ASSESSING THE EFFECTIVENESS OF SIMULATOR-BASED TRAINING

Richard A. Thurman

Ronald D. Dunlap

Air Force Research Laboratory

Aircrew Training Research Division
Mesa, AZ

The digital revolution has sparked a worldwide movement toward the use of simulators to enhance training and accelerate learning. As simulator-based training grows, the concern of trainers should increasingly turn to determining if effective training is taking place, rather than merely using simulators more extensively. We reviewed the simulator training literature to see just what literature exists as well as the effect simulation is having in terms of training effectiveness. The review concentrated on the literature from several different domains/perspectives, including the NASA space program, commercial aviation training, medical procedures training, and nuclear power plant operation training. The objective of the review was to focus on prototypical studies which showed utility in determining the effects of simulator-based training of highly complex tasks. Unfortunately, our review showed that little attention is being directed toward determining the effectiveness of these training devices and research on the effective tactics and strategies for utilizing simulation are almost nonexistent. We then put forward a brief explanation for the lack of motivation to assess simulator-based training, along with a plea to move forward in this area. Finally, we review a model, first outlined by Lewis (1996), for assessing the effectiveness of simulator-based training.

ABOUT THE AUTHORS

Dr. Richard A. Thurman is a Research Psychologist at the Warfighter Training Research Division of the Air Force Research Laboratory located at Williams Gateway Airport (formerly Williams AFB), Mesa, Arizona. He has managed research projects concerned with the application of instructional simulation techniques and technologies for aircrew training, virtual environment-based training, and the use of artificial intelligence technologies for agent-based training.

Dr. Ronald D. Dunlap is a Research Psychologist at the Warfighter Training Research Division of the Air Force Research Laboratory located at Williams Gateway Airport, Mesa, Arizona. He is currently responsible for evaluating the benefits of distributed mission training techniques and technologies on operational aircrews. His research interests include individual and team training, training technologies for distributed systems, and applying theories of learning and memory to problems in training.

ASSESSING THE EFFECTIVENESS OF SIMULATOR-BASED TRAINING

Richard A. Thurman

Ronald D. Dunlap

Air Force Research Laboratory

Aircrew Training Research Division
Mesa, AZ

The digital revolution has sparked a worldwide movement toward the use of simulator-based training to enhance training and evaluation processes and to effectively accelerate learning. As the use of simulation-based training increases, the concern of simulation users should increasingly turn to determining if effective training is taking place, rather than merely using simulators more extensively. For example, there is a tendency for training managers to make the assumption that all simulators are optimally designed and used, so simulator training effectiveness is primarily a matter of how much the device is used. Such is not the case, however, as it is possible to use any simulator extensively, while at the same time to use it ineffectively (Caro, 1988).

The training literature is full of testimonials as to the power of simulation – its ability to free users from current restraints and restrictions, the performance assessment opportunities it provides, the ability to evaluate new tactics, etc. However, despite tremendous engineering innovations and a plethora of rhetoric concerning the value and utility of simulation as a training strategy, public perception has it that considerably less attention is actually directed toward determining the effectiveness of this training medium. A recent Inspector General's report adds credence to this perception by indicating that after spending over 1.6 billion dollars on large-scale, distributed simulation, the DOD has failed to prove that simulators enhance training at all – and calls on them to do so (as reported in Clark, 1997).

The purpose of this paper is to ameliorate this perceived shortfall by reviewing quantitative and quantifiable evaluations of simulation-based training. The effort reviewed training literature from four specific domains/perspectives which have the public perception of being on the cutting edge of simulation-based training: the NASA space program, commercial aviation training, medical/surgical procedures training, and nuclear/fossil power plant operation training.

METHODS FOR EVALUATING SIMULATION-BASED TRAINING

There are many ways to assess the effectiveness of a training simulation. Bell & Waag (1998) discuss three general approaches for estimating the training effectiveness of flight simulations. They include utility evaluations (collecting opinions of effectiveness from relevant parties), in-simulator learning (assessing simulator-based performance improvement), and transfer of training (seeing if the simulation affects subsequent performance). To date, the most prescribed method for assessing effectiveness is the transfer of training study. Indeed, some training researchers believe that this is the only sufficient method for determining simulator training effectiveness (Bell & Waag, 1998). This method seeks to determine the effects of practicing one simulated task on the learning or performance of another separate, but similar task. Generally, assessing training effectiveness via the transfer of training study takes the form of an experiment where two groups of participants receive different kinds or amounts of simulation-based training (Boldovici, 1987). Both groups are then tested on the task under study. Their proficiency at the task may be evaluated in terms of the amount, quality, or cost of learning or performing the task. As a final step in the experiment, the two groups' proficiency scores are compared with each other. According to Boldovici (1987), transfer can be:

Positive: Learning is improved due to training via the simulator.

Negative: Training in the simulator somehow interferes with performing the new task.

Neutral: Training in the simulator has no discernible effect on performing the task.

Some tasks are not amenable to true transfer studies. They may be too dangerous, too costly, or so rare that it is not recommended that trainees be allowed to perform them in the actual setting until all doubt about their proficiency has been removed. (Would one really want nuclear power plant trainees, student anesthesiologists, or uncertified airline pilots to have access to operational equipment before one is absolutely sure they are ready?) For this reason, some training transfer studies do not use the actual task. Instead, they may require the trainee to demonstrate proficiency via another simulator. Others have termed these kinds of evaluations “quasi-transfer studies” (Lintern, Roscoe, & Sivier, 1990).

Kirkpatrick (1959a; 1959b; 1960a; 1960b) delineates a four-step training evaluation plan, the first three of which parallel Bell and Waag's (1998) categories. The first step, according to Kirkpatrick (1959a) is to determine the trainees' reaction, or "how well the trainees liked a particular training program." p.4. The second step is to measure "learning" (Kirkpatrick, 1959b) by comparing those who received training to those who did not. Evaluation of performance (i.e., behavior) is the third step (Kirkpatrick, 1960a). It is measured by comparing before and after performance on the part of trainees. The fourth step is to measure "results" in terms of reduced costs, increased productivity, and so forth (Kirkpatrick, 1960b).

WHAT EVIDENCE DO WE HAVE THAT SIMULATORS ARE EFFECTIVE: A REVIEW OF THE LITERATURE

In this review we concentrated upon finding published evaluations of simulation-based training which purported to show a training effect. The effort involved reviewing training literature from several domains including medical procedures training, nuclear power plant operation, commercial aviation, and the NASA space program.

Medical/Surgical Training

Until about 1993, there were only a handful of centers with a realistic, high fidelity training simulator in use (Gaba, 1996). Today, the medical community is in the middle of a period of greater acceptance and use of simulation, and there are now over fifty simulator centers worldwide (Gaba & Small, 1997). The majority of them have taken on the challenge of providing high-fidelity training in anesthesia care. Indeed, it seems that anesthesiologists have been the leaders within the medical community in applying simulation technology. They have also developed curricula for simulation-based training for a number of nonanesthesia health care professionals, such as; emergency room personnel, acute care nurses, medical specialists, and community science students (Murray & Schneider, 1997).

At this point in time, high fidelity medical simulators are still relatively costly. A commercial simulator costs nearly $200,000 and operational costs may exceed $500 an hour (largely due to personnel costs). It behooves the medical training community to use these devices efficiently and effectively so that they can produce an acceptable return on investment (Murray & Schneider, 1997).

Our search of the medical training literature revealed 73 published reports where simulator-based training was the topic of study. Of those 72 articles, 61 were simple descriptions of a simulator-based approach to training and had no data upon which to assess training utility. No articles reporting a true transfer of training study could be found, but three articles described the results of several quasi-transfer studies.

Garfield, Paskin, and Philip (1989) investigated the effects of a microcomputer-based simulation (called Gas Man) designed to teach the principles of anesthetic uptake and distribution. The authors compared student’s knowledge of the patient prior to training with their post-simulation scores. Eight weeks after the participants had finished the simulator training their knowledge of uptake and distribution had significantly improved. The authors also report that students wished to continue using the simulation, implying that it is a “pleasant and satisfying way to learn.” They conclude by writing that computer simulation has the potential to be an important teaching tool in an anesthesia residency training program.

Holzman, et al. (1995) document a study in which they tried to assess the utility of simulation for training anesthesia crisis resource management (ACRM). In this study 68 anesthesiologists and 4 nurse-anesthetists participated in an ACRM training course held over a 2 ½-month period. The anesthesia environment was a re-creation of a real operating room utilizing available standard equipment and simulations. The task was to perform as close to possible to actual clinical interventions. The crisis scenarios included overdose of anesthetic, oxygen source failure, cardiac arrest, malignant hyperthermia, tension pneumothorax, and complete power failure. Following the scenarios, participants were given detailed questionnaires to assess the training value of the setup. Over half of the participants felt that the course should be taken every 12 months, while another third felt that the course should be repeated every 24 months. Participants rated the potential benefit of this course for anesthesiologists to practice ACRM in a safely controlled simulated environment "very highly."

Cain and Shirar (1996) examined the transfer effects of training for repair of second-degree perineal birth trauma. While this repair is one of the most common surgical techniques in pregnancy care, the three-dimensional nature of the required procedure makes it hard for students and residents to learn. Cain and Shirar report that using a three-dimensional model of the tear and allowing trainees to practice the repair procedure via the model had positive results. Participants tested eight months after the initial training sessions were able to consistently perform the procedure (via simulation) and rated the model-based training to be superior to learning from textbooks.

Nuclear Power Plant Training

The first use of simulation in the nuclear power plant industry occurred in June 1968, with the inauguration of the Dresden 2 Simulator. According to Mills (1992), at that time the Atomic Energy Commission (AEC) required each operator to participate in at least five nuclear plant startups or shutdowns every year in order to remain "current." The plant used a five-shift crew setup, and it became readily apparent that the AEC's requirement would mean that the new nuclear reactors would be spending a lot of their time starting and stopping just for training purposes. The goal of the simulator was to provide training which the AEC would accept in lieu of actually shutting down the nuclear plant on a continual basis. The AEC unofficially indicated that they would approve the idea if the simulator was "good enough." Mills (1992) relates how the simulator received its approval:

> … The AEC inspector arrived. He played with the simulator one morning. In the afternoon we let him "pull critical" with us. Pulling critical in real time with real simulated instruments was so much fun it was almost irresistible. We had to repeat it for every visiting engineer from San Jose, and of-course we did it for the AEC man. Each of us, including the AEC man, felt as if he were Enrico Fermi himself under the bleachers in Chicago in 1942 … The simulator was pronounced "good enough." Ignominious as it may seem, this pronouncement was, and still is, the foundation of our simulator industry.

According to Baudhuin (1987), the Nuclear Regulatory Commission (NRC) requires that all nuclear power facilities must have a "full-scope simulator" for initial and refresher training for all control room operator personnel. The simulator industry has responded by designing simulators which, " for all practical purposes, are exact replicas of the actual control rooms they simulate." (p. 232) Baudhuin continues with the NRC's justification, "The NRC believes that individuals who have been trained on simulators have a better grasp of the variety of potential normal and abnormal operating conditions which may occur in the designated plant. In short, the simulator must reproduce the operating characteristics of the actual plant….Transfer of training is indeed implied but appears not to be verified here." (p. 233)

Our search of the nuclear/fossil power plant training literature revealed 23 published articles where simulator-based training was the topic of study. However, none of those articles contained any transfer of training data upon which to base the value or utility of simulation in this field.

While a bit disconcerting, the lack of published transfer data is actually predictable. After all, one would assume that much of simulator-based training within the power plant industry would revolve around critical incident and emergency training. Lack of published transfer data suggests that there simply are not that many "incidents" with which one can assess the degree of training benefit. In addition, from a public relations standpoint, it is probably not in the best interest of the industry to provide too much data on the capabilities of its training or of its operators. Perhaps the feeling is, that some data are better kept internal – and not released for public consumption. Baudhuin (1987) justified the lack of nuclear control room simulator training evaluations by referring to the relative "newness" of the business. However, a decade later (with still no published evaluations), Mills, (1997) justified the dearth of training effectiveness studies in the following manner:

> To justify that operators should be trained shouldn't need statistics. We wouldn't want American Airlines to run an experiment to test how safe jumbo jets are with untrained pilots. Similarly, neighbors of nuclear power plants would be unhappy if we stage meltdowns from time-to-time for practice. The point is we never get the opportunity to gather comparison statistics for operation using untrained operators … Simulators generally justify themselves economically by displacing training exercises using the real plant by simulated training sessions. They further justify themselves by simulating malfunction and emergency conditions you would be unable to reproduce in the real plant. First set your training standards. What skills and knowledge are required? Then look at the effectiveness of providing equivalent training by various means, including simulators. Factor in the costs and risks of staging emergency drills in the real process for training purposes. If on-the-job training provides every operator with sufficient exposure to startup/shutdown, malfunctions, and emergencies, then so be it.

Commercial Aviation Pilot Training

Testifying before the 1997 White House Commission on Aviation Safety and Security, Douglas Schwartz, Director of Standards at Flight Safety International, related; "Simulator-based training has proven to be one of the most dramatic sources of improvements in aviation safety. The period of sharp decline in the accident rate in the mid '60s and early '70s is in part attributable to the growing use of flight simulators for training that occurred at that time. Today, the presence of flight simulators has become so commonplace that it is easy to take this powerful training tool for granted" (Schwartz, 1997).

Despite such glowing testimony, our search of the commercial aviation literature failed to find any published studies assessing the effectiveness of simulators for flight training. Instead, every paper, technical report, or conference presentation we reviewed was either a simple description of the simulation system and capabilities, an engineering description of how the system was developed, or a general description of decisions which were made concerning the simulation's instructional capabilities. We were unable to find any published article that purported to have even performed a utility evaluation study.

This finding is perplexing, given the assumption on our part that flight simulation has been one of the most active areas of simulation R&D. We are aware of many studies and reviews, performed by federal and academic laboratories which address the effects of flight simulator training (e.g., Bell & Waag, 1998; Carretta & Dunlap, 1998; Hays, Jacobs, Prince, & Salas, 1990; Lintern & McMillan, 1993).

Perhaps the commercial aviation community is relying on the findings of laboratory research as a validation of the simulation approach and do not feel a need to publish their assessments of simulator- based training. Perhaps the fact that simulator-based training is mandated by the Federal Aviation Administration for many commercial aircraft training programs has an impact on the need to perform transfer studies. Or, perhaps the use of simulators has had such an overwhelming impact on commercial aviation, as Schwartz (1997) indicates, that any proof of transfer is deemed unnecessary.

NASA

Our search of the astronaut training literature revealed 8 published reports where simulator-based training was the topic of study. Of those 8 articles, 7 were simple descriptions of a simulation-based approach to training and did not contain data which could be used to assess training transfer. However, one article did describe the results of a utility evaluation study.

In this study, Loftin and Patrick (1995) trained over 100 members of the ground-support team for the Hubble Space Telescope (HST) repair mission. The objective of the training was to familiarize ground-support controllers, engineers and technicians with the location, appearance, and operability of HST components in the space shuttle payload bay. It was hoped that this experience would instill mental models of system components and correct interrelationships, provide experiential knowledge of task procedures, and enhance the ability of ground-based flight controllers to interact effectively with the crew during the mission. Trainees were given 121 episodes where they were placed in an “immersive virtual environment” simulation and instructed to perform the same kinds of tasks the repair crew were going to perform. Each training episode lasted on average 100 minutes (i.e., 40 minutes devoted to an immersive experience and 60 minutes of “over-the-shoulder” observation). At the end of the HST repair mission, each participant was given a “post-flight” evaluation instrument. Loftin and Patrick (1995) reported that members of the ground support team believed that the training had a positive effect on their performance during the mission.

MISSING TRANSFER STUDIES

Our literature review was an attempt to search out studies where simulator-based training was the topic of discussion. We focused our search on the four domains where public perception has it that simulators are used extensively in training (nuclear power, medicine, commercial aviation, and NASA). Our search lead us to 103 published articles documenting the benefits of simulator-based training. Yet, a closer examination revealed that the majority were meant to be descriptions of a new or improved simulator. Of the 103 articles, only four actually supplied any assessment data (and they were limited to simple reactions and opinions of trainees and/or trainers).

This review of the training literature indicates that evaluations of simulator-based training, other than reactions of participants, are not being attempted. Or, if they are being attempted, they are not being reported. It would seem that beyond the level of utility evaluations, formal evaluations of simulator-based training are done infrequently, and by Kirkpatrick's (1959b, 1960a, 1960b) or Bell and Waag's (1998) standard, incompletely.

The lack of formal assessments of simulator-based training is not new. Indeed, Bell and Waag (1998) found that combat simulators are rarely evaluated to ascertain how much transfer there is. "Our review of the available literature found very limited data regarding the value of simulation for air combat training. Although a fair amount of opinion data exists that suggests there is training potential in using simulation, actual transfer data are extremely limited." (p.232) Of the entire literature on air-combat simulation, they could find only nine studies that addressed training transfer -- seven studies showed positive results, two did not. Likewise, Carretta & Dunlap (1998) reviewed 67 articles, conference papers, and technical reports purporting to evaluate flight simulator-based training. They reported that only 13 articles attempted to address transfer of training. The view coming back from the simulator-based training field is that except for opinion data, formal assessments of simulator-based training are few and far between.

The lack of transfer studies is not limited to simulator-based training. In their survey of other training practices in U.S. companies, Sarri, Johnson, McLaughlin and Zimmerele (1988) found that, "In general there is limited evidence of systematic evaluations of training by U. S. companies. To the extent that evaluations are conducted, the primary method used is evaluation forms administered after program participation. . . This study indicates that evaluations of management training, other than reactions of participant following program attendance are not evident in U. S. companies." (p.741) Carnevale, Gainer, and Villett (1990) came to the same conclusion stating, "Current evaluation methods consist mostly of informal worksheets measuring a trainee's reaction to the program. Follow-up sessions may occasionally be held to see if training was useful for trainees. This type of evaluation is usually very subjective." (p. 156). Moller and Mallin (1996) report results of a survey of U. S. corporate trainers where 90% of respondents indicate that the companies they work for "place significant barriers to evaluation" and over 30% describe their workplace as "having a culture that resists evaluation." It would seem then, that the lack of transfer data is not simply limited to domains that rely highly on simulator-based training, such as medicine, aviation, or nuclear power, but is endemic to much of training in general.

WHY THE LACK OF TRAINING TRANSFER DATA?

Assessing simulator-based training via transfer experiments is usually very hard to do. The very same issues that make field-based training so difficult also crop up when it is time to evaluate trainees' proficiency. For example, many of the safety constraints that trainees are placed under cannot and should not be lifted until after they have demonstrated their ability to handle emergency and/or dangerous situations. In other cases, the reason to use simulator-based training is based on the fact that there is no other way to perform the task(s). That is, in some situations neither trainees nor operators are allowed to operate equipment in a certain manner unless they find themselves in an emergency situation. Simulation is the only option. But how does one assess simulation's training effectiveness?

Sometimes transfer studies are simply not seen as being an effective way to assess training. Newstrom (1985) identified nine barriers to performance transfer, few of which are actually controlled by training. Similar findings are reported by Rummler and Bracke (1995) who identified six factors essential for effective performance, only one of which is directly influenced by the trainer (the design and delivery of training). It is only natural then, for training managers to place little faith in transfer studies when the outcomes are based upon so much more than the fact that a simulator was used. And, in their minds it becomes resource prohibitive to take into account all the myriad factors which may influence training performance. The main point of this discussion is that oftentimes, the transfer effectiveness of simulator-based training is not assessed due to the high costs associated with performing the study, in terms of safety, restrictions, or cost versus perceived benefits.

Perhaps the strongest reason why transfer studies are not used for assessing simulator-based training is the tenuous relationship some feel exists between training and subsequent proficiency. Training is, after all, not the only factor that contributes to a person's ability to perform. The unique role of training would need to be isolated first, before its effects could be ascertained. Even then, performance itself and how to measure it remains an issue. Vreuls and Obermayer (1985), as well as Mohs, MacDiarmid amd Andrews (1988) remind us that performance measurement remains a contentious and unsettled aspect of training research. The confluence of these two factors (ability vs. training, and performance measurement) serve to illustrate that the connection between simulator-based experiences and trainee proficiency is often not a simple case of cause and effect. Is it the simulator, or is it the knowledge, skills, and attitudes unique to the trainee that makes the difference?

Brinkerhoff (1995) goes further and contends that, to some extent, evaluations are a futile way to assess training:

> To set out to try to prove that training makes a difference is as ludicrous and useless as the automobile engine production department spending its precious resources on a study to try to prove

> how much it has caused overall business performance. First, it is virtually an impossible task to get an answer to the question; the methodology of causal analysis and statistical proof is extremely complex and fraught with tenuous assumptions. Second, it is just not a relevant question, and certainly not worth spending good money to find out. A far more worthwhile aim is for training evaluation to seek to find out whether training customer needs are being met, and met well. (p.395)

Brinkerhoff's position is a bit provocative. But he does point out the mindset of many corporate training departments and why large segments of the simulation community seem to be reluctant to perform training evaluations.

GETTING PAST TRANSFER AS A WAY TO ASSESS SIMULATOR-BASED TRAINING

In a review of the training transfer literature, Baldwin AND Ford (1988) suggest that there are large knowledge gaps on the part of training researchers with respect to transfer. They contend that a major flaw in the use of transfer as a way to assess proficiency is the use of "single input factors." They contend that transfer type assessments have not acknowledged that most training exercises consist of composites of interconnected factors -- all of which need to be taken into account. These authors challenged the training field to, "take a more eclectic orientation toward transfer by focusing on a number of other literatures neglected … by researchers." (p.98) Thomas, Anderson, Getahun AND Cooke (1992) chastise training evaluators further and contend that transfer studies should get past the "surface feature" mentality. They should instead work toward determining "intermediate level knowledge" on the part of trainees. That is, training assessment should be concerned with measures that "reflect general knowledge and principles that are embedded in domain-specific tasks." Thomas et al. (1992) see the end result of such activity as the ability to assess "deep feature transfer" -- the ability for trainees to generalize and respond to situations which don't exactly match those they were trained for.

These comments should not be construed as a denigration of the Kirkpatrick model of evaluation. Instead, they point to the idea that simulator-based training assessment should improve upon the model and get past transfer as the sole criteria for determining if the training is "good." For example, both the nuclear power industry and commercial aviation seem to be well beyond defensively trying to prove that training works. For them, training is a federally mandated activity (Lewis, 1996) which must be accomplished. Instead of concerning themselves with proving transfer occurs, they understand that they cannot afford the consequences of not training.

A POST-KIRKPATRIC MODEL OF SIMULATOR-BASED TRAINING

Lewis (1996) proposes a model for thinking about training evaluation that may hold promise for conceptualizing and assessing simulation-based training. The model is predicated on the idea that training (and its assessment) is much broader than a limited technical definition may connote:

> Training has a narrow, circumscribed meaning. One is trained to do this or that. Animals can be trained. Thus, the term "training" has appeared to this author to be inappropriate with reference to people. "Training" does not connote reflection (see Schon, 1983). One acts the way one was trained to act. It implies that there is one way.… Concepts and desired skill clusters such as problem solving, lifelong learning, communicating, being flexible, ability to work in teams, and learning how to learn, do not square with the concept of "training." …Being of short duration (a week of training would be long), they perforce can present only a small aspect of a larger picture, and they must do so under cram conditions. There is little time for reflection. What is learned is deemed to be of value only if it is obviously applicable back at the worksite. There must be "transfer." But transfer has come to be taken too literally. (p.10)

Lewis goes on to say that reconceptualizing the idea of what "training" should accomplish will lead to a better understanding of what many users need:

> A rival term to training is "education," which connotes reflection, breadth of view, possibilities, and ability to transfer (in the fullest sense of the term). Since education cannot be applied as an instrument (get it today, apply it tomorrow), adopting the term will immediately force rethinking of evaluation. Education does not necessarily require an identified workplace "need" as its starting point. It does aim at transfer as an end, but not transfer that's conceived as transporting what is unchanged back into the workplace. (p. 10)

His evaluation model is an attempt to help researchers as well as evaluators think about training in a broad, overarching way. It attempts to address Thomas et al. (1992) admonition to get past the "surface feature" mentality of current training effectiveness studies. It also addresses Baldwin and Ford's (1988) concern about assuming a single factor as the causative agent in complex training environments. And, finally it concedes that the outcomes of training are dependent not only on the goals of the corporation, but of the goals of the trainee as well.

The model has 11 elements -- the first eight address "context and process" while the last three elements address "outcomes." The context and process elements are:

Basis of Training. This refers to the basic reason for the training. The reason can be either proactive or reactive. Proactive training "shifts the discourse on evaluation from a defensive posture, characterized by fear that training must show results 'or else,' to one that places emphasis on well conceived training." (p. 13) Reactive training is based upon remediating some deficiency -- a pre-identified lack of skill, attitude, or knowledge.

Purpose of Training. Related to basis of training, purpose refers to the "nature of the reason for training." Purpose can be either intrinsic or instrumental. An intrinsic purpose may not have an immediate productivity goal. Instead, it accommodates trainees who are in an attitude of learning. An instrumental purpose, on the other hand, has a "clear targeted behavior change in mind." It is training aimed at solving a specific problem.

Unit of Productivity. Is the training aimed at one person, a team, or an organization? As the unit of productivity changes the nature of the evaluation must change. Team proficiency is assessed differently than individuals.

Time Horizon. This refers to when the assessment should take place (immediately, or short, medium and long term). With some training, its effect may not show up for months, while other training will show immediate results. Often, training is assessed on too short of a time scale. According to Lewis, "… time is a major variable to consider when thinking about the evaluation of training." (p. 14)

Data Sources. The kinds of data needed to determine training effects varies according to the kind of assessment undertaken. Types of sources include, but are not limited to: artifactual, anecdotal, archival, observational, perceptual, and experimental.

Kind of Training. Is the training formal, informational, or supplemental? While we tend to look only at formal training (i.e., structured, conventional), there are other ways trainees gain information. Informational training refers to the opportunities novices take to learn from more experienced peers. The expert may be helping out the novice on an informal basis. This form of "training" should not be overlooked. Supplemental training may consist of simply providing access to information sources (manuals, instructions, etc.) and letting "trainees" pursue them a their own pace. Each of these three kinds of training will require a different assessment approach.

Who is Trained. It is obvious that different personnel (with differing job descriptions) will need different kinds of training. Their training assessment will differ as well, according to the purposes of the training.

Costs of Training. Some costs are obvious (e.g., trainer salaries, equipment purchases). While some are hidden (e.g., foregone productivity). This is a classic area of evaluation research and not much needs to be said concerning cost/benefit analyses.

The three outcome elements are:

Trainee Personal Outcomes. Kirkpatrick's evaluation model includes checking for the initial reaction of trainees. While this check is useful, personal outcomes can go much deeper. Trainees, should feel that they benefit personally from the intervention. Lewis posits two kinds of personal outcomes, vocational and consumptive. Vocational outcomes refer the belief on the part of trainees that their worth and value as an employee is increased. This can be measured in terms of organizational commitment, career commitment, job involvement, and job satisfaction.

Consumer/Customer Outcomes. One of the main reasons for training is to produce a more capable, competent employee. Yet it is the consumer/customer which should be the true beneficiary of the training. Training assessment should take into account the results of training such as safe products, safe processes, or thoughtful, reliable and competent practitioners.

Corporate Outcomes. Lewis posits that there are two types of corporate outcomes accessible for assessment, tangible and intangible. Tangible outcomes are usually directly measurable and constitute the corporate bottom line (higher quality product, increased sales, more product produced per hour, etc.). Intangible outcomes seem to be measurable only indirectly. Lewis includes examples such as, "increased commitment to the organization (staying late, coming in on Saturdays) increased willingness to learn, greater concern for the team, commitment to lifelong learning, safe customers and satisfied customers." (p.17)

CONCLUSION

Assessment of learning and/or training has been of interest to researchers and policy makers since Ed Link created the first flight simulator. The intent of this paper has been to first, present an overview of how simulator-based training has been evaluated in areas other than military aircrew training. Since the majority of simulator-based training research is found in the military domain, and there are ample reviews of that literature base. However, even those reviews found relatively few empirically sound evaluations of simulator based training (Bell & Waag, 1998; Carretta & Dunlap, 1997; Hays, et al., 1990; Lintern & McMillan, 1993). Instead we intended to review simulator-based training assessments from commercial domains such as the medical community, NASA, nuclear power, and commercial aviation. However, upon investigation, we found very few published articles that actually attempted to assess the value or utility of simulator-based training. The vast majority of the literature consisted of simple descriptions of a simulator's features, engineering descriptions of how a particular simulator was developed, or general articles extolling the virtues of simulation-based training (without supporting data to back up the claim).

We then turned our attention to reasoning why there seems to be reluctance (even passive resistance) on the part of training practitioners to perform and document simulator-based training effectiveness studies. Some of the reasons seem to be: (a) a perception (perhaps based on experience) that assessing simulator-based training via formal experimentation is costly and difficult to conduct (e.g., Bell & Waag, 1998); (b) that this approach is not viewed as the most effective way of assessing simulator-based training (e.g., Newstrom, 1985; Rummler & Bracke, 1995); (c) that it is hard to assess performance (e.g., Vreuls & Obermayer, 1985); and (d) that even if you could assess performance, there are so many intervening variables that it is difficult to show that a simulator-based intervention is the cause of performance improvement (e.g., Brinkerhoff, 1995).

Finally, we outlined an approach for improving the current shortfall of training assessment by referring to Lewis' (1996) model for thinking about the evaluation of training. The hope is that by so doing we will help simulator-based training practitioners, researchers, and students see beyond current practice and begin to create more robust assessments -- assessments which shift evaluations from a defensive stance (typified by the concern that training must demonstrate positive results "or else") to one that places more emphasis on well conceived simulator-based training.

REFERENCES

Baldwin, T. T. & Ford, J. K. (1988). Transfer of training: A review and directions for future research. Personnel Psychology, 41, 63-105.

Bell, H. H., & Waag, W. L. (1998). Evaluating the effectiveness of flight simulators for training combat skills: A review. International Journal of Aviation Psychology, 8(3), 223-242.

Boldovici, J. A. (1987). Measuring Transfer in military settings. In S. Cormier & J. Hagman (Eds.), Transfer of Learning: Contemporary Research and Applications (pp.239-260). San Diego: Academic Press.

Brinkerhoff, R. O. (1995). Using evaluations to improve quality of technical training. In Leslie Kelly (Ed.), The ASTD technical and skills training handbook (pp.385-407). New York: McGraw Hill.

Cain, J. J., & Shirar, E. (1996). A new method for teaching the repair of perineal trauma of birth. Family Medicine, 28(2), 107-110.

Carnevale, A. P., Gainer, L. J., & Villett, J. (1990). Training in America: The organization and strategic role of training. San Francisco: Jossey-Bass.

Caro, P. (1988). Flight training and simulation. In E. L. Weiner & D. C. Nagel (Eds.), Human factors in aviation. New York: Academic Press.

Carretta, T. R., & Dunlap, R. D. (1998). A brief review of research in flight simulation training and transfer: 1986 to present. (Unpublished manuscript).

Clark, C. (1997). Simulation buying spree needs restraint. Defense Week, 18(23).

Gaba, D.M. (1996). Simulators in anesthesia. In C. Lake (Vol. Ed.), Advances in Anesthesia: Volume 14 (pp. 55-94). St. Louis, IL: Mosby-Year Book, Inc.

Gaba, D., & Small, S. (1997). How can full environment-realistic patient simulators be used for performance assessment? American Society of Anesthesiologists Newsletter, 61(10), 4-7.

Garfield. J. M., Paskin, S., & Philip, J. H. (1989). An evaluation of the effectiveness of a computer simulation of anaesthetic uptake and distribution as a teaching tool. Medical Education, 23(5), 457-462.

Hays, R. T, Jacobs, J. W., Prince, C., & Salas E. (1990). Requirements for future research in flight simulation training: Guidance based on a meta-analytic review. The International Journal of Aviation Psychology, 2, 143-158.

Holzman, R. S., Cooper, J. B., Gaba, D. M., Philip, J. H., Small, S. D., & Feinstein, D. (1995). Anesthesia crisis resource management: Real-life simulation training in operating room crises. Journal of Clinical Anesthesiology, 7(8), 675-687.

Kirkpatrick, D. L. (1959a). Techniques for evaluating training programs. Journal of the American Society of Training Directors, 13(11), 3-9.

Kirkpatrick, D. L. (1959b). Techniques for evaluating training programs: Part 2- Learning. Journal of the American Society of Training Directors, 13(12), 21-26.

Kirkpatrick, D. L. (1960a). Techniques for evaluating training programs: Part 3-Behavior. Journal of the American Society of Training Directors, 14(3), 13-18.

Kirkpatrick, D. L. (1960b). Techniques for evaluating training programs: Part 4-Results. Journal of the American Society of Training Directors, 14(4), 28-32.

Lewis, T. (1996). A model for thinking about the evaluation of training. Performance Improvement Quarterly, 9(1) pp.3-22.

Lintern,, G., & McMillan, G. (1993). Transfer for flight simulations. In R. Telfer (Ed.), Aviation training and instruction. Aldershot, United Kingdom: Ashgate.

Lintern, G., Roscoe, S. N., & Sivier, J. E. (1990). Display principles, control dynamics, and environmental factors in pilot training and transfer. Human Factors, 32, 299-317.

Mills, D. (1992). Nostalgia: The first full scope simulator, Dresden II. In A. Sharon (Ed.), Simulators IX: Vol. 24. Proceedings of the 1992 Simulation Muticonference. Dan Diego: Society for Computer Simulation.

Mills, D. (1997). "Essay on how to cost justify a training simulator." Available, World Wide Web: http://www.albany.net/~dmills/events.htm#wanted

Mohs, B., MacDiarmid, W. R., & Andrews, D. (1988). A review and annotated bibliography of training performance measurement and assessment literature. (DTIC Publication No. AD-A207-258). Alexandria, VA: Defense Technical Information Center.

Moller, L. & Mallin, P. (1996). Evaluation Practices of Instructional Designers and Organizational Supports and Barriers. Performance Improvement Quarterly, 9(4) 82-92.

Murray, W., & Schnneider, A. (1997). Using simulators for education and training in anesthesiology. American Society of Anesthesiologists Newsletter, 61(10), 2-4.

Newstrom, A (1985). Journal of Management Development, 5(5), 51-65.

Rummler, & Bracke, (1995). Improving performance: How to manage the white space on the organization chart. San Francisco: Josey Bass.

Sarri, L. M., Johnson, R. R., McLaughlin S. D. & Zimmerele, D. M. (1988). A survey of management training and education practices in U. S. companies. Personnel Psychology, 41(4), 731-743.

Schon, D. (1983). The reflective practitioner – How professionals think in action. New York: Basic Books.

Schwartz, D. (1997). Standards at Flight Safety International. Proceedings of the White House Commission on Aviation Safety and Security.

Thomas, R., Anderson, L., Getahun, L. & Cooke, B. (1992). Teaching for transfer of learning. Berkley, California: National Center for Research in Vocational Education.

Vreuls, D.V., & Obermayer, R. W. (1985). Human-system performance measurement in training simulators. Human Factors, 27(3), 241-250.