TRANSFER OF TRAINING FOR AEROSPACE OPERATIONS: HOW TO MEASURE, VALIDATE, AND IMPROVE IT

Malcolm M. Cohen, Ph.D.
NASA-Ames Research Center
Moffett Field, CA 94035-1000

ABSTRACT

It has been a commonly accepted practice to train pilots and astronauts in expensive, extremely sophisticated, high fidelity simulators, with as much of the real-world feel and response as possible. High fidelity and high validity have often been assumed to be inextricably interwoven, although this assumption may not be warranted. The Project Mercury rate-damping task on the Naval Air Warfare Center's Human Centrifuge Dynamic Flight Simulator in Warminster (Johnsville), Pennsylvania, the shuttle landing task on the NASA-Ames Research Center's Vertical Motion Simulator at Moffett Field, California, and the almost complete acceptance by the airline industry of full-up Boeing 767 flight simulators for transition training of airplane captains, are only a few examples of this approach. For obvious reasons, the classical models of transfer of training have never been adequately evaluated in aerospace operations, and there have been few, if any, scientifically valid replacements for the classical models. This paper reviews some of the earlier work involving transfer of training in aerospace operations, and discusses some of the methods by which appropriate criteria for assessing the validity of training may be established.

Introduction

Effective functioning of aerospace systems critically depends on how well the operator can be trained to perform his relatively complex tasks under the unique environmental conditions encountered in space operations. Human factors considerations in the design of aerospace systems, while acknowledged to be extremely important, have frequently taken a back seat to the adaptability and the great capacity of the operators to learn how to control complex systems. For example, the space shuttle is operated, and must function effectively, during launch, orbital flight, re-entry, and landing. Although the control characteristics of the vehicle change dramatically under these different segments of the flight profile, the operator must be trained to make the system function effectively under all conditions. Extensive, and expensive, training has traditionally been used to help the operator learn how to perform appropriately.

Training and Transfer of Training

Probably, the most important aspect of any training program that should be evaluated to determine its efficacy is the phenomenon known as "transfer." Transfer of training occurs whenever the performance on one task has an effect, either beneficial or detrimental, on the performance of another task that is performed subsequently. Positive transfer results when performance on the initial task leads to improved performance on a subsequent task; negative transfer results when performance on the first task has a detrimental effect on performance of the second task. (1,2)

To quantify the amount of transfer between two tasks, researchers generally obtain a score based on the initial performance of the second task for those individuals who had previously practiced the first task, and compare the score with the score of initial performance on the second task obtained from individuals who did not practice on the first task. A criterion for the amount of practice, or the degree of mastery, on the first task is usually specified in advance. Traditionally, the scores are based on the amount of practice needed to reach the criterion, e.g., speed of performance, accuracy of performance, a combination of speed and accuracy, or some other stable measure that can be used to characterize the performance on each task. (3)

Measuring Transfer of Training

Two classical means of specifying the amount of transfer involve the concepts of savings and transfer effectiveness. For example, consider a case where individuals require an average of ten hours of actual flight time to achieve adequate proficiency for them to fly their first solo. If one hour of practice in a ground-based simulator (i.e., first task) allows a similar group of individuals to solo after only 8 hours of flight time (second task), we see that there is a savings of 2 hours of flight time.

Expressed quantitatively,

$$S = T_{(t2:t1=0)} - T_{(t2:t1)}$$

Where S is the savings, $T_{(t2:t1=0)}$ is the time required to master task 2 given no practice on task 1, and $T_{(t2:t1)}$ is the time required to master task 2 given previous practice on task 1. The effectiveness of transfer between training in the ground-based simulator versus the actual aircraft can be expressed as a ratio of the difference between flight time needed for the control group (flight practice only) and the training group (simulator and flight practice), relative to total simulator time.

Expressed quantitatively,

$$TE = [T_{(t2:t1=0)} - T_{(t2:t1)}] / T_{(t1)}$$

where TE is the training effectiveness ratio, $T_{(t2:t1=0)}$ is the time required to master task 2, given no practice on task 1, $T_{(t2:t1)}$ is the time required to master task 2, given practice on task 1, and $T_{(t1)}$ is the time actually spent on task 1.

In the example given, we have 10 hours of flight time needed for the control group [$T_{(t2:t1=0)}$] minus 8 hours of flight time needed for the group who had trained on the simulator [$T_{(t2:t1)}$], divided by 1 hour of time in the simulator $T_{(t1)}$. This yields a transfer effectiveness ratio of 2, which means that the one hour spent in the simulator provided training that was as effective as two hours in the actual aircraft. From a practical perspective, this makes good sense, because much time is often wasted in the aircraft before it is can be used for training. For example, to learn techniques for recovery from a stall, the student pilot must be at sufficient altitude over an appropriate practice area, and it takes time to get there in an aircraft.

Because gains in performance that are achieved from practice usually decrease over time (i.e., learning is a negatively accelerating function), the transfer effectiveness ratio also decreases with increasing time spent in practice. When the transfer effectiveness ratio declines to 1.0, there is no training advantage to be obtained from additional use of a simulator or training device, although there still may be other significant advantages.

If, for example, training in an aircraft costs three times as much as training in a simulator, there will still be a financial advantage in using the simulator until the training effectiveness ratio declines to 1/3.

Predicting Transfer of Training

As a general rule, the more similar two tasks are, the more likely it is that they will interact with one another. Further, the beneficial or detrimental nature of the resulting transfer between two tasks usually depends on the similarity of the displays (stimulus conditions) and on the similarity of the controls (required responses) in the two tasks. (4) Four cases may be distinguished: (5)

Case 1 - HiHi Where the displays and controls on both the initial and the subsequent tasks are so similar that they are practically indistinguishable from one another, transfer will usually be both large and positive; learning to perform the first task can provide the equivalent of an opportunity to practice on the second task. For example, learning to fly in a particular aircraft, and then attempting to fly another aircraft of the same type would have extremely high positive transfer. Another example would be to fly a high-fidelity simulator of the aircraft as the first task.

Case 2 - LoLo Where the displays and controls on the initial and subsequent task differ dramatically from each other, there is generally little transfer of training between them. For example, learning how to play a piano probably will not help someone to learn how to fly an airplane.

Case 3 - LoHi Where the displays are different, but the controls on the two tasks are similar, transfer of training is usually positive (but much less effective than in Case 1). This would generally be the case where one initially learns to fly in one type of aircraft, and subsequently attempts to fly a different type of aircraft.

Case 4 - HiLo This case is somewhat more complex than the other three cases cited. Where the displays are similar, but the controls are different, either weak positive transfer or negative transfer may occur. The weak positive transfer could result when the displays are highly similar, but the controls are so different that confusion between them would be very unlikely; an example would be that occasioned by a flashing red light when driving an automobile (apply the brakes), or a flashing red light indicating the failure of a landing gear to have locked (recycle the lowering of the gear). The major advantage of training in this type of situation may lie in having the individual learn to pay careful attention to the appropriate stimulus display. In contrast, if the controls are not only different, but conflicting, negative transfer would be expected. This would be the case where two aircraft have similarly appearing control levers that are placed in the same location in the cockpit, but with different resulting functions (e.g., flaps and throttle levers in reversed positions in two different aircraft). Learning to fly the first aircraft could interfere with subsequent flying of the second aircraft (and it could lead to a major accident).

The Use of Simulators in Training

Simulators, and other training devices, have been widely used throughout the aviation industry, but their use in the space program had been different, at least until recently. Generally, commercial pilots and aircrew were given extensive opportunity to practice their skills in the operational system before they were officially required to perform in actual operations. Further, transfer of training could be determined in checkout flights, and the efficacy of specific training programs could be evaluated in depth. Recently, however, the use of high-fidelity simulators for training has received such wide acceptance by the aviation industry that, following an authorized training program on a high-fidelity simulator of some new commercial passenger aircraft, the very first flight of a pilot-in-command can often be a revenue flight. This relatively recent development in commercial aviation parallels the training of astronauts, which usually demands that the first flight after training be an operational mission, providing little or no opportunity for additional training. The aviation industry is now using a technique that was originally developed in the space program.

If we consider the conditions under which astronauts are generally expected to perform in space missions, we find that both the displays and controls for training specific operational tasks can often (but not always) be made to be highly similar. This type of condition corresponds to Case 1 - HiHi, discussed previously, and yields a high degree of transfer of training.

A particularly relevant application of this training paradigm was used in the early days of manned space flight, when the original astronauts of project Mercury experienced realistic acceleration profiles, and performed control tasks in a Mercury capsule that was mounted in the (Johnsville) Naval Air Development (now, Naval Air Warfare) Center's human centrifuge. (6,7) The centrifuge was used as a dynamic trainer for a re-entry rate damping task, largely because it added realistic acceleration cues to the instrument displays; it also was used to train the astronauts in sequence monitoring and emergency procedures during simulated launch and re-entry profiles. Following their high-fidelity simulation training, the astronauts mastered the necessary skills, and were considered to be well prepared to function in the actual operational environment.

Similarly, the shuttle landing simulations conducted at Ames Research Center over the past several years (8) have also taken advantage of this general approach to training. A realistic mock-up of the shuttle cockpit, mounted in the Vertical Motion Simulator (VMS), was used to train astronauts to land the shuttle under various conditions, including reduced visibility approaches, high cross-winds, and steering mechanism failures upon landing. Before any shuttle pilots ever performed an actual landing in the shuttle itself, they had already experienced several landing scenarios in the VMS. As a result of their performance on the VMS, they were regarded as well prepared to perform effectively in the actual shuttle landings.

Cost versus Validity of Simulators

The major cost of striving to attain a large degree of transfer of training in high-fidelity simulators is actual financial cost. Since we know that high display and control similarity leads to the best transfer of training, we sometimes go overboard in insisting that a training simulator must have high face validity or unnecessarily high fidelity in representing the actual vehicle. In addition, the financial payoffs for the simulator manufacturers lie in providing the most advanced state-of-the-art devices. The pull of the user community for more and more sophisticated simulators as training devices, coupled with the push of the manufacturer to provide all of the "bells and whistles" often combine to drive simulator costs ever higher. Although true validity of training and high fidelity of training devices are often related, they are definitely separable. This is an area where considerable research needs to be done, both to reduce costs, and to establish how much high fidelity is actually needed to produce the best transfer of training.

Limitations of Ground-based Simulation

If one wishes to train an astronaut to function effectively in the space environment, it is not always possible to make both displays and controls (i.e., stimulus conditions and required responses) sufficiently similar here on Earth to expect a high degree of positive transfer. For example, attempting to don or doff a space suit on Earth versus in orbit probably involves both similar and different stimulus conditions that are coupled with similar and different motor responses. The issue of donning or doffing a space suit on Earth and in orbit leads to obvious questions regarding how one should go about training astronauts to perform effectively in different environments.

Skylab experiment M-151 provides an excellent case in point. The time required to don a space suit on Earth initially was between 900 and more than 1400 seconds. With practice, this time was reduced to between 800 and 850 seconds. The time required to don a space suit in orbit initially shows a dramatic increase when compared against the preflight times following practice on Earth, with an initial value for the first donning in orbit around 1000 to 1100 seconds. Subsequent attempts in orbit lead to significantly improved performance, and eventually, some of the astronauts even perform better in orbit than they ever did on Earth, with times as low as 669 and 740 seconds. (9)

Although this study was not directed towards evaluating transfer of training between the terrestrial and orbital environments, it is clear that the initial *apparent* disruption of performance in orbit could have been due to three of the four possible cases of transfer of training that we discussed previously. First, the task of donning a space suit in orbit could have been disrupted by attempts to don the suit on the ground; i.e., there was negative transfer between the two tasks such that the techniques acquired on Earth interfered with the techniques required to don the suit in orbit (Case 4, with negative transfer). A second possible explanation is that the two tasks were so very different from one another that donning the suit on the ground had no effect on donning the suit in orbit and, had the subjects never practiced on the ground, the same results would have been obtained in orbit; i.e., there was no transfer between the two tasks (Case 2). A third possible explanation is that there was positive transfer of training from ground-based results to orbital donning of the suit, and that the apparent disruption in orbit would have been significantly greater than that actually obtained if the terrestrial practice in donning the suit had not been undertaken (Case 3).

If control groups were used, and donning times were obtained in orbit for individuals who did not practice on the ground, it would have been possible to evaluate the alternatives discussed above. From a practical point of view, however, it is not likely that mission planners would have an astronaut don a space suit for EVA for the very first time in orbit without any opportunity to practice the task before he/she gets there. As a result, this issue may have to remain unresolved for some time. Nevertheless, because there was an apparent initial disruption of performance in orbit, it is clear that the task of donning the space suit on the ground is not the same as that of donning the suit in orbit (i.e., not Case 1).

To describe a complex task, such as donning a space suit on Earth, and then to compare it with the task of donning the same suit in microgravity, requires the specification of differences in both stimuli and responses (displays and controls) under both terrestrial and space conditions. Although the complexity of the task, and the lack of a theoretical model with which to characterize the relationships between stimuli and responses make the solution of this problem extremely difficult, the general approach remains feasible.

A similar approach that has been used frequently is to attempt to create a training environment that is similar to the operational environment in which the subjects are expected to perform. Under such conditions, a high degree of stimulus similarity can be expected, and if the training task itself is similar to the operational task, then a high degree of transfer might also be expected. Unfortunately, as the next example illustrates, this approach does not always work as well as might be desired.

The value of underwater training to simulate the effects of microgravity had been strongly supported for EVA space assembly tasks such as those initially proposed for Space Station Freedom; further, the Soviet cosmonauts who were to perform EVAs in missions on Mir were also given extensive underwater training. It was believed that this approach, coupled with work on air-bearing platforms, would be appropriate for training astronauts in the satellite recovery tasks such as those that were required for STS-49. Apparently, the inertial and viscous damping characteristics of the underwater environment, and the limited degrees of freedom for movements on the air-bearing platforms, were sufficiently different from the conditions encountered in space that the training was not fully adequate to prepare the astronauts for their tasks in STS-49. For example, continuous reactive rotational movements of the astronauts are difficult to initiate under water, but they are relatively easy to stop; in space these movements are extremely easy to initiate, but they are very difficult to stop. Similarly, air-bearing platforms will not respond to forces that are applied orthogonally to their surfaces; no such "null-planes" are present in space. The delays and difficulties in retrieving the satellite appear to have been at least partially due to the inadequacy of the underwater training.

In fact, based on the available evidence, the actual value of the underwater training for STS-49 cannot be fully determined. As was previously noted in the context of Skylab experiment M-151, we have no means to establish how good the astronauts' performance would have been without the underwater training. Thus, although it is clear that underwater training on the satellite recovery task did not completely prepare the astronauts for performing the task in orbit, it could have been of value, or it could have actually provided negative transfer by having the astronauts master techniques that relied on viscous damping and other characteristics of the underwater environment that were not available in orbit. Nevertheless, the flexibility and the perseverance of the astronauts allowed for successful completion of their mission despite possible inadequacies in their training.

Space in Little Pieces: The Value of Training in Microgravity

For the training of astronauts, there is probably no better way than parabolic flight in the KC-135 or similar aircraft to familiarize them with the kinds of problems that they will encounter in space.

Although the duration of any single parabolic phase is too brief for the conduct of a complex task, such as donning a space suit, parabolic flight does provide a close approximation of the microgravity conditions encountered in space. Further, the appreciation of the value of parabolic flight as a technique for training astronauts to function appropriately in the microgravity environment of space is a relatively recent phenomenon. At Johnson Space Center, the KC-135 aircraft has been used to fly "zero-gravity" parabolas. The astronaut trainees have taken advantage of the 20 to 30 seconds of "weightlessness" that are produced in this aircraft to practice on critical tasks that they are expected to perform in orbit. It is now generally believed that, if practice is restricted to the hypogravity phases of parabolic flight, the astronauts will be able to master new techniques that would have been extremely difficult, if not impossible, to learn on the ground. (10)

Similarly, operational tasks to be performed in orbit, such as obtaining blood samples, conducting surgical procedures, operating control devices, attaching sensors, implementing various experimental protocols, or donning a space suit, are now practiced, part by part, during the brief periods of hypogravity achieved in parabolic flight. By performing small individual segments of the task in hypogravity, and combining segments across multiple parabolas, astronauts are believed to achieve a high degree of proficiency before they actually enter orbit. Thus, the cumulative effects of training across several hypo-gravity phases of a series of parabolas probably can be extremely useful, particularly if the training is done systematically.

Validating and Improving Transfer of Training: A Reasonable First Step

Unfortunately, as was the case with the centrifuge simulations of the Mercury launch and reentry profiles at Johnsville, the shuttle landing simulations on the VMS at Ames, and the underwater training in the WETF at JSC, there has been little scientific evaluation or documentation of the value of the training methods described above. Despite the adoption of these general techniques throughout the aerospace industry, the choice and use of simulations remains more art than science.

Recently, and particularly within NASA, it has become a popular dictum that it is impossible to manage something unless one is able to measure it. As far as transfer of training is concerned, we know how to measure it. Nevertheless, as a rule, we do not do so. Without measures of transfer of training, there are no ways to validate the training or to improve it. Thus, one of the most important tasks that currently lies before the training community is to provide measures of transfer of training that can be used for validating and improving the training.

An important, and reasonable, first step in substantiating the already common use of parabolic flight in training astronauts could be to revisit Skylab experiment M-151. That study was discussed here as a demonstration of the value of microgravity in training; it should be repeated, but with a new twist. In this case, we could have astronauts train in donning a space suit either on the ground or in the hypogravity phase of parabolic flight. If the previously reported disruption occurs following ground-based training, but not following parabolic flight training, the practical value of parabolic flight for training purposes will have finally been appropriately documented.

References

1. Holding, D.H. (1977) Transfer of Training. In: R.A. Schmidt and E.A. Fleishman (Eds.) Section IX: Psychomotor learning and performance. In: B. B. Wolman, (Ed.) International encyclopedia of neurology, psychiatry, psychoanalysis and psychology. New York: Van Nostrand.

2. Holding, D.H. (1981) Human Skills. New York: John Wiley & Sons.

3. Holding, D.H. (1987) Concepts of training. In: G. Salvendi (Ed.) Handbook of Human Factors. New York: John Wiley & Sons. pp. 939-962.

4. Osgood, C.E. (1949) The similarity paradox in human learning, a resolution. *Psychological Review.* 56:132-143.

5. Holding, D.H. (1965) Principles of Training. Oxford, England: Pergamon Press.

6. Voas, R.B. (1961) Project Mercury: Astronaut Training Pro-gram. In: B.E. Flaherty (Ed.) Physiological Aspects of Space Flight. New York: Columbia University Press. pp. 96-116.

7. Chambers, R.M. & Nelson, J.G. (1961) Pilot performance capabilities during centrifuge simulations of boost and reentry. *American Rocket Society Journal.* 31:1534-1541.

8. Nguyen, V.H., Nishimi, J.T., Payne, T.H., & Woosley, E.W. (1990) Space shuttle descent flight verification by simulation: A challenge in implementing flight control system robustness. AGARD Conference Proceedings CP-489: Space vehicle Flight Mechanics. Neuilly sur Seine, France: North Atlantic Treaty Organization. pp. 9-1 - 9-17.

9. Kubis, J.F., McLaughlin, E.J., Jackson, J.J., Rusnak, R., McBride, G.H., & Saxon, S.V. (1977) Task and work performance on Skylab missions 2, 3, and 4: Time and motion study - Experiment M151. In: R.S. Johnson & L.F. Dietlein (Eds.) Biomedical results from Skylab. Washington: National Aeronautics and Space Administration. pp. 136-154.

10. Cohen, M. M. Perception and action in altered gravity. *Annals New York Academy of Sciences,* 1992, 656:354-362.