Journal of Air Transportation Vol. 7, No. 3 – 2002

THE RELATIONSHIP BETWEEN FIDELITY AND LEARNING IN AVIATION TRAINING AND ASSESSMENT

Cliff Noble
Embry-Riddle Aeronautical University
Daytona Beach, Florida

ABSTRACT

Flight simulators can be designed to train pilots or assess their flight performance. Low-fidelity simulators maximize the initial learning rate of novice pilots and minimize initial costs; whereas, expensive, high-fidelity simulators predict the real-world in-flight performance of expert pilots (Fink & Shriver, 1978; Hays & Singer 1989; Kinkade & Wheaton, 1972). Although intuitively appealing and intellectually convenient to generalize concepts of learning and assessment, what holds true for the role of fidelity in assessment may not always hold true for learning, and vice versa. To bring clarity to this issue, the author distinguishes the role of fidelity in learning from its role in assessment as a function of skill level by applying the hypothesis of Alessi (1988) and reviewing the Laughery, Ditzian, and Houtman (1982) study on simulator validity. Alessi hypothesized that there is a point beyond which one additional unit of flight-simulator fidelity results in a diminished rate of learning. The author of this current paper also suggests the existence of an optimal point beyond which one additional unit of flight-simulator fidelity results in a diminished rate of practical assessment of nonexpert pilot performance.

INTRODUCTION

Fidelity is a concept that expresses the degree to which a simulator or simulated experience imitates the real world. It has been viewed as a critical variable in the design of both mechanical simulators and computerized simulation experiences. For years, the aviation-training community has held fast to the belief that a high level of fidelity is required to produce the highest level of transfer of learning to the actual equipment. This concept was driven by intuitive appeal, as exemplified by Klauer (1997) in the following text: “The closer a flight simulator corresponds to the actual flight environment (i.e., high physical fidelity), the more skills will transfer to the aircraft” (p. 13). This current paper provides evidence that supports

---

Cliff Noble currently teaches aviation English and conducts human-factors research at Embry-Riddle Aeronautical University in Daytona Beach, Florida. Cliff holds an EdD in Technology and Learning and master’s degrees in Teaching English to Speakers of Other Languages and Business Administration.

the viewpoint that this common belief may not always be true for all learners in all cases involving training and assessment in flight-simulation devices. Furthermore, it distinguishes the level of fidelity required of simulation devices designed to optimize the transfer of learning throughout the training cycle from that required of simulation devices designed to assess performance in the actual aircraft.

The total-fidelity concept may be most appropriate for the training and assessment of expert pilots who readily identify and process all the visual, aural, and other contextual cues of real-world aviation tasks. Novice pilots can become overwhelmed with total fidelity. Implemented in part-task emergency trainers, however, high levels of fidelity that are limited to actual equipment in the cockpit, excluding the fidelity of the real-world, out-of-the-cockpit environment, may be quite effective for novice pilots. This is because, initially, novice pilots must first familiarize themselves with the look, shape, location, and feel of the actual devices in the cockpit to aid in the memory and execution of emergency procedures. The procedures should become second nature for survival. Most flight-training programs require memorization of these procedures before actual flight, reflecting motivation, safety consciousness, and good piloting habits. An example might be extinguishing a simulated engine fire or responding to a simulated engine failure. If the novice pilot cannot perform the maneuver on the ground, then grave difficulty performing the procedure in the air, where dynamic situations require more attentional resources, can be expected.

The fidelity of a real-world flight environment may detract from, rather than enhance, the performance of a novice pilot (Miller, 1974). This stands to reason because, when it comes to assessment in the real world, expert pilots are expected to react accurately and efficiently, whereas novice pilots are expected to make frequent mistakes in the learning process. It can therefore be deduced that high fidelity is desired in simulation-based assessment devices that propose to predict expert pilot performance in real-world situations; however, the same may not hold true for the practical assessment of pilots with skill and experience levels falling between novice and expert. Moreover, high levels of total fidelity may be of little value for enhancing the transfer of learning of novice pilots, except for limited procedural checklists in part-task trainers. With part-task trainers, novice pilots can build confidence in procedural knowledge, while enhancing safety and learning from mistakes (Feifer, 1994).

Empirical evidence on the relationship between the degree of flight simulator/simulation-device fidelity and learning transfer and learning rate can be misleading if the reviewer fails to carefully scrutinize the learning stage of the participants in the experiments. Failure of researchers to consider the learning stage(s) of the sample population may corrupt simulator/simulation-fidelity studies that (a) propose to predict participant performance on the operational equipment in the real world, or (b) propose to measure the ability of a simulation device to transfer learning to actual operational equipment in real-world operations.

FIDELITY RESEARCH

Alessi (1988) and Valverde (1973) outlined some of the major studies providing empirical evidence on the relationship between the degree of flight-simulator fidelity or simulation-device fidelity and learning rate and transfer of learning. For example, Wolfe (1978) found that medium fidelity is better for learning than low fidelity in business simulations where the degree of complexity of the business simulation concurrently represents the degree of fidelity. Roscoe (1971, 1972) and Povenmire and Roscoe (1973) found that initial training in a flight simulator was more efficient than in actual aircraft, up to a point, after which transfer of learning began to decrease.

Cox, Wood, Boren, and Thorne (1965) and Grimsley (1969) discovered there was no difference in learning rate and transfer of learning in mechanical flight simulators with different degrees of fidelity. Similarly, Hopkins (1975) discovered that motion fidelity in mechanical flight simulators had no significant effect on learning. Koonce (1974), however, clarified that motion fidelity in mechanical flight simulators holds a measure of importance for expert pilots, but no value for novice pilots. These few studies are examples demonstrating the importance of confirming the learning stage of each study participant before generalizing the findings of any research or specific relationships between degree of fidelity and learning, transfer of learning, and the ability of a simulator/flight-simulation device to predict performance in the real world on actual operational equipment. It is important that such verification is specifically addressed in the findings of any related study.

General-Aviation Trainer Effectiveness

Povenmire and Roscoe (1973) conducted research on the effectiveness of the incremental transfer of a ground-based general-aviation trainer (GAT)-the Link GAT-1. The study sought to assess the cost effectiveness of training novice student pilots for private-pilot certification in the Piper Cherokee PA-28-140B trainer aircraft. The practical issue of determining the amount of training that would be required on a low-fidelity ground-based simulator to reach a marginal rate of return and training effectiveness, in terms of time and cost through student achievement of private-pilot certification, was addressed. The ultimate purpose of the study was to determine the point beyond which the Link GAT-1 ground-simulation device became inefficient in terms of cost for optimizing transfer of learning for private-pilot certification of novice pilots in the Piper Cherokee PA-28-140B aircraft. Low fidelity characterized the Link GAT-1 trainer, which was used to transfer the skills of novice pilots to high-fidelity operational equipment in real-world airspace. Consequently, the issue of fidelity and the associated transfer of learning for novice pilots was a focus of the research. However, no further generalizations or applications to intermediate or expert pilot-skill level should be gleaned from the Povenmire and Roscoe study.

The sample population of the Povenmire and Roscoe (1973) research consisted of 65 inexperienced student pilots who completed a private-pilot Aviation 101 course at the university serving as the study site. The gender and age of the participants were not disclosed. They had no prior flight instruction and were considered novice student pilots. The study population was divided into one control group and three experimental groups. Participants within the three experimental groups were selected from six regularly scheduled flight-operation class periods offered by the institute of aviation within the participating university. They were then randomly assigned to primary flight instructors and to one of the three experimental groups—Group 3, Group 7, or Group 11—also referred to as the *transfer groups*. The control group received no training in the Link GAT-1 simulation device. The experimental groups received 3, 7, and 11 hours of training, respectively, in the simulation device. Both the control group and the transfer groups received routine training in the Piper Cherokee PA-23-140B aircraft until completion of their flight training. Only data from the 65 participants who successfully completed flight training were used to determine the effectiveness of the transfer from the Link GAT-1 simulation device to the Piper Cherokee PA-23-140B aircraft. Transfer effectiveness was determined by comparing the total time required to train each participant within both study groups.

The routine flight syllabus used in the Povenmire and Roscoe (1973) study was characterized by incremental 10-hour flight evaluations in the Piper Cherokee PA-23-140B aircraft, as well as a final recommendation by the primary and secondary flight instructors confirming the readiness of the student for the private-pilot check-ride. The primary and secondary instructors would typically fly together with the student on a single flight to assess the suitability of their joint recommendation. The instrument used to evaluate flight performance was the Illinois Private Pilot Performance Scale (Povenmire, Alvarez, & Damos, 1970). This scale is claimed to have an observer-to-observer reliability of .80 (McGrath & Harris, 1971; Selzer, Hulin, Alvarez, Swarzendruber, & Roscoe, 1972). The instrument was used in conjunction with the Federal Aviation Administration Practical Test Guide for Private Pilot Certification (U.S. Department of Transportation, Federal Aviation Administration, 1970). Ten maneuvers were scored for each student participating in the Povenmire and Roscoe study who was recommended for the check-ride. The maneuvers were scored at each incremental, 10-hour stage check. Equal weight was assigned to each maneuver on each check. Maneuver performance measures were based upon four to six variables that could be quantified by the primary flight instructor on all of the stage checks. For example, if the student deviated beyond the maximum 10-degree-of-heading parameter, the maximum deviation would be recorded.

The instructors pooled the scores of each maneuver from the preterminal recommendation flight for all those confirmed as ready for the terminal flight. Passing scores were tallied by the maximum amount of deviation made from the predefined parameters for each of the 10 maneuvers. From this pool of scores, a standard deviation was calculated for each maneuver. Subsequently, a modified *z* score was assigned for each maneuver by dividing the deviation criteria by the standard deviation established from the pool of passing scores. The mean *z* score was then calculated for each incremental, 10-hour check flight, for each student, up to the final check flight. The *z* scores of each 10-hour stage checkpoint were plotted on a chart in a straight line; specifically, between each 10-hour, 20-hour, 30-hour, and final check flight for each student. From this chart, a straight line was calculated for each member of the pool. The average of all the scores of the recommended students from the control group and the three experimental groups was used as the private-pilot flight criterion.

Table 1 reveals the number and distribution of students who completed fight training in the Povenmire and Roscoe (1973) study. Table 2 reveals the specific amount of flight time required for students within both the control group and experimental groups to pass their terminal flight checks for certification as private pilots. Table 3 reveals the flight time in hours that students in both study groups accumulated to attain the private-pilot proficiency criterion (i.e., the *z* score). Table 4 reveals the results of the analysis of variance (ANOVA) determining the number of hours of flight time the successful students of both groups accumulated to pass their terminal check-rides. This analysis was conducted independently (i.e., the control group without Link GAT-1 training and the three experimental groups with 3, 7, and 11 hours of training in the Link GAT-1 simulation device, respectively) with unequal numbers of students. Table 4 also reveals that the average flight times at which participants in both study groups passed their terminal flight checks differed both orderly and

Table 1. Flight-Training Completion Rates

Group	*Total students*	*Students passed*	*Students failed*	*Percentage passed*
Control	20	14	6	70
3 hours in GAT-1	14	13	1	93
7 hours in GAT-1	14	9	5	64
11 hours in GAT-1	17	10	7	59
Totals	65	46	19	71
Nonexperimental	20	17	3	85
All students	85	63	22	74

Note. GAT = ground-based general-aviation trainer. Adapted from "Incremental Transfer effectiveness of a Ground-Based General Aviation Trainer," by H. Kingsley Povenmire and Stanley N. Roscoe, 1973, *Human Factors*, *15*(6), p. 537. Copyright 1973 by the American Psychological Association. Adapted with permission.

Table 2. Flight Hours Needed to Pass Final Check and Summary of Resulting Transfer Measures

Data Type	*Control Group*	*Transfer Groups*		
	0	*3*	*7*	*11*
Hours in GAT-1	0	3	7	11
Hours in Cherokee	41.3	44.8	42.7	37.3
	45.6	44.8	42.7	37.5
	48.0	47.5	40.2	40.7
	49.0	44.3	43.3	39.6
	46.0	40.6	42.5	34.8
	43.3	25.6	42.8	35.8
	43.7	32.4	35.8	40.1
	53.7	43.2	35.0	37.1
	41.2	36.8	28.2	34.8
	41.6	39.3		41.6
	51.2	39.0		
	38.0	40.1		
	50.8	45.0		
	42.5			
N	*14*	*13*	*9*	*10*
$\bar{X}$	45.42	40.26	38.62	37.93
σ	4.51	6.00	5.07	2.45
Cumulative Savings		5.16	6.80	7.49
Incremental Savings		5.16	1.64	0.69
Transfer (%)		11.00	15.00	16.00

Note. GAT = ground-based general-aviation trainer. Adapted from "Incremental Transfer effectiveness of a Ground-Based General Aviation Trainer," by H. K. Povenmire and S. N. Roscoe, 1973, *Human Factors*, *15*(6), p. 538. Copyright 1973 by the American Psychological Association. Adapted with permission.

reliably, as indicated by a probability factor (p = .0014) based upon a summary of the ANOVA for the independent groups. Table 5 reveals the results of the ANOVA determining the flight times at which successful students in both the control and experimental subgroups achieved the private-pilot performance criterion. The difference between the mean times calculated for the participants of the control and transfer groups to reach the performance criterion and pass the terminal check-ride was not statistically significant.

Table 3. Flight Hours Needed to Reach Proficiency Criterion and Summary of Resulting Transfer Measures

Data Type	*Control Group*	*Transfer Groups*		
	0	*3*	*7*	*11*
Hours in GAT-1	0	3	7	11
Hours in Cherokee	29.54	47.59	33.78	35.62
	47.23	39.88	41.52	30.55
	42.64	60.00	41.94	43.76
	42.26		38.88	34.45
	37.71	45.54	57.74	33.99
	34.32	23.56	47.56	28.93
	45.46	25.74	37.70	34.46
	40.48	38.82	25.87	59.27
	50.40	41.54	19.46	32.33
	46.15	38.65		
	52.24	34.48		
	39.41	36.75		
	70.56	46.29		
	42.5			
N	*14.00*	*12*	*9*	*10*
$\bar{X}$	44.49	39.90	38.27	37.30
σ	9.64	9.76	11.28	8.82
Cumulative Savings		4.59	6.22	7.19
Incremental Savings		4.59	1.63	0.97
Transfer (%)		10.00	14.00	16.00

Note. GAT = ground-based general-aviation trainer. Adapted from “Incremental Transfer Effectiveness of a Ground-Based General Aviation Trainer,” by H. K. Povenmire and S. N. Roscoe, 1973, *Human Factors, 15*(6), p. 539. Copyright 1973 by the American Psychological Association. Adapted with permission.

Table 4. Analysis of Variance in Times Final Flight Check Passed

Sources of variance	*df*	*MS*	*F*	*p*
Hours in GAT-1 (Groups: 0, 3, 7, 11)	3	141.97	6.19	.0014
Participants/Groups	42	22.93		
Total	45			

Note. GAT = ground-based general-aviation trainer. Adapted from "Incremental Transfer Effectiveness of a Ground-Based General Aviation Trainer," by H. K. Povenmire and S. N. Roscoe, 1973, *Human Factors, 15*(6), p. 539. Copyright 1973 by the American Psychological Association. Adapted with permission.

Table 5. Analysis of Variance in Times to Reach Private-Pilot Performance Criterion

Sources of variance	*df*	*MS*	*F*	*p*
Hours in GAT-1 (Groups: 0, 3, 7, 11)	3	124.82	1.29	.2914
Participants/Groups	42	96.95		
Total	45			

Note. GAT = ground-based general-aviation trainer. Adapted from "Incremental Transfer Effectiveness of a Ground-Based General Aviation Trainer," by H. K. Povenmire and S. N. Roscoe, 1973, *Human Factors, 15*(6), p. 539. Copyright 1973 by the American Psychological Association. Adapted with permission.

Implications. There were no standardized instructor lesson plans for the Link GAT-1 simulation device documented in the Povenmire and Roscoe (1973) study. The implication here is that, due to a lack of well-defined, standard operational procedures for the type and quality of training in the Link GAT-1 simulation device, the data collected in the study may have been compromised. This is because the data may have reflected the degree of instructor effectiveness with the students in the experimental groups, rather than the degree of simulator effectiveness and transfer of learning from the Link GAT-1 simulation device to the Piper Cherokee PA-23-140B aircraft. This may account for the inverse relationship between the percentages of students who passed within each experimental group and the number of hours each experimental group was exposed to the Link GAT-1 simulation device (see Table 1). The researchers commented that a chi-square test indicated a probability coefficient of 0.5 for the differences in the success ratios among the control group and the three experimental groups; however, this observation is moot. The important point is that, if the type of treatment received by each experimental group in the Link GAT-1 simulation device had been controlled, the probability of differences in success ratios among the control group and the three experimental groups as factors of chance may have been reduced. Furthermore, most instructors are not taught how to use the simulator as an effective instructional tool, which could have also affected the results.

Accurate assessment of the amount of student learning that transferred from the Link GAT-1 trainer to the Piper Cherokee PA-23-140B aircraft was partially dependent upon the point at which student-learning curves intersected precalculated criterion levels of private-pilot performance. It was also determined by the number of flight hours required to pass the terminal check-ride in the Piper Cherokee PA-23-140B aircraft. The reason the two measures were proposed was because of the varying learning rates among the participants. The implication is that this variation must be controlled to ensure that the findings measure transfer of learning rather than learning rate. To account for this variable, a least-squares criterion straight line was "fitted to all the check scores each student received on the Illinois Private Pilot Performance Scale throughout training" (Povenmire & Roscoe, 1973, p. 537).

Finally, a gradual reduction in the effectiveness of the Link GAT-1 simulation trainer used in the Povenmire and Roscoe (1973) study, in terms of transfer of training, was evident. The implication is that, as the skill level of the learner improves, low-quality fidelity devices become less effective in terms of the funds invested to build them versus their training efficiency. The larger implication here, however, is that the level of fidelity in flight-simulator devices built to transfer learning to real-world operational tasks in real-world operational airspace may need to be adjusted to the learning stage of the respective pilot for optimal transfer. Furthermore, degree of fidelity, learning stage of the student, and the goals of the training device are not mutually exclusive.

It is evident from the Povenmire and Roscoe (1973) research that the learning stage of students must be clarified, controlled, and monitored throughout an experiment before applying the evidence to practical use in training pilots or assessing pilot performance on flight-simulation devices. Without this understanding, unsound generalizations can potentially be made that could result in impractical expense, especially due to the high cost of fidelity (Miller, 1953). Furthermore, the potential for implementation of unsound generalizations may radiate to professional educators, psychologists, and cognitive engineers who could mistakenly apply such findings to the learning/assessment process. Therefore, the learning stages of student pilots must be clearly distinguished when comparing empirical evidence on studies proposing relationships between the degree of fidelity in flight-simulator devices and transfer of learning, learning rate, and prediction validity of student performance on actual operational equipment in the real world.

Training fallacies. Schneider (1985) provides empirical evidence and an excellent overview of the training fallacies that can potentially result from unsound generalizations. One of these fallacies is that practice always makes perfect. This is not always true. For example, in the flight domain, novice pilots must develop time-sharing skill, which allows them to efficiently divide their limited attentional resources to the many tasks encountered both inside and outside the cockpit. By optimizing performance of a single task, novice pilots can inadvertently fixate on a single component of a time-shared task. This, in turn, can inhibit the division of attention that is required of a time-shared task (e.g., scanning instruments during a flight maneuver). In this manner, a negative transfer of learning can occur, in terms of the critical time-shared skills novice pilots must develop to achieve acceptable levels of flight proficiency.

Antithetically, the fallacy that total-task training is required for maximal transfer of learning may be true for the expert pilot who is familiar with high-fidelity environments and who can only improve his or her learning level through challenging and somewhat unfamiliar flight scenarios accompanied by demanding flight tasks (Schneider, 1985). Transfer of learning through these high-skill tasks may lead to automation and further reduce workload. However, the same might not be true for the intermediate pilot who is more unfamiliar with such tasks. In this case, the intermediate or novice pilot could falter in performance (Wiggins, 1997).

Another training fallacy is that extrinsic motivators for expert pilots inhibit concentration (Schneider, 1985). This suggests that external stimulus will always interfere with experts who perform tasks requiring heavy concentration. Conversely, boredom sometimes accompanies tasks of repeated concentration because the task is a familiar one. An interfering stimulus may provide the extra level of difficulty that sparks a challenge within the expert, along with greater and more efficient concentration. Schneider also revealed the fallacy that the primary goal of skill training is accurate performance. This cannot be true of air-traffic controllers who must focus their attention on the general separation of aircraft while concurrently attending to the accuracy of pilot readbacks while the pilots are flying on final approach to landing. Although it is important for the expert controller to attend to the accuracy of pilot readbacks, the real-world mission is to ensure the separation of aircraft.

Another fallacy is that the conceptual understanding of systems that is acquired in the classroom will develop needed performance skills within the flight domain. Although the conceptual understanding of systems obtained within a training program may enhance procedural knowledge, the time-sharing skills required of pilots can only be developed via hands-on experience. The fallacies documented by Schneider (1985) should always be considered when conducting experiments where degree of fidelity, learning rates, and learning stages are pivotal factors.

The Alessi Hypothesis

From an intuitive viewpoint, it would appear that the higher the level of fidelity in flight simulators and in flight simulation, the higher their prediction validity would be for pilot performance on operational equipment in real-world airspace (i.e., check-rides in actual airspace on operational equipment). It is tempting to carry this hypothesis one step further and deduce that the closer mechanical and computerized simulators can emulate the real world, the more efficiently they can train and aid in the transfer of learning to the actual equipment in real-world operational scenarios. Such a deduction, however, could be misleading without consideration of the stage of the learner. The learner is an integral part of the machine-environment system.

Alessi (1988) clearly illustrated the role of fidelity within different learning stages. He hypothesized the existence of a marginal rate of return on learning and fidelity based upon the stage of the learner. The law of diminishing returns states that a point exists beyond which one additional unit of simulation fidelity results in a diminished rate of return on investment. Figure 1 prompts the following question: How much fidelity should be programmed into a simulation experience or built into a mechanical simulator? Alessi (1988) proposes that the degree of fidelity on a computerized simulation experience should match the goal and the training stage of the learner. Miller (1953) originated this viewpoint and his original terminology for fidelity was *degree of simulation*. He hypothesized the existence of a relationship between the degree of learning transfer, cost, and engineering simulation. He recognized that the higher the degree of fidelity, the higher the cost of the training device. Furthermore, Miller recognized that the more familiar students became with a simulator or simulation device designed for transfer of learning, the greater amount of fidelity they needed to sustain adequate transfer-of-learning rates (i.e., positive transfer). Hays and Singer (1989) pointed out that “task types and the trainee’s level of learning, as well as other variables, interact with Miller’s hypothesized relationships” (p. 31). The viewpoint espoused by Alessi (1988) is that fidelity is only critical in terms of how much should be used in flight-simulation experiences, not necessarily that high amounts of fidelity are needed for all learners in all cases. Students may benefit from increased amounts of fidelity as their training progresses.

Alessi and Trollip (1991) proposed the following four stages of effective instruction: presentation, guidance, practice, and assessment. Each stage of instruction should present increasing degrees of simulation fidelity

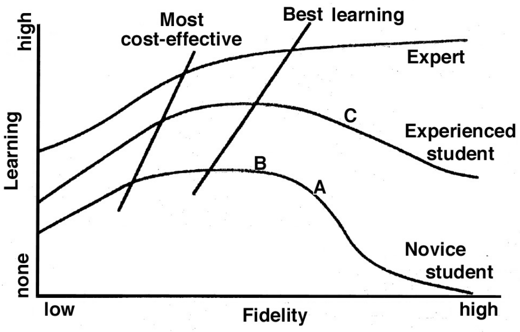

Figure 1. This illustration displays the relationship between degree of fidelity and learning for novice, experienced learners, and expert learners. (Alessi, 1988, p. 42).

constrained by return on investment and the stage of the learner. Regarding the assessment stage, Gagne (1954) suggested that the highest level of fidelity should be reserved for measuring (i.e., assessing) performance of expert pilots. He recognized the existence of diminished rate of return on the learning rate of expert pilots with increased fidelity alone and, furthermore, that expert pilots require high levels of fidelity and difficult tasks to enhance transfer of learning. The implication here is that there is a point beyond which training devices fail to sufficiently motivate experts, even with high degrees of fidelity, if the design of the simulator and/or simulation device fail to sufficiently challenge the ability of the individual to handle novel tasks of increasing difficulty. On the other hand, novice pilots may be overburdened or confused by excess fidelity and/or a training task that is overly difficult for learning or assessment purposes. Therefore, interface designers, educators, experimental psychologists, cognitive engineers, and other aviation experts must weigh the state and training stage of the learner when determining the extent of fidelity to program into mechanical simulators and computerized simulation devices (Flach, Hancock, Caird, & Vicente, 1995).

The information that expert, intermediate-level, and novice pilots process is not always the same; therefore, the spare capacity of limited attentional resources for each piloting-skill level will not be the same for all tasks. What is overwhelming for the novice pilot may be handled with ease by an expert who will have more spare capacity to attend to other tasks upon completion of a given task or set of tasks. Antithetically, the novice pilot may fail to process certain visual and aural cues that would induce added workload for the intermediate- or expert-level pilot. This could prove disastrous in situations that require accurate and efficient processing of critical information for flight safety. Clearly, simulators and simulation devices must be designed for the learning stage of the learner. Consequently, it is imperative to distinguish the roles of fidelity for training and assessment by the goals of the simulator or simulation device *and* the stage of the learner when using such devices as learning and assessment tools. Finally, the concept of learning and assessment must be viewed as complementary to training.

The Training Cycle and the Learner

A clear understanding of the general relationship between training, instruction, and performance assessment is necessary before comprehension of the specific differences between the role of fidelity in flight instruction and performance assessment is possible. As described earlier, Alessi and Trollip (1991) viewed the relationship between fidelity, stage of the learner, and task difficulty in the following four proposed stages of instruction: presentation, guidance, practice, and assessment. Assuming that these four stages of instruction are increasingly demanding, requiring the learner to expend added attentional resources with tasks of increasing difficulty (Kahneman, 1973; Norman & Bobrow, 1975), the state of learners and their stage of training become mandatory considerations in determining the amount of fidelity to use in simulators and simulation devices.

Figure 2 illustrates the major subsystems within the training cycle. It identifies instruction as a component of training and assessment as a component of instruction. It illustrates that, in addition to the state of the learner, training goals, objectives, and tasks must be considered during the needs-assessment stage of any training program. Assessment is embedded within the development and implementation stage of any training-program design.

Similar to the design cycle of computer products, the final stage of the training cycle provides feedback for practical issues such as cost, time to train, and assessment accuracy. This process can be applied to simulators and simulation devices used to transfer learning to actual equipment, as well as to devices implemented to assess (i.e., measure) terminal performance (i.e., check-rides). The concept of training effectiveness emerges from these relationships. This concept requires measurement of the transfer of learning to real-world equipment to achieve positive results. It also requires measurement of the ability of any specific device to predict trainee performance in the real world (i.e., prediction validity). There is a distinct difference in these two measures because a training device is not

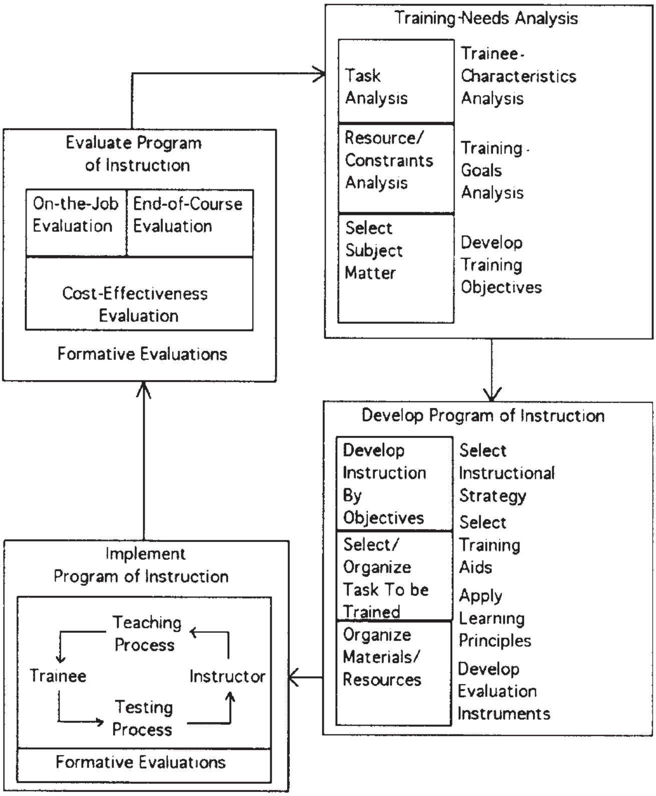

Figure 2. This figure illustrates the role of assessment within the training cycle. (Hays & Singer, 1989, p. 8).

required to be validated to aid in the transfer of learning; however, a device that predicts performance *does* require validation. Therefore, the concepts of learning and performance assessment are separate, yet inseparable, elements of training.

Transfer of Learning Versus Prediction of Performance

Laughery et al. (1982) conducted a study that distinguished between the discrete characteristics of simulation devices built to ensure a high transfer of learning and those built to predict real-world performance. The research demonstrated that a KC-135 Boom-Operator Part-Task Trainer (BOPTT), configured to simulate refueling of an F4 fighter-jet aircraft, was an effective simulation device for ensuring higher transfer of learning for an experimental group of boom operators qualified to refuel the B-52 aircraft in their initial training. Despite the high transfer of learning from the simulation trainer, the device was not reliable in predicting student performance for the F-4 refueling categorization.

The control group in the Laughery et al. (1982) study, comprised of participants with no training in the C-5 (large cargo aircraft) or F-4 configuration versions of the simulation trainer, scored *100% qualified* on an actual C-5 in-flight refueling experience. These individuals had also previously qualified in the refueling of the B-52 aircraft. This demonstrated that initial qualification for refueling the B-52 aircraft was similar to the refueling experience of the C-5. Prediction validity of the BOPTT was 100% for the refueling of the C-5 for both the control and experimental groups who were all initially qualified to refuel the B-52 aircraft. Therefore, although a training device of high fidelity can aid in the transfer of learning, the concept of transfer cannot be generalized to assessment or to prediction validity of a simulation device without considering the state of the environment (e.g., refueling the C-5 versus the F-4).

The purpose of the Laughery et al. (1982) study was twofold: (a) to measure and compare the transfer of learning for both groups to determine if a part-task trainer could optimize operational costs and training time, and (b) to determine if BOPTT was a valid predictor of performance for refueling fighter jets and cargo aircraft using operational aircraft in operational airspace. The research was divided into two phases. The first was conducted at a California Air- Force base and involved simulation training on the BOPTT and categorization briefings for the C-5 and F-4 aircraft. The second phase was conducted at the home squadrons of the study participants and involved actual in-flight evaluations on refueling the C-5 and F-4 aircraft. In Phase 1, 30 student boom operators, who were initially qualified to refuel the B-52 bomber aircraft in flight, were divided into a control group and an experimental group. Five students were selected each month over a period of 6 months from six separate classes of flight-line-designated trainees. Initially, six students were to comprise the control group; however, after half of the experiment was completed, it was decided that 10 students should comprise the control group and 20 should be assigned to the experimental group to increase the accuracy of the learning transfer.

The second phase of the Laughery et al. (1982) study measured the number of real-world flights required for the students in both study groups to qualify on fueling operations for both the C-5 cargo aircraft and F-4 fighter-jet aircraft. The students in both the control and experimental groups had received the same training, using the same training syllabus, up through their graduation and solo flight. Prior to evaluating the ability of both groups to refuel the C-5 and F-4 aircraft, the groups received separate research treatment. The control group received what was termed *Treatment A*, and the experimental group received *Treatment B*. Treatment A consisted of separate categorization briefings for the F-4 fighter aircraft and the C-5 cargo aircraft. Treatment B consisted of Treatment A, plus two one-hour simulation experiences on refueling a C-5 cargo plane and three one-hour simulation experiences on refueling the F-4 fighter aircraft. The device used to deliver the simulation experiences to the experimental group was the KC-135 BOPTT, which could be configured for novice, intermediate, or expert pilots (Clapp, 1985).

The BOPTT was built with a student station and boom-operator pallet with window operator controls and indicators (Laughery et al., 1982). To simulate oncoming aircraft needing refueling, a model of a C-5 or F-4 aircraft, which was scaled down 100 times, was viewed outside the training device. The BOPTT housed a 20-inch aerial refueling boom, and the 1/100 scale model of the C-5 or F-4 was mounted on a gimbal that delivered pitch, roll, and yawing moments. A closed-circuit video displayed the appropriate aircraft onto a cathode-ray-tube screen 20 inches outside the student's window. The model boom was located between the window and the model aircraft. The student could manipulate the boom mechanism by extending it and simulate connecting it to the aircraft. Environmental features, such as clouds, were visible on the cathode-ray tube. Engine noise and noise from operation of the boom could be heard through speakers inside the boom operator's station. The BOPTT simulation device showed the approaching C-5 cargo aircraft or the F-4 fighter-jet aircraft from a simulated 1.5-mile distance up to the refueling point from the window of the KC-135 aircraft. The device allowed for manipulation of independent variables such as turbulence, trajectory of the oncoming aircraft to the refueler aircraft, and refueling speed and altitude. It was also able to simulate five piloting-skill levels.

Procedures. Following simulation training in C-5 and F-4 fueling on the BOPTT, the students in the experimental group of the Laughery et al. (1982) study were evaluated on a real-time refueling-assessment flight for the C-5 cargo aircraft and on another for the F-4 fighter-jet aircraft. Students participating in the control group received no refueling-simulation training on the BOPTT for either C-5 cargo or F-4 fighter aircraft. The evaluation proposed to measure the ability of the students in both study groups to refuel both the C-5 and the F-4 aircraft in actual operational airspace. A training and evaluation squadron located at the base serving as the study site conducted the experiment with a sister squadron located at another training facility within the state of California.

The Boom-Operator Qualification-Performance Measurement Form was used as the criterion for both study groups in the Laughery et al. (1982) research. The instrument measured student execution of critical procedures, communications, boom control, and boom operation. The experimental group was evaluated on each simulation experience and on each actual flight for the C-5 and F-4 aircraft. The control group was evaluated on actual flights only. Following data collection at the participating base, a questionnaire was subsequently distributed to the home squadrons of the participants to gather data on the actual amount of time they required to qualify in the refueling of the C-5 and F-4 aircraft.

Table 6 illustrates the design of the Laughery et al. (1982) study. Table 7 reveals the number of flights required for the student participants to qualify in the refueling of the C-5 and F-4 aircraft at their home squadrons. A one-way ANOVA on the number of flights required for the students in the control group, experimental group, and their classmates to qualify in the C-5 and F-4 aircraft indicates a significant difference among the three research treatments among the control and experimental groups ($p < .05$). This suggests that significant savings can be realized with categorization training via implementation of the BOPTT.

Table 8 provides the number and percentages of students from both of the study groups who were found either qualified or unqualified to refuel the F-4 fighter aircraft from the KC-135 aircraft at the end of the applied treatments. Although all of the participants in the experimental group were found qualified to refuel F-4 fighters in the BOPTT, only half of them qualified on the actual equipment in the air. Only 20% of the control group qualified for refueling the F-4 fighter aircraft at the end of their respective

Table 6. Study Design: Categorization Research

Groups	*C5 CAT Training*	*Evaluation*	*F4 CAT Training*	*Evaluation*
Control (10 Subjects)	CAT briefing	Actual C-5 air refueling	CAT briefing	Actual F-4 air refueling
Experimental (20 Subjects)	Two 1-hour BOPTT missions	Actual C-5 air refueling	Two 1-hour BOPTT missions	Actual F-4 air refueling
	CAT briefing		CAT briefing	

Note. CAT = categorization; BOPTT = boom-operator part-task trainer. Adapted from “Differences Between Transfer Effectiveness and Student Performance Evaluations on Simulators: Theory and Practice of Evaluations,” by K. R. Laughery, J. L. Ditzian and G. M. Houtman, 1982, Proceedings of the I/ITEC Interservice Industry Training Equipment Conference, USA, p. 219. Copyright 1982 by ITEC. Adapted with permission.

Table 7. Flights Required to Qualify

	No. of Flights				
Groups	*One*	*Two*	*Three*	*Four*	*Five*
Control	3 (37.5%)	1 (12.5%)	2 (25.0%)	1 (12.5%)	1 (12.5%)
Experimental	6 (42.85%)	6 (42.85%)	2 (14.3%)		
Other[a]	5 (17.9%)	11 (39.3%)	8 (28.6%)	2 (7.1%)	2 (7.1%)

Note. Adapted from "Differences Between Transfer Effectiveness and Student Performance Evaluations on Simulators: Theory and Practice of Evaluations," by K. R.Laughery, J. L. Ditzian and G. M. Houtman, 1982, *Proceedings of the I/ITEC Interservice Industry Training Equipment Conference, USA*, p. 220. Copyright 1982 by ITEC. Adapted with permission.

[a] These individuals did not participate in the test program but were classmates of the participating members.

research treatment. This clearly indicated that the BOPTT was useful in transferring learning to the F-4 aircraft; however, it was not a valid predictor of performance on real-world F-4 refueling operations.

Table 8. In-Flight Performance for Fighter Category (F-4) Qualification

Group	*Considered qualified in aircraft*	*Considered unqualified in aircraft*
Control	2 (20%)	8 (80%)
Experimental	10 (53%)	9 (47%)

Note. Adapted from "Differences Between Transfer Effectiveness and Student Performance Evaluations on Simulators: Theory and Practice of Evaluations," by K. R. Laughery, J. L. Ditzian and G. M. Houtman, 1982, *Proceedings of the I/ITEC Interservice Industry Training Equipment Conference, USA*, p. 220. Copyright 1982 by ITEC. Adapted with permission.

Table 9 displays data indicating that all of the students participating in the Laughery et al. (1982) study-in both the control and experimental groups-were found to be qualified in the refueling of the C-5 aircraft from a KC-135 aircraft platform. It was impossible to determine if the BOPTT was a valid predictor of performance on the C-5 refueling operations because all of the student participants were considered qualified to refuel this aircraft from the KC-135 aircraft following their respective research treatments. Consideration should be given to the fact that all of the students in both study groups had been initially qualified to refuel the B-52 aircraft from the KC-135 before the research treatments were received; therefore, it could be concluded that the fueling operations of B-52 and C-5 aircraft are very similar.

The implication here is that near transfer of learning must be distinguished from far transfer of learning when making generalizations

Table 9. In-Flight Performance for Cargo Category (C-5) Qualification

Group	*Considered qualified in aircraft*	*Considered unqualified in aircraft*
Control	8	0
Experimental	13	0

Note. Adapted from "Differences Between Transfer Effectiveness and Student Performance Evaluations on Simulators: Theory and Practice of Evaluations," by K. R. Laughery, J. L. Ditzian and G. M. Houtman, 1982, *Proceedings of the I/ITEC Interservice Industry Training Equipment Conference, USA*, p. 220. Copyright 1982 by ITEC. Adapted with permission.

related to learning transfer and prediction validity of simulation devices (Osgood, 1949). Self-transfer is the improvement or decrement of the learner that results from repeated practice of the same event. Near transfer is the improvement or decrement that results from repeated practice of different, but very similar events. Far transfer is the improvement or decrement that results from repeated practice of dissimilar events in a similar domain. All three types of transfer must occur for optimal effectiveness of training and evaluation. Theoretically, each type of transfer should precede the other in the learning process because learning is a cumulative process. Likewise, in the training and assessment of novice, intermediate, and expert pilots under training, methods of learning and assessment should be consistently aligned with the appropriate level of learning taking place throughout the training cycle.

Implications. Laughery et al. (1982) demonstrated the basic difference between the relationship of near transfer of learning with far transfer of learning as it relates to the discovery of prediction validity and transfer of learning from simulation devices to actual equipment in real-world operations. Some simulators may propose to accurately assess student performance in real-world operational aircraft, while others may propose to measure transfer of learning only. Those such as the BOPTT when it is in the B-52 configuration, claim to do both in environment-specific configurations. The BOPTT did not, however, prove to be a good predictor for aircraft categorization assessment, even though it was indeed an excellent tool for improving learning rate in refueling the F-4 aircraft.

A clear understanding of the definition of terms is critical when making generalizations from experimental studies. For example, *transfer* is defined by Gick and Holyoak as "the change in the performance of a task as a result of the prior performance of a different task" (cited in Cormier & Hagman, 1987, p. 10). Osgood (1949) defines transfer as the ability to perform the same task in the same environment. It involves the ability of a student to demonstrate skills learned from practice on a training device to performance on the actual operational equipment. Apparently, refueling of the C-5 aircraft was similar enough to the refueling of the B-52 aircraft that slight environmental changes did not affect performance on the same category of aircraft. Consequently, under the definition of transfer provided by Osgood, prediction validity of the BOPTT in the B-52 configuration would be high for the actual refueling of the C-5 cargo aircraft. However, this was not true for F-4 fighter-jet refueling operations because the environment was sufficiently different. The definition of transfer provided by Gick and Holyoak would require a different target situation (cited in Cormier & Hagman, 1987). Learning from one situation could be transferred to a new situation with some environmental differences.

The *prediction validity* of a simulation device is the expression of its ability to accurately assess the flight performance of a student on real-world equipment in real-world airspace. If the performance scores attained on a simulation device closely match scores on the same tasks with actual operational equipment, then the simulation device is said to have high prediction validity. However, this does not mean that simulation devices that are valid predictors of the performance of real-world tasks are effective for transfer of learning. Similarly, devices that prove to be effective for transfer of learning may not be valid predictors of performance on actual equipment (see Table 8). For a simulation device to be a valid predictor of performance, the device itself must be validated. However, as demonstrated by Laughery et al. (1982), the same requirement is not always necessary to maximize transfer of learning, or even to realize a positive transfer of learning.

CONCLUSION

In conclusion, the degree of fidelity and the learning stage of the learner are mutually interdependent variables that must be considered when designing flight simulators intended for transfer of learning or performance assessment. It is important to recognize the similarities and differences between simulators designed for performance assessment and those designed for transfer of learning. The environment of the target skills is also a pivotal component. When all these elements are considered it becomes apparent that degree of fidelity, learning stage of the learner, learning rate, and the environment are not mutually exclusive. Further research is necessary to discover if there is a point beyond which one additional unit of fidelity will result in a diminished rate of practical (i.e., cost-effective) assessment for pilots who are between the novice and expert stages of learning. What must be considered, however, is that optimal performance assessment and transfer of learning in flight training is best served with shared goals and aligned values and expectations by all pilots, instructors, training departments, examiners, and licensing authorities (Telfer & Moore, 1997).

REFERENCES

Alessi, S. M. (1988). Fidelity in the design of instructional simulations. *Journal of Computer-Based Instruction, 15*(2), 40-47.

Alessi, S. M., & Trollip, S. R. (1991). *Computer-based instruction: Methods and development*. Englewood Cliffs, NJ: Prentice Hall.

Clapp, R. E. (1985, March). Visual display requirements for boom operator training in aerial refueling. *Proceedings of the Conference on Simulators: Society for Computer Simulation, USA, 16*(1), 188-192.

Cormier, S. M., & Hagman, J. D. (Eds.). (1987). *Transfer of learning: Contemporary research and applications*. San Diego, CA: Academic.

Cox, J. A., Wood, R. O., Boren, L. M., & Thorne, H. W. (1965). *Functional and appearance fidelity of training devices for fixed-procedure tasks* (NASA Tech. Rep. No. 65-4). Washington, DC: George Washington University.

Feifer, R. (1994). Cognitive issues in the development of multimedia learning systems. In S. Reisman (Ed.), *Multimedia computing: Preparing for the 21st century* (pp. 289-291). Harrisburg, PA: Idea Group.

Fink, C., & Shriver, E. (1978). *Simulators for maintenance training: Some issues, problems and areas for future research* (Tech. Rep. No. AFHRL-TR-78-27). Lowery Air Force Base, CO: Air Force Human Resources Laboratory.

Flach, J., Hancock, P., Caird, J., & Vicente, K. (Eds.). (1995). *Global perspectives on the ecology of human-machine systems* (Vol. 1). Hillsdale, NJ: Lawrence Erlbaum.

Gagne, R. (1954). Training devices and simulators: Some research issues. *American Psychologist, 9*, 95-107.

Grimsley, D. L. (1969). *Acquisition, retention, and retraining: Group studies on using low level training devices* (HumRRO Tech. Rep. No. 69-4). Alexandria, VA: George Washington University.

Hays, R. T., & Singer, M. J. (1989). *Simulation fidelity in training system design: Bridging the gap between reality and training*. New York: Edwards.

Hopkins, C. O. (1975). How much should you pay for that box? *Human Factors, 17*, 533-541.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

Kinkade, R., & Wheaton, G. (1972). Training device design. In H. Van Cott & R. Kinkade (Eds.), *Human engineering guide to equipment design* (pp. 668-699). Washington, DC: Department of Defense.

Klauer, K. M. (1997). Examining the concept of total fidelity simulation. *Gateway, 8*(2), 13-15.

Koonce, J. (1974). *Effect of ground-based aircraft simulator motion conditions upon prediction of pilot proficiency. Part I*. Savoy, IL: University of Illinois at Urbana. (NTIS No. AD-783 256)

Laughery, K. R., Ditzian, J. L., & Houtman, G. M. (1982). Differences between transfer effectiveness and student performance evaluations on simulators: Theory and practice of evaluations. *Proceedings of the I/ITEC Interservice Industry Training Equipment Conference, USA*, 214-222.

McGrath, J. J., & Harris, D. H. (1971). Adaptive training. *Aviation Research Monographs, 1*(2), 1-130.

Miller, G. (1974). Some considerations in the design and utilization of simulators for technical training (Tech. Rep. No. AFHRL-TR-74-65). Brooks Air Force Base, TX: Air Force Human Research Laboratory.

Miller, R. (1953). *Handbook on training and training equipment design* (WADC Tech. Rep. No. 53-136). Wright-Patterson Air Force Base, OH: Wright Air Development Center.

Norman, D., & Bobrow, D. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7*, 44-64.

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

Povenmire, H. K., Alvarez, K. M., & Damos, D. L. (1970, October). *Observer-observer flight check reliability* (Tech. Rep. No. LF-20-2). Savoy, IL: University of Illinois, Institute of Aviation, Aviation Research Laboratory.

Povenmire, H. K., & Roscoe, S. N. (1973). Incremental transfer effectiveness of a ground-based general aviation trainer. *Human Factors, 15*(6) 534-542.

Roscoe, S. (1971). Incremental transfer effectiveness. *Human Factors, 13*, 561-567.

Roscoe, S. (1972). A little more on incremental transfer effectiveness. *Human Factors, 14*, 363-364.

Schneider, W. (1985). Training high performance skills: Fallacies and guidelines. *Human Factors, 27*, 285-300.

Selzer, U., Hulin, C. L., Alvarez, K. M., Swartzendruber, L. E., & Roscoe, S. N. (1972). Predictive validity of ground based flight checks. *Proceedings of the 16th Annual Meeting of the Human Factors Society, USA*, 318.

Telfer, R. A., & Moore, P. J. (1997). The roles of learning, instruction and the organization in aviation training. In R. Telfer & P. Moore (Eds.), *Aviation training: Learners, instruction and organization* (pp. 1-13). Aldershot, England: Ashgate.

U.S. Department of Transportation, Federal Aviation Administration. (1970). *Private pilot practical test standards for airplane single-engine land and multi-engine land* (FAA Publication No. 8081). Washington, DC: Author.

Valverde, H. H. (1973). A review of flight simulator transfer of training studies. *Human Factors, 15*, 510-523.

Wiggins, M. W. (1997). Expertise and cognitive skills development for ab-initio pilots. In R. Telfer & P. Moore (Eds.), *Aviation training: Learners, instruction and organization* (pp. 54-66). Aldershot, England: Ashgate.

Wolfe, N. E. (1978). *Instructional systems utilizing modeling and simulation for technician training in the area of measurement and control* (Report No. IR 008452). Akron, OH: University of Akron. (ERIC Document Reproduction Service No. ED188598)