FIELD TEST OF VIDEO GAME TRAINER

Sandra G. Hart and Vernol Battiste
NASA-Ames Research Center
Moffett Field, CA

A field study was conducted at the US Army Aviation Center to determine whether workload-coping and attention-management skills developed through structured video game experience would generalize to flight training. Three groups of 24 trainees were compared: (1) One received 10 hours of training on an IBM-PC version of Space Fortress, replicating an earlier study; (2) The second played a commerical video game (Apache Strike) for 10 hours which also required tracking, monitoring, situation assessment, and memory; (3) The third matched group receive no game training. Flight school records were monitored during the next 18 mos to compare performance of the three groups during initial flight training. Check ride ratings began to show an advantage for the group trained with Space Fortress by the Instrument stage of training, as predicted. Furthermore, attrition rates were lower for this group, replicating the results of an earlier study conducted by Gopher (1990) in the Israeli Air Force Flight School.

INTRODUCTION

As the complexity of military and civilian aircraft and the missions they are expected to perform have increased, the time and cost of training pilots to fly them have escalated. The capabilities and availability of flight simulators have made simulator training an attractive alternative to traditional methods. Although it is clear that not all skills require the same level of physical fidelity for acquisition and transfer, nor the same training philosophy, decisions about which methods are most appropriate have been based on tradition rather than scientific principle (Gopher, Weil, & Bareket, 1991). Until recently, physical fidelity has been the guiding principle (i.e., the more closely a simulator resembles the real system, the better it will serve as a training device). However, the constraints and limitations of physical fidelity as a guiding principle have become clear, and the cost of trying to achieve it prohibitive. Furthermore, the relationship between physical fidelity and skill transfer has never been established.

The rich, colorful and challenging environments offered by computer games provide powerful tools with which the foundations of a new approach might be studied and tested. Exposure to a carefully constructed "game" that is structurally, rather than physically, similar to the task of flying might be used to develop skills (e.g., efficient management of attention and coping with very high workload), which are normally acquired during flight training. This rationale led to the development and evaluation of a computer-game trainer in the Israeli Air Force (IAF) Flight School (Gopher, 1990; this volume; Gopher, Weil, & Bareket, 1991). The game, Space Fortress II (SF-II), was based on a complex experimental task developed by participants in the DARPA Learning Strategies Project (Donchin, 1989). The goal of this project was to compare different methods of improving the effectiveness of unstructured practice in low-cost, part-task simulators. One of the more successful methods manipulated subtask priorities within the context of the whole task (Gopher, Weil, & Siegel, 1989). It fostered specific patterns of behavior, encouraged exploration and acquisition of different response strategies, and maximized the match between individual capabilities and the demands of the task. Dr. Gopher proposed that, not only did his training methodology improve performance on the target task (i.e., Space Fortress), but fostered the development of attention-management skills and the ability to cope with very high workload in such a way that these skills would transfer to the performance of other complex, demanding tasks. Furthermore, he hypothesized that these skills should improve an individual's ability to take advantage of instruction provided during the process of learning another complex task.

He tested this idea by introducing SF-II during the first 18 hours of flight training in the IAF Flight School. Here, he found that some of the skills acquired during 10 hours of experience with SF-II did indeed transfer to performance in flight school; trainees with game experience performed significantly better on many of the flight tasks evaluated on flights 10-18 than did a matched control group, who had no game experience. Individual tutoring on the game, which significantly improved performance on the game itself, was not necessary to achieve flight school improvements. It appeared that the game experience promoted better coping strategies for high workload situations and more efficient attention-allocation strategies for difficult, complex tasks, thereby improving either: (1) the trainees' abilities to benefit from instruction in flight school or (2) specific skills that are essential to both good performance on SF II as well as in flight. The benefits of game training continued to increase as the complexity and difficulty of flight training increased and resulted in a higher percentage of graduates. As a result, Space Fortress II was incorporated into the IAF training program.

In 1990, the NASA researchers who funded the IAF study, were asked by the Commander of the US Army Aviation Center to evaluate the potential of special purpose, as well as commercial video games for improving the performance of Army aviators during Initial Entry Rotary Wing (IERW) Training. Helicopter flight requires the effective management of information from competing visual and auditory sources and making concurrent discrete and continuous manual and verbal responses while performing a variety of cognitive activities including memory, information processing, spatial transformations, calculation, etc. Pilots must schedule, prioritize, and integrate these requirements appropriately to successfully complete even the simplest mission. This requires effective workload coping strategies and attention management skills. It is possible that these "generic" skills could be developed through extensive interaction with ANY complex, highly demanding task that imposes an appropriate combination of sources and levels of demands. If so, then casual video game play (or structured practice with a specific game) might promote skills that would generalize to flight. Although there is anecdotal evidence of positive transfer between video games and flight, no well-controlled studies have been conducted to demonstrate such a relationship. Thus, a survey was conducted to evaluate the relationship between causal video game experience (i.e., types of games played, frequency of use, expertise) and flight school performance at the US Army Aviation Center (Hart & Battiste, 1991a; 1991b). This survey provided demographic information about casual video game experience for a "typical" IERW class and aided in the selection of a commercial video game to be used in the field test. It also demonstrated a significant correlation between causal video game experience and success in flight school.

The current study had two goals: (1) To replicate the critical elements of the IAF study with a US population and (2) Determine whether a commercial video game with similar properties might offer similar training benefits. To accomplish these goals, performance during the initial phases of flight training was compared for three groups of trainees: (1) A control group that received no game training, (2) A second group that played a commercial video game, Apache Strike, for 10 hours, and (3) A third group that played Space Fortress for 10 hours. There were several differences between the populations of students in the IAF Flight School and the US Army Aviation Center that were potentially relevant: (1) Average age entering flight school (IAF: 18 years; USA: 24 years), (2) Education (IAF: High school; USA: Many with college degrees), (3) Causal video game experience (IAF: Little or none; USA: Most have some experience), and (4) Differences in training vehicle (IAF: fixed wing; USA: rotary wing). Thus, the three groups were matched on the basis of age, education, and video game experience as well as flight aptitude test scores. Because the manual-control demands of helicopter flight are so much greater than those imposed by the fixed-wing trainers used in the IAF study, additional flight tasks, such as navigation and communications are not introduced until later (helicopter students are still learning stick-and-rudder control after 18 flight hours, which was the point at which data were taken in the IAF study). Thus, we anticipated that the skills targeted by SF-II would not emerge until much later in rotary wing training (e.g., during Basic and Advanced Instruments) and continued to track their performance until the end of basic flight training, (i.e, approximately 80 flight hr).

METHOD

Subjects

From the pool of 109 student aviators who were expected to begin flight school between July 5 (Class 90-19) and July 19, 1990 (Class 90-20), 42 Warrant Officer Candidates (WOCs) and 42 Officers were selected and assigned to one of the three experimental groups. Although it was anticipated that the participants would begin flight training within 2-4 wks of completing video-game training, Operation Desert Shield/Storm delayed flight training for many of the study participants. Thus, actual class membership ranged from 90-18 (3 weeks post training) to 91-1 (21 weeks post training). Due to this delay, and normal attrition, the data from only 33 WOCs and 37 officers were available at the end of the study.

Apparatus

Screening tests and game training were conducted in a large room with visual separation between experimental stations. The Aiming Task, used for screening and group assignment, and SF-II were programmed on an IBM-compatible microprocessor. The game was displayed on a 13-in color monitor (Figure 1). Spaceship control and actuation of discrete functions were performed with a two-axis CH Flight Stick in the subject's right hand and a three-button mouse in the subject's left hand. The controls were mounted on a lab board. A more complete description of the game may be found in Gopher (1990; this volume), although their version of the game was programmed on a different computer.

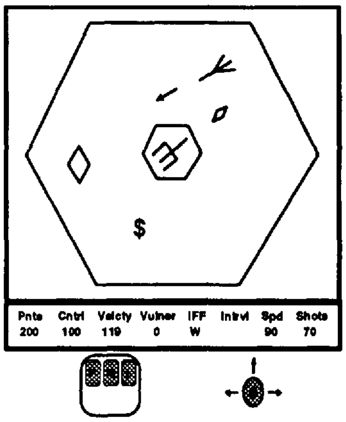

Figure 1: Space Fortress Display

Apache Strike, made by Silicon Beach Software, was selected as the commercial video game. Similar to SF-II, it required visual scanning, performing multiple simultaneous tasks, maintaining situational awareness, and imposed relatively high workload. The game was displayed on a 9-in black and white Apple monitor (Figure 2). Helicopter control and actuation of discrete functions was performed with a one-button Apple mouse.

Experimental Design

The study was a 2 (WOC, Officers) X 3 (Space Fortress, Apache Strike, Control) between-groups design with 14 student pilots assigned to each block. Phase of training (Primary 1, Primary 2, Basic Instrument, Advanced Instrument) was the only within-group variable.

Procedure

More than 100 WOCs and officers waiting to enter flight school were briefed about the study, completed a Video

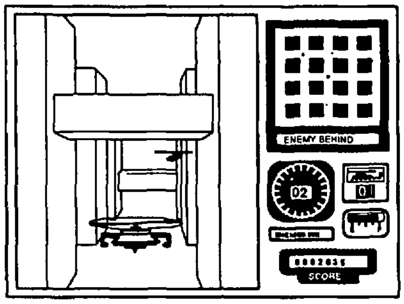

Figure 2: Apache Strike Display

Game Survey, and performed a screening test as a group. They were told that the study was designed to evaluate a candidate flight-school selection test and that their performance would have no bearing on their careers. If they were selected for participation in the study, they were instructed to refrain from talking about their experimental group assignment with instructor or evaluation pilots. The video game questionnaire was administered to determine the extent of prior video game experience (e.g., types of games played, frequency of play, level of performance achieved). It was developed by NASA to assess the relationship between casual video game experience and performance in flight school in a prior study. The Aiming Task, which is a 10-15 min psycho-motor screening task that has been used in previous studies, was administered individually.

From the pool of potential participants, three groups of 14 WOCs and 14 Officers each were selected, matched on the basis of: (1) Prior video game experience, (2) FAST/Multitrack Scores, (3) Background Data (e.g., education, prior flight experience, age), and (4) Aiming Task score. The Control Group received no further training and did not interact again with the experimenters. The Apache Strike and Space Fortress Groups initiated 10 daily, 1-hour period of video game training.

Four students were trained at a time by an experimenter and two assistants - - two with SF-II and two with Apache Strike. Each student was scheduled for no more than one hour per day with no more than three days between successive training sessions. For both groups, initial instructions were given in writing, with verbal clarification, when necessary. Progress within a training perior was controlled by the subject, with no intervention by the experimenter, who monitored each subject's performance on remote displays.

The Space Fortress Group performed eight sessions per day for the first 5 days, and 10 per day for the last 5 days. The instructional method used was similar to that of the "partial training" group in the IAF study (see Gopher, this volume). The playing time of each session lasted 3 min, although the actual duration varied somewhat depending on the frequency with which visual feedback was provided after specific events (e.g., killing the Fortress). Initiation of successive sessions and knowledge of results were controlled by the computer. The Apache Strike Group simply played the game for 45 min. The duration of each game depended upon the trainees' skill. Early in training, the players were "killed" more frequently and played a number of short games. As training progressed, however, players improved and began to move on to higher (and more difficult) levels within the same game. Thus, the number of games played per day decreased, and a very successful player might not finish even one game by the end of a training period. In this case, the session was arbitrarily terminated after 45 min.

Measures of Performance

For the Space Fortress Group, 15 measures of performance were recorded (e.g., total score, total points, mine handling, fortress destructions, velocity control, etc.). Summary performance feedback was displayed to subjects after each session. In addition, subjective ratings were obtained after each session.

For the Apache Strike Group, performance measurement was more difficult. Since it is a commercial game, and not a research instrument, measures of performance were not recorded automatically, although summary statistics were displayed to the subject after each game. This information was written on forms by both the subject and the experimenter, and later entered into a database. Some of the measures recorded included total score, estimated duration and difficulty, and number of tanks, helicopters, etc destroyed.

Performance during the initial (Common Core) phase of training was evaluated similarly for all three groups of trainees using the ratings normally provided by instructor pilots before each check ride (termed "put up grades") and by evaluation pilots after each of four check rides: (1) Primary Stage 1, (2) Primary Stage 2, (3) Basic Instruments, and (4) Advanced Instruments. On average, the students' cummulative flight/simulator hours when they reached these points in training were: (1) Primary Stage 1: 18/0, (2) Primary Stage 2: 57/0, (3) Basic Instruments: 57/15, and (4) Advanced Instruments: 76/30. In addition, a detailed analysis of component ratings was performed for the Advanced Instrument Check Ride, which was the last phase of training included in this study.

RESULTS

Game Performance

As might be expected after approximately 10 hr of training, performance on both games improved dramatically. For Space Fortress, the average number of points earned per session improved significantly from -500 to 1600 ($F=1148$, $p<.0001$). These data are roughly similar to those obtained in the IAF study, although initial performance was somewhat better in the current study and performance on the last day was halfway between that of the the "full" and "partial" training groups. Overall, officers' performance was significantly better than that of the WOCs ($F=5.50$, $p<.02$). For Apache Strike, a similar improvement in performance was found: the total number of points earned per training session increased significantly from 797 to 11,805 ($F=49.57$, $p<.0001$). The number of games played per training session

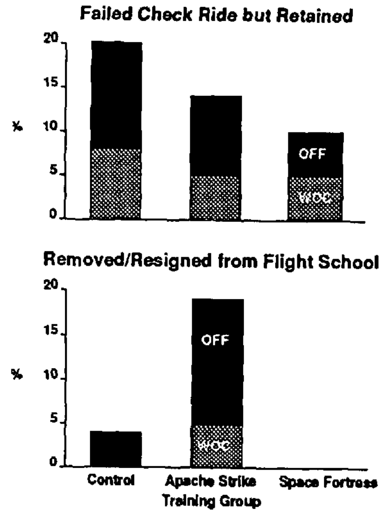

Figure 3: Students with problems

decreased significantly from 7 to 3 (F=31.2, p<.004). Although the subjective workload associated with playing the game changed significantly from the first to the last day (F=2.87, p<.02; from 2.9 to 2.7 on a scale of 1-5), the trend was non-linear. For the Apache Strike Group, WOCs outperformed officers, but the difference was not significant (5605 *vs* 4681), and they rated their workload to be significantly lower (F=102.23, p<.0001; 2.65 *vs* 3.25).

Flight School Performance

Because so many of the trainees did not start flight training immediately after completing game training, as planned, linear regressions were computed for each Group (and across groups) to determine whether there was a consistent relationship between the length of the delay (which ranged from 3 to 21 wks, with an average delay of 8.6 wks) and performance in flight school (Advanced Instruments grade). The differences between groups ranged nonsignificantly from 8.1 to 8.8 wks. None of the individual or combined regressions were significant. Thus, whatever benefits game training offered were stable over time.

Of the 84 WOCs and officers selected to participate in the study, only 33 WOCs and 37 officers had actually entered flight school by January 1, 1991. The total number remaining in each experimental group were: Control: 26; Apache Strike: 22; Space Fortress: 22. Of these trainees, the percent that failed at least one of the four check rides and either remained in flight school, resigned, or eliminated is depicted in Figure 3. As you can see, the percentages are considerably lower for the Space Fortress Group than for either of the other Groups.

The ratings given by evaluation pilots were used as the primary measures of flight school performance. For the purpose of evaluating the potential benefits of computer-game trainers, the rating given the *first* time a check ride was taken were used in the analyses. Other data were available as well. For example, the "put-up grades", given by a student's instructor pilot immediately before each check ride, were compared to those given by the evaluation pilot. The correlation between these two grades was statistically significant (rxy = 0.68), although the put-up grades were usually considerably higher. In addition, ratings of factors termed "basic qualities" (e.g., motivation, attitude, cross check, coordination) and performance on component tasks were analyzed for the Advanced Instrument phase. Space precludes a detailed review of those results, although the Space Fortress Groups' scores were generally higher for those factors that might have been expected to improve as a result of attention/workload management training (e.g., planning, control, coordination, crosscheck), whereas performance was the same or worse on others that were related to other factors (e.g., teamwork, motivation, and aptitude).

Overall, grades decreased significantly from Primary Stage 1 to Advanced Instruments (F=6.02; p<.001), reflecting the increasingly demanding nature of the process. 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. For example, there was only one point difference in mean ratings between the three groups for Primary Stage 1. However, for Advanced Instruments the range had increased to 5 points (between 77 and 82, see Figure 4). Furthermore, the distributions of grades were were quite different for the three groups (Figure 5), again suggesting the growing impact of SF-II training on flight-school performance.

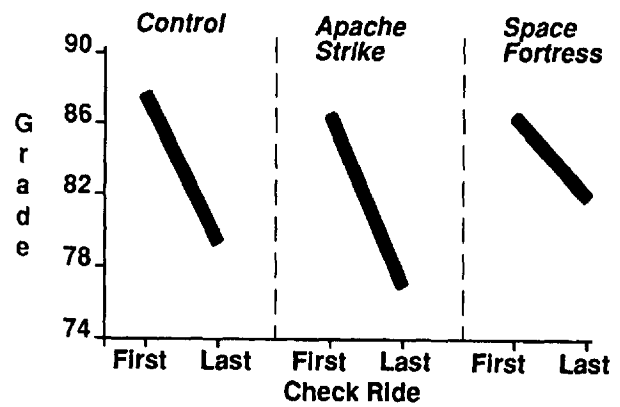

Figure 4: Check ride ratings early and late in training

DISCUSSION

Although the results of this study were not conclusive, taken together with those of the earlier study, they suggest that SF-II shows promise as a low-cost method of improving pilots' abilities to cope with the very high workload and competing attention demands typical of flight training. There was no evidence to suggest that a commercial video game could offer the same benefit - - performance by the Apace Strike Group was even worse than it was for the Control Group who received no special training. In addition, this finding eliminates the possibility that the differences found in the IAF study represented some sort of Hawthorn effect. Finally, as with the IAF study, differences between students who received SF-II training, and those that did not, emerged during the later, more complex phases of training.

These results lend support to the concept of developing computer-game trainers to foster other skills. These systems

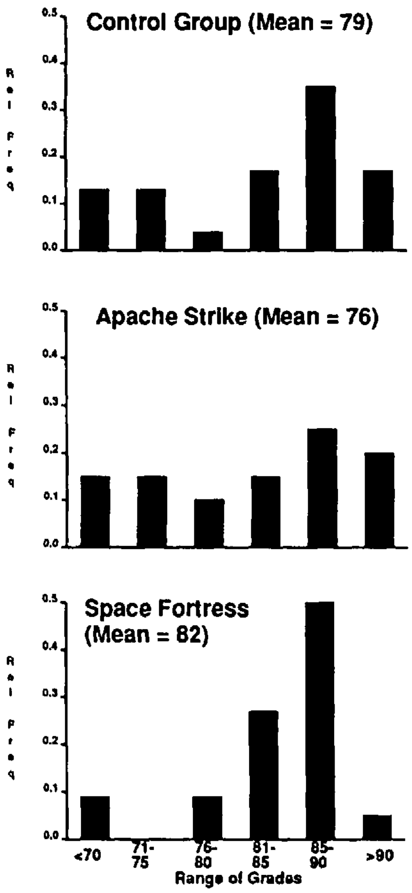

Figure 5: Distributions of Advanced Instruments ratings

might be based upon commercially available video games that have been found to encourage specific behaviors or special purpose tasks that just appear to be video games. The format and content of these training devices could emerge from human factors research in a variety of areas (e.g., sensor imagery interpretation, geographical orientation, and decision making), and serve as a mechanism by which the human factors community could "transfer" their "technology" to other communities. The content and realism of visual displays, sequence of instructional elements, presence of explicit or implicit embedded instruction, trainee interaction, response devices, etc. needed to develop target skills must be determined empirically at first, although general principles should emerge as this approach achieves broader application.

ACKNOWLEDGEMENTS

We wish to thank several people whose patience and support made this study possible: Dr. James Dees (DOTD), Dr. Gabriel Intano (ARI), and MAJ John Barron (DPT). We wish to give special thanks to CPT Mark Reisweber (now an instructor at the US Military Academy). Without his leadership skills, first-hand knowledge of Army aviation, and human factors training, this study would not have happened. We are also grateful for the consultation provided by Dr. Daniel Gopher.

REFERENCES

Donchin, E. (1989) The learning strategies project: Introductory remarks. *Acta Psychologica, 71*, 1-16.

Gopher, D. (1990, July). The Skill of Attention Control: Acquisition and Execution of Attention Strategies. Ann Arbor, Michigan. To appear in *The XIV Symposium of the International Association for the Study of Attention and Performance*. NJ: Lawrence Erlbaum Associates.

Gopher, D. Weil, M., & Siegel, D. (1989) Practice under changing priorities: An approach to training of complex skills. *Acta Psychologica, 71*, 147-178.

Gopher, D., Weil, M., & Bareket, T. (1991). *Fidelity revisited: The transfer of skill from a computer game trainer to actual flight.* (Final report for NAGW 1012). Haifa, Israel: Technion - Israel Institute of Technology.

Hart, S. G. & Battiste, V. (1991a, August). Transfer of Training: Video Games to Flight. Paper presented at the Ninety Ninth Annual Convention of the American Psychological Association, San Francisco, CA.

Hart, S. G., & Battiste, V. (1991b, April). Use of Video-Game Trainers in Military Flight Training. Paper presented at the Sixth International Symposium of Aviation Psychology, Columbus, OH.