HUMAN FACTORS, 1994, 36(3), 387–405

Transfer of Skill from a Computer Game Trainer to Flight

DANIEL GOPHER,[1] MAYA WEIL, *and* TAL BAREKET, *Technion-Israel Institute of Technology, Haifa, Israel*

An experimental study was conducted to test the transfer of skills from a complex computer game to the flight performance of cadets in the Israeli Air Force flight school. The context relevance of the game to flight was argued on the basis of a skill-oriented task analysis, using the framework provided by contemporary models of the human processing system. The influence of two embedded training strategies was compared, one focusing on the specific skills involved in performing the game, the other designed to improve the general ability of trainees to cope with the high processing and response demands of the flight task and teach better strategies of attention control. Efficient control and management of attention under high task load are argued to be skills that can improve with proper training and generalize to new situations. Flight performance scores of two groups of cadets who received 10 h of training in the computer game were compared with those of a matched group without game experience. Both game groups performed significantly better than the no-game group in the subsequent test flights. The results are discussed with reference to the theoretical framework within which the context relevance of the game was supported. Also considered are the effects of the different training schedules and the significance of the data to the study of attention control. The game has now been incorporated into the regular training program of the Air Force.

INTRODUCTION

In training research, there has been a call for a systematic development of guiding principles for the design of training simulators. Such principles, anchored in human performance and learning theories, are targeted to replace the long-prevailing physical fidelity approach, which has been enshrined by its compelling appeal to folk psychology (see Baudhuin, 1987; Donchin, 1989; Flexman and Stark, 1987; Gopher, Weil, Bareket, and Caspi, 1988; Lintern, 1991; Roscoe, 1980).

Contemporary technology underscores the importance of these efforts. The increased sophistication of engineering systems, their much-enhanced performance envelopes, and extreme operational environments (e.g., air, space, underwater, and nuclear) preclude on-the-job training and also render high-fidelity simulations difficult, impossible, or

[1] Requests for reprints should be sent to Daniel Gopher, Industrial Engineering and Management, Technion-Israel Institute of Technology, Haifa, 32000, Israel.

prohibitively expensive. Consequently most of the existing training simulators represent a compromise in physical fidelity, the extent and influence of which is hard to evaluate.

Modern microprocessor technology and the development of rich, colorful, and challenging computer games provide powerful tools with which a new approach to training can be studied and tested. Indeed, this is the rationale that has guided an international research collaboration directed at the development of training strategies embedded in a complex computer game named Space Fortress (Donchin, Fabiani, and Sanders, 1989). Our group at the Technion-Israel Institute of Technology was part of this project, and the work reported in this article is an outgrowth of that effort.

We describe the results of a field study conducted at the Israeli Air Force flight school to investigate the transfer of skill from a complex computer game to flight performance. Our leading claim was that instructed practice in this game and the use of a training strategy directed to improve attention management skills would enhance the ability of cadets to cope with the high demands of flight. We further argued that although the elements and parameters of the computer game were physically remote from those of the flight situation, the game would provide a useful training context for developing flight-relevant skills, particularly those related to the control of attention and coping with high load.

The foregoing argument is a composite of three separate claims. One is that attention control is an important element in the acquisition of flight skills. A related claim is that such control can be treated as a skill and can be improved by training. The third claim is that the context provided by the selected computer game is relevant to the training of flight skills. In the following sections, each of these claims is briefly examined.

Attention Demands in Flight

Piloting a high-performance aircraft, particularly in military aviation, stretches human capabilities to their limit. High attention demands are a major concern in all aspects of human factors work in aviation, including design, training, operational procedures, and safety regulations (O'Hare and Roscoe, 1990; Wiener and Nagel, 1988). The influences of high load and heavy attention demands are most apparent in flight training, during which many trainees are overwhelmed by the concurrent demands of the flight task. Indeed, a review of students' files in Israeli flight school, conducted prior to the present experiment, revealed that difficulties associated with attention control and inability to deal with the load of flight were among the problems most frequently cited by instructors and were a common cause of candidates' washout. Substantial research efforts have been directed to the development of selection tests that identify individuals with higher attention capabilities (Gopher, 1982; North and Gopher, 1976). However, experienced flight instructors are well aware that continuous selection and washout of individuals who encounter problems is a limited and costly solution. Consequently, testing the claim that coping with high load can be improved through training is of much interest and daily concern to both instructors and trainees.

The Skill of Attention Control

A study of attention control from a skill perspective has been the subject of a series of experiments conducted in our laboratory (see Gopher, 1992; Gopher, Weil, and Siegel, 1989). Two major questions were addressed in this research: "How able and efficient are humans in the control of attention under high concurrent demands?" and "Can attention control improve with proper training and generalize to new situations?" We have proposed the idea that when attempting to cope with the high demands of tasks composed of multiple and dynamically changing subelements (e.g., driving a car or flying an airplane), the human operator is required to adopt attention-allocation strategies corresponding to the priority of each of the covarying elements. For example, a pilot is required simultaneously to control the airplane, monitor the outside environment, scan the instrument panel, and handle radio messages. Each of these covarying components would benefit most if the pilot could fully attend only to it. As this is not possible, pilots are required to adopt strategies of attention allocation and change the priorities of attending to task elements during different segments of their mission.

Accordingly, an attention-control strategy is defined as a vector of emphasis levels, or attention weights, over the subelements of a complex and demanding task. The weights (and hence the strategy) may change as the pilot switches, for example, from a navigation task to a sequence of aerobatic maneuvers. Thus attention strategies may be adopted and changed dynamically in the course of performing a complex task.

Our work, as well as the work of others, has shown that when subjects develop their experience by trial and error under high load, they tend to adopt suboptimal attention strategies. Moreover, these strategies tend not to change or be replaced by better ones with the progress of experience (Foss, Fabiani, Mane, and Donchin, 1989; Gopher et al., 1989). In contrast, by a systematic manipulation of emphasis on different task subelements, subjects were led to explore a wider range of attention strategies and improved their ability to cope with the high load of tasks. This ability was further shown to generalize to new task situations (Gopher, 1992). Therefore theoretical and empirical evidence exists in support of the argument that the control of attention and the ability to establish better attention-management strategies can develop with training.

Flight-Relevant Context: The Modified Space Fortress Game

A key challenge when experimenting with a low-physical-fidelity trainer is the justification of its relevance to the performance of the intended operational task. The original Space Fortress (SF) game was developed at the Cognitive Psychophysiology Laboratory of the University of Illinois to simulate a complex and dynamic flight environment (Mane and Donchin, 1989). It was then reevaluated and adopted as the common task for the abovementioned collaborative study of learning strategies (Donchin et al., 1989). The modified version of the game employed in the present work introduced a few changes, following an analysis of the Israeli flight training program, while preserving the major characteristics and parameters of the original SF game.

A detailed description of the game is presented in the method section (see also Figure 1). The game required subjects to control the movement of a spaceship while aiming and firing missiles to destroy the space fortress. Subjects had to protect the ship from hostile elements and manage resources. The game components include high visual monitoring and scanning demands, a difficult manual control, discrete and timed motor responses, short-term memory load, long-term memory of procedures and cost functions, and research management considerations. Several game elements are present concurrently and vary dynamically under severe time pressure, resulting in high load and a mandatory requirement to develop attention management strategies.

The conceptual framework that has guided the development of the game is common to many contemporary models of human information processing. It is equally prevalent in human performance research and in the analysis of mental load (Gopher and Sanders, 1984; Wickens, 1992). The framework is also easy to apply to describe the skill components involved in flight training. Our analysis of the flight training program at the early stage at which the SF game was introduced identified six major topics of training:

1. Familiarization with airplane controlling and maneuvering: use of primary and secondary control surfaces; effects of speed and angle change.
2. Introduction to the cockpit: use and interpretation of instruments and controls.
3. Spatial orientation: plane orientation in space relative to the horizon; ground reference to determine position.
4. Procedures: aircraft performance and safety envelope; routine and emergency takeoff; enroute and landing procedures.
5. Airport traffic patterns: departure from and entry into the airport control area.
6. In-flight communication: dialogue with instructor; in-flight declarative statements of performed and intended activities; debriefing and radio communication.

The general similarity between the demand elements in the game and those in flight training is readily apparent: Both include continuous and discrete manual control, visual and spatial orientation, procedural knowledge involving long- and short-term memory information, and high attention demands under severe time constraints. Verbal communication was also introduced into the game to simulate these demands in the flight situation. We argue that attention-control skills and efficient allocation strategies develop within the context of these dimensions of task demands (e.g., trading manual control, visual orientation, and memory requirements). This is also the context profile that set the boundaries for generalizing the acquired skill to different tasks and other situations.

Selection of a Training Strategy

The objectives of the original Space Fortress project were to examine and compare the contributions of different training approaches to performance of the Space Fortress game. Two contrasting approaches proved to be most beneficial: the emphasis-change approach proposed by our group (Gopher et al., 1989) and a hierarchical part-task approach proposed by Fredriksen and White (1989). Subjects trained under the emphasis-change approach practiced the whole game at all times, but they were led through instructions and auxiliary feedback indicators to vary their focus of attention in different game trials on different aspects of the game. Under this method, subjects were always exposed to the full load of the task and were taught alternative ways for coping with it. In contrast, the hierarchical part-training method was more specific and prescriptive in the training of components. The whole task was decomposed, and before subjects were introduced to the full standard game, they were led through a sequence of simplified part games, which gradually became more integrative and complex. Difficulty and load were much tempered under this method, compared with the emphasis-change approach. In addition, subjects were given verbal tips on recommended behavior, based on performance analysis of expert Space Fortress players.

From the present vantage point of testing transfer from Space Fortress training to actual flight, and in light of the general context similarity between the two environments, we had no clear criteria with which to determine whether one of the training approaches, or some combination of the two, would be best to employ. Given the availability of only one experimental unit and the limitations set by the flight school on time and access to flight cadets, it was resolved to experiment with only two game training groups. One was given a schedule based on the combined power of emphasis change and hierarchical-part-task method. The second was given a more constrained schedule and trained only on the whole task under emphasis-change and attention-management procedures.

METHOD

Apparatus and Task

Space Fortress II. The main elements of the modified Space Fortress (SF) game are depicted in Figure 1. The game display was presented by an HP 1351A graphic system, on an HP 1310 monitor with a 45-cm (21-inch) diagonal CRT display. The task was governed by a Data General PDP-11/23 microprocessor. Sound effects were produced by a Votrex voice system. Subjects played the game using a game board equipped with a two-axis, right-hand game stick (WICO command control) and a left-hand customized, three-push-button panel.

The task required the control of a spaceship moving in a frictionless, hostile environment where it was threatened by a space fortress, located in the center of the screen, and by mines that were actively chasing it. To achieve a maximum score, subjects had to destroy the fortress, defend themselves, destroy all mines, manage their resources of missiles and point bonus, and avoid being hit by either fortress or mines.

To destroy the fortress, subjects first had to hit it 10 times with missiles, firing at a rate slower than one missile every 250 ms, and then demolish it with a fast double shot. Mines appeared every 4 s for a total duration of 10 s or until destroyed. They actively chased the ship to damage it. During mine presence, the ship missiles were ineffective against the fortress. There were two types of mine, friends and foes. Mine type was identified by one letter from three letter memory sets, displayed in the IFF indicator on the instrument panel at the bottom of the screen. The regular missiles could be used to hit friend mines. To destroy foe mines, subjects had to change weapons by pressing twice on the IFF button of the control panel, with an interpress interval of 250–400 ms. The actual interpress interval was displayed in the INTRVL counter.

A game trial begins with a total of 100 missiles, but players could renew their supply or, alternatively, receive extra points by monitoring the random symbols displayed below the fortress for two consecutive presentations of a dollar sign. The SHOTS counter at the bottom of the screen informed the player of the remaining number of missiles.

The ship moved in a frictionless environment, and its control was based on difficult and demanding dynamics: clockwise and counterclockwise rotations were produced by right and left movements of the stick, using velocity dynamics. Acceleration in the direction of pointing was produced by fore movements of the stick. The area on the display between the two hexagons was the recommended area for ship movement.

Additional performance feedback and score counters were located on the instrument panel: VEL presented a score proportional to the time the ship's motion was slowed below a given velocity, CONT presented a score proportional to the subject's ability to navigate the ship in the area bounded by the two hexagons, and SPEED reflected the subject's efficiency in detecting and destroying mines. These counters were part of the emphasis change manipulation.

Part tasks. Apart from the whole game, there were seven possible part games. The following is a brief description of these games.

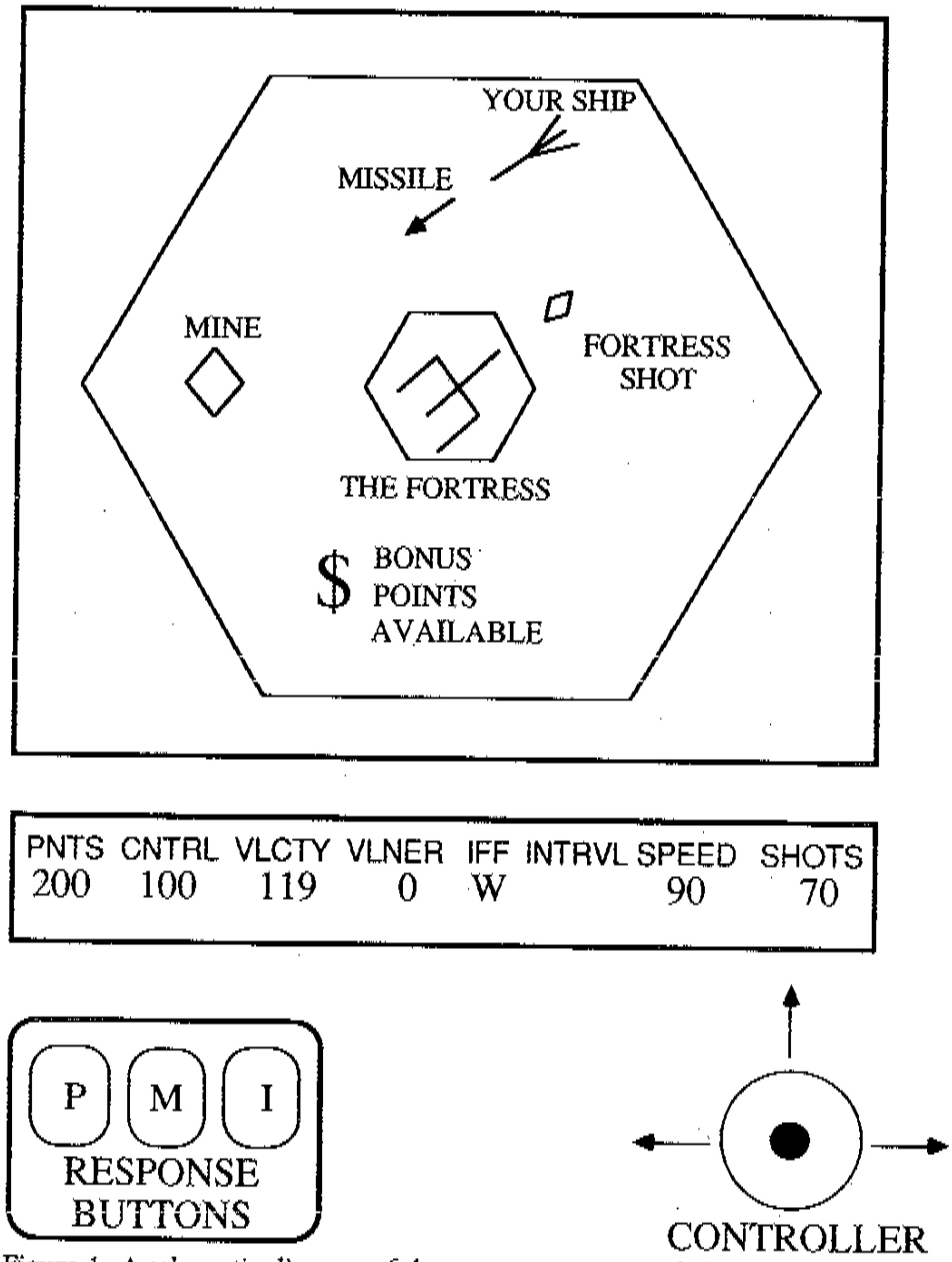

Figure 1. *A schematic diagram of the computer screen and the controllers of the modified Space Fortress game (see text for explanation of elements).*

All were part of the battery used by Fredriksen and White (1989).

1. Aiming: The ship was displayed at a fixed position in the center of the screen but could rotate around its center. Subjects were required to aim and fire missiles to hit static mines that appeared every 4 s at random locations on the screen.
2. Control of ship motion: A moving ship entered the screen. Subjects had to bring it to a halt by reversing the direction and velocity of its motion vector. A new trajectory and velocity were introduced each time the subject was able to halt the ship for 3 s.
3. Trajectory control: Only the ship and the hexagons were presented on the screen. Subjects were asked to fly the ship within the area bounded by the hexagons.

4. Change of weapon system and hit mines: The situation was as in Part Game 1 above, but mines could be friend or foe, as indicated by a letter appearing below the ship. Subjects had to change weapon systems and kill the mine, or fire regular missiles, depending on the mine's identity.
5. Dynamic mine handling: The situation was as in Part Game 4, but both the ship and the mines were in motion.
6. Fortress kill: The ship, fortress, and hexagons appeared on the screen, but there were no mines. Subjects had to circle the fortress slowly, moving within the hexagons while firing at the fortress and avoiding its shots.
7. Full game with friend mines: The full game was played, but only friend mines appeared.

Training Schedules

Two groups of flight cadets were trained for 10 1-h sessions. A training session consisted of 10–14 practice trials of 2 or 3 min each, totaling 38 min net game time per session.

One group, the *full training* (FT) group, spent about 18% of its total training time practicing the seven part games (during the beginning of each of the first six training sessions). During the remaining time, subjects practiced the full game under the emphasis-change manipulation. Subjects in this group were debriefed and given feedback on their performance at the end of each 3-min game trial as well as at the beginning and end of each session. They were also given verbal tips based on the original tips proposed by Fredriksen and White (1989). Tips pointed out recommended ways of performing different task segments. Among the tips were "Hold the stick lightly by your fingertips," "Move slowly in circles facing the ship, continuously aiming at it," and "Do not leave your path to chase mines; wait for them to approach you." To simulate the demands of handling verbal communication and radio messages in flight, subjects in this group were asked during Sessions 7–9 to talk their way through the game. They reported, in real time, specified game events or intended actions, such as the appearance of a foe mine, an intended double shot to destroy the fortress, or a command to renew missile supply. Verbal reports were monitored and recorded by the experimenter.

The *emphasis only training* (EOT) group practiced the whole game throughout training and was not given any part task. Subjects were given the same display and information counters as those in the FT group. It should be noted that the major elements of the emphasis change manipulation were embedded in the display information counters VEL, CONT, and SPEED. Hence, although not directly instructed to do so, subjects in this group were able to monitor tradeoffs between task components. They received only general information on their performance levels at the end of each session, were not debriefed on their errors after individual trials, and were not given any verbal tips.

The control group was without game experience. It included trainees matched to the game groups in their basic abilities and light aircraft flight scores. Ability scores were based on the air force flight selection test battery and included the overall regression score, intelligence measures, psychomotor ability, and life and interest scale (Gopher, 1982). If we had had sufficient funds and laboratory equipment, we would have run two additional control groups: one on a completely stripped-down version of the game, without any auxiliary information and emphasis instructions, and another trained on a markedly dissimilar computer game. These aspirations, however, could not be realized under the existing conditions.

Training Procedure

Based on the skill-oriented task analysis of the training program, it was decided to introduce the game at an early stage of flight training, when basic habits are formulated and elementary response patterns emerge.

Accordingly, we selected the period immediately following the completion of 10 h of flight on a light aircraft and preceding the transition to the high-performance jet trainer.

Subjects participated in one training session per day, with a maximum interval of two or three days between successive sessions. Because of scheduling constraints, the 10 sessions of training were broken into two segments: Five hours were given immediately after completing training on the light aircraft, the other five were administered during the two weeks of ground school that preceded the transition to the jet trainer. There was a three-month interval between the two segments, during which subjects underwent basic military training. No flights took place during this period.

Selection of Flight Validation Criteria

Transfer effects from the game training to actual flight were tested during eight flights (45–60 min each) of the transition stage to the high-performance jet trainer. This period normally served also for evaluation and decision on candidates' assignment to future training programs. The first 4 of the 8 flights focused on practicing basic, individual flight elements, such as climbs, descents, and turns. During the last 4 flights, trainees practiced more complex combined maneuvers based on the previously learned elements.

Each flight was scored by the flight instructor on several criteria and was also given a general performance score on a scale of 4–10 (10 was best). Formal check rides were conducted during Flights 4 and 8, each resulting in two general scores: One summarized performance over the relevant flight segment (1–4 or 5–8); the other conveyed the instructor's estimate of the candidate's likelihood of completing flight training. As a routine, instructors were changed after the first four flights, to ensure that each candidate would be taught and evaluated by at least two instructors. Six of the available standard scores were selected for the present study: the score of Flight 4, as well as the summary and prediction scores based on the first four flights, and the score for Flight 7 and the summary and prediction scores based on Flights 5–8. The score of Flight 7 was selected rather than that of Flight 8 because, although both flights included the same components, Flight 7 was not a test.

As indicated, standard flight scores were all based on an overall estimate of an entire flight or several flights together. Also, as in many other scoring systems, the actual effective range of scores shrank to 5–8. To augment flight evaluation, we developed, together with the school training squadron, three special evaluation forms. These were used to evaluate three individual maneuvers judged to be most typical of the flight demands at this stage. The forms were completed by the instructors during flight. The three maneuvers were 30-deg and 45-deg standard turns, included in Flights 4 and 6, respectively, and departure from practice area, which was introduced in Flight 7. The two standard turns were each rated on 10 different performance elements, and the departure maneuver was rated on 7 aspects (Figure 4). To avoid bias, the special forms were filled for all candidates in the class, not only the subgroup participating in the experiment.

Cover Story

To help eliminate potential bias of experimental subjects, the project was formally introduced to cadets and flight instructors as a study in individual differences to be used in the development of a computer-based flight selection battery. Trainees were told that their participation would not influence their own course in school.

Subjects

A total of 58 cadets participated in the experiment: 16 in the FT group, 17 in the EOT group, and 25 in the control group. All subjects were 18–20 years old, from two consecutive flight classes. To ensure equal proportions of high, medium, and low potential candidates in each group, they were matched individually and assigned to groups by the school psychologist. The matching scores included the integrated index of the flight selection test battery (Gopher, 1982) and the check-ride score of Flight 10 in the light aircraft. Ability and flight scores of individuals were not disclosed to the experimenters until the completion of the experiment.

Initially, an equal number of subjects, in matched pairs, were assigned to each of the two game groups (FT and EOT). However, because of medical problems and personal difficulties unrelated to flight performance 7 subjects in the FT group and 6 in the EOT group were washed out during the three months of basic military training. They were therefore unable to complete the second segment of game training. The dropout from two groups was roughly similar in terms of matching scores and did not change the initial balance between groups. Cadets assigned to the control group with no game experience were selected and matched in their ability and light aircraft scores to the remaining subjects in each of the two game groups, only at the point of entering the eight transition flights. Hence, the matching between the three groups was not impaired by the washout during basic military training. This conclusion was confirmed by the results of the analysis of variance conducted on two matching scores. For the index of the selection test battery, $F(2,53) = 1.19, p > 0.31$. For the light aircraft check-ride scores, $F(1,53) = 0.86, p > 0.43$.

To reduce a possible confounding influence of cadets' basic abilities on the judgment of flight instructors during the eight test flights, cadets were assigned to instructors in ability-matched triads, one from every experimental group. Because of scheduling problems and the initial washout in the game groups, several triads included only a single subject from a game training group and two from the control group. This procedure added several subjects to the control group, without changing the matching of the groups in terms of prior ability, as evident from the statistical analysis reported above. Flight instructors had no knowledge of the prior experience and group assignment of their trainees.

RESULTS

The presentation of results focuses on the transfer of training from SF to flight performance. However, before these results are presented, we describe briefly the learning curves and final game scores of the full-training (FT) and emphasis-only-training (EOT) groups.

Space Fortress Performance with Full and Emphasis-Only Training

Subjects in the FT group obtained significantly higher final game scores on all measures of game performance, compared with the EOT group. Figure 2 depicts the learning curves of the two groups as reflected in the overall point score. Although both groups progressed monotonically with training, the FT group showed a clear advantage. Note also that neither group suffered a sizable performance loss attributable to the three-month break between Sessions 5 and 6.

The average 3-min game scores for the two groups at Session 10, the last session, on the three major performance measures are presented in Table 1. All differences were significant at $p < 0.0001$ in analyses of variance.

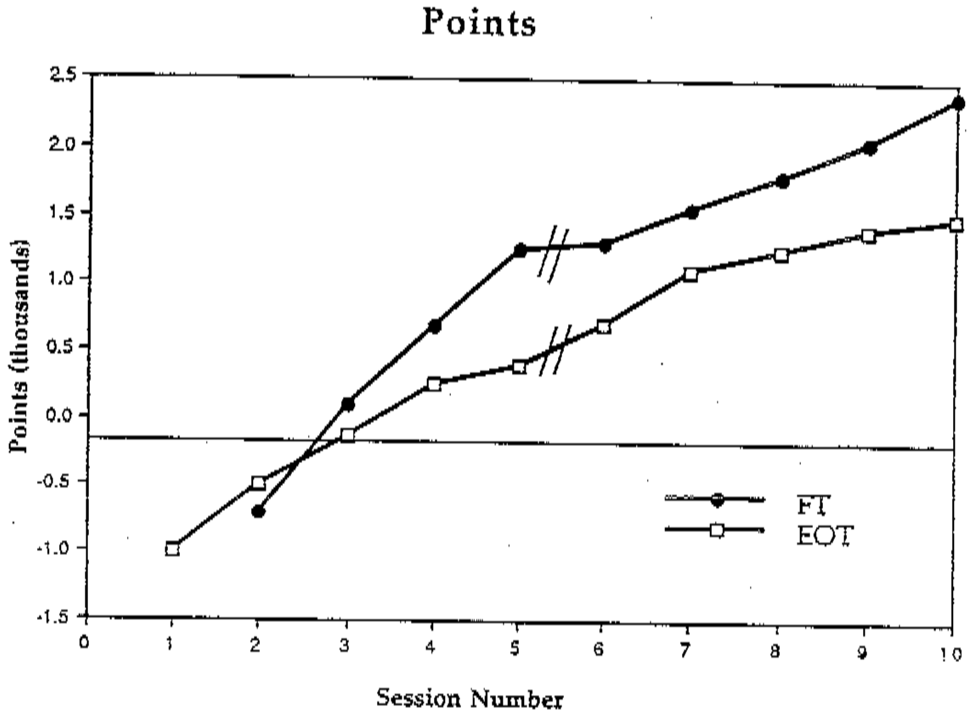

Figure 2. *Learning curves for the FT and EOT groups over 10 1-h training sessions (average general point scores for 3-min trials).*

The respective F ratios (1,38) were 19.89, 31.41, and 25.38.

Flight Performance

Altogether there were 33 flight scores, including the 6 standard scores and 27 scores obtained from the three special rating forms for individual maneuvers. Recall that the special forms were experimental in nature and had not been used before.

Differences between Game Groups in Flight Performance

Despite the large differences in the final game scores, the FT and EOT groups did not differ in subsequent flight performance. Both groups obtained similar average scores on the six standard flight measures and on the scores of the three special rating forms. The separate MANOVAs that were conducted on the six standard scores and the special forms of Flight 4 (10 scores) and Flight 7 (17 scores) confirmed this observation. None of the three analyses was statistically significant ($F < 1, p > 0.50$ in all cases). It was hence concluded that the two groups did not differ in their actual flight performance. Their data were combined for comparison with the control group of matched subjects who did not receive game training. The resultant combined group is hereafter named the *game* group (FT + EOT), and the other is labeled the *no-game* group.

TABLE 1

Average Game Scores for the FT and EOT Groups on the Tenth Training Session

Scores	*FT*	*EOT*
Total points	2251.70	1350.80
Fort destructions	12.80	6.30
Mines handling	1215.40	615.50

Influence of Game Training on Flight Performance

The combined game group had a total of 33 subjects, and the no-game group included 25 subjects. To get a preliminary, first-cut impression on the overall pattern of the results, we conducted a MANOVA on all 33 scores, comparing the flight performance of the two groups. We also examined the frequency distribution of the direction of differences between groups in their flight scores. These analyses were followed by separate MANOVAs of the six standard scores and the Flight 4 and Flight 7 measures.

The overall MANOVA showed that the game group was significantly better in its flight performance, $F(33,18) = 2.18$, $p < 0.05$. Five single flight scores were shown to significantly differ between the groups. Two additional scores had marginal significance. Table 2 lists the seven scores, their average values, and the outcomes of the statistical analysis.

Three of the five significant scores listed in Table 2 belong to the group of six standard flight measures. Each one of them represents a summary score over several flights (summary of Flights 1–4, summary of Flights 5–8, and the overall success predictor given after Flight 8). The statistical significance of the observed differences on these three was reconfirmed by the separate, and statistically more meaningful, MANOVA conducted on only the six scores. The overall effect was significant, $F(5,47) = 2.44$, $p < 0.05$, and the same three scores were significant contributors with somewhat enhanced significance levels, as compared with those presented in Table 2: summary of Flights 1–4, $F = 6.58$, $p < 0.014$; summary of Flights 5–8, $F = 5.73$, $p < 0.02$; success predictor for Flights 5–8, $F = 4.33$, $p < 0.04$; *d.f.* 1,52 in all cases.

Of the 7 scores presented in Table 2, 6 belong to Flights 7 and 8. This general pattern of an increased flight advantage of the game group in the more advanced flights can be clearly observed in Figures 3a–3d, which depict the differences between the two groups on the 33 flight scores. To simplify the presentation, the average difference on each score is displayed as the net difference in favor of one group or the other (recall that all actual scores were on a scale of 4–10).

Of the 33 flight scores, the game group was higher on 25 and the no-game group was higher on only 8 ($\chi^2 = 8.75$, $p < 0.01$). However, the distribution of the 8 reversals was such that 6 belonged to Flight 4 and only 2 to Flights 7 and 8. This is from the total of 13 scores available for Flight 4 and the 20 scores

TABLE 2

Comparative Averages and Significance Levels for Flight Scores, Discriminating Significantly between the Game and No-Game Groups in the General MANOVA ($N = 52$)

Variables	*Game*	*No Game*	*F*	*df*	*p<*
Flight 7 45 deg					
a. Looking into turn	6.65	6.23	4.46	1,50	0.04
Flight 7 Depart from area					
b. Time to prepare departure	6.97	6.42	7.43	1,50	0.001
Standard scores					
c. Summary Flights 1–4	6.93	6.61	6.28	1,50	0.016
d. Summary Flights 5–8	6.70	6.38	5.80	1,50	0.02
e. Success Predictors 5–8	6.60	6.09	4.07	1,50	0.05
Flight 7 45 deg					
f. General score	6.62	6.52	2.83	1,50	0.099
Flight 7 Depart from area					
g. General score	6.80	6.43	3.02	1,50	0.088

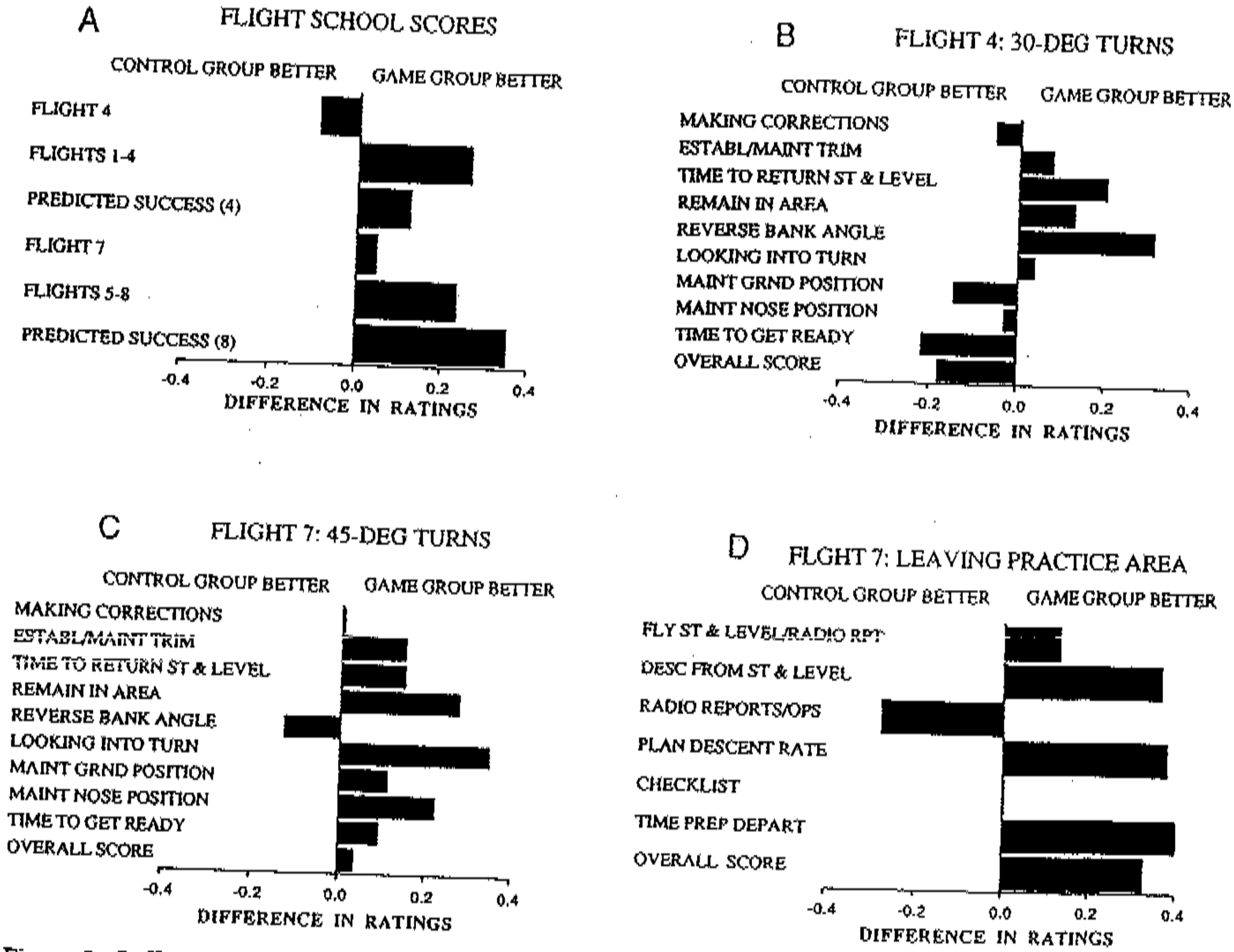

Figure 3. *Differences between the game and no-game groups on the 33 flight measures; each bar graph depicts the magnitude and the direction of difference on one particular measure.*

of Flights 7 and 8. Although it can be argued that the 33 scores were not completely independent from one another in the strict statistical sense of the term, it should be remembered that each of them, including the summary scores, represented an individual instance of evaluation. In addition, there were 12 flight instructors, each grading only three cadets, and every cadet was graded by 2 different instructors. Hence, the trends are revealing. The same pattern of results emerged also in the separate MANOVAs conducted on the special forms obtained at each flight. The MANOVA for Flight 4 was not significant, $F(10,45)$, $p < 1.0$, whereas the analysis of the 17 scores of Flight 7 showed a marginally significant overall effect, $F(17,38) = 1.7$, $p < 0.08$. The four individual scores of this flight that emerged in the general MANOVA (Table 2) reappeared in this analysis with similar significance levels.

Canonical regression and analysis of variance. A combined index of flight performance comparing the game and no-game groups was computed using only the five flight scores that were found to distinguish significantly between the two groups in the separate and overall MANOVAs. The canonical regression analysis method provided the best regression equation to predict group membership based on the joint contribution of the five scores. It also computed the significance of the difference between groups in terms of the resultant joint index scores. Equation 1 presents the obtained regression formula and the outcomes of the analysis of variance.

$$\begin{aligned} CFS = {} & 0.43(a) + 0.81(b) + 1.82(c) \\ & + 0.16(d) + 0.09(e) \end{aligned}$$

$$F(1,48) = 11.38, p < 0.0015, \quad (1)$$

where CFS is the combined flight score (variables *a–e* are listed in Table 2).

The average combined flight score for the game group, based on Equation 1, was 17.12 (*SD* 1.01), whereas the combined score for the no-game group was only 16.13 (*SD* 1.03). Figure 4 presents the frequency distribution of the two groups on the combined flight index. The advantage of the game group is apparent. About one-third of the subjects in the game group were included in the highest score category, whereas only 3.40% were in the lowest category. In comparison, none of the no-game subjects was included in the high score category, whereas 28.6% were in the lowest category.

Distribution of the success predictor score. Our final analysis compared the distribution of the game and no-game groups on the Flight 8 success predictor score. Instructors evaluated on a scale of 4–10 the overall likelihood that a candidate would complete flight training. Flight school statistics show the following correspondence between these scores and actual percentage who graduated: 9% for a score of 5, 22% for 6, 56% for 7, and 75% for 8. Table 3 presents the distributions for the game and no-game groups.

About 66% of the candidates in the game group were included in the 7–8 score category, while only 33% of the no-game group were included in this category. Given the correlation of this score with actual success probability, the implication is that, on the average, the game group increased its probability of graduation by 30%.

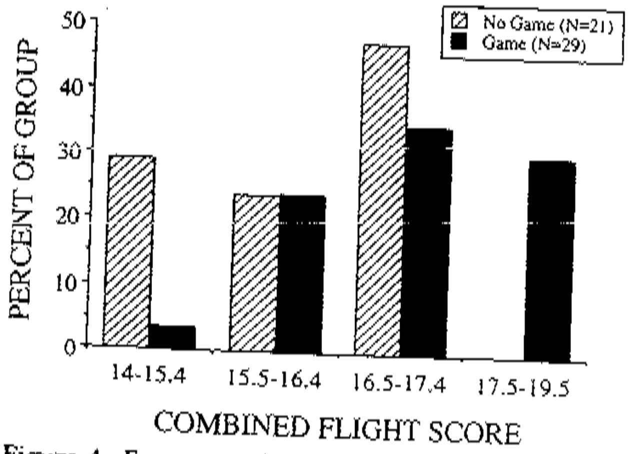

Figure 4. *Frequency distributions for the game and no-game groups on the combined index of flight performance; higher scores imply better flight performance.*

DISCUSSION

Flight performance showed a clear advantage of the game group over the no-game group, which included cadets with matched ability who did not receive SF training. The best overall representation of this advantage is the frequency distribution of the two groups on the combined flight performance index (Figure 4), whereas the largest differences on single scores were revealed on the standard (Flights 5–8) summary and success-predictor scores. These are also the main scores that are used by the school for evaluating students' performance at this stage. Similar differences, with somewhat reduced magnitude, were also found on the special

TABLE 3

Distribution of the Game and No-Game Groups on the Success Predictor Score, Flights 5–8 (N = 58)

	Score Categories		
Groups	*5–6*	*7–8*	*Totals*
Game	12	21	33
No game	16	9	25
Totals	28	30	58

scores of the two maneuvers in Flight 7, but not on those obtained during Flight 4.

Several factors may have contributed to the greater sensitivity of standard scores to the effect of game training. Each of the three significant scores was an integrative estimate based on four flights. Flight school statistics have repeatedly shown that these scores have higher predictive validity and better reliability than do single-flight ratings. It is possible that judgments of ability and performance levels are more valid and stable if based on a group of flights. The only summary score that did not show a significant difference between the game and the no-game group was the success predictor given after the first four flights. This finding is consistent with the overall pattern showing lack of differences between the two groups in early flights and their emergence in advanced flights. The weaker effects revealed on the special rating forms are not surprising. These forms were exploratory in the selection of maneuvers, the rated behaviors, and their mode of administration. Given these constraints, it is significant that the overall trend was consistent with the standard scores, including the distinction between early and advanced flight maneuvers. It points to a clear potential for developing an alternative rating system with more research.

Further support for the contribution of SF training to flight performance came from a follow-up study that was carried out by the school psychologist after the conclusion of the formal study. Although the exact numbers and percentages are classified, we were informed by the school command that at graduation (18 months later), the percentage of graduates from the game group was twice that of the no-game group. A replication study with some modifications was conducted at the U.S. Army helicopter flight school (Hart and Battiste, 1992). It obtained similar results, including the lack of difference in early flights and their emergence in advanced flights. This study included only one SF group, following the training schedule of the EOT group in the present study. It also included an additional control group that practiced for 10 h on an off-the-shelf Apache Strike Force computer game. This practice showed no transfer to flight.

The positive transfer from SF training should be considered side by side with the seemingly conflicting finding that the full training and emphasis-only training groups, although markedly and significantly different in their final game achievements, did not differ in flight performance. If SF training augments flight, why were FT trainees not better than the EOT group?

One possibility to consider is the so-called Hawthorne effect (Homans, 1958)—namely, that the advantage shown by the game group did not stem from the game experience itself but from the fact that subjects in the game group behaved differently, were more motivated, or received special treatment by their flight instructors. Several factors render this possibility unlikely, however. The flight school is prestigious and very selective; washout rates are high. All cadets are volunteers, are highly motivated, and make their best effort to compete and complete training. Participation in the experiment was a small and unimportant fraction of their daily activities; they complied with the experiment but did not welcome it with enthusiasm. SF training sessions were regarded as an extra load rather than a privileged engagement. Subjects were called out from their regular ground school classes and on many occasions expressed concern about the lost material.

The official cover story, which introduced the experiment as a study of individual differences, was another factor that reduced the likelihood of a Hawthorne effect. Also, flight training and check rides were conducted in a different squadron, at a different location, and with different personnel than was SF training. The original groups were distributed among many instructor pilots and were part of a much larger class in which all cadets underwent the same flight evaluation process. Recall also that within the flight sequence itself, instructors were changed after four flights. The school report on the persistence of the effects over the long duration of the training program, with its intensive pace and change of environments, also negates a simple Hawthorne effect interpretation. Finally, the lack of transfer reported by Hart and Battiste (1992) from 10 h of training on the Apache Strike Force game clearly overrules a Hawthorne interpretation. Also refuted are suggestions that practice on any computer game, even one with high face validity, would necessarily have a similar impact on flight performance.

The key to the lack of difference in flight performance between FT and EOT groups, despite the large difference in their SF scores, should lie in the main features of the training that distinguished the two groups. To recapitulate, both groups received the emphasis-change training elements, following Gopher et al. (1989). In addition, the FT group was given part tasks, verbal tips, and individual coaching, as recommended by Fredriksen and White (1989). A possible interpretation of the pattern of results obtained is that emphasis-change elements promoted the development of skill components that could be transferred and generalized to flight, whereas part tasks and verbal tips contributed only to skill elements that were exclusively relevant to SF performance. This suggestion is supported by the results of two laboratory studies that were conducted in parallel with the present work.

Fabiani et al. (1989) conducted an experiment to compare the hierarchical part-task (Fredriksen and White, 1989) and the emphasis-change (Gopher et al., 1989) methods. Performance levels of matched groups, each trained with one method, were compared at the end of 10 h of training on SF and in five following transfer sessions during which SF was performed concurrently with a battery of secondary tasks. At the end of training, both methods led to better performance than was seen in a control group that practiced without instruction. However, hierarchical part-task subjects obtained better scores than the emphasis-change group. In contrast, during transfer sessions the differences reversed. Emphasis-change subjects were better able to adjust to the dual-task demands and maintain performance in the new conditions of increased load, whereas performance levels of subjects in the hierarchical part-task group deteriorated considerably. The same pattern of results repeated in a study conducted at the Technion-Israel Institute of Technology (Bareket, 1990). This study showed a clear advantage in transfer for emphasis-change training over the hierarchical part-task approach, in particular when difficult new elements were added or when the operational rules of already-trained elements were radically changed (e.g., when ship control dynamics were changed from velocity to acceleration).

The conclusions from both studies are that part tasks and verbal tips that are based on the analysis of expert behavior appear to focus the trainee's attention on elements and modes of behavior that are specific to the performance of the trained task. In contrast, emphasis change, which is practiced on the task as a whole, leads to the development of more general skills and response strategies that are less dependent on the specific peculiarities of the task. Such strategies maintain their relevance and are easier to generalize when variables are changed or new tasks with a similar context are encountered. Another difference between the methods is that under the part-task procedure, trainees are not exposed to the full load of the task until they pass a sequence of simplified task configurations. Under the emphasis-change method, subjects practice and progress while continuously experiencing the full load of the task and learning to cope with it.

In light of these differences between the methods, the apparent contradiction in the pattern of results obtained in the present flight school study can be reconciled. The FT and EOT groups had about equal exposure to the whole task with emphasis-change information. The distinguishing features were the limited practice on part tasks and the ongoing coaching and verbal tips given to the FT group. These elements enhanced game performance in modes that were too specific to generalize to flight. Hence the two groups were equally prepared in terms of transfer to the actual flight situation, even though their game performance differed. In this regard, it seems reasonable to consider the results of the present experiment as another replication of the pattern of differences between the two training approaches that was observed in the laboratory studies.

What was acquired during game practice under the emphasis-change manipulation that was so useful and relevant to flight training? We propose that these skills are related to the experience of coping with the attention demands and high load of the game. In other words, 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. The logic and concepts that were presented in the introduction to describe the skill of attention control, and that have also led to the development of the emphasis-change method, may provide the framework for delineating a general category of skills that are relevant both to the game and to the flight situation. To recapitulate, the key construct in describing the skill of attention control is *attention strategy*, defined as a vector of the attention weights assigned to each of the covarying elements of a complex task (Gopher, 1992). The efficiency of attention control depends on the number, diversity, and availability of strategies, as well as on the quality of the match between the individual's unique capabilities and task demands. Training should lead performers to develop, explore, select, and stabilize attention strategies (Gopher et al., 1989).

To evaluate the foregoing suggestions, let us review again the initial claim for the contextual relevance of SF training to flight performance. We discussed the similarities between the two environments in terms of contemporary models of the human processor (see Card and Moran, 1986; Wickens, 1992), highlighting common features on such dimensions as continuous and discrete motor control, visual scanning and monitoring, short- and long-term memory requirements, time pressure, and attention load. It was already concluded that the equal levels of transfer to flight that were gained by full- and part-task training indicate that the effects cannot be interpreted as a simple product of enhanced performance on these components. Another possibility is to view the components as elements of the demand profile of the task, in the framework of attention theory and resource limitations. Most relevant are those models proposed by researchers who advocate a multiple resource approach (e.g., Gopher, Brichner, and Navon, 1982; Wickens, 1984) or processing stage analysis (e.g., Sanders, 1980). These approaches distinguish between the resource demands of perceptual analysis, central processing, and response selection that can be easily linked with the components of our task analysis. Development of attention control strategies implies that trainees had to cope with, allocate to, and balance their resource investments among variants of the demand profile of these components.

Within this framework, there are at least two possible explanations for the transfer of training from the SF game to flight. One explanation is that the functional and demand characteristics of the two situations were sufficiently similar that attention strategies that had been developed while practicing the game were also efficient in actual flight. Another possible explanation is that when practicing the SF game, trainees learned the value of exploring alternative response modes and developing attention strategies in a context that bears general relevance to the flight situation. This experience motivated them to explore similar modes of behavior when confronting the demands of flight. The first explanation thus assumes a stronger and more direct transfer of acquired attention strategies, whereas according to the second explanation, transfer was less direct and limited to the use of similar approaches and methods in coping with the demands of a new environment. In the data of the present experiment, we do not have evidence to favor one explanation over the other. Although both are based on the conceptual framework of attention control, they differ in the level and nature of the transferred experience. To get better closure on the determinants of transfer, we should conduct experiments varying the context similarity between the training and transfer situations. Some experiments along these lines are being conducted at Brooks Air Force Base as part of the Schooling and Training Approaches Measurement Project (STAMP; Shebilske and Regian, 1992; Shebilske, Regian, Arthur, and Jordan, 1992).

An interpretation of the transfer effects based on attention control concepts is consistent not only with the equal success of the FT and EOT groups in their flight performance but also with the fact that the main effects of game training were observed in the more progressive and advanced flights, when concurrent attention and coordination demands were much elevated. These results were replicated in the study conducted by Hart and Battiste (1992). Despite the consistent and converging evidence in support of an attention-control interpretation, we regard our evidence at this stage as suggestive and inconclusive. More control conditions and direct manipulations of attention control are required to substantiate the claim that the improved attention component is to be credited for the transfer effects from SF to actual flight. Such research is being conducted both in the United States and in Israel. Because the Israeli Air Force has now incorporated the game into its flight training program, we may be allowed access to a large database, which will allow finer examinations and more conclusive analyses.

Our concluding comments are on the topic of fidelity. This subject has been the backbone of all the work reported here. The study as a whole serves to demonstrate a different approach to fidelity. From this vantage point, the question of whether improved attention control was indeed the true source of the observed transfer effects is secondary to proving that such transfer did occur, and it has already been replicated in a second, independent study. A detailed discussion of the fidelity topic is beyond the scope of this paper. We consider here only two points that emerge from the results of the present study. One involves the conceptual framework that has been utilized to claim context similarity between the game and the flight situation. The other reiterates the implications of the differential efficiency of the two training schedules (FT and EOT) during the stages of training and transfer.

Our results suggest that a conceptual

framework that is deeply rooted in human performance theory and research was sufficiently powerful to provide a context within which an operational skill trainer can be developed. This framework points the way to the development of alternative approaches to task fidelity analysis. Unlike the physical fidelity approach, which relies on the surface properties of the situation, this approach is based on the deep structure of the task in terms of its processing, response, and resource demands. In addition, the differences between the FT and EOT groups underscore the importance of evaluating the influence of the selected training schedule, beyond the physical context and the properties of the trainer. Although previous work has addressed the harmful consequences of gaps between simulators and operational environments on these latter elements, training strategies were not considered (Lintern, 1991; Roscoe, 1980). Our results suggest that within the same physical configuration (the SF game), one method may improve performance by focusing on task elements that are too specific to generalize to the operational environment, whereas another concentrates on aspects that are equally relevant to both situations. Thus one can argue that the first method reduces fidelity and the other increases it. The term *fidelity* is used here in the sense that one framework acts to increase the gap between two situations, whereas the other brings them closer together by carefully constraining itself to dimensions and aspects that are relevant to both.

These results show the delicate balance between the aims of providing a specific, simple, well-structured, and supportive learning environment and the task of focusing exclusively on elements that may bridge the gap between the training and the operational situations, thus providing better fidelity. At this stage we can offer neither a model nor systematic criteria to indicate when the scale has been tipped to one side or the other. However, we should keep this concern in mind. Many training situations present performers with complex and demanding tasks in which one may be tempted to propose specific solutions to obtain fast improvement in a trainee's performance. Alas, the long-term effect of such leads may be costly.

ACKNOWLEDGMENTS

The present work was supported by a grant from NASA-Ames Research Center, Aerospace Human Factors Division. The scientific monitor of this grant was Sandra Hart. The drawings used for Figures 1 and 3 were adopted from briefing material prepared by Sandra Hart. We thank Ayala Cohen for her statistical advice and would like to acknowledge the constructive discussions and comments made by Chris Wickens, Gordon Logan, Gaven Lintern, Sandra Hart, and two anonymous reviewers on earlier versions of this manuscript. Last but not least, we are indebted to the cooperation and help of the staff and students at the Israeli Air Force flight school, in particular to the invaluable contributions of Zeev Raz, Ami Manor, Doron Karni, Eyal Goshen, and Udi Zohar of the flight school staff. Without their continuous support, intellectual curiosity, and daily help, the conduct of this study would not be possible.

REFERENCES

Bareket, T. (1990). *The influence of training methods and verbal coaching on the training and transfer of a complex skill.* Unpublished master's thesis, Technion-Israel Institute of Technology, Haifa, Israel.

Baudhuin, E. S. (1987). The design of industrial and flight simulators. In S. M. Cormier and J. D. Hagman (Eds.), *Transfer of learning.* San Diego, CA: Academic.

Card, S. K., and Moran, T. P. (1986). The model human processor. In K. B. Boff, L. Kaufman, and J. M. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2, pp. 1–35). New York: Wiley.

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

Donchin, E., Fabiani, M., and Sanders, A. (1989). The learning strategies program: An examination of the strategies in skill acquisition. *Acta Psychologica, 71,* 1–311.

Fabiani, M., Buckley, J., Gratton, G., Coles, M., Donchin, E., and Logie, R. (1989). The training of complex task performance. *Acta Psychologica, 71,* 259–300.

Flexman, R. E., and Stark, E. A. (1987). Training simulators. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 1012–1038). New York: Wiley.

Foss, M. A., Fabiani, M., Mane, A., and Donchin, E. (1989). Unsupervised practice: The performance of the control group. *Acta Psychologica, 71,* 23–52.

Fredriksen, J. R., and White, B. Y. (1989). An approach to training based upon principle task decomposition. *Acta Psychologica, 71,* 89–146.

Gopher, D. (1982). A selective attention test as a predictor of success in flight training. *Human Factors, 24,* 173–183.

Gopher, D. (1992). The skill of attention control: Acquisition and execution of attention strategies. In S. Kornblum and D. Meyer (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience*. Cambridge, MA: MIT Press.

Gopher, D., Brichner, M., Navon, D. (1982). Different difficulty manipulations interact differently with task performance: Evidence for multiple resources. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 146–157.

Gopher, D., and Sanders, A. F. (1984). S-Oh-R: Oh stages! Oh resources! In W. Printz and A. F. Sanders (Eds.), *Cognition and motor processes*. Berlin: Springer-Verlag.

Gopher, D., Weil, M., Bareket, T., and Caspi, S. (1988). Fidelity of task structure as a guiding principle in the development of skill trainers based upon complex computer games. In *Proceedings of the Human Factors Society 32nd Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

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

Hart, S. G., and Battiste, V. (1992). Flight test of a video game trainer. In *Proceedings of the Human Factors Society 36th Annual Meeting* (pp. 1291–1295). Santa Monica, CA: Human Factors and Ergonomics Society.

Homans, G. C. (1958). Group factors in worker productivity. In E. Maccoby, T. Newcombs, and E. Hartley (Eds.), *Readings in social psychology*. New York: Holt.

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

Mane, A., and Donchin, E. (1989). The Space Fortress game. *Acta Psychologica, 71*, 17–22.

North, R. A., and Gopher, D. (1976). Measures of attention as predictors of flight performance. *Human Factors, 18*, 1–14.

O'Hare, D., and Roscoe, S. (1990). *Flightdeck performance*. Ames, IA: Iowa State University Press.

Roscoe, N. S. (1980). *Aviation psychology*. Ames, IA: Iowa State University Press.

Sanders, A. F. (1980). Stage analysis of reaction processes.In G. E. Stelmach and J. Requin (Eds.), *Tutorials in motor behavior*. Amsterdam: North-Holland.

Shebilske, W. L., and Regian, J. W. (1992). Videogame training and investigating complex skills. In *Proceedings of the Human Factors Society 36th Annual Meeting* (pp. 1296–1300). Santa Monica, CA: Human Factors and Ergonomics Society.

Shebilske, W. L., Regian, J. W., Arthur, W., Jr., and Jordan, J. A. (1992). A dyadic protocol for training complex skills. *Human Factors, 34*, 369–374.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman and D. R. Davis (Eds.), *Varieties of attention*. Orlando, FL: Academic.

Wickens, C. D. (1992). *Engineering psychology* (2nd ed.). New York: HarperCollins.

Wiener, E. L., and Nagel, D. C. (Eds.). (1988). *Human factors in aviation*. San Diego, CA: Academic.

Date received: March 31, 1992
Date accepted: October 1, 1993