N93-30684

# 10. TRAINING EFFECTIVENESS ASSESSMENT: METHODOLOGICAL PROBLEMS AND ISSUES

KENNETH D. CROSS

If any one of you has talked with a person who has recently examined the literature on helicopter simulator training effectiveness, I'll bet you dollars to donuts that they were positively shocked by the small amount of research that has addressed this important topic. The persons I have talked with ask me, "How can it be that the military has invested enormous sums in helicopter simulators without having solid empirical data on how effective they are and how they should and should not be used?"

Although there is a host of reasons for the lack of data on helicopter simulator effectiveness, it is my contention that one of the most important is the lack of an evaluation methodology that yields comprehensive and valid training-effectiveness data in a timely manner at an affordable cost. Accordingly, my comments today are aimed at identifying some of the methodological problems encountered in assessing the training effectiveness of helicopter simulators and some of the issues that must be addressed in developing solutions to these problems.

Before proceeding, it is important to acknowledge that my comments reflect the perspective of a behavioral sciences researcher (table 1). It is also important to acknowledge that my views have been greatly influenced, and perhaps biased, by my experience in considering the training needs and problems of Army aviators. I have attempted to make all of my comments relevant to civilian aviation, but I cannot promise that I have been completely successful.

Because time is short, I have limited the focus of my comments. The methods I discuss are ones that I consider suitable for assessing the cost and training effectiveness of a new, production-model simulator for initial skill-acquisition training. These methods may or may not be suitable for collecting the data needed to support the simulator design decisions that must be made in the early design phase of a simulator development effort. Similarly, the methods may or may not be suitable for assessing a simulator's effectiveness for skill-sustainment training.

Table 1. Perspective and scope

| |
|---|
| Perspective |
|     Behavioral sciences research |
|     Army aviation |
| Focus |
|     New production model simulator evaluation |
|     Initial skill acquisition (basic/transition) |
| Important topics not addressed |
|     Predicting training effectiveness from engineering data |
|     Utility of simulators for proficiency checking |
|     Utility of simulators for skill sustainment training |

Three important topics that I have not attempted to address except in passing include the feasibility of using engineering data to predict training effectiveness, the utility of simulator for proficiency checking, and the utility of simulators for skill-sustainment training.

I will commence with a brief description of what I refer to as the "classic" transfer-of-training methods and an illustration of the types of data generated by them. Then, I will describe what I consider to be the key shortcomings of these methods. Finally, I will describe a methodological approach that, in my view, is more effective and efficient than the classic approach.

It is important to emphasize that the approach I propose does not eliminate the need to measure empirically the extent to which training in the simulator transfers to the parent aircraft. Rather, the approach is intended to insure that the simulator is functioning optimally and that the simulator training method is near optimal before an expensive transfer-of-training study is performed. Believe me, a researcher's worst nightmare is to complete a transfer-of-training study costing hundreds of thousands of dollars, only to discover that the simulator was not functioning properly or that the trainees were given the wrong kind or amount of training in the simulator.

It is also important to emphasize that many of the methods and ideas I discuss are not new. If anything I have to say is truly a novel idea, it is the sequence in which the methods are used and the specific purposes for which they are used.

Figure 1 illustrates the classic transfer-of-training research design. One group of trainees—the control group—receives no simulator training. The purpose of the control group is to provide information about the amount of time required to achieve proficiency through aircraft training alone. In addition to the control group, there are one or more groups of trainees who receive some amount of training in the simulator before being trained to proficiency in the aircraft; these groups are referred to as experimental groups. This illustration assumes that there are five experimental groups that differ only in the number of hours of training they they receive in the simulator—5 hours, 10 hours, and so on. All groups are trained to the same level of proficiency in the aircraft, and the number of aircraft hours required to reach proficiency is recorded.

A simulator is training-effective to the extent that simulator training reduces the amount of aircraft training required to achieve proficiency in the aircraft. In short, a simulator is training-effective to the extent that simulator training hours replace aircraft training hours. The hypothetical data presented in figure 2 illustrate the well-established relationship between the amount of simulator training the trainees receive and the amount of training required to achieve proficiency in the aircraft. The control group trainees, who receive no training in the simulator, require an average of 50 hours in the aircraft to reach proficiency; trainees who receive 5 hours of simulator training require only 40 hours in the aircraft to reach proficiency. This negatively decelerating monotonic function illustrates the simple fact that each increment in simulator training time yields progressively less savings in aircraft training time. Data of this type are interesting, but are not sufficient to determine what amount of simulator training is optimal.

Cost data must be brought to bear in deciding how much simulator training is enough. Figure 3 shows the relationship between the amount of simulator training and total training costs, or, its mirror image, cost savings. In producing this figure, I used the hypothetical training-effectiveness data shown in figure 3, along with the Army's current estimates of the cost of an hour of Blackhawk simulator time and the cost of an hour of Blackhawk aircraft time. As you see, the simulator and aircraft costs are $338 and $1,424 an hour, respectively. The cost curve shows that very little cost reduction is realized from simulator training beyond 10 hours. If cost is the prime consideration, total cost can be minimized by giving each trainee 15 hours of training in the simulator. However, if aircraft are unavailable for training, as many as 25 hours of simulator training can be given without increasing total training cost appreciably.

So, how can one find fault with a method that yields data like these? Let's consider some of the problems.

Table 2 lists some of the key shortcomings of the classic transfer-of-training method. First, the method yields only a composite measure of training transfer. This would not be a problem if the simulator were equally effective for training every maneuver. However, what is more likely is that training transfer for some maneuvers will be large and positive whereas training transfer for other maneuvers will be negligible or even negative. If this is indeed the case, the composite measure of training transfer is an underestimate of the simulator's optimal training effectiveness. Stated differently, the cost effectiveness of the simulator could be increased by eliminating training on those maneuvers for which training transfer is negligible or negative.

Second, the relatively high cost of transfer-of-training studies prevents the use of this method for optimizing the various components of the training system. When the first version of a production simulator appears on the scene, there are going to be many uncertainties about how best to set it up and use it. For example:

1. Are all components of the simulator functioning as they were designed to function?
2. Are there ways the simulator components can be adjusted or modified to increase the simulator's training effectiveness?
3. What maneuvers should be trained, in what order should the maneuvers be trained, and how much training should be given on each maneuver?
4. What is the best method or procedure for training a given maneuver?
5. What is the best way to employ the instructional support features available on the simulator?

Although these questions are of critical importance, it would be prohibitively costly to answer them through classic transfer-of-training studies. Another more efficient method is required for this purpose.

The third shortcoming is that transfer-of-training methods are not suitable for assessing some simulator training applications. Although a simulator may be highly

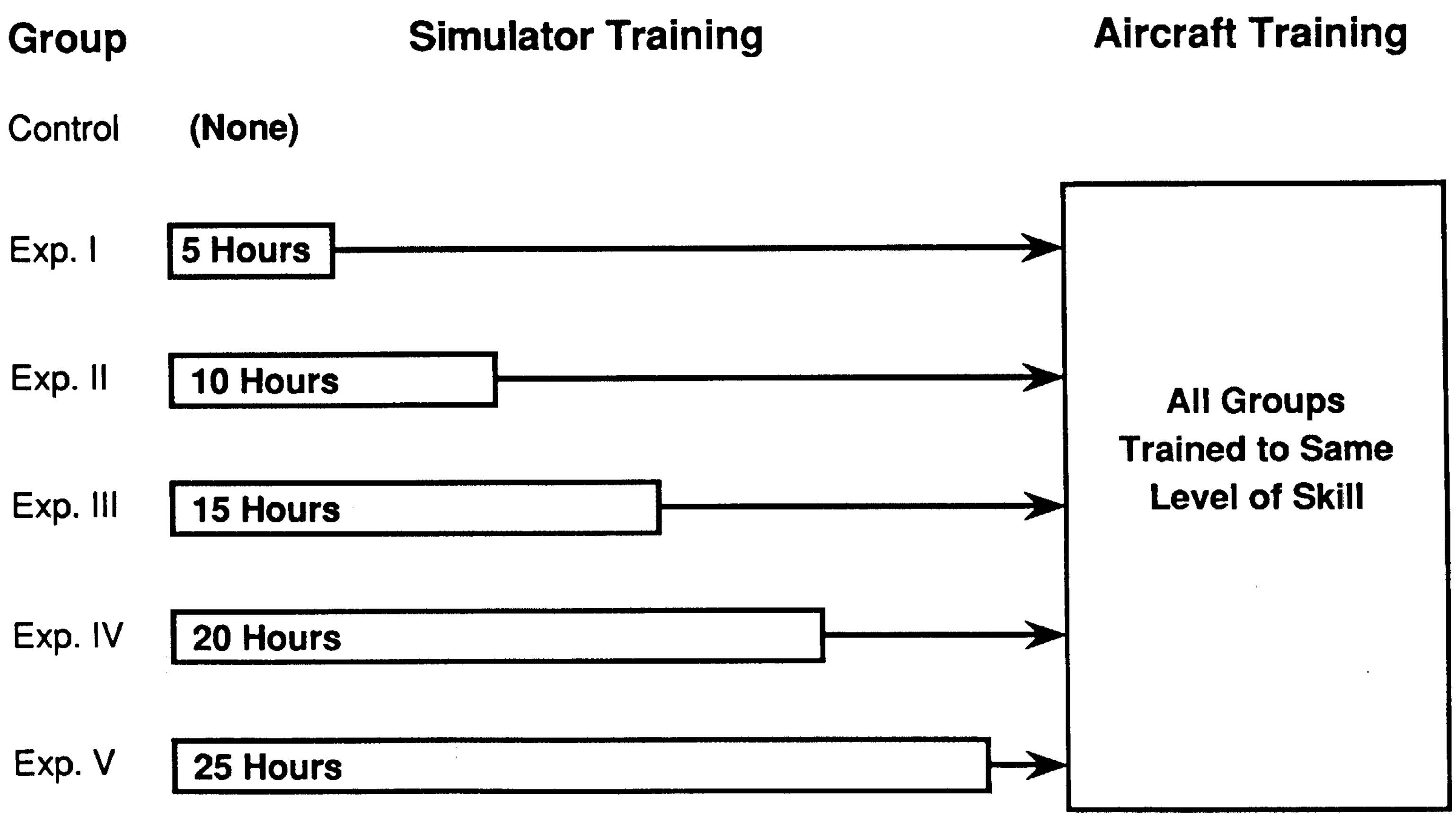


Figure 1. Classical TOT research design.

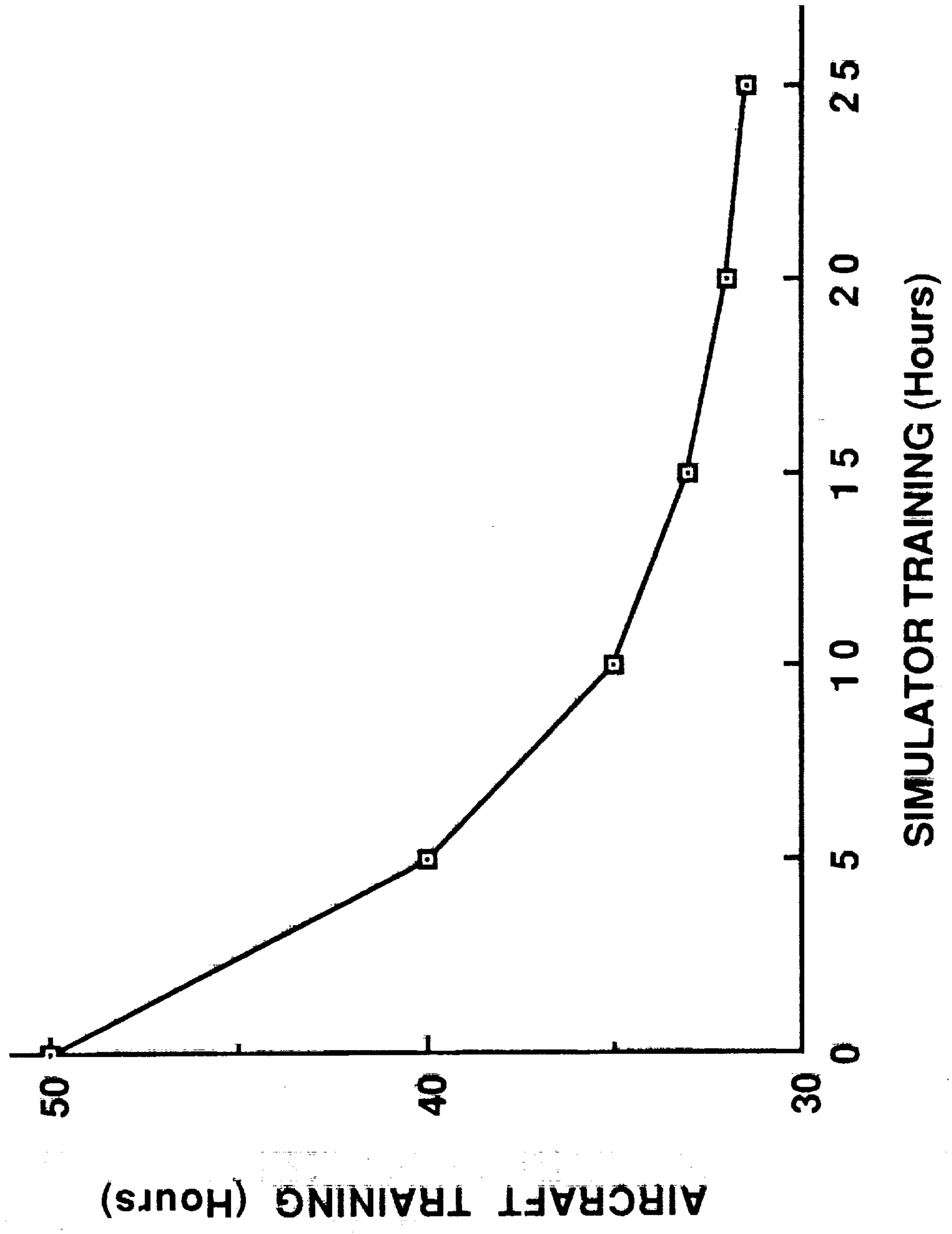


Figure 2. Aircraft training hours to criterion as a function of prior simulator training (hypothetical data).

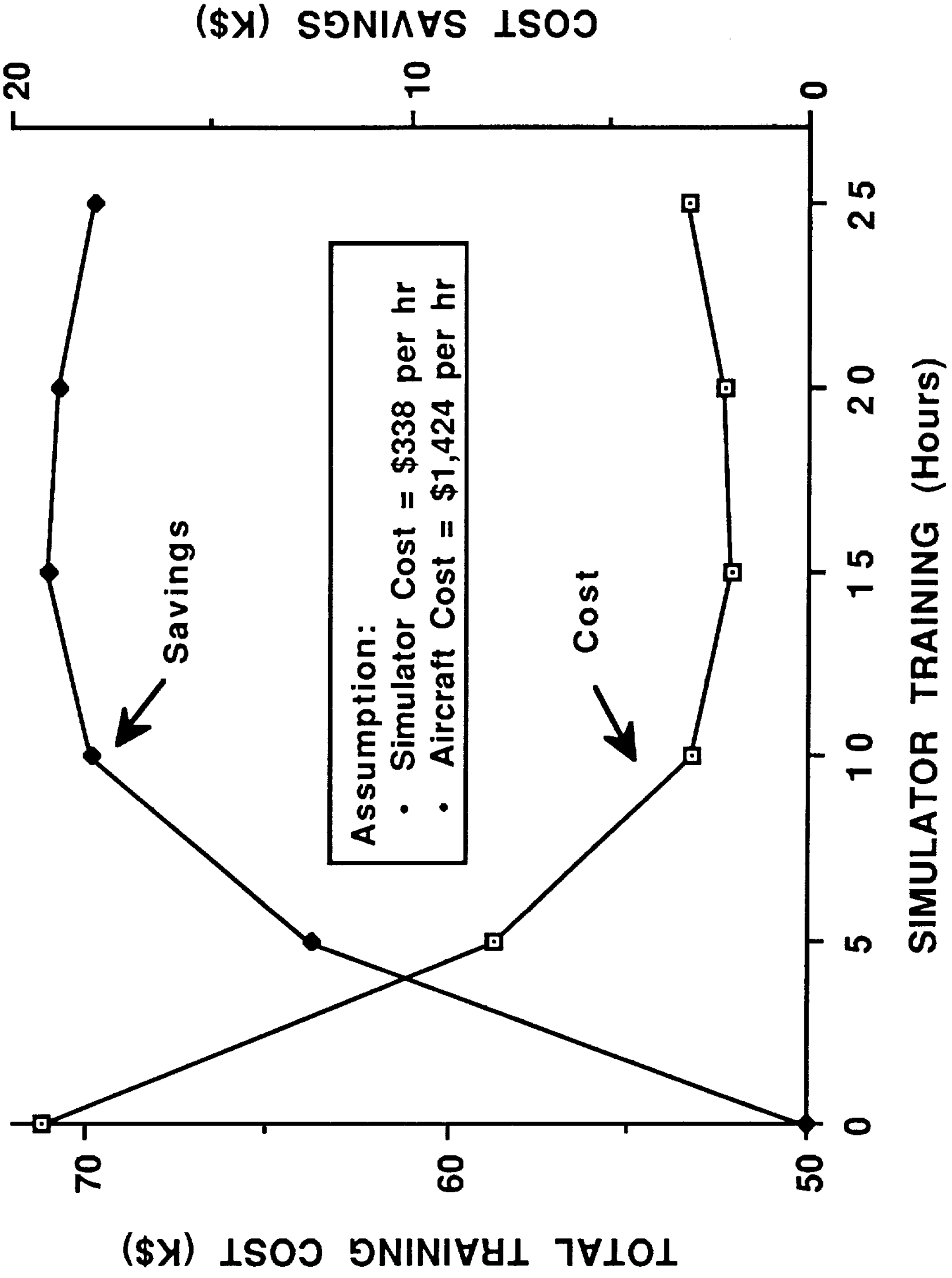


Figure 3. Total training costs/savings as function of simulator training hours (hypothetical data).

Table 2. Key shortcomings of classic transfer-of-training method

1. Yields only a composite measure of training transfer
2. High cost prevents use for optimizing training system
    Simulator set up and functioning
    Type and sequence of maneuvers
    Training method and instructional support features
3. Unsuitable for assessing some simulator training applications

effective for training maneuvers that are too hazardous to perform in the aircraft, it is not possible to measure the extent to which such training transfers since it is not possible to measure how well pilots can perform these hazardous maneuvers in the aircraft. For example, it is probably too hazardous to measure in the aircraft a pilot's ability to recover from such emergencies as a brown-out or white-out, a dual engine failure, a complete loss of tail-rotor effectiveness, or a severe wind sheer.

There are other maneuvers and conditions for which proficiency measurement in the aircraft is excessively costly, even if the risk is acceptable. For instance, measuring pilots' ability to perform takeoffs and landings at high surface elevation may be costly if the research is not conducted at a location that is close to mountainous terrain. Also, because visibility conditions in the real world cannot be controlled, it may be excessively costly to measure pilots' ability to perform maneuvers under specific degraded visibility conditions.

The flow diagram shown in figure 4 illustrates my views about the type, sequence, and purpose of research studies that, together, may eliminate some of the shortcomings of the classic transfer-of-training methods. This approach to simulator evaluation is the result of a large amount of thought and a small amount of data collection, so it is not presented here as a proven research method. Although my colleagues and I believe the approach is workable and sensible, I invite all of you to critique the approach and to let me know what doesn't make sense to you.

The four small shadowed boxes in figure 4 identify four types of research studies that I consider necessary for the efficient assessment of a simulator's training and cost effectiveness; the boxes with the rounded corners identify the purpose served by each of the four types of studies.

As you can see in the upper left corner, the purpose of the analytical studies is to identify maneuvers for which training transfer cannot be assessed either because the maneuver clearly cannot be trained in the simulator, or a pilot's proficiency on the maneuver cannot be measured in the aircraft without unacceptable risk or cost. For obvious reasons, these maneuvers must be excluded from a transfer-of-training study. The purpose of the next two types of studies is to insure that the simulator and the simulator training are near optimal before a transfer-of-training study is commenced. Because of the limited amount of time available, I will not comment further on the analytical studies. Instead, I will use the time I have left to discuss the rationale and procedures for the three remaining studies: backward transfer, in-simulator skill acquisition, and modified transfer of training.

The idea behind a backward-transfer study is a simple one (table 3). If forward transfer is the extent to which training in a simulator transfers to the parent aircraft, backward transfer must be the extent to which training in the parent aircraft transfers to the simulator. If the skills required to perform a maneuver in the parent aircraft are the same as the skills required to perform that maneuver in the simulator, one would expect a high degree of backward transfer. If backward transfer is not high, it is reasonable to assume that something about the simulator is not right. In short, the fundamental premise is that a low backward transfer indicates one or more important shortcomings in the simulator. About 30 years ago, Jack Adams and his colleagues at the University of Illinois considered the feasibility of using measures of backward transfer to predict the degree of forward transfer. Although backward transfer may indeed be a reasonably valid predictor of forward transfer, it is important to emphasize that predicting forward transfer is not the purpose for which backward-transfer studies are proposed here.

The procedure for conducting a backward transfer-of-training study is simple and straightforward. The first step is to select pilots who are highly experienced in the parent aircraft and who have had little or no experience in simulators, especially in the simulator being evaluated. The next step is to evaluate each pilot's proficiency in the aircraft for each maneuver to be evaluated in the simulator. The third step is to measure the pilots' initial proficiency on each maneuver in the simulator. Initial proficiency refers to how well the pilots perform on no more than the

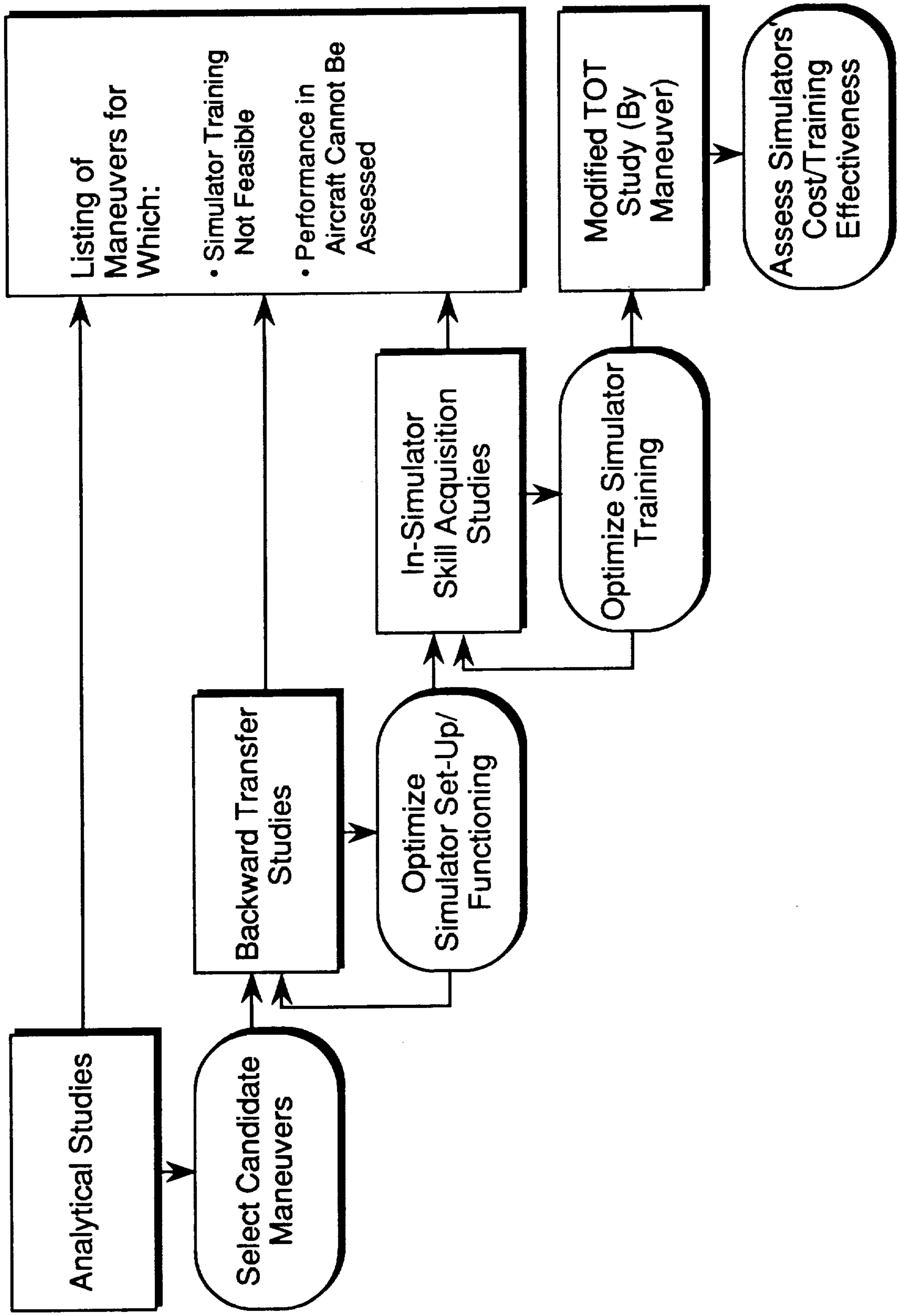


Figure 4. Type/sequence/purpose of evaluation studies.

Table 3. Backward-transfer studies

- Concept
  - Measure aircraft-to-simulator transfer (experienced aviators)
- Premise
  - Low backward transfer indicates simulator shortcomings
- Procedure
  - Select pilots with long aircraft experience and no simulator experience
  - Assess task proficiency in aircraft (desirable)
  - Measure initial task proficiency in simulator (one to three iterations)
  - Assess backward transfer
  - Interview pilots
- Benefits
  - Efficient (time and cost)
  - Yields diagnostic data about simulator shortcomings

first three attempts. There is a substantial amount of evidence that indicates that experienced pilots are able to adapt very quickly to even substantial differences between the aircraft and the simulator; as a result, a pilot's performance may quickly become contaminated by simulator-specific learning. The fourth step is to assess the degree of backward transfer by comparing simulator performance with aircraft performance, published performance standards, or both. The final step is to question pilots about the reasons for any poor performance in the simulator.

If the results reveal simulator shortcomings that can be eliminated completely or in part, the simulator can be modified and backward transfer can be measured again for the maneuvers that were performed poorly.

Backward-transfer studies have two important benefits. First, they are highly efficient in terms of both cost and time. If necessary, further cost reductions can be realized by eliminating proficiency measurement in the aircraft. The results of our backward-training research indicate that proficiency measurement in the aircraft is useful but not essential. Second, backward-transfer studies yield data that are useful in determining the reasons for poor simulator performance. In addition to the judgments of the participating pilots, much can be learned about simulator shortcomings by studying the types of errors made in performing a maneuver and the manner in which simulator performance differs from aircraft performance.

Figure 5 presents an example of the kind of results that can be expected from a backward-transfer study. The study was the first step in evaluating the effectiveness of the AH-1 Flight and Weapons Simulator for sustaining proficiency on emergency touchdown procedures. The 15 pilots who participated in the study were highly experienced AH-1 instructor pilots. The solid bars show the mean ratings for performance in the aircraft; the cross-hatched bars show the mean ratings for the first attempt to perform the same maneuvers in the AH-1 simulator. A rating of 1 indicates clearly unacceptable performance—a crash, a hard landing, landing short, and so on. A rating of 7 indicated the level of performance that the evaluators expected of the average AH-1 instructor pilot.

The ratings of aircraft performance indicated that the various emergency touchdown procedures differ in their inherent difficulty—the simulated anti-torque failure appears to be the most difficult maneuver, and the shallow approach to a running landing appears to be the least difficult maneuver. You can see that the ratings of simulator performance are far lower than the ratings of aircraft performance. More important, there is little correlation between the simulation ratings and the aircraft ratings. For instance, although most aviators performed standard autorotations very proficiently in the aircraft, no aviator received a rating higher than 1 on a standard autorotation in the simulator.

Although these results are not definitive proof that the AH-1 simulator is ineffective for training emergency touchdown procedure, they leave no doubt that the simulator and the aircraft differ in ways that may have a major influence on training effectiveness. In truth, it is not possible to examine these findings without worrying about negative transfer.

Table 4 shows a tally of the IP's spontaneous comments about the factors that contributed to the poor performance in the simulator. It can be seen that most of the IPs attributed their poor performance, in part, to the lack of visual cues needed to operate near the ground. The

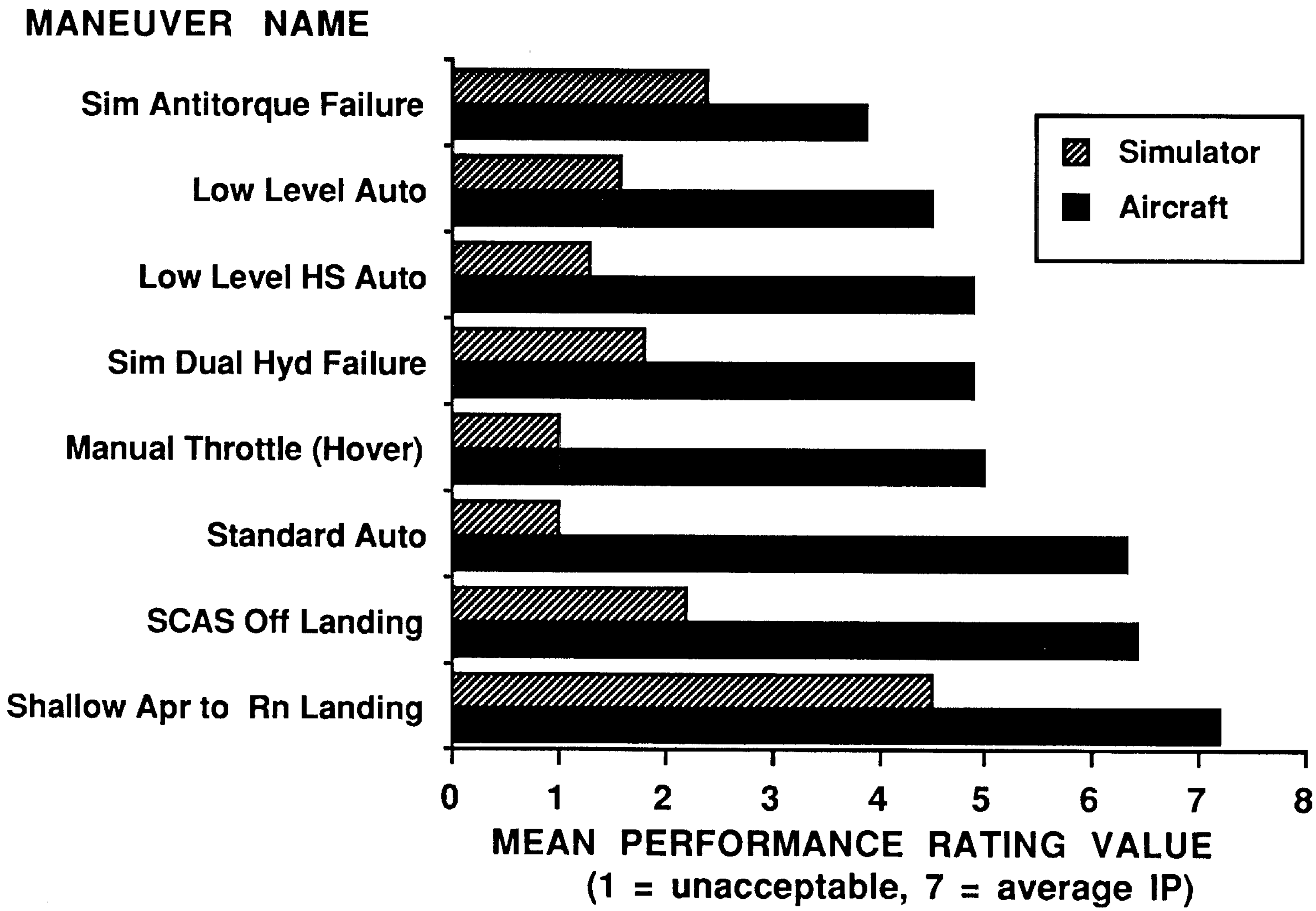


Figure 5. Example of backward transfer results.

C2

Table 4. Factors contributing to low backward transfer[a]

| | |
|---|---|
| Lack of visual cues | |
| | Visual display blurred near ground (100%) |
| | Unable to judge altitude near ground (94%) |
| | Insufficient visual cues to maintain hover (87%) |
| | Entry point difficult to judge (81%) |
| | Lack of peripheral cues (69%) |
| Unrealistic response to control inputs | |
| | Response to collective inputs (75%) |
| | Response to cyclic inputs (63%) |

[a]N = 15 for all percentages.

study was conducted in one of the early AH-1 simulators that was equipped with a camera-model-board visual system. The comments of the IPs are consistent with the results of tests that have shown that the camera-model-board system has poor focus and resolution when the probe is located very close to the model board. Table 4 also shows that most of the IPs identified unrealistic response to collective and cyclic inputs as an important contributor to poor performance in the simulator.

Although pilot judgments have not always proved to be highly reliable sources of information about simulator functioning, it would be foolish to ignore judgments that are as consistent as the ones shown here.

As I define the term, an in-simulator skill-acquisition study is a study performed to determine (1) how much simulator practice is required to gain proficiency on a given maneuver, and (2) the maximum level of proficiency that can be achieved (table 5). The recommendation to conduct skill-acquisition studies is based on two premises. The first premise is that the cost effectiveness of a simulator can be degraded significantly by inefficient simulator training. Inefficient simulator training may be the result of such factors as (1) too much or too little simulator training, (2) the use of inefficient training methods, and (3) the expenditure of an excessive amount of time on training maneuvers for which skill acquisition is very slow. The second premise is that skill acquisition data can be used to optimize simulator training.

Before proceeding, I would like to comment briefly on a couple of issues. The first is the importance of determining the optimal amount of simulator training for each maneuver. It is obvious that money is wasted when training on a maneuver is continued beyond the point at which performance asymptotes. What is not so obvious is that overtraining on a maneuver may actually reduce training transfer. Jack Dohme, an Army Research Institute researcher at Fort Rucker, has shown me unpublished data that strongly suggest that too much simulator training on a maneuver can, in fact, reduce training transfer.

On the other hand, there are reasons to believe that too little simulator training on a task may create problems of a different kind. The problems stem from the fact that some minimum level of proficiency on some maneuvers is

Table 5. In-simulator skill-acquisition studies

| | | |
|---|---|---|
| Premises | | |
| | Cost effectiveness of simulator degrades by training inefficiencies | |
| | | Too much/little simulator training |
| | | Ineffective training methods |
| | | Time spent training maneuvers for which skill acquisition is slow |
| | Simulator training can be optimized using skill acquisition data | |
| Procedure | | |
| | Select trainees (novice and experienced aviators) | |
| | Measure practice-iterations/time-to-criterion as function of maneuver type/sequence, training procedures | |
| Benefits | | |
| | Yields data with which to specify near-optimal training | |
| | | Maneuver sequence |
| | | Practice iterations |
| | | Training procedures |
| | Efficient (time and cost) | |
| | Identifies maneuvers that should be excluded from simulator training | |

required to learn other, more complex maneuvers efficiently. For example, instructor pilots claim that efficient learning of out-of-ground-effect hover is not possible until a student is reasonably proficient at performing in-ground-effect hover.

The second issue is the importance of establishing optional training methods. Many persons believe that all simulator training should be conducted in the context of a training scenario that approximates an aircraft training flight. Training in the context of a scenario of this type invariably wastes a lot of time in traveling from one point to another. For instance, training on approaches and landings in a simulator need not require the trainee to fly the entire traffic pattern in order to get the needed practice on the final approach and landing. Using the simulator's "initial condition set" to place the simulated aircraft on the final approach leg can greatly increase the number of practice iterations that can be accomplished during a training period. Although training method is certain to have a major effect on training efficiency, few studies have been conducted to assess the relationship between training method and rate and level of skill acquisition in the simulator.

Now let us discuss the procedures for conducting skill acquisition studies (refer to table 5). The procedures are simple. The first step is to select the pilots who are to participate in the study. Normally, the study would be conducted only with novice aviators who have no experience in the simulator. However, we have found it useful also to investigate the skill acquisition of pilots who are highly experienced in both the simulator and the parent aircraft. The use of experienced aviators is an efficient way to determine the maximum level of proficiency that is possible for a given maneuver.

The second step is to measure the number of practice iterations and the amount of training time required to reach a prescribed level of performance on each maneuver. Since the purpose of the skill-acquisition study is to optimize training methods, the practice iterations and training time would be measured as functions of such independent variables as type of maneuvers, the sequence in which maneuvers are trained, and the training procedures used.

Skill-acquisition studies have three kinds of benefits. As I have already mentioned, the main benefit is that the data can be used to specify a near-optimal training method before a transfer-of-training study is commenced. The second is that skill-acquisition studies are very efficient relative to transfer-of-training studies. A third benefit is that the data can be used to identify maneuvers that should be excluded from simulator training because skill acquisition in the simulator is slow or nonexistent.

I would like to take a few minutes to show you the results of a skill-acquisition study we performed on the AH-1 Flight and Weapons Simulator (fig. 6). The ultimate objective of the study was to assess the utility of the simulator for sustainment training, so we measured the simulator skill acquisition of experienced AH-1 pilots rather than trainees. Because we had not conducted skill-acquisition studies before, we assumed that experienced pilots would require no more than 10 practice iterations to reach proficiency on any task. So, the entire schedule was set up to obtain data on only 10 iterations. This assumption turned out to be grossly incorrect. In fact, more than 10 iterations were required to reach proficiency on most maneuvers. As a consequence, it was necessary to use regression analysis to project the number of practice iterations required to reach proficiency. Figure 6 shows projected iterations to proficiency for each of 15 maneuvers. For three maneuvers, there was no measurable learning during the first 10 iterations, so no projections could be made for the maneuvers. For the remaining maneuvers, the projected numbers of iterations to proficiency varied from 9 to 27.

Results such as these are useful for making decisions about the kinds of maneuvers that should be trained in the simulator and the amount of simulator time required to accomplish training on each maneuver. In addition, such results lead to some interesting questions about the design and function of the simulator. For instance, why do skilled aviators require so many trials to master normal approaches and hover tasks in the simulator?

The final and most critical study in the sequence is a transfer-of-training study. Table 6 shows my views about ways in which the classic transfer-of-training method can be modified to produce more useful data. Some involve changes in the simulator training and some require changes in the aircraft training.

There are three ways in which simulator training should be changed. First, I believe that all trainees should be trained to a prescribed level of proficiency in the simulator rather than receive some pre-defined amount of simulator training. Second, the amount of simulator training should be varied by varying the number of maneuvers trained rather than spreading fewer and fewer hours of training over some fixed number of maneuvers. And third, I believe that good estimates of cost effectiveness are possible only if the researcher is careful to record the nonproductive training time spent in the simulator. The

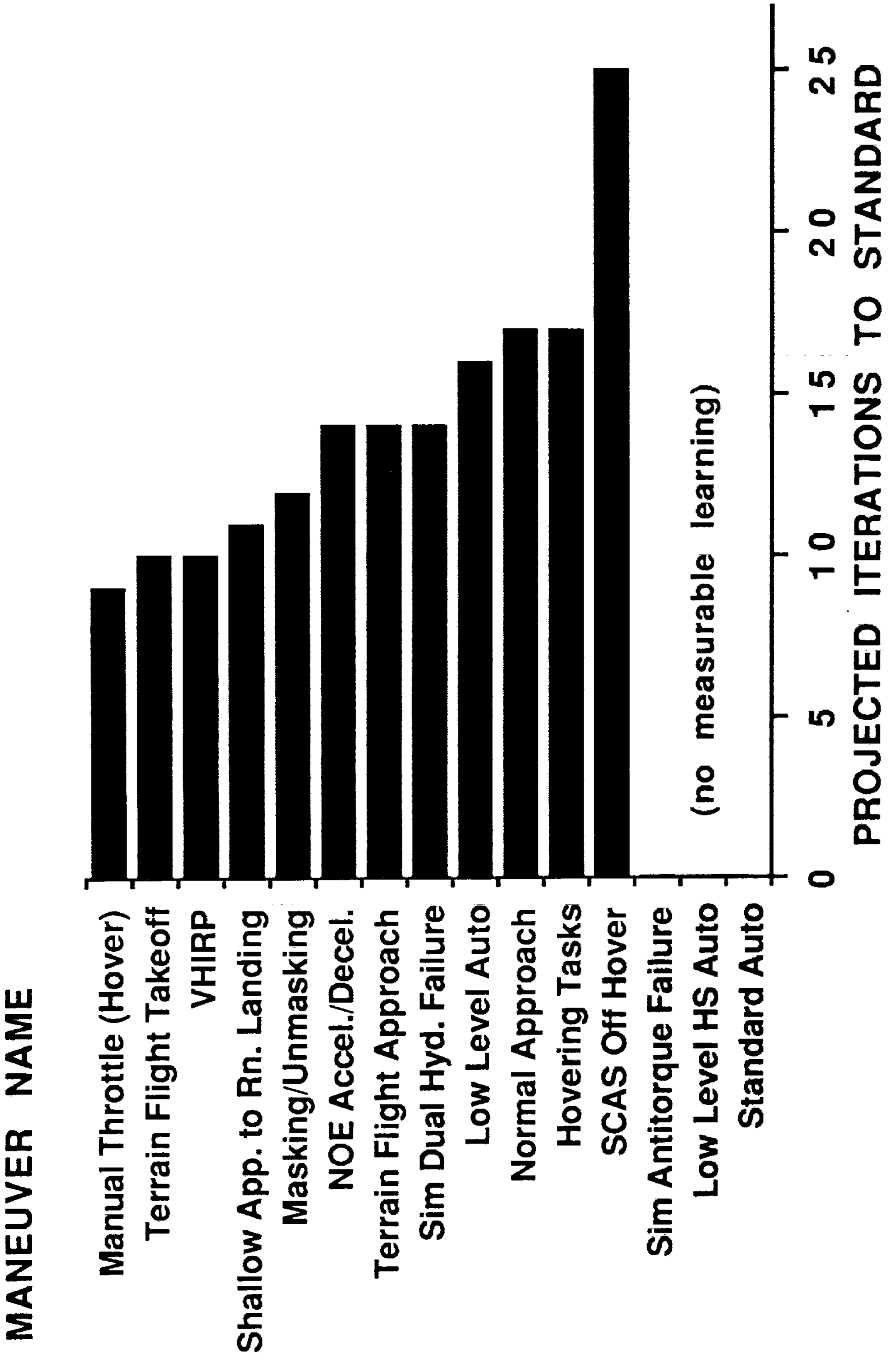


Figure 6. Example of in-simulator skill acquisition results.

Table 6. Modified transfer-of-training study (key differences from classic TOT study)

- Simulator training
  - Train to proficiency on each maneuver
  - Record nonproductive training time
    - Crash re-set
    - Repair
    - Procrastination, etc.
- Aircraft training
  - Record iterations to proficiency for each maneuver
  - Record nonproductive training time
    - Transit and refueling
    - Performing maneuvers already mastered
    - Procrastination, etc.

apparent cost of simulator training can be increased substantially by such extraneous events as crashes, simulator failures, and procrastination by instructors or students.

Next, consider the aircraft training procedure. I think it is essential to monitor and record iterations-to-proficiency on each maneuver trained in the aircraft. During aircraft training, a trainee simply cannot avoid performing certain maneuvers even though they already have been mastered. For instance, a trainee cannot accomplish a training flight without performing at least one takeoff and one landing. Hence, the total number of maneuver iterations performed during aircraft training is not determined by a trainee's training needs alone. In short, the effect of simulator training on the amount of aircraft training required cannot be determined without knowing the point at which the trainee reached proficiency on each maneuver.

As was true for simulator training, I believe it is necessary to record nonproductive training time for aircraft training. The quality of the aircraft cost data can be improved by subtracting from total aircraft hours the amount of time spent traveling between training sites, the time spent refueling, the time spent performing maneuvers already mastered, the time wasted because of procrastination, and so on.

A transfer-of-training study with the changes recommended here should provide the data needed to determine transfer-of-training by maneuver and by blocks of maneuvers. Moreover, the cost effectiveness of a simulator can be computed as a function of the specific maneuvers trained in the simulator. Finally, the cost-effectiveness estimates will not be confounded by unproductive time spent in the simulator, or in the aircraft, or both.

That concludes my remarks about training effectiveness assessment. Before inviting questions I would like to thank the sponsors of the workshop for giving me an opportunity to test my views before such a large body of experts. And, I would like to thank those of you in the audience for your kind attention.

MR. McGOWAN: On these backward-transfer-of-training studies, how do you account for a situation in which a maneuver, let's say AFCS-off flight in a helicopter simulator, may actually be easier in the simulator than it is in the aircraft, and how would you catch that in such a study? Does that question make sense?

DR. CROSS: Yes, Greg, your question certainly does make sense. And you have pointed out one shortcoming of backward-transfer studies. The results of a backward transfer study enable you make a one-sided decision. If you have a high degree of positive transfer you cannot conclude that everything is right with the simulator. It is possible that a task is so easy to perform in the simulator that it doesn't even come close to representing its corresponding task in the aircraft. In the example you gave, I don't know exactly why AFCS-off flight in a simulator is easier. I don't remember that our results show that to be the case.

MR. McGOWAN: No, I am not saying that is the case. I am just saying that could be the situation.

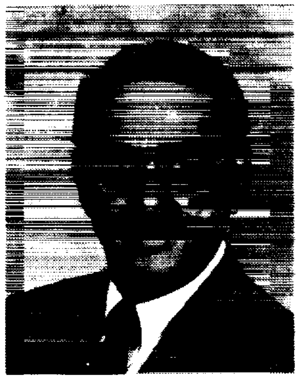
DR. CROSS: Oh, I see. My answer is still relevant. If you have a task that is unrealistically easy to perform in the simulator, it is unlikely that the simulator would provide effective training on that task. Unfortunately, backward-transfer studies are not effective in identifying that kind of problem. Such a problem might be revealed by in-simulator skill-acquisition studies, and most certainly would be revealed by transfer-of-training studies.

MR. HART: You used the Huey simulator, which apparently only poorly duplicates the helicopter. If you did the same study, let's say with a more modern simulator, would you get similar results? It seems to me that the problem in backward transfer has to do with the lack of authenticity of the simulator itself. Is that accurate?

DR. CROSS: No, it is not. I may have said Huey; if I did, I apologize. The backward-transfer and the in-simulator skill-acquisition studies were conducted in the AH-1 flight simulator, which is far more sophisticated than the old Huey simulator.

MR. HART: But again, wouldn't the results vary significantly as the quality of authenticity improves? Is that not an accurate statement?

DR. CROSS: It is perfectly accurate. That is the fundamental premise underlying all these kinds of studies that I have discussed today.



Kenneth Cross has been engaged in human factors research for 27 years. He has been at Anacapa Sciences since 1970 and now serves as Anacapa's president. Dr. Cross has formal training in research psychology and statistics; he received his doctorate degree from Kansas State University. Before joining Anacapa Sciences, he was research coordinator at the Naval Missile Center's Human Factors Laboratory. His research has dealt mainly with human performance in complex military systems. Much of Dr. Cross's time over the last ten years has been spent conducting studies of Army helicopter training at the U.S. Army Aviation Center, with emphasis on assessing the effectiveness of training conducted in helicopter simulators.