N93-30687

13. TRANSFER OF TRAINING AND SIMULATOR QUALIFICATION OR MYTH AND FOLKLORE IN HELICOPTER SIMULATION

JACK DOHME

I noticed something yesterday—perhaps others in the audience did too. Have you seen any young kids come up here and address this body? It would be, perhaps, impolite to note that gathered here are the grand old men of the field, that is, considerable experience is represented. So...I got to wondering why Bill Larsen asked me to speak. I'm not a test pilot; I'm not a graduate engineer; I don't have 10,000 hours experience beating the air into submission. However, it occurred to me that I wear trifocals so I'm certainly not a kid. I carry an AARP (American Association of Retired Persons—minimum age 50) card in my pocket so I guess I'm old enough. And we do have a perspective at "Mother Rucker" (The Army Aviation Center at Fort Rucker, Alabama) that may be worth sharing. So, I changed the title of my presentation this morning to "Myth and Folklore in Helicopter Simulation." This presentation has a second author, my boss, Chuck Gainer, and I should note that he contributed ideas but did not suggest the mode of this presentation. In summary, I suspect the real reason that Bill Larsen asked me to address this august body is for comic relief.

Table 1 is intended to present some political stuff, to stir up trouble and to get people to think about the issues in that X-rated document we've been asked to read, the "AC 120-XX." I thought I would begin by listing three of what I'm calling "myths" in the field of helicopter simulation. In discussing these myths, the "straw man" that I'm attacking wears green and I think that's fairly safe in this audience.

Let's look at myth 1: "A Simulator Should Look, Taste, and Smell Like a Helicopter." An IP (Army Instructor Pilot) once kidded me that, "If it don't smell like JP-4 (jet fuel), it couldn't be no good." Well, what is the objective of simulation? Is it to look, taste, smell, and feel like a helicopter? Let me answer that question with three examples: The Crew Station Research and Development Facility (CSRDF) at NASA Ames is an engineering simulator, right? Could you train somebody

Table 1. Myths in helicopter simulation: myth 1

Simulator should look, taste, and smell like a helicopter
Must determine the objective of the simulation
Crew station design/man-machine interface; CSRDF
Combat training technology/user requirements: SCTB
Primary training technology/train neophytes: UH-1TRS
Must define fidelity to meet the objective
AGARD Working Group (Key 1980)
1. "Objective fidelity" – simulator reproduces measurable aircraft states or conditions
2. "Perceptual fidelity" – degree to which Ss perceive the simulator to duplicate aircraft states or conditions
STI definition (Heffley et al. 1981)
"Simulator fidelity is the degree to which characteristics of perceivable states induce correct psychomotor and cognitive control strategy for a given task and environment"

in it? Sure you could, but that's not its purpose. It was created to address questions in crew station design. The Simulator Complexity Test Bed (SCTB at ARI, Fort Rucker) is a $24 million toy that is coming to Fort Rucker this year. Initially configured as an Apache, it has red-station/blue-station training capability beyond any helicopter simulator in existence. It is an ideal device for developing advanced combat training. It is a trainer, but it is more of a training research tool. It is not an engineering simulator, not directly.

Moving from the sublime to the ridiculous, how about "Cheap Charlie," the UH-1TRS? It is a trainer, pure and simple. You cannot start it, you cannot fly an ILS with it, it does not use fuel, that is, we don't currently drive the fuel gauge. But, it trains "hands and feets," neophytes, kids off the street. In other words, it has evidenced significant positive TOT (transfer of training) to the UH-1 aircraft using neophyte pilots as research subjects. I think we should keep the objective of a given simulation in mind as we review our ideas today.

Once we have decided on the objective of a given simulator, an associated issue is the question of simulator fidelity. I brought some of my favorite definitions that I think are worth reviewing (table 1). Dave Key, who was in the audience yesterday, was the key player, no pun intended, in the AGARD working group in 1980, when they distinguished between "perceptual versus objective fidelity." The issue here is, do you want to measure what the simulator does and compare it with the aircraft, or do you want to measure what the "bus driver" does and compare pilot responses from the simulator to the aircraft? I think the latter is more appropriate, at least from a trainer's perspective.

The definition we most commonly use at ARI is that set forth by Heffley and a cast of thousands at Systems Technology Incorporated (STI). STI did a report for us that defined fidelity as "the degree to which characteristics of perceivable states induce correct psychomotor and cognitive control strategy for a given task and environment." Although I worry about the word "correct," I think this definition is worth considering; it focuses on the bus driver and not on the bus.

While we are reviewing the issues involved in simulator fidelity, I think it is worthwhile to reconsider Vernon Carter and Clarence Semple (table 2). When I first read their definition of "error fidelity," I thought, what kind of nonsense is that? Any good psychologist knows about error-free learning. But then, I thought the definition and saw that it has several important advantages. Looking at the error distribution that students make in a simulator and in the aircraft places the focus on the behavior of trainees, with the ultimate goal being "good" performance in the aircraft. Although this definition is specific to training simulators and not engineering simulators, it does suggest a metric for simulator evaluation...training errors.

At the bottom of table 2, I've included a reminder from Ed Eddowes and Wayne Waag: "There is no compelling relationship between training effectiveness and fidelity/realism." That's the kind of statement I'd like to use as a final examination question. We could ask the students to react to it as either true or false and then write a short essay to support their choice. The students could get 100% credit for agreeing or disagreeing, depending on the strength of their arguments. I think I would disagree because training effectiveness is a practical definition of fidelity. If a simulator trains, it has fidelity...who cares what it looks like?

Table 2. Myths in helicopter simulation: more myth 1

Carter and Semple (1976)

"Error fidelity" – assumes objective is training

1. Trainees make same errors in simulator and aircraft
2. Relative frequency distribution of errors same in both simulator and aircraft
3. Effect of trainee errors on system performance is same in both simulator and aircraft

Advantages of concept

1. Focus on behavior of trainees
2. Recognizes ultimate goal – performance in aircraft
3. Suggests a metric – training errors

A reminder (Eddowes and Waag 1980)

"There is no compelling relationship between training effectiveness and fidelity/realism"

Table 3 suggests a second myth: "The engineering test pilot knows best." In this myth, the bad guys are not people like Roger Hoh, they are the green-suited simulator test pilots. We all know how the Army goes about accepting a helicopter simulator for training. There's a procedure called Operational Test-2 (OT-2) in which a would-be expert, usually a senior warrant officer with a lot of time in the airframe, is assigned to make subjective judgments regarding the simulator's handling qualities. I don't necessarily mean Cooper-Harper ratings but something more subjective than that. Then the software is "tweaked" to satisfy the judgment of the "expert pilot." This is the way simulators are accepted into the Army inventory.

Is there anything wrong with this approach? Yes there is! The smart folks at STI, Hogue, Jex, and Magdelano evaluated the Army's UH-60 simulator. The UH-60 simulator has a six degree of freedom (DOF) synergistic motion base, but the STI report noted that as a result of the OT-2, two of the degrees of freedom were "tweaked" entirely out of existence! Specifically, the simulator has only pitch, roll, yaw, and heave. It has no measurable sway or surge. The Army owns 18 UH-60 simulators, 17 in the field and one at the factory in Binghampton. And none of them exhibits more than four DOF. Is that what improving simulator fidelity is all about? It doesn't make sense to me. But, if we're going to attack this green straw man, let me offer an alternative.

Yesterday, Ken Cross (Anacapa Sciences) offered "backward transfer" as an empirical yardstick with which to evaluate existing simulators. Senior aviators performed emergency touchdown maneuvers in the AH-1 Cobra aircraft until they met published criteria. Then they flew the same maneuvers in the AH-1 flight simulator: 58% failed one or more maneuvers. The backward transfer ratios were relatively low, ranging from 0.16 to 0.43. Since the aviators had been qualified in the aircraft within the past few days, it is unlikely that they "forget" how to accomplish the maneuvers. It is more likely that the skill requirements in the simulator and the aircraft are not the same. As Ken Cross noted, the existence of positive TOT data does not necessarily mean that the simulator is effective. The OT-2 report on the AH-1 simulator (by Bridgers, Bickley, and Maxwell) cited some evidence of positive transfer to the aircrarft and yet look at the results of the backward transfer study. Positive TOT alone may simply reflect some procedural transfer to the aircraft while obscuring a substantial aerodynamic deficit that will limit the overall training efficacy of the simulator.

Can we improve on the subjective pilot opinion method of evaluating a simulator's effectiveness? I think so. Let's look at Stan Roscoe's transfer effectiveness ratio (TER) (table 4). As an example from our Cheap Charlie research, we took a random sample of 10 Army officer trainees and dragged them kicking and screaming into the UH-1TRS where we substituted 9 hours of simulator time for 9 hours of aircraft time. We trained them to published criteria in the simulator (three successive maneuver iterations that met the Flight Training Guide standard) and then we employed the same criteria on the flight line in

Table 3. Myths in helicopter simulation: myth 2

"The engineering test pilot (or SIP or Eagle Scout) knows best"

Army acceptance test procedure

1. Assign an "expert"
2. Subjective judgment of handling-qualities/training features
3. "Tweak" the software

Outcome (Hogue, Jex, and Magdaleno 1982)

1. UH-60FS has six DOF synergistic motion base
2. Only four DOF (no sway or surge)
3. Army has 17 fielded UH-60FSs with four DOF motion bases

Alternative approach: empirical yardstick to evaluate existing device – backward transfer

Example (Kaempf and Blackwell 1990)

1. Trained to criterion in AH-1 Cobra (ETMs)
2. Flew AH-1FWS: 58% failed one or more maneuvers
3. Backward transfer ranged from 0.16 to 0.43
4. Demonstrates skill requirements different in aircraft and simulator

the aircraft. We compared them with a control group of students who did not have simulator training. I've included a kind of "middling" example, traffic pattern flight. We found that the control group required about 21 maneuver iterations to meet the standard whereas the experimental (simulator-trained) students required about 13. That savings of about eight maneuvers on the flight line can be divided by the "cost" of producing the savings: about 13 iterations in the simulator. Thus, the TER for that manuever is 0.60. This could be interpreted as meaning that the simulator was about 60% as effective as the aircraft, using the aircraft as the criterion measure. This metric has the advantage of measuring "in vivo" training effectiveness of actual flight students embedded in the Initial Entry Rotary Wing (IERW) training program.

Let's look at the final myth: "The more features the better" (table 5). Here, we tread on some hallowed ground. My favorite example, after 14 years at Fort Rucker, is in the area of motion-base requirements. I recall the Singer-Link folks telling the Army that the cost of a simulator motion base adds only 2% to the total device cost. To evaluate that assertion, I'd like to develop ROC (Required Operational Characteristics) requirements for the LHX simulator specifying no motion base and then, on the day of the best and final offer, add 2% to the contract and say we changed our minds!

Anyway, the draft Advisory Circular 120-XX that Dean Resch and I talked about requires a motion system for acceptance, even for level A. Is there any evidence that motion even contributes to training, let alone is required for training? We've done two small-number empirical evaluations at ARI using neophyte trainees, one in 1984 using five students on motion and five without motion, and one in 1990, with six on motion and six off. All students were strapped in the simulator; we erected the motion base in every case and students were not informed (nor did they guess) that we were evaluating the effects of motion on training. In both experiments, the nonmotion students outperformed the motion students. Now that evidence only pertains to Army *ab initio* (that's a Latin phrase for "kids off the street") trainees learning basic hovering and traffic pattern skills. However, our

Table 4. Myths in helicopter simulation: more myth 2

To evaluate developing technology: transfer effectiveness ratio (TER)
Measure transfer of training "in vivo" – embedded in training program
Random sample of trainees
"Blind" evaluation of flight line – same criterion
Calculate TER:
$TER = (C_a - E_a)/E_s$
Example from UH-1TRS – traffic pattern:
$TER = (20.7 - 13.2)/12.6 = 0.60$

Table 5. Myths in helicopter simulation: myth 3

"The more features the better the simulator"
Motion base – the 2% myth
Draft AC-120-XX requires a motion system, even for level A
Small N research suggest motion may inhibit training
Instructional support features – unused/unusable
Auto co-pilot
Auto check ride
Recorded demonstrations
AAA reviews (1982, 1985)
Insufficient training data to justify acquisitions
Recommended training requirements – empirical basis
Identified no "blade hour" savings

research agrees with the literature in finding no significant training advantage for a motion base.

Another example would be Instructional Support Features (ISFs). I'm short on time and won't discuss these but table 5 lists three examples from the 2B24 Huey instrument flight simulator that either don't work, are virtually never used, or have been recently taken off-line by the Army. Couldn't we have based the simulator features on a research evaluation of the requirements instead of just buying all the bells and whistles the manufacturer could offer?

My third example of simulator features requires that I bend logic a bit. In 1981, the Army Audit Agency (AAA) came to Fort Rucker and evaluated simulator utilization. Their 1982 report noted that the written premise for procuring flight simulators had been "blade hour savings." The folks from the AAA looked around Rucker and couldn't find the money! The Command Group's answer was that there was no intent to reduce flight hours but that simulators were training multipliers. There's nothing essentially wrong with viewing simulators as adjuncts to "blade hour" training, except perhaps the inherent dishonesty. The AAA made two recommendations: first, that Fort Rucker needs more training data to justify further simulator acquisitions and second, that something as expensive as Army aviation training should have an empirical basis. Actually, the AAA said that Fort Rucker can have simulators to experiment with in "the schoolhouse" but that procurements of simulators for the field would be carefully scrutinized for appropriate training requirements analyses and for empirical means of establishing simulator effectiveness. I think it's embarrassing to have a bunch of auditors come around and tell the trainers how to do their business. But it makes the point that simulators should be designed, evaluated, and procured for effectiveness and not for a bunch of "gee whiz" features.

So, what would we propose as an alternative? Again, if you want to stir up a hornet's nest, you'd better have a bug bomb. The philosophy behind our suggested approach is to do a thorough, boring, tedious front-end analysis to determine the training requirements based on the ultimate criterion of mission readiness in the field (table 6). At an initial level, that's not all that difficult to accomplish since the Army's Directorate of Evaluation and Standardization (DES) sends flight-skill evaluators worldwide for no-notice evaluations. Thus, it should be relatively easy to determine where the basic mission-readiness training deficits are. Then, a cost-of-training-effectiveness-analysis (CTEA) could be used to compare the training cost of simulation, blade time, or a combination. In summary, if you don't have a problem training the maneuver or mission in the aircraft, don't design a simulator to train it.

We can also try to design our simulators to be more flexible...to anticipate future requirements. We don't want to perpetuate the Army way: procure by publishing requirements, discover that the requirements won't get the training job done before the device is even fielded, initiate a Product Improvement Program (PIP) to modify the device to do what you originally intended (but didn't ask for). The PIP system makes the Army look dumb and the contractors look wealthy. We should be able to do better. Can't we develop requirements with an eye to the future? Can we design part-task trainers and modular simulator designs in place of plenary simulators that are designed as aircraft replacements? Can we do CTEAs to estimate the effectiveness of simulators before we buy them?

Table 6. Myths in helicopter simulation: more myth 3

Alternative philosophy
1. Perform a front-end analysis Training requirements: assess mission readiness in field CTEA: Compare aircraft and simulator efficiency
2. Design/construct modular simulations – flexibility to meet changing requirements Design for spare capacity – hardware/software Use TOT evaluations of training effectiveness; iterate design
3. Design/construct part-task trainers instead of plenary simulators Design to meet training requirements Iterate design

So let me go to my last point. In our shop, we call our simulator Cheap Charlie because we don't want to be taken too seriously, but also because we want to emphasize that it's a low cost training tool and not a surrogate aircraft. In a similar vein, I'll call our approach the ACME "Fly by Night" Simworks to try to keep our attention directed to doing useful and meaningful research related to our charter; low-cost entry level helicopter flight training (table 7). I apologize for the pedestrian acronym, ACME, but it may serve to keep our attention focused on our research goals. Perhaps it has value to other simulator designers, researchers, and users as well.

Table 7. Myths in helicopter simulation: still more myth 3

ACME "Fly-by-Night" Simworks and Oyster Bar
Analyze – does it meet requirements? (CTEA)
Combat – does it address Army mission?
Modular – is the design flexible?
Evaluate – does it train? (TOT)

My wife and I don't watch much television but we have come afficionados of the network show, Twin Peaks. My hero, Special Agent Cooper, has a new enamorate...a woman recently released from a convent, that is, an ex-nun. Given her status, he decided to woo her with a joke about penguins. There were two penguins on an ice flow in Antarctica and one turned to the other and said, "You look like you're wearing formal evening wear." The other penguin said, "Maybe I am." The connection to fidelity in simulation is obvious, right?

MR. DAVE GREEN: Just a quick observation with which you can agree or disagree regarding your comments about motion. I think what we say is that bad motion is worse than no motion. When somebody tweaks a machine to make motion, it was probably pretty bad motion. When you get the kind of training you get by taking motion out, it is because motion was a negative training feature. Would you agree or disagree?

MR. DOHME: Well, I would pass the baton. The question is, regarding our getting worse training with motion than with no-motion, Mr. Green is saying that the issue is probably that bad motion is worse than no motion. I would agree that perhaps bad motion is worse than no motion at all. However, we probably had a most thorough evaluation of the motion system on the UH-1TRS by the University of Alabama Flight Dynamics Laboratory (FDL). The FDL engineers analyzed and tweaked our motion system and wrote a thorough report on their efforts and I would refer you to that report since I'm not an engineer.

The FDL engineers were convinced that our motion base was doing as well as it could, given the limitations inherent in simulating the motions of flight. For example, the issue of washout. Is it subliminal or not? It wasn't that we were naive regarding the issue of motion base fidelity, and we did have simulator-experienced engineers develop and tweak our motion-base equations as best they could. I would be happy to provide a copy of the report; I think it was done right.

MR. FRANK CARDULLO: I would like to follow up on that comment a little bit further. Virtually every transfer-of-training study that has been done about motion has indicated that there has been no transfer. Unfortunately, though, just about every transfer study on training of motion has been done on bad motion systems. You admitted yourself there were two degrees of freedom missing.

MR. DOHME: No, not on the motion system we used with the UH-1TRS. All five were working; as a matter of fact, we had sway, which, it turns out, the original 2B24 doesn't have.

MR. CARDULLO: But, nevertheless, that one is a fairly archaic motion system and the performance is poor, and the cueing-out rhythms are poor. That has been virtually true of all the motion-transfer-of-training studies. I think good motion-transfer-of-training studies should be done, and I wish the impetus would come from the Army or from your organization in particular to do a good transfer-of-training study on a good motion system.

MR. KATZ: Good suggestion, Frank. I am not here to comment on the work that the laboratory did previously, but again, along this same vein, because obviously your talking invites these comments, let me first of all note that you did not say anything abut the effect of motion on backward transfer. And you see you had the problem of backward transfer, I assume, with motion.

DR. DOHME: Yes.

MR. KATZ: And then you had a problem with forward transfer with the motion so it invites the hypothesis that the bad motion as a matter of fact caused this. And the thing that I think ought to be studied is to see if the backward transfer would also improve by eliminating this motion. And then I would make the hypothesis that if you get your engineering work up to the level where the backward transfer would be good with the motion that in this case also the forward transfer would be good with that motion.

MR. DOHME: Interesting hypothesis. Those, of course, are different vehicle we used for those two studies. The backward-transfer work was done in Germany. It's difficult to do that kind of in vivo testing in an active military unit, but it is a good idea: A motion versus no-motion backward transfer study.

MR. GERDES: My background is about 25 years of simulation at Ames, ever since we had first fixed-base and then motion-based simulators. And I have extensive experience on our five, six and three degree-of-freedom simulators. I'm only saying this to give you some qualifications for what I am about to say.

First of all, I agree very, very highly that no motion is better than bad motion. That is what we have been saying for years. Second, motion comes into play or is useful in an engineering simulator, perhaps more so than in a training simulator, where you are looking for, say, the six and one half boundary, the boundary where controllability or emergency control of the aircraft or helicopter is important. Then motion feedback to the pilot is extremely important for the engineering pilot to assess what the control problem might be. Third, about five years ago, I participated in a simulation on our VMS, which has plus or minus 30 feet of vertical travel. It is a six degree-of-freedom and we did an autorotation simulation. I think it was for this particular theme we are looking at, but for the Army; in other words, are simulators useful for training? And the autorotation maneuver was critical, extremely hard to perform and learn and so forth. That one simulation was probably the one that stands out most in my mind as to where motion, and it was good motion, played a very, very important part in this training business.

I was able, with practice, to make a whole series of autorotations down to a fairly reasonable area and this is a vertical motion simulator. So you have this stress that others here have talked about. There is a simulator you can break, so you try very hard. With the sound system we had, we were able to give the pilot cueing for the rotor sounds. When we pitched up to flare, we got the motion travel to give us the deceleration and we had to doctor up the visuals a little bit. We had to put in a couple of vertical towers for visual height perception.

We did have a fourth window, a chin bubble, we could see through as you could on a Huey. We could do some fantastic things as far as accurate touchdowns are concerned. This was not training, this was an engineering simulation in which we varied disk loading, weights, winds, all of these things. We did a whole matrix of autorotations under difficult conditions, and all of them turned out really well and defined boundaries and so forth.

I am saying all this because motion, when properly used, is very good for training, as well as for engineering simulations.

Jack A. Dohme has a B.A. degree in psychology from Baldwin-Wallace College, and M.S. and Ph.D. degrees in experimental psychology from the University of Arizona. He has taught undergraduate psychology at the universities of Arizona and Maryland and at Concordia College and Grinnell College, and has worked as a senior research consultant at Stanford Research International. His primary research interests are aviator training, flight simulation, and human performance measurement. Dr. Dohme is a licensed fixed-wing and helicopter pilot.