Flight lesson number *n*	2.71	10.3	$<10^{-6}$	*
Pilot Ability score *a*	0.46	-4.82	$3 \cdot 10^{-6}$	*
Instructor *u*	-4.17	-2.15	0.033	*
Instructor *v*	-0.61	-0.40	0.69	

The results in table 1 show that overall model utility is high ($F(6, 181)=26.5$, $p<10^{-6}$), but with a coefficient of determination R^2 of only 0.48, such that 52 percent of the variance in performance scores (see also figures 6(a) and 7) is left unexplained by the model of equation (1). The intercept β_0 is not significantly different from zero.

Most importantly, table 1 reveals that the experimental ground-training had no significant effect on in-flight aerobatic performance, when compared to the control condition. Thus, neither the ground-training with the standard PC-configuration for the S-group, nor the ground-training with the PC-configuration with extra features for the X-group had any effect, according to the linear regression model of equation (1).

The covariates reveal the following:

- The number of flight lessons *n* had a significant effect on performance, with an estimated increase of 2.71 percent per flight lesson.
- Also, the pilot ability score *a*, is a significant predictor of in-flight aerobatic performance. A percent increase in score *a* is estimated to yield a 0.46 percent increase in aerobatic performance.
- Finally, one of the instructors (instructor U) had a significant effect on in-flight performance, which is negative in comparison with the other two instructors.

We subsequently investigated the interactions between independent variables, but these were not significant.

3.6 Group learning curves according to instructor ratings

Instructors agreed to use the binary criteria of appendix A, and rated each maneuver during the flight. When we consider the learning curves on the basis of instructor ratings as depicted in figure 8, two characteristics attract attention. First, all ratings are generally higher than in figure 6, an indication that the instructors were generally more tolerant in the judgement of the criteria than justified by the in-flight measurements. Second, already in the second flight lesson the X-group receives higher ratings than the S-group. The higher ratings persisted throughout the remainder of the training. These differences will be further analyzed in the next section.

3.7 Regression analysis of the instructor ratings

We carried out an additional regression analysis to investigate the instructor ratings. In this analysis, performance score (*P'*, 0-100%) on the basis of instructor ratings is now the dependent variable. As in the previous analysis, the experimental treatment (C, S or X) is the independent variable, and the same covariates (*n*, *a*, *u*, *v*) are used. Hence, the model for performance *P*' on the basis of instructor ratings is identical to the model of equation (1).

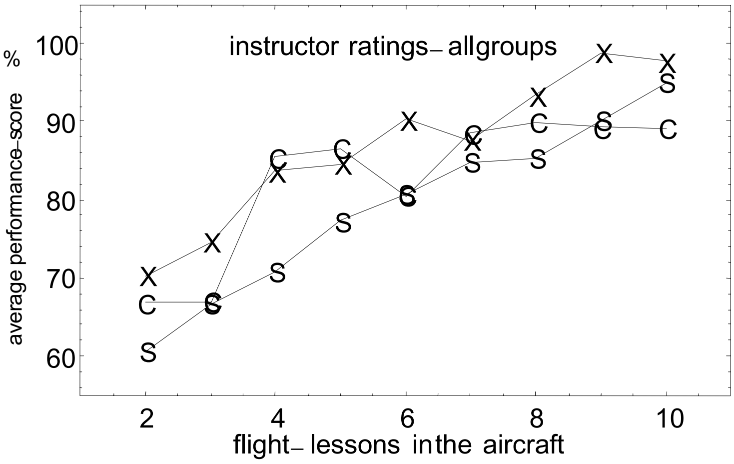

Figure 8: in-flight learning curves based on instructor ratings for all three groups; the control group (C), the standard PC-group (S) and the extra PC-configuration –group (X).

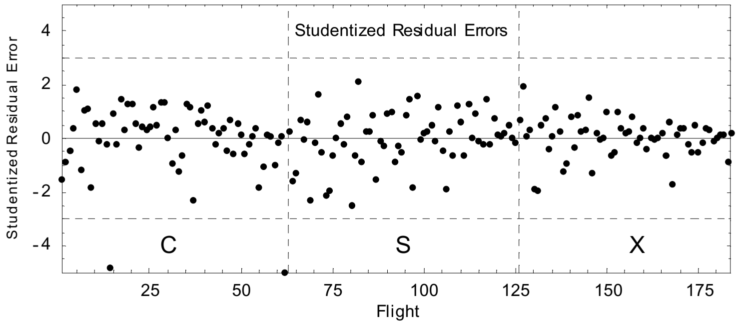

Figure 9: Plot of studentized residuals for the fit of equation (1) to instructor ratings. The vertical dashed lines separate the flights of the three different groups. The horizontal dashed lines denote the criterion for outliers. Two outliers (both for the C-group) are detected.

A plot of studentized residual errors for the best fit of this model to the instructor ratings is given in figure 9. It reveals two outliers in the instructor ratings; This concerns unusual low scores during two flight lessons of C-group trainees. These outliers have been removed. Estimates of the coefficients β_0'..β_6' are summarized in table 2.

Table 2
Multiple regression analysis results of instructor ratings for aerobatic performance in nine flight lessons (2-10). Significant effects (at the 5 percent level) are denoted by an asterix (*).

Model for Aerobatic Performance – Instructor Ratings
$R^2 = 0.59$, $F(6, 180)=42.8$, $p<10^{-6}$

Variable	Coefficient β'	T-value	p-level	
Constant (intercept)	41.8	4.69	$5\cdot10^{-6}$	*
Standard PC-configuration *s*	-1.42	-0.79	0.43	
Extra PC-configuration *x*	5.63	3.21	$2\cdot10^{-3}$	*
Flight lesson number *n*	3.20	11.6	$<10^{-6}$	*
Pilot Ability score *a*	0.25	2.50	0.013	*
Instructor *u*	-10.0	-5.04	$1\cdot10^{-6}$	*
Instructor *v*	4.06	2.53	0.012	*

As in the previous analysis, based on the flight data recordings, the current results of table 2 show that overall model utility is high ($F(6, 180)=42.8$, $p<10^{-6}$). The coefficient of determination R^2 is 0.59, which is higher than in the previous analysis (0.48).

3.8 Comparison between instructor ratings and in-flight data

In contrast to the analysis of in-flight data, the PC-configuration with extra features for the X-group contributed positively and significantly to aerobatic performance, but the ground-training with the standard PC-configuration for the S-group had no significant effect, according to instructor ratings. Table 2 further reveals that the intercept of the model for instructor ratings is significantly larger than zero and is estimated at 41.8 percent, indicating a large overall bias in instructor ratings when compared to the analysis of in-flight data in table 1.

Additionally, table 2 reveals the following:

- The number of flight lessons *n* contributes significantly to performance rated by the instructors, with an estimated magnitude of the effect of 3.2 percent per flight lesson, which is slightly larger than the effect of *n* (2.7 percent per flight lesson) in the previous analysis of in-flight data.
- As with the analysis of in flight-data, pilot ability score *a* has a significant effect on instructor ratings.
- Finally, different instructors have significantly different effects on the ratings. Instructor U gave significantly lower ratings, and instructor V gave significantly higher ratings than instructor W.

The main result of this analysis is that there is no conformity between instructor ratings and in-flight data. The two scores are in disagreement with respect to the contribution of the PC-configuration with extra features. Moreover, there are significant differences between instructors.

3.9 Analysis of the differences in instructor ratings

To clarify the general differences between instructors-ratings and the in-flight recorded data, we consider how the instructors shared the flights of the three different groups between them (see table 3).

Table 3
Distribution of experiment flights over instructors

	Group C	**Group S**	**Group X**
Instructor U	4.8 %	36.5 %	28.6 %
Instructor V	54.0 %	39.7 %	34.9 %
Instructor W	41.2 %	23.8 %	36.5 %

Table 3 reveals that the distribution of the three instructors over the flights of the C-group was not balanced. For the S- and X-groups, the balance was better, but not perfect. Thus, because the three instructors rated differently and since the flights were not evenly distributed over the three instructors, the differences in rated performance between groups are comprehensible.

However, the data in table 3 does not explain why the X-group was rated significantly higher than the C-group, while the relative share in flights by the ‘negative’ instructor U (as apparent from the analysis of in-flight data) was much higher during the period in which the X-group was trained. This leads to the assumption that the instructor-ratings hide interaction effects between the instructors and groups.

To investigate this assumption we extend our model for performance score *P’* on the basis of instructor ratings with four interaction terms (with coeffcients β_7’- β_{10}’) to become:

$$\begin{aligned} P_i' &= \beta_0' + \beta_1' s_i + \beta_2' x_i + \beta_3' n_i + \beta_4' a_i + \beta_5' u_i + \beta_6' v_i \\ &+ \beta_7' u_i s_i + \beta_8' u_i x_i + \beta_9' v_i s_i + \beta_{10}' v_i x_i + e_i. \end{aligned} \tag{2}$$

Estimates of the coefficients are summarized in table 4.

Table 4
Multiple regression analysis results of instructor ratings for aerobatic performance in nine flight lessons (2-10), including instructor-group interactions. Significant effects (at the 5 percent level) are denoted by an asterix (*).

Model for Aerobatic Performance – Instructor Ratings (equation (2))

$R^2 = 0.62$, $F(10, 176) = 28.6$, $p < 10^{-6}$

Variable	**Coefficient β'**	**T-value**	**p-level**	
Constant (intercept)	39.4	4.46	$1 \cdot 10^{-5}$	*
Standard PC-configuration s	-0.55	-0.17	0.85	
Extra PC-configuration x	3.37	1.24	0.21	
Flight lesson number n	3.18	11.3	$<10^{-6}$	*
Pilot Ability score a	0.29	2.91	0.004	*
Instructor u	-28.2	-5.02	$1 \cdot 10^{-6}$	*
Instructor v	4.58	1.85	0.066	
Interaction $u * s$	17.6	2.73	0.007	*
Interaction $u * x$	21.8	3.41	0.001	*
Interaction $v * s$	-3.69	-0.92	0.36	
Interaction $v * x$	1.22	0.33	0.75	

We tested whether the inclusion of four interaction terms (equation (2)) provide better predictions for instructor ratings than the model of equation (1), which is significant ($F(4, 176) = 3.6$, $p < 0.05$).

In contrast with the results in table 3, table 4 reports that, in the presence of significant interactions, neither the ground-training for the S-group nor that for the X-group had a significant effect on instructor ratings.

However, table 4 indicates that instructor U gave significantly lower ratings than the other instructors, that is, to the C-group. Moreover, considering the interactions between S- and X-groups and instructor U, that is, $u * s$ and $u * x$, the model indicates that instructor U gave significantly higher ratings to the experimental groups S and X than did the other instructors. Possible differences between instructors V and W are not significant. Thus, during the course of the experiment, which started with the flights of the C-group and subsequently those of the S-group and X-group, instructor U significantly changed his rating behavior relative to the other two instructors, to the advantage of the S- and X-groups.

4 Discussion of results

In this research we analyzed learning curves - on the basis of in-flight recorded data - of three different groups which received training in manual flying skills (aerobatics). Each group was subjected to a different type of ground-training preceding each flight lesson, that is, no simulation for the C-group, standard PC-simulation for the S-group, and PC simulation with extra features for the X-group.

We analyzed the accuracy of flight profiles for all aerobatic maneuvers during 189 flight lessons (all flight lessons except the first introductory lesson for each trainee) on the basis of 25 pre-defined criteria for the maneuvers to be flown, in terms of altitude, acceleration, speed, roll-angle, pitch-angle and heading. All three groups demonstrated a comparable level of skill at the beginning of the training and comparable progress per lesson of approximately 3 percent per lesson of the maximum attainable score. Despite the 500 minutes extra ground-training of the S- and the X-group, no significant increments or decrements in the level of manual flying skills were found as a result of the skills that were acquired with the simulation on the ground. Thus, there was neither negative transfer nor positive transfer of manual flying skills learnt during the simulation lessons.

Measurable aerobatic skills were significantly determined by three other factors: flight-time in the aircraft, pilot ability as determined pre-experimentally by Pilot Ability score *a* and the presence of specific flight instructors.

4.1 Instructor ratings

An analysis of in-flight instructor ratings, which were based on the same set of criteria, initially suggested a significant advantage for the X-group, i.e. positive transfer from the simulation lessons with the PC simulation with extra features. However, the analysis also revealed significant differences in the rating behavior of the three instructors. An additional analysis included the interactions between the instructors and the three groups. The change in rating behavior of one instructor explained an important difference between the instructor-ratings and the flight data recordings. It explained why the average ratings of the X-group were significantly higher than those of the C-group, whereas this was not found in the flight data.

Thus, notwithstanding the use of clear rating criteria and standards, unreliable instructor ratings could not be prevented, as became apparent from the analysis of the recorded in-flight data and an in-depth analysis of instructor ratings.

The present study is unique in that it used equipment on board the aircraft to record flight data throughout the training of the three groups in order to evaluate flight-performance. Ten other transfer-of-training studies found in the open literature, which also dealt with low-fidelity/PC-based simulation, are based entirely on instructor judgements or instructor ratings. All ten studies established positive transfer of certain skills from the simulation to real flight. On the basis of these studies the question that arises is why the present study failed to find transfer of aerobatic skills from PC-based simulation.

4.2 Transfer of manual flying skills

One could argue that transfer effects for manual flying skills are very specific, i.e. that transfer can only take place for specific component skills, under specific conditions, and that transfer effects must be sought at a lower level of task performance. In our case, this could mean that no transfer was found because we measured skill level by an aggregated performance score based on 25 binary criteria. More detailed analysis of the flight data recordings of flight maneuvers could possibly reveal positive transfer for certain component skills and negative transfer for other component skills, such that the net result is zero transfer.

This possibility could serve to motivate further exploration of the flight data and might provide more insight into transfer-of-training. However, in the current investigation all the 25 criteria together were generally agreed by aerobatics experts to reflect acceptable performance in the sequence of five maneuvers. Thus, only an aggregated performance score based on these criteria represents ‘manual flying skills’, and our hypothesis was that ‘manual flying skills’ transfer from PC-based simulation to real flight.

4.3 Transfer of landing skills

The transfer-of-landing-skills studies by Lintern, Roscoe, Koonce & Segal (1990), and Lintern, Taylor, Koonce, Kaiser & Morrison (1997) reveal that ‘landing skills’, as an aggregated whole, did transfer from the simulation to the aircraft. In both cases, the dependent variable was the number of attempted landings by the trainee prior to release for solo flight, i.e. the instructor judged skill level in the aircraft, and on that basis the trainee was sooner or later released for solo flight. The trainees who received simulation training required significantly fewer pre-solo landings. The simulator training in these two studies was performed on a digital flight trainer for light aircraft (the so-called ILLIMAC), which may be more advanced than a regular PC-based simulator. However, the studies are interesting for the present discussion because they incorporate intentional deviations from realism during training. The earlier study employs adaptive flight path guidance during training and the later study employs the same type of guidance in combination with an impoverished visual scene.

A third study of landing skills by Koonce, Moore & Benton (1995) used the same dependent variable and trainees also demonstrated significantly fewer pre-solo landings. However, since the special training of the experimental group in the latter study also includes other types of computer-based training, it is unclear what parts of the training package promoted the transfer of landing skills.

In fact, none of the three landing studies is conclusive with respect to the type of ‘landing skills’ that transfer from the simulator to the actual landing. The only dependent flight variable is the number of required pre-solo landings (or pre-solo flight time) as judged by a flight instructor. In principle, the type of skills that transferred to the real aircraft could have been manual flying skills, but they could also have been skills related to attention, time management or communication. It may even be that other characteristics transferred from the simulation to the real flight, such as a certain attitude or motivation. However, additional quasi-transfer tests (transfer tests in a more realistic simulator, rather than in the real aircraft) in the experiment of Lintern, Taylor, Koonce, Kaiser & Morrison (1997) suggested improved manual control, in particular for trainees who were trained with adaptive guidance and an impoverished visual scene. Manual control accuracy during the quasi-transfer tests in the simulator was measured in the final approach to landing from 2425 m to 606 m from the runway aim point, and expressed as the logarithm of the variance in lateral deviations from runway line-up and the logarithm of the variance in altitude deviations from the descent-path.

A transfer study by Dennis & Harris (1998) was undertaken to investigate the transfer of basic flying skills (straight-and-level flight and standard turns) from PC-based simulation to the in-flight situation. The experimental design was similar to the current study, i.e. a group with no simulation and two groups with simulations that differed in fidelity. Results were based on instructor ratings and subjective workload measures. The authors remarked: ‘The results suggest that PC-based flight simulators do not aid in the psychomotor skills required to fly a light aircraft. Their benefits lie elsewhere’.

4.4 The quality of the simulation and transfer

There might have been systematic qualitative differences in the type of training provided in our experiment with that provided in studies that *did* report a positive transfer-of-training effect, for example in terms of fidelity of the simulation and/or validity of the skills being learnt. There is no way of ruling out this possibility, since there are few valid theoretical concepts on the basis of which it can be determined what the effect will be of deviations from full fidelity on skill transfer. Considering that others did find positive transfer from PC-based simulation, we need to consider how they dealt with the issue of fidelity and how critical elements for transfer were identified.

Transfer from Space Fortress to real flight. The most extreme deviation from fidelity is encountered in the study by Gopher, Weil & Bareket (1992, 1994). On the basis of instructor ratings, these researchers found that 10 hours of practice on the Space Fortress game improved performance in the pilot training of initial flight cadets in the Israeli Air Force (IAF).

The Space Fortress game did not look like an aircraft cockpit and it did not have an out-of-the-window view as one would see from an aircraft. Gopher et al. argued that in their experiment transfer was not the result of practice with this specific game, but rather it was the result of an appropriate training method or instruction strategy. In this study, trainees in the experiment group were trained to change their focus of attention to different aspects of the game during different game trials. According to this method, trainees were always exposed to the full load of the task and were taught alternative strategies for coping with it. This training method leads to the development of more general skills and response strategies that are less dependent on the specific peculiarities of the task. Gopher et al. state:

> *“.. such strategies maintain their relevance and are easier to generalize when variables are changed or new tasks with a similar context are encountered. We suggest that the attention control skills that were developed in the context of SF training could be generalized to the flight situation and that the similarity between the two environments was sufficient for such generalization to occur.”*

Thus, Gopher et al. argued that, by using the appropriate instruction strategy, ‘attention control skills’ could transfer from a low-fidelity simulation to the flight situation in order to enhance flight performance. With this approach, Gopher et al. refuted the notion of direct transfer of specific skills from the low-fidelity simulation to the in-flight situation, and favored the notion of indirect transfer of more general skills and response strategies.

However, the study raises some methodological questions. First, the experiment started with three groups of trainees, one control group (25 trainees) and two groups (23 trainees each) that trained with Space Fortress. Trainees in the groups were matched on the basis of ability scores. However, there were 13 drop-outs during the experiment, but only in the two Space Fortress groups, because of medical problems and personal difficulties. No drop-outs in the control group were reported. This raises the question whether the significant differences between the control group and the Space Fortress groups can be attributed to differences in attrition during the course of the experiment.

Second, one of the Space Fortress groups was trained with a different instruction strategy than the other group. Because analysis revealed no difference between the two Space Fortress groups, both groups were taken together and further treated as one experimental group. This weakens the arguments of Gopher et al., namely that transfer from the game is merely the result of the appropriate instruction strategy.

Third, the conclusions are based on relatively small differences in instructor ratings for the two remaining groups. Instructors rated 33 flight measures on a scale ranging from 4 to 10. Inspection of these measures reveals that even the largest differences in favor of the Space Fortress group on these measures were 0.4 or lower (Gopher et al., 1994, fig 3, page 398), i.e. less than 7 percent of the full scale. Among the highest ranking effects were such measures as ‘looking into a 45 degrees turn’ and ‘time needed to prepare departure from the practice area’.

Fourth, the significant effects of the Space Fortress game as reported by Gopher et al. could not be replicated in a study by Hart and Battiste (1992). These authors conclude their results with the statement: ‘Although the differences among the three groups were not statistically significant, the trends suggest that differences were beginning to emerge as training progressed, as was observed in the IAF study'. Thus, there are several reasons for questioning the validity of this study. Nevertheless, according to Gopher et al. (1994), the IAF incorporated the game into its flight program.

Transfer from a PCATD to real flight. The PCATD study provides a further example of how the issue of simulation fidelity is dealt with. In 1997 the US Federal Aviation Administration (FAA) allowed 'PC-based aviation training devices' (PCATDs) to be used for a maximum of 10 hours in the instrument training of pilots, whereas previously these 10 hours had to be trained in a more expensive simulator or in the real aircraft.

Taylor, Lintern, Hulin, Talleur, Emanuel & Phillips (1999) studied the transfer-of-training from such PCATD to real flight in a formal training program. On the basis of this well-controlled study it was concluded that: 'transfer savings were generally positive and substantial when new tasks were introduced but low when tasks already learned in previous lessons were reviewed'.

To qualify as an FAA-approved PCATD, the PC-based simulation had to provide a training platform for at least the procedural aspects of flight relating to an instrument training curriculum (FAA, 1997, p1). Required features include (a) a displacement yoke or control stick, (b) self-centering rudder-pedals (c) a physical throttle lever, and (d) 12 additional physical controls for aircraft systems (e.g. flaps, propellers, radio, etc.). Koonce & Bramble (1998) sought references from the literature and from the FAA to support the need for the 12 additional physical controls for aircraft systems but were unable to determine the empirical basis for these features.

Could these transfer savings as reported by Taylor et al. be attributed to improved flying skills? Instrument flying has certainly more procedural components than mere visual flying. However, some of the maneuvers, such as the steep turn on instruments, clearly call for manual flying skills.

Two of the ten flight lessons in the Taylor et al. experiment were fully dedicated to training steep turns. The control group, which only received training in the aircraft, needed on average 3.83 steep turns to reach acceptable performance. The experimental group, after being trained with the PCATD, needed on average 3.40 steep turns in the aircraft. There was no significant difference (0.43 trials) in the number of trials that the control group and the experimental group needed to achieve the criterion performance level.

However, when expressed in flight time, the control group needed on average 1.52 flight hours to demonstrate acceptable steep turns and the experimental group needed on average only 0.95 flight hours to demonstrate acceptable steep turns, a significant difference of 0.57 flight hours. Thus, surprisingly, the control group needed an extra 0.57 flight hours (34 flight minutes) for an extra 0.43 steep turns, whereas the net time for a maneuver such as a steep turn may be a minute or so. The only possible conclusion, which is acknowledged by the authors, is that the control group needed more 'non-specific flying activities'. Thus, the significant advantage for the experimental group, in the case of training for steep turns, is caused by less non-specific flying time, or rather, by more efficient use of flight time.

However, since non-specific flying activities are not under the control of the experimenter, one wonders whether this 34 minute reduction of non-specific flying time should be regarded as 'transfer-savings' induced by the experimental manipulation.

In the current study, we also found significant advantages for both experimental groups in terms of less non-specific flying time. However, it could not be demonstrated that the more specific practice led to a measurable performance increment.

<u>Transfer from ELITE to real flight.</u> Three more transfer-of-training experiments concerned with the transfer of basic instrument flight skills from a commercially available PC-based instrument training system (ELITE) to real flight were reviewed. These three studies were also based on instructor judgements. Two of these studies (Phillips, Hulin & Lamermayer, 1993, Ortiz, Kopp & Willenbucher, 1995) compared two groups, one group trained with the PC-based system and one group trained with a more expensive ground-based flight simulator. As with the Taylor et al. study, the groups received the training in the context of a complete instrument flying course. The Philips et al. study found a significant advantage for the group trained with the PC-based system. The Ortiz et al. study found no significant difference between the two groups. However, both studies lack a control group (i.e. a group that only received training on the aircraft) and do not provide information on the skills being learnt on the ground.

The third transfer study (Ortiz, 1994) compared a control group, which was taken directly to the aircraft, with a group that was trained with ELITE before flying. None of the trainees had any previous aircraft piloting experience. The task to be mastered was to fly a basic square pattern within a required accuracy (rated by an instructor). Ortiz claims a significant advantage for the ELITE group: one hour of practice on ELITE should save up to 29 minutes in the aircraft.

However, this claim is based entirely on an average training time of 20:23 (minutes:seconds) in the aircraft for the control group, and an average training time of 12:23 min in the aircraft for the ELITE group. The ELITE group received an additional ground-training of (on average) 16:48 min. Thus, the ELITE group needed 8 (!) minutes less flight time, which is thought to be the result of 16:48 min of ELITE training. Ortiz's claim that an hour of practice on ELITE saves up to 29 minutes in the aircraft is an extrapolation of these figures. The claim, however, is not supported by the experimental procedure, or by the statistics provided.

4.5 Overall perspective on transfer from PC-based simulation

We found no evidence for the transfer of manual flying skills from PC-based simulation to real flight in the aerobatics experiment. Evidence is also hard to find in the ten relevant transfer studies that were reviewed, with the possible exception of the ILLIMAC landing studies by Lintern and his co-workers (Lintern, Roscoe, Koonce & Segal, 1990, and Lintern, Taylor, Koonce, Kaiser & Morrison, 1997). These studies suggest transfer of flight path control skills during final approach.

With regard to the theoretical relationship between simulator fidelity and skill transfer, the latter two studies are embedded in a theoretical framework that is consistent with E.J. Gibson's (1969) view of perceptual learning and Lintern's (1991) view of skill transfer. The other eight studies are not based on a clear theoretical agenda with respect to fidelity and skill transfer, nor was the present experiment. The research question was predominantly based on the availability of PC-based technology.

In accordance with the PCATD study by Taylor et al. (1999) the present study suggests an advantage in terms of routine for trainees that used PC-based simulation. This routine allowed trainees to execute more maneuvers and spend less time on non-specific flying. Moreover, there was a significant increase in this advantage with flight lessons for the S-group over the C-group and a significant increase in this advantage for the X-group over both the C-group and the S-group.

However, such a reduction in non-specific flying between maneuvers was not under control of the experiment and therefore cannot be called transfer-of-training as a result of the simulation. Moreover, in the present case we established that this procedural advantage for trainees that used PC-based simulation did not result in a measurable improvement of manual flying skills.

A marginal advantage of PC-based simulation, observed in the current training program, is that trainees needed less briefing time from the instructor after every 50 minutes of simulation. Since trainees reviewed the maneuvers of the previous flight lesson and prepared the maneuvers for the next flight lesson with the aid of simulation, briefing times went down from approximately 15 minutes for the C-group trainees to approximately 5 minutes for the S-group trainees and to almost zero briefing time for the X-group trainees. This indicates that PC-based simulation serves as a kind of automatic briefing tool which saves flight-instructor time. Obviously, this was an observed side effect and not within the research objectives of the current study.

5 Acknowledgements

This research was performed with funding from the EUCLID CEPA RTP 11.1 project MASTER (Military Application of Simulation and Training concepts based on Empirical Research). Initial results based on instructor ratings were previously reported in Roessingh & Chlapowski (1997). Full results, based on both instructor ratings and in-flight data recordings, were previously reported in Roessingh (2002). The author would like to thank Ronald van Gent (NLR) for providing some of the basic ideas for this research, Bert Huizenga (Wings over Holland) for helping to set up and realize the training-program and Staszek Chlapowski for helping with experiment preparation and data collection. Jan Koenderink, Astrid Kappers (Universiteit Utrecht), Brian Hilburn and Harrie Bohnen (NLR) commented on a previous version of this paper. Furthermore, the author is greatly indebted to all who helped to realize the experiment: Hans Brugman, Piet Werkhoven, Jaap Jan van den Bosch, Rolf Zon, Martin Joosse, Bas Kuijpers, Esmyralda Eveleens, Piet Hoogeboom and Hugo van Dort (all NLR), staff from Wings Over Holland and trainees from the Internationale Luchtvaartschool, Lelystad.

6 References

Dennis, K.A. & Harris, D. (1998). Computer-based simulation as an adjunct to ab initio flight training. The International Journal of Aviation Psychology, 8(3), 277-292.

Donchin, E. (1989). The Learning Strategies Project, introductory remarks. Acta Psychologica, Vol. 71, 1-15.

Federal Aviation Administration (1997). Qualification and approval of personal computer-based aviation training devices (Advisory Circular No. AC61-126, 5/12/97). Washington, DC: Department of Transportation.

Foss, M.A., Fabiani, A., Mané, A.M. & Donchin, E. (1989). Unsupervised practice; the performance of the control group. Acta Psychologica, Vol. 71, 23-51.

Galanis, G., Jennings, A. and Beckett, P. (1998). A mathematical model of glide-slope perception in the visual approach to landing. The International Journal of Aviation Psychology, 8(2), 83-101.

Gibson, E.J. (1969). Principles of perceptual learning and development. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Gopher, D., Weil, M. & Bareket, T. (1992). The transfer of skill from a computer game trainer to actual flight. In: Proceedings of the Human Factors Society, 36th annual meeting.

Gopher, D., Weil, M. & Bareket, T. (1994). Transfer of skill from a computer game trainer to flight. Human Factors, 36(3), 387-405.

Hart, S.G. & Battiste, V. (1992). Field test of video game trainer. In: Proceedings of the Human Factors Society, 36th Annual Meeting.

Koonce, J.M. & Bramble W.J. (1998). Personal Computer-based flight training devices. The International Journal of Aviation Psychology, 8(3), 277-292.

Koonce, J.M., Moore, S.L. & Benton, C.J. (1995). Initial validation of a basic flight instruction tutoring system (BFITS). In R.S. Jensen & L.A. Rakovan (Eds.) Proceedings of the Eighth International Symposium on Aviation Psychology (pp. 1037-1040). Columbus: The Ohio State University.

Lintern G. (1991). An informational perspective on skill transfer in human-machine systems. Human Factors, 33(3), 251-266.

Lintern G. (1992). Flight simulation for the study of skill transfer. Simulation/Games for Learning, 22(4), 336-350.

Lintern G. (1995). Flight Instruction: The challenge from situated cognition. The International Journal of Aviation Psychology, 5(4), 327-350.

Lintern G. (2000). An analysis of slant for guidance of landing approaches. The International Journal of Aviation Psychology, 10(4), 363-376.

Lintern, G. & Garrison, W. (1992). Transfer effects of scene content and crosswind in landing instruction. The International Journal of Aviation Psychology, 2(3), 225-244.

Lintern, G. & Koonce, J.M. (1992). Visual augmentation and scene detail effects in flight training. The International Journal of Aviation Psychology, 2(4), 281-301.

Lintern, G. &Liu, Y. (1991). Explicit and implicit horizons for simulated landing approaches. Human Factors, 33, 401-417.

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

Lintern, G., Taylor, H.L., Koonce, J.M., Kaiser, R.H. & Morrison, G.A. (1997) Transfer and quasi-transfer effects of scene detail and visual augmentation in landing training. The International Journal of Aviation Psychology, 7(2), 149-169.

Looking Glass Technologies (1995). Flight Unlimited – Operators Manual – A Looking Glass Technologies Production, Cambridige, Ma, US.

Ortiz, G.A. (1994). Effectiveness of PC-based flight simulation. The International Journal of Aviation Psychology, 4(3), 285-291.

Ortiz, G.A., Kopp, D. & Willenbücher, T. (1995). Instrument training using a computer-based simulation device at Lufthansa Airlines. In R.S. Jensen & L.A. Rakovan (Eds.), Proceedings of the Eighth International Symposium on Aviation Psychology (pp. 1037-1040). Columbus: The Ohio State University.

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

Reisweber, D.A. & Lintern, G. (1991). Visual properties for the transfer of landing skill. In R.S. Jensen (Ed.), Proceedings of the Sixth International Symposium on Aviation Psychology. Columbus: The Ohio State University.

Roessingh, J.J.M. (2002). The acquisition of complex skills. Ph.D. thesis, Dept. of Physics and Astronomy, Universiteit Utrecht.

Roessingh, J.J.M. & Chlapowski, S.F. (1997). Advanced manoeuvre flight training on PC's and transfer to the real aircraft. In K.-P. Holzhausen (Ed.), Proceedings of the Europe Chapter of the Human Factors and Ergonomics Society Annual Conference. Bochum: Fachhochschule Bochum.

Taylor, H.L., Lintern, G., Hulin, C.L., Talleur, D.A., Emanuel, T.W. & Phillips, S.I. (1999). Transfer of training effectiveness of a personal computer aviation training device. The International Journal of Aviation Psychology, 9(4), 319-335.

Wewerinke, P.H. (1980). The effect of visual information on the manual approach and landing. Technical Publication NLR MP 80019 U. National Aerospace Laboratory NLR, the Netherlands.

Appendix A Criteria used for the rating of aerobatic maneuvers

Pass/fail criteria for the task consisting of a sequence of aerobatic maneuvers.		
Loop	1	Entry speed should be between 130 and 150 mph.
	2	Acceleration at pull-up should be 3.2 to 4.2 g.
	3	Roll angle in the top should be between 175 and 185 degrees.
	4	Heading at entry should equal (within 10 degrees) the heading at exit.
	5	Altitude at entry should equal (within 100 feet) the altitude at exit.
Slow Roll	1	The roll-in rate should be less than 30 degrees per second.
	2	The roll-out rate should be less than 30 degrees per second.
	3	The variation in roll rate should be less than 10 degrees per second.
	4	Heading at entry should equal (within 10 degrees) the heading at exit.
	5	Altitude at entry should equal (within 100 feet) the altitude at exit.
Inverted Flight	1	Entry speed should be between 120 and 140 mph.
	2	Duration of actual inverted flight should be at least 10 seconds.
	3	While inverted, the variations in roll angle should be less than 10 degrees.
	4	While inverted, average altitude loss/gain should be less than 10 feet/second.
	5	Heading at entry should equal (within 10 degrees) the heading at exit.
Immel-Mann	1	Entry speed should be between 150 and 170 mph.
	2	Roll angle at entry should be smaller than 5 degrees in absolute value.
	3	Acceleration at pull-up should be 3.2 to 4.2 g.
	4	Altitude change during roll-out should be less than 100 feet.
	5	Heading at entry should be opposite to the heading at exit (within 10 degrees).
Split-S	1	Entry speed should be less than 100 mph.
	2	Altitude change during roll-in should be less than 100 feet.
	3	The acceleration in the half-loop should be 3.2 to 4.2 g.
	4	Exit speed should be less than 160 mph.
	5	Heading at entry should be opposite to the heading at exit (within 10 degrees).