EVALUATION OF THE SYNTHETIC FLIGHT TRAINING SYSTEM (DEVICE 2B24) FOR MAINTAINING IFR PROFICIENCY AMONG EXPERIENCED PILOTS

D. O. WEITZMAN, M. FINEBERG, H. OZKAPTAN
US Army Research Institute for the Behavioral and Social Sciences
and
CW4 G. L. COMPTON, US Army
Fort Campbell, Kentucky

BACKGROUND

Since the early days of flying, flight simulators have been in use and their value has been amply demonstrated. Flight simulators have evolved from the Link Trainer, which was widely used for pilot training in the Second World War, into precisely engineered devices capable of accurately computing the aerodynamic responses of an airplane to control inputs, and of reproducing realistic cockpit instrument indications for all flight situations. It was realized, as Adams (1957) points out, that flight simulators have many advantages over the operational situation. First, the simulator provides its users with greater control over ambient conditions. Whereas the "real" world is subject to unpredictable variations, a simulator can provide planned variation of various elements of the real situation with unessential elements in the real situation omitted. Second, the simulator can represent dangerous elements in flight more safely. Emergency procedures that would be too dangerous to teach in the air may be taught safely in ground-based simulators. Third, a major advantage offered by simulators is their low operating costs in comparison with the costs of operating aircraft. For these reasons, simulators continue to play an important role in pilot training during initial acquisition, transition training, and for maintenance of established flying skills. The importance attached to simulators in meeting training goals is, of course, predicated on the assumption that training given in the simulator will transfer to the aircraft.

Although there has never been a serious challenge concerning the value of simulators to pilot training, the utility of simulator training for maintaining the proficiency of experienced pilots has not been adequately established. The various studies that have assessed the training effectiveness of a specific flight simulator for maintaining pilot proficiency among experienced pilots have sometimes produced contradictory results (Caro, 1971; Crook, 1965; American Airlines, Inc., 1969; TransWorld Airlines, 1969). For example, whereas Caro reports positive transfer from the simulator to the aircraft among experienced pilots, Crook reports less conclusive findings. Three somewhat different explanations for these discrepant findings have been suggested. First, the variance in flying skill among experienced pilots is often greater than among student pilots who bring little, if any, initial skill to the task (McGrath and Harris, 1971). In performance training it is well known that the beneficial effects of training varies as a function of experience or skill level (Briggs, 1957; Roscoe, 1971); that is, the same training simulator may exhibit different effectiveness functions for different levels of pilot experience or skill. To the extent that experienced pilots are not carefully matched in terms of flying skill, the significance of assessing what can be learned in the simulator is diminished. Second, most performance measures used were judgmental in nature and, thus, highly subjective evaluation instruments. Third, instructional techniques may vary widely depending upon the ability and attitudes of the instructor pilots. In considering the possibilities of such instructional differences, it is essential to determine the utility of simulator training independent of the instructor variable.

Currently, there is a need to determine the training value of Device 2B24, a high-fidelity simulator, for maintaining instrument flight (IFR) proficiency among experienced Army helicopter pilots in the UH-1H aircraft. This paper reports on the transfer effectiveness of Device 2B24 for use in maintaining IFR skills among experienced pilots as compared to inflight training. As will be seen below, this study extends previous studies by more adequate control of pilot and instructor variability and through more objective measures of pilot performance obtained both before and after training in Device 2B24.

METHOD

Pilots. The pilots participating in this study were 36 fully qualified, combat ready Army aviators. The sample was comprised of both warrant and commissioned officers, all of whom had between 400-600 hours of rotary wing flight experience. Two groups of aviators of 18 pilots each were formed on the basis of a composite score on three IFR performance measures. A written test on instrument procedures (I), a standardized IFR check ride in the UH-1H helicopter (H), and a standardized IFR check ride in Device 2B24 (S) were the three measures making up the composite score (P). The a priori scoring formula reads: $P = .5 I + S + H$. The three performance measures were administered to each of 51 experienced pilots assigned to the 101st Aviation Group, Ft. Campbell, Kentucky. The best 18 and the poorest 18 scores were used for group matching and for control of differences in initial ability. The two groups are referred to as "high" and "low," respectively.

Following the selection of each group, pilots from each group were quasi-randomly assigned to one of three IFR training modes under the following conditions:

a. Simulator training with Device 2B24.

b. Inflight training in the UH-1H.

c. A combination of simulator and inflight training divided equally.

A total of 12 pilots, 6 from each group, were assigned to each training mode. This procedure provided an opportunity not only to evaluate training modes but to evaluate them with respect to initial skill level among pilots. An analysis of variance failed to show reliable differences between pilots assigned to each training mode ($F(2,30) = .31$), but revealed a reliable difference between the pilot groups assigned within each mode ($F(1,30) = 17.91$, $p < .001$) as constituted. Thus, pilots appeared to be effectively matched among training modes while differentiated in terms of initial IFR skills.

Procedure. The pilots in each of the three training modes were given a total of 12 hours of IFR training during 9 months of testing in each of the three training modes--that is, in the SFTS, inflight, or combined SFTS and inflight training.* The 12 hours of training were divided into 4 hourly increments taken quarterly. At the end of each quarterly (3 month) period, pilots were required to take a standardized IFR check ride in the UH-1H, identical in all its essentials with the IFR check ride given before training.

Prior to testing, an objective inventory for pilot evaluation was developed at the Ft. Campbell facility. The decisions to include items were based upon desired terminal flight behavior for experienced Army pilots. Evaluated tasks were arranged into a standard flight sequence, including both objective and subjective items. The inventory evaluated performance on each of five quantitative variables with one to four items for each variable scored on an appropriate scale. Equal weights were given to each of the objective variables measured; namely: (1) flight planning, (2) departure and enroute procedures (e.g., navigation and tracking, radio reports), (3) holding and arrival procedures (e.g., missed approach, time, track, airspeed, etc.), (4) ATC procedures, and (5) aircraft control (e.g., airspeed, altitude, heading, etc.). All scored items were completed during flight. A score of 100 represented "errorless" performance.

To minimize the effects of instructor differences, six instructor pilots (IP) were given familiarization training on the use of the flight inventory before each check ride. The order in which IP's were assigned to provide check rides was varied systematically. In addition, knowledge of an individual pilot's ranking and group assignment was withheld from the IP's to prevent such information from influencing their judgment. Nevertheless, because of the relatively small pilot sample, some IP's were aware of a particular pilot's group assignment.

Instrument flight training began soon after the two groups of pilots were formed. Training sessions were 1 hour in length and four sessions completed a quarterly training period. At the end of each quarter of training, an instrument check ride in the UH-1H was given to each pilot using the flight inventory for scoring pilot performance. The present paper reports on some initial findings of the transfer effectiveness of Device 2B24 among experienced pilots.

RESULTS

Because of conditions not related to the experimental variables, two of our pilots assigned to the "low" group, one in the SFTS training mode and the other

*The experiment was originally designed to provide 16 hours of training divided into four quarterly increments. We were unable to complete the last quarterly test period due to the temporary transfer of the pilots to a new duty station.

in the combined training mode were unable to complete the experiment during the third quarter testing period. The third quarter results and the statistical analyses are based then on an unequal number of pilots among blocks.

The main results deal with the differences in check-ride performance following training in the three training modes and the percent change in performance representing the differences between check-ride performance obtained before and after training. The curves in Figure 1 compare mean inflight performance for the three training modes over the three test periods with baseline performance plotted separately. The curves show that over test periods the changes due to training were not the same for each group, although the relative level of performance for the different training modes persisted throughout training. That

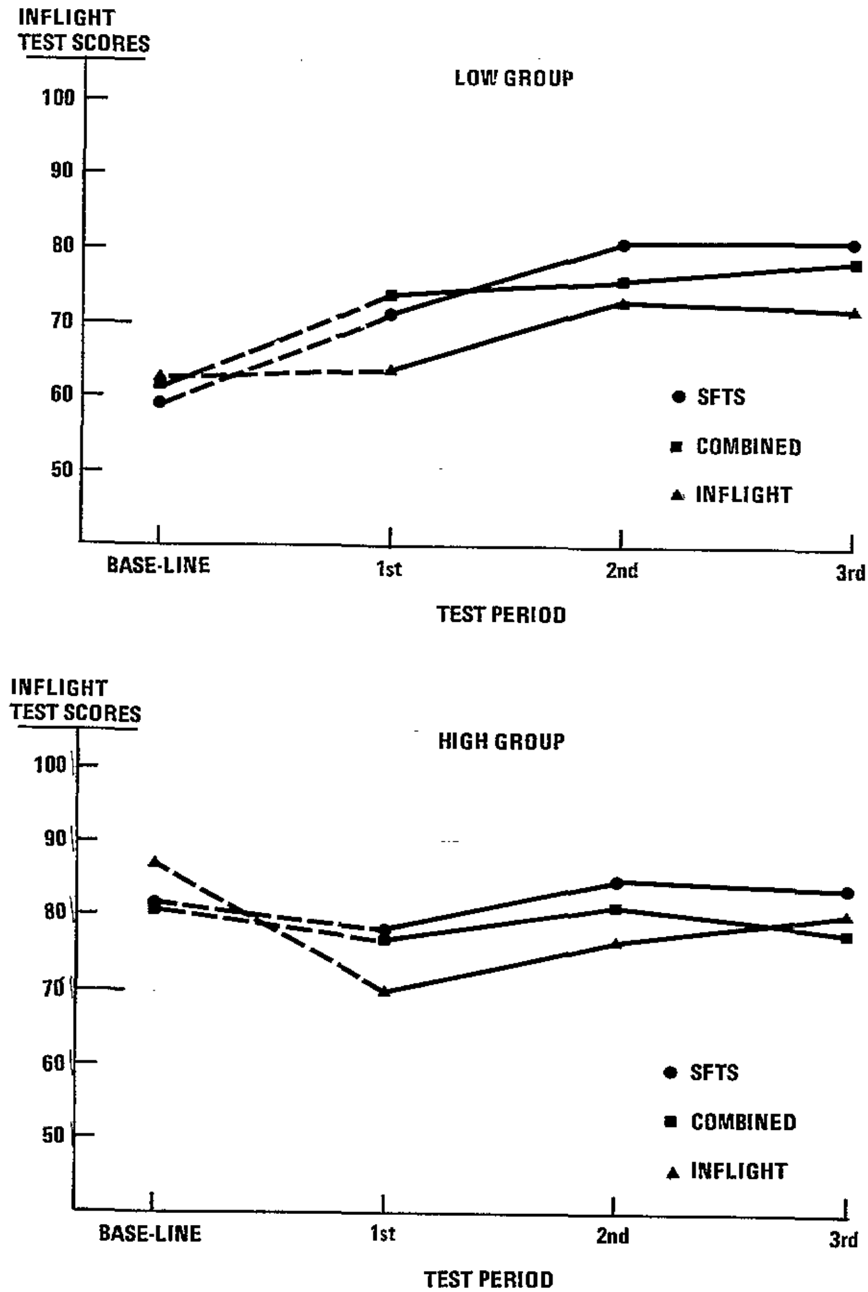

Figure 1. Learning Curves of Check-ride Performance for the Three Training Modes and Two Pilot Groups.

is, while the three curves for the "low" group showed marked learning effects over test periods, the"high" group gave very little evidence of learning. Both groups however, showed that check-ride performance was, in general, best for Device 2B24 training and poorest for inflight training. Analysis of variance for repeated measures indicate that variation in performance due to test periods was significant, $F(2,46) = 17.9$, $p < .01$, for the "low" group, although not a significant source of variance for the "high" group. It is also important to note that the analysis revealed that the differences between pilots assigned to groups were no longer statistically significant following training. Further, although Device 2B24 training appeared to result in better performance throughout the training period, F tests of the training mode means for each group were not significant. However, Duncan multiple range tests (Duncan, 1955) were carried out on the training mode means of the last training period to determine which, if any, of the differences *between the means are significant and* which are not. Results of these tests are summarized in Table 1. Common lines underlying two or more means indicate that these means are not significantly different from one another. The results on means for the third training period indicate that for the low group the mean differences between inflight training and Device 2B24 training were significantly different. All other comparisons were not significant. Thus, the value of Device 2B24 is particularly apparent for the low group.

Table 1

RESULTS OF RANGE TEST OF THE DIFFERENCES BETWEEN PERFORMANCE SCORE MEANS FOR TRAINING MODE AND PILOT GROUP

Group	SFTS	Combined	Inflight
High	84	78	81
Low	81	79	72

Note: Significant at the .05 level. Common lines underlying two or more means indicate that these means are not significantly different from one another.

Percent change scores for the three *training modes are given by the curves of* Figure 2. These curves plot the difference, in percent, between check-ride performance before and after training in each mode and each test period. The curves show again that the level of change in check-ride performance were quite distinct, with the SFTS mode producing the highest degree of positive transfer for both groups. The

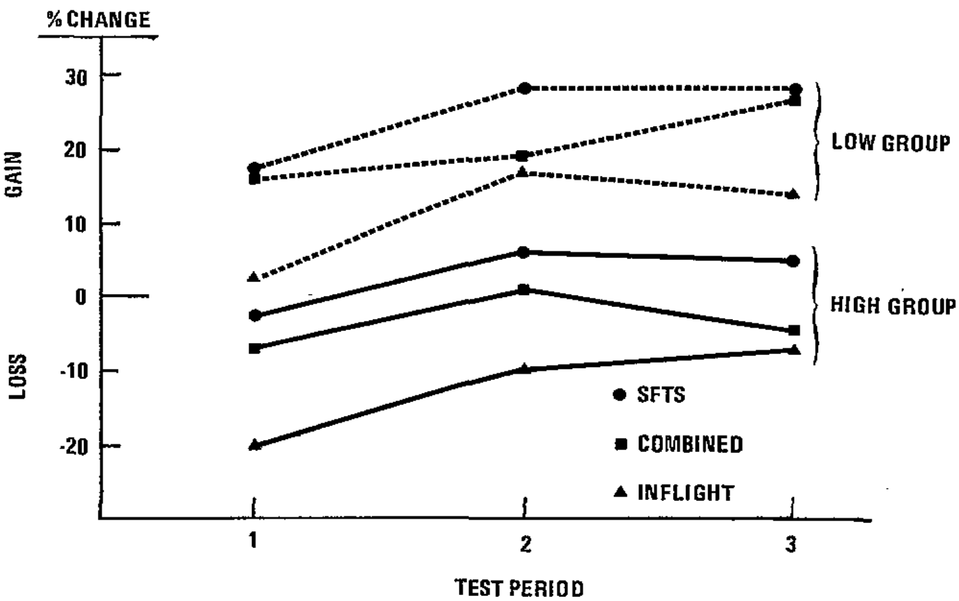

Figure 2. Change Scores in Check-ride Performance for the Two Pilot Groups as a Function of Training Mode and Test Periods.

plot of these test period scores show that most of the performance changes that occurred took place by the second test period. It may be noted, for example, that all of the transfer that occurred for Device 2B24 training occurred by the second test period; that is, after 8 hours of training in Device 2B24. Thus, the loss of the fourth test period did not appear to be crucial to the present interpretation of our results.

A three-way analysis of variance, encompassing groups, training modes and test periods, was carried out on the change score data. Results of the analysis related to the principal variables, and their first order interactions indicate that only the variations due to groups and test periods were statistically significant at the .025 and .01 levels, respectively.

Duncan multiple-range tests were again carried out on the training mode change scores for the third measurement period. The results are shown in Table 2.

Table 2

RESULTS OF RANGE TESTS OF THE DIFFERENCES BETWEEN PERCENT CHANGE SCORE MEANS FOR TRAINING MODE AND PILOT GROUP

Group	SFTS	Combined	Inflight
High	5	-4	-7
Low	28	27	14

Note: Significant at the .05 level. Common lines underlying two or more means indicate that these means are not significantly different from one another.

Negative values in Table 2 indicate a percentage decrement in performance associated with training. This decrement, we believe, is in accord with normal variation in performance that may be expected with a pilot group already at a high level of IFR proficiency prior to training. The results on change scores indicate that Device 2B24 training and inflight training show the greatest number of significant differences consistent with the evidence of beneficial effects for SFTS training when compared to inflight training.

DISCUSSION

The objective of this research was to determine the extent to which IFR skills acquired during simulator training in Device 2B24, or inflight, or in combination, maintained or even enhanced subsequent performance in the UH-1H among experienced pilots. In this research the transfer effectiveness of the simulator is measured by determining the extent to which performance of Device 2B24 trained groups are equal to or exceed the performance of the inflight trained group on the criterion check ride.

The results of this study indicate substantial positive transfer of training from Device 2B24 to the UH-1H helicopter as evidenced by the fact that simulator training produced as good or better overall performance as inflight training. Thus, Device 2B24 training appeared to be as much a factor in maintaining IFR skills among experienced Army pilots as inflight training for the periods of time tested. These findings were characteristic of both groups of pilots. Further, the results showed that the beneficial effects of training were more often observed among pilots who demonstrated the least IFR skill level as measured by our criterion check ride. The consequences of our training procedures were that the initial differences in IFR performance between our two groups of pilots were no longer significant.

The fact that IFR training in Device 2B24 transfers to the UH-1H is not surprising since it is a high-fidelity simulator of the UH-1H helicopter. These results are also consistent with the record of commercial aviation which has shown that simulators can provide an effective means of providing training for highly experienced pilots.

It should be emphasized, however, that the effectiveness of any training device depends upon how it is used; it is influenced by all the well-known variables concerning conditions favorable and unfavorable to learning. Prophet (1966) has pointed out that a simulator is a training tool and, as such, is going to produce results no better than the quality of the training program of which it is a part. The evidence of transfer in the present study may well be attributed to the training program rather than simulator design alone. It is possible, for example, that if inflight training was conducted differently--that is, with a different training program--some evidence of additional benefits might be found.

The present findings complement past results which demonstrate the operational utility of simulator training (Caro, 1972; Woodruff and Smith, 1974). In particular, this preliminary analysis of the data indicate that much of the instrument training now conducted in rotary wing aircraft can be conducted more efficiently on the ground. Although with Device 2B24 pilots may require approximately the same number of hours training as in the aircraft, the total time required is actually reduced by reason of the greater availability of the simulator and the elimination of much of the aircraft preparation time. The safety associated with elimination of much of the inflight training time, the release of aircraft for operational flights, and decreased cost of operating the simulator rather than an aircraft for training, all add up to a highly positive evaluation of Device 2B24 for the maintenance of IFR pilot proficiency. A reasonable conclusion from this study is that some evidence was found to support the assumption that simulator training administered in Device 2B24 can maintain, and perhaps improve, subsequent pilot performance in a tactical instrument situation. Whether or not the attitudes of the pilots who served as our subjects are in accord with this interpretation, is presently under review.

ACKNOWLEDGMENTS

The research presented in this report was requested and supported by Major Matthew Kambrod, Deputy Chief of Staff for Operations, Aviation Office. We gratefully acknowledge his assistance as well as those of the Aviation Group, 101st Airborne Division, Ft. Campbell, Kentucky.

REFERENCES

Adams, J. A. Some considerations in the design and use of dynamic flight simulators. (AFPTRC-TN-57-51) Lackland Air Force Base, TX: Air Force Personnel and Training Research Center, 1957.

American Airlines, Inc. Flight Training Academy. Optimized flight crew training, a step toward safer operations. American Airlines, Inc. Forth Worth, Texas, 1969.

Briggs, C. E. Retroactive inhibition as a function of the degree of original and interpolated learning. Journal of Experimental Psychology, 1957, 53, 60-67.

Caro, P. W. An innovative instrument flight training program. New York: Society of Automotive Engineers, Paper No. 710480, 1971.

Caro, P. W. *Transfer of Instrument Training* and the synthetic flight training system. Alexandria, VA: Human Resources Research Organization, Professional Paper 7-72, 1972.

Crook, W. G. Experimental assessment of ground trainers in general aviation pilot training. (FAA-ADS-67-5) Federal Aviation Agency, Washington, D. C., April, 1965. (AD 652371).

Duncan, D. B. Multiple range and multiple F tests. Biometrica, 1955, 11, 1-42.

McGrath, J. J. and Harris, D. H. Adaptive training. Aviation Research Monographs, 1971, 1(2).

Prophet, W. W. The importance of training requirements information in the design and use of aviation training devices. Human Resources Research Organization, Professional Paper 8-66, 1966.

Roscoe, S. N. Incremental transfer effectiveness. Human Factors, 1971, 13, 561-567.

TransWorld Airlines. Flight simulator evaluation. Kansas City, MO: Trans World Airlines, Flight Operations Training Department, June, 1969.

Woodruff, R. R and Smith, J. F. T-4G simulator and T-4 ground training devices in USAF undergraduate pilot training. Air Force Human Resources Laboratory, Brooks AFB, Texas, Rep. No. AFHRL-TR-74-78, Nov. 1974.

ABOUT THE AUTHORS

DR. DONALD O. WEITZMAN is a senior research psychologist at the US Army Research Institute. His research interests include perception, sensory-motor integration, and information processing. He received his Ph.D. in experimental psychology from the University of Maryland in 1965.

DR. MICHAEL L. FINEBERG is a senior research scientist with the US Army Research Institute. He is currently involved in training support research, conducting field studies in aircrew performance measurement and laboratory work in visual performance under reduced light levels. He has worked as a research psychologist in behavioral assessment of nuclear weapons effects, and also as an engineering psychologist in the field of control/display design in Naval aircraft systems. Dr. Fineberg received the B.A. degree from Rutgers University, the M.A. degree in Human Factors, and a Ph.D. in applied Experimental Psychology from Catholic University.

MR. HALIM OZKAPTAN is a principal scientist and work unit area leader with the US Army Research Institute, and is currently responsible for laboratory and field studies in the area of human performance enhancement. He was the Director, Manpower Systems Research Division for the Navy Personnel Research and Development Laboratory where he was responsible for research and development programs in the areas of manpower utilization and human performance. Mr. Ozkaptan was Chief of Human Factors for the Martin-Marietta Corporation in Orlando, Florida; Head of F-111B Personnel Subsystems for Grumman Aircraft; a research scientist at Republic Aviation; and a branch head at the Naval Training Equipment Center. He received a Masters degree from Fordham University in Experimental Psychology, and has over thiry publications in the area of Human Factors.

CW4 GEORGE L. COMPTON, US Army, is presently Officer-in-Charge of the Army's Synthetic Flight Training System at Ft. Campbell, Kentucky. He is presently assigned to the Aviation Group, 101st Airborne Division and serves as an instrument flight examiner and standardization instrument pilot in rotary wing aircraft. Mr. Compton entered the US Army in 1954 and has over 20 years of Army flight experience. He received a B.S. degree from Troy State University in 1972.