Cue Fidelity Evaluation: A Requirements-driven Approach to Training Effectiveness Evaluation

Laura M. Milham, Meredith B. Carroll, David L. Jones, Susan E. Dean, Dervon Chang*

Design Interactive, Inc.
Oviedo, FL

Laura, Meredith, David, Susan@designinteractive.net

*now at Basic Commerce and Industries, Inc., Dahlgren, VA

ABSTRACT

Although the task of performing an empirical Training Effectiveness Evaluation (TEE) is not trivial in any domain, the nature of military training makes it a particularly difficult arena. Specifically, in operational environments, the lack of experimental control, limited participants, and extreme expense and resource requirements all lead to challenges for evaluators investigating the degree that the training system addresses targeted goals.

A requirements-driven approach to TEE aimed at facilitating a less resource intensive evaluation has been designed. Advantages of this approach include the capability to be conducted early in the design cycle, allowing results to be fed back into the iterative design of the system. This approach leverages Sensory Task Analysis (STA) outputs to identify the multisensory information and interaction capabilities necessary to support the task, which are then used to evaluate the degree that the system supports these requirements. From this, redesign recommendations are developed, identifying specific system upgrades that would allow a greater percentage of training objectives to be targeted. This approach has been used successfully in the evaluation of multiple operational United States Marine Corps training systems including a desktop Fire Support Team (FiST) team training system, a high fidelity Forward Area Controller (FAC) training system, and a mixed reality infantry patrol trainer. This paper will discuss the method of conducting requirements-driven TEEs and initial validation results.

ABOUT THE AUTHORS

Laura M. Milham, Ph.D. is the Director of Training Systems at Design Interactive, Inc. She received her doctorate from the Applied Experimental and Human Factors Psychology program at the University of Central Florida in 2005. She has worked as Principle Investigator in support of the development and assessment of the effectiveness of training systems design. For this, she has performed task analyses; assessed design and training needs for VE systems; performed usability evaluations; developed design recommendations for system HCI; developed research plans; conducted Training Effectiveness Evaluations (TEEs) and Transfer of Training (TOT) evaluations. She has also been responsible for the design of training content, from scenario development to human performance assessment toolkits for individuals and teams. In addition she has developed guidelines for training tactical teams.

Meredith B. Carroll is a Senior Research Associate at Design Interactive, Inc. and is currently supporting training system design and evaluation for the office of Naval Research's Human Performance, Training and Education (HPT&E) Program. Her work focuses on Task Analysis, HCI design of VE training environments, design of Performance Measurement and Training Management tools and design and conduct of training effectiveness and transfer evaluations and experimentation. Her research efforts have been directed towards human/team performance and training in complex systems in aviation and military domains, with focuses on performance measurement and virtual training technology. She received her B.S. in Aerospace Engineering from the University of Virginia in

2001, her M.S. in Aviation Science from Florida Institute of Technology in 2003 and is currently a Doctoral Candidate in Human Factors and Experimental Psychology at the University of Central Florida.

David L. Jones is a Senior Research Associate at Design Interactive, Inc. His primary research focus is on designing multimodal interactive systems. He has a Masters in Industrial Engineering from the University of Central Florida, with a focus on human-computer interaction (HCI) and usability. He has a Bachelors in Human Factors Psychology from Embry-Riddle Aeronautical University, with a focus on human factors and computer science.

Dervon Chang is a Human Factors Engineer at Basic Commerce & Industries, Inc. (Dahlgren, VA) supporting the Human Systems Integration (HSI) division. She currently works on subcontracted projects for the USMC and DHS developing HSI inputs for requirements documents, including efforts toward a Program of Record (PoR) for Defense Acquisition. Past work includes evaluating operator-system interfaces, conducting usability testing, developing HSI redesign recommendations, and building training effectiveness evaluation (TEE) approaches suited for military training programs.

Cue Fidelity Evaluation: A Requirements-Driven Approach to Training Effectiveness Evaluation

Laura M. Milham, Meredith B. Carroll, David L. Jones, Susan E. Dean, Dervon Chang*

Design Interactive, Inc.
Oviedo, FL

Laura, Meredith, David, Susan@designinteractive.net

*now at Basic Commerce and Industries, Inc., Dahlgren, VA

INTRODUCTION

Training Effectiveness Evaluation (TEE) is a method of assessing the degree to which a system facilitates training on targeted objectives and the degree that training transfers to the operational environment. Organizations can utilize TEEs to better understand the overall value of an existing or newly implemented training program by illustrating strengths and weaknesses of the training program that should be either maintained, further developed, or improved upon to benefit the organization's performance as a whole. Although the inherent value of a TEE is undeniable, instantiating this concept has been elusive.

One challenge of conducting TEEs in operational settings is due to the difficulty in capturing whether learning has occurred and, as a result, is operationally relevant. Traditional TEEs have primarily been designed based on Kirkpatrick's (1959) four level model of training evaluation: reactions, learning, behavior, and results. While Kirkpatrick's approach does provide a structural framework, many TEEs have often fallen short of evaluating beyond trainee reactions and declarative learning, failing to capture training transfer behaviors in the operational environment and overall organizational effects of training. However, learning metrics focusing on declarative knowledge do not capture whether the system leads to the development and utilization of targeted skill sets. Next, in cases in which training and transfer measures are targeted, the logistical constraints present a challenge which limits both feasibility of successful completion and utility of the resulting data. A third challenge of operational TEE's is that they are focused on evaluating a training system after it has been fully developed and instantiated into a curriculum, limiting their impact on improving the system after the evaluation had been conducted. Each of these challenges will be addressed below, a method for conducting 'training effectiveness in the wild' will be put forth, and initial work using this method will be discussed.

Challenge 1: Capture of Meaningful Metrics

Operational TEEs are typically limited to trainee reactions to the training and the impact the training has on declarative knowledge, due to the ease of collecting this type of data. There is equivocal evidence supporting existence of a relationship between trainee reactions and learning or any other training outcome (Alvarez, Salas, & Garofano, 2004). However, when not supplemented with other metrics, they may not provide sufficient information regarding the utility of the system. Trainees may find a training program or system enjoyable or have the impression of utility but the training may have little or no impact on actual operational performance (Salas, Milham, & Bowers, 2003). Target operational performance is often skilled demonstration of performance of procedural and perceptual skills consolidated into effective task performance. These aspects of performance cannot be captured via a knowledge test.

Challenge 2: Logistical Constraints

As military training is instantiated in operational units, there are several factors inherent to the domain (e.g., lack of experimental control in operational environments, limited participants), which threaten the validity of inferences made based upon the training results (Boldovici, Bessemer, & Bolton, 2002). For example, often training cannot be withheld from a group of soldiers. In these cases, soldiers in pre-deployment training programs will receive all possible training intervention opportunities available to ensure their safety and that of their teammates. As such, experiments comparing a control group (either with or without alternative training) to an experimental group may be infeasible. Given that inferences made from performance results cannot be unequivocally attributed to the training itself without a baseline control group against which to compare, evaluation techniques must be extended to include quasi-experimental design methods. Other limiting factors include having access to only a limited numbers of participants and rigid scheduling issues, both of which require that evaluations are designed to be performed during trainees traditional training courses.

Challenge 3: Pre-instantiation Evaluation

Traditionally, TEEs are performed after a training system is fully developed and instantiated in the training curriculum. Although post-instantiation TEEs provides data regarding the utility of a training system or approach, the late evaluation limits the impact that results have on updating the system.

A lifecycle approach to TEEs is designed to follow the development lifecycle in order to provide high-level input as the training system is initially being developed and more precise guidance as training system releases are made. The most evident advantage to this approach is that the input provided by the TEE can be used to shape the future development of the training system. This type of approach allows evaluation and redesign with multiple system drops (i.e. spiral development), giving training system designers and evaluators the opportunity to evaluate whether or not the changes that are instantiated are effectively increasing the utility of the training system that is under development and the training program as a whole.

Addressing the Challenges

In order to address these challenges, a requirements-driven approach to TEE has been designed which is 1) grounded in training science on fidelity to allow effective evaluation of a system's ability to target training objectives and transfer to operational performance, 2) aimed at facilitating a less resource intensive evaluation of training systems which does not interfere with operational training and 3) performed early enough in the design cycle to effectively impact design. This approach leverages Sensory Task Analysis (STA) data to identify the required multisensory information and interaction capabilities necessary to support task performance (Milham, Carroll, Stanney, & Becker, in press). These cue fidelity and interaction capability requirements are then used to guide the evaluation of how well a training system supports these requirements. From this, redesign recommendations are developed, identifying specific system upgrades that would allow a greater percentage of training objectives to be targeted. With this approach to TEEs, high-level input is provided throughout the development lifecycle. The most evident advantage to this approach is that the input provided by the TEE can be used to shape the future development of the training system, a much more cost efficient option than later redesign. This approach has been used successfully in the evaluation of multiple operational United States Marine Corps training systems including a desktop Fire Support Team (FiST) team training system, a high fidelity Forward Air Controller (FAC) training system, and a mixed reality infantry trainer. To illustrate the utility and outcomes of this approach, the methodology will first be described, then several case studies will be discussed to show how outcomes can be used to affect the effectiveness of training systems

REQUIREMENTS-DRIVEN TEE APPROACH

A requirements-driven approach to TEE can be performed by conducting a cue fidelity evaluation combined with an interaction capability evaluation. This requirements-driven TEE method attempts to answer the questions: *Does the system environment afford learning: Can learning occur?* These methods ensure that even the earliest system drops incorporate environmental cues and capabilities to support training objectives and effectiveness, before trainees have utilized the system.

Prior to introducing trainees to the system, data from a sensory task analysis is used to derive system requirements (i.e. level of cue fidelity required for each environmental cue to allow practice on target training objectives). Next, the evaluation examines the degree to which the system includes the cue fidelity requirements identified in the STA as necessary to address goal accomplishment (Herbert & Doverspike, 1990); identifying gaps between requirements driven by the TA and actual system specifications. Next, it is determined if the training system supports operational functionalities and coordination requirements.

Sensory Task Analysis

At the heart of the requirements-driven TEE process is the Sensory Task Analysis (STA). STA involves determination of multisensory information requirements necessary to facilitate learning of these targeted objectives. A sensory task analysis extends a traditional task analysis beyond task breakdown, flow, sequence, etc. to include identification of information an operator must receive to successfully complete a task, specifically, the critical multimodal cues (visual, auditory, haptic) and associated fidelity level required. More specifically, sensory task analysis is conducted to determine how trainees gather information from the operational environment and how they act upon the environment in the real world. For each task and subtask, one must identify the multimodal cues (visual, auditory, haptic, etc.) that the operator relies upon to perceive and comprehend the surrounding environment in order to successfully complete the task (Milham, Carroll, Stanney and Becker, in press). From the tap on the shoulder from a teammate, to the geometry of an incoming aircraft, to the crunch of the ground beneath a tiptoeing enemy's foot, relevant multimodal information requirements must be identified for each training objective. Knowing the multimodal cues on which a performer depends is not enough to ensure a training system facilitates training of target tasks. It is also necessary to deduce which aspects of the cues are relied upon and how the cues are used. For example, do personnel on the ground rely on merely a spot of black in the sky to detect incoming aircraft, or do they have to be able to see the wing positions in order to make fine discriminations of aircraft dynamics to assess if the aircraft is pointed at the correct target? It is important to define the task at this level of detail in order for to ensure cues are presented at an appropriate level of fidelity to allow successful performance of tasks and effective learning, without unnecessary technology costs.

Additionally, STA identifies the interactions of the user during task performance (e.g., physical tool use, communication with team members). To support transfer, systems should facilitate performance of the actions and procedures they will be required to execute in the field. These functionalities are identified through the STA, which facilitates systematic identification of the system functionalities/capabilities required for an operator to successfully complete each task (e.g. system requirement for buttonology issues to temporal realism). For example, given the task of using a laser designator to mark a target, in order to facilitate practice of the skills required to perform this task, some representation of the tool functionality must be present, the required fidelity of which is dependent on whether the goal is to train the cognitive or physical aspects of the task. To conduct the sensory task analysis, working from the task/subtask framework, information can be extracted both from training observation as well as a structured interview or questionnaire to probe an SME or instructor.

Evaluation

From STA outcomes, an evaluation is then performed to identify the gaps between requirements driven by the sensory TA and actual system specifications. To perform this evaluation, it is necessary to have an SME who is familiar with the task walk through the tasks being targeted by the training system. If this SME is not extremely familiar with the training system, it is necessary to have them view or actually perform in the training system. To extract information regarding cue fidelity and capabilities, the SME is walked through all cues and interactions requirements for each task and probed regarding the presence, characteristics and utility of each. For the cue fidelity, the SME rates cues based on a scale that defines the degree of match between the real world and the training system.

Two main products results from this evaluation. The first is a listing of what can and cannot be trained with the system. From this, training system effectiveness can be quantified based on the number of training objectives, tasks or subtasks supported. This allows a comparison of overall training effectiveness between systems (e.g. which system has a better training value for a specific training need) as well as a qualification of which systems can best train which tasks/training objectives. The second product consists of redesign recommendations that detail improvements to the training system that are necessary to allow targeting of specific training objectives and tasks.

Utility of Evaluation

The utility of the requirements driven cue fidelity method, specifically, "what the method gives you" is substantial. As discussed in the previous section, results indicate both a quantifiable effectiveness level and recommendations for achieving high levels of training effectiveness, but there are added benefits as well. With the availability of extremely high fidelity training systems it is important to understand the degree to which such fidelity aids in achieving target training objectives. There are both cues for which high fidelity presentation is critical to ensure training effectiveness and cues for which fidelity will provide little training benefit for added cost. This method facilitates cost benefit analysis of which recommendations, if implemented, would result in the greatest impact on training for the optimal cost. Such analyses can be performed iteratively to allow improvements to prototypes throughout the system build lifecycle to be quantified in terms of training effectiveness. Hence, the utility of the requirements driven cue fidelity evaluation extend to support in:

1. Determining quantifiably how well existing simulators are addressing task training objectives/training needs.

2. Determining optimal fidelity requirements necessary to achieve the highest degree of training effectiveness.
3. Determining the most cost effective means of meeting task training objectives/training needs.

REQUIREMENTS-DRIVEN TEE APPLICATION

To illustrate the utility of this method, several case studies will be discussed. The evaluation and outcomes for several training systems are described, including USMC individual trainers (e.g. Forward Air Controllers) and team trainers (e.g. Fire Supports Teams (FiST) and USMC Infantry Patrol.)

FiST Trainer

Iterative cue fidelity evaluations were performed on a USMC FiST training system across three versions released by the development team (V1-V3). As a first step, a training needs analysis was conducted to determine the training goals from the Fire Support Team (FiST) team domain. Next, those goals were mapped to the multimodal environmental cues required to support those goals. Finally, evaluations were conducted to evaluate the degree to which system specifications matched these requirements and results and redesign recommendations were provided to the design team.

Results

Outcomes from this process led to the identification of how early versions of the trainer could be redesigned to address gaps in tasks targeted and tasks effectively addressed. This led to specific redesign recommendation, which were incorporated into subsequent drops of the system. This iterative process led to a ***34% increase in the number of total training goals that were targeted*** between V1 and V3. This increased training utility and user interaction represented a substantial and quantifiable gain in overall training system effectiveness.

Infantry Patrol Trainer

A limited scope cue fidelity evaluation was performed on the USMC mixed reality training system used to train infantry pre-deployment training objectives ranging from MOUT room clearing to cultural interactions. The goals was to examine the degree that current hardware and software installations support targeted training goals as listed in the USMC Pre-deployment Training Package (PTP). For eight high priority training objectives, cue fidelity requirements were identified, and an evaluation was conducted via SME ratings during a walk through of the facility

Results

Results indicated that six out of eight (75%) high priority training objectives were addressed by the system at the time of evaluation. For targeted training objectives that could be improved, redesign recommendations were developed, detailing fidelity requirements for multimodal cues to bring the system to 100% capability for addressing high priority training objectives.

Table 1: TEE Validation Results

Training Task	Cue fidelity TEE results (Build 1)	Operational TEE results (Build 1)	Follow up Operational TEE (Build 2)
Mission Briefing/ Communication	System provided communications did not support natural interaction and timing	Transfer improvement after using Build 1 **not** present.	Transfer improvement after using Build 2.
Mission Execution	All cues required to perform general mission execution available at required levels.	Transfer improvement after using Build 1.	
Locating Aircraft	All cues required to locate aircraft available at required levels.	Transfer improvement after using Build 1.	
Final Attack Heading (FAH) Evaluation	All cues required to evaluate whether aircraft is outside designated FAH available at required levels.	Transfer improvement after using Build 1.	
Wings Level Evaluation	All cues required to evaluate whether aircraft is flying wings level on final approach available at required levels.	Transfer improvement after using Build 1.	
Aircraft Maneuvering Evaluation	Aircraft flight paths (models) integrated into Build 1 did not accurately depict the associated aircraft maneuvers.	Transfer improvement after using Build 1 **not** present.	Transfer improvement after using MSAT Phase 2B.
Evaluating SEAD effectiveness	All cues required to evaluate SEAD effectiveness available at required levels.	Transfer improvement after using Build 1.	
Evaluating Battle Damage Assessment	Ordnance effects not accurately visually depicted within Build 1.	Transfer improvement after using Build 1 **not** present.	Transfer improvement after using MSAT Phase 2B.

Individual FAC Trainer

A cue fidelity evaluation was performed on a high fidelity Forward Air Controller (FAC) trainer used to support schoolhouses by creating an environment to practice perceptual, procedural, team performance (via instructor playing supporting aircraft roles) and spatial skill sets required for the FAC task. Based on an in depth sensory task analysis, cue fidelity and capability requirements were identified and the degree to which the system specifications met these requirements was determined by a SME very familiar with the system.

Additionally, a series of operational TEEs were performed on this trainer in the field which facilitated partial validation of the requirements driven (cue fidelity and capability) TEE method. Working with schoolhouse instructors, the degree to which the FAC trainer impacted live fire performance on target training objectives was assessed. Upon each assessment, design recommendations were provided based on the theoretical TEE that was performed. Performing the TEE method described within this paper multiple times and performing an operational TEE after each provided an opportunity to compare the cue fidelity TEE results to empirical TEE results as well as track the utility of the effects of the recommended changes on training transfer capability.

Results

In both comparative evaluations that were carried out the results provided by the requirements-driven TEE closely matched those provided by the operational TEE. Specifically, as can be seen in Table 1, the results of the cue fidelity driven TEE (see column 2), suggested that the cues required for the tasks of mission briefing/communication, aircraft maneuvering evaluation, and battle damage assessment were not presented at the level required to effectively practice the tasks. The results of the operational TEE (see column 3) mimicked these results, validating that the approach was effective in predicting the tasks that could effectively be trained using the system.

After the evaluation and feedback was provided to the system design team, the suggested redesigns associated with the TEE results were implemented to improve system utility. A final operational TEE was performed on the system to determine if the system modifications led to increased levels of transfer on the tasks described above. As can be seen in the final column of Table 1, each of the improvements integrated by the design team led to additional training transfer by the training system, ultimately leading to increased overall performance levels of trainees.

DISCUSSION

In sum, the presented method provides an initial solution to collecting TEE data within the bounds of operational environments. By using STA methods, fidelity requirements are used to evaluate the degree that the environment can support targeted training objectives. This method does not require a large number of participants (often just a few SMEs), and can be conducted on partially developed systems, allowing early evaluation. As such, findings can be incorporated back into the development, allowing responsive and rapid integration of TEE data into the system development cycle. Resultant systems have training effectiveness built into the system from the ground up.

This method has been used across a variety of USMC training systems, resulting in early identification of potential issues, with quick mitigation and solutions to those issues. Although initial data is favorable, next steps may be to more thoroughly evaluate the degree that this method correlates with traditional methods of TEE.

REFERENCES

Alvarez, K., Salas, E., & Garofano, C. M. (2004). An Integrated Model of Training Evaluation and Effectiveness. *Human Resource Development Review*, *3*(4), 385-416.

Herbert, G.R. & Doverspike, D. (1990). Performance appraisal in the training needs analysis process: A review and critique. Public Personnel Management, 19, 3, 253-270.

Milham, L. M., Carroll, M. B., Stanney, K.M., & Becker, W. (in press). Training System Requirements Analysis. In D. Schmorrow, J. Cohn, & D. Nicholson (Eds.), *The Handbook of Virtual Environment Training: Understanding, Predicting and Implementing Effective Training Solutions for Accelerated and Experiential Learning*. Aldershot, Hampshire, UK: Ashgate Publishing.

Salas, E., Milham, L., & Bowers, C. (2003). Training evaluation in the military: misconceptions, opportunities, and challenges. *Military Psychology* 15(1) 3-16.