

Using an Analytical Approach to Understand Integrated Training Environment Utility

**LTC Glenn A. Hodges Ph.D.
Deputy Division Chief
Human Dimension Division, ARCIC
FT Eustis, Virginia
glenn.a.hodges.mil@mail.mil**

ABSTRACT

This paper discusses the development and application of an analytical assessment methodology anchored in systems engineering principles, affordance theory, and human abilities, to measure the potential of an integrated training environment's (ITE) ability to effectively support training. An integrated training environment is defined as any human in-the-loop training system that includes live, virtual, constructive, or game supported training aids, devices, simulators, or simulations (TADSS) alone or in combination, that support the deliberate practice of skills for defined mission tasks. Empirical investigation of ITEs is costly, lacks formal guidance, and is therefore often unreliable. If conducted, ad hoc studies, commissioned by individual organizations, constitute the current state of Army ITE evaluation. These assessments are often entirely based on subjective opinions gained through surveys, which produce results that are at best indirectly and loosely linked to the ITEs themselves. What is required is a repeatable, inexpensive, analytical approach to ITE assessment that bounds the potential of a given system to the support it provides to the deliberate practice of specific tasks. The results of this research include the development and use of the integrated training environment assessment methodology (ITEAM). ITEAM was used to evaluate the ability of several ITEs to support the deliberate practice of specific tasks during training. During application, ITEAM consistently predicted where training was supported by an ITE and generally how well. ITEAM is offered as a tool for assessing ITE utility post fielding and as a guide for defining and verifying ITE requirements during the systems engineering development process.

ABOUT THE AUTHOR

LTC GLENN A. HODGES U.S.A. works as the Deputy Division Chief in the Human Dimension Division (HDD) at the U.S. Army's Capabilities Integration Center (ARCIC). Prior to this assignment he was the Lead Human Performance and Learning Scientist in ARCIC HDD. During his career as an Armor Officer, LTC Hodges has served in various command and staff positions. In 2001 he was selected for and designated as a Simulation Operations Officer (FA57). In this capacity LTC Hodges has worked with and employed various training aids, devices, simulators, and simulations (TADSS) and mission command systems. LTC Hodges holds a Bachelor of Science degree in Business Administration from Old Dominion University, a Master of Science degree and Ph.D. in Modeling, Virtual Environments, and Simulation (MOVES) from the Naval Postgraduate School.

Using an Analytical Approach to Understand Integrated Training Environment Utility

**LTC Glenn A. Hodges Ph.D.
Deputy Division Chief
Human Dimension Division, ARCIC
FT Eustis, Virginia
glenn.a.hodges.mil@mail.mil**

INTRODUCTION

The value of human-in-the-loop (HITL) simulation primarily comes from its ability to offer practice opportunities in environments that replicate important features of the real world (Salas, Rosen, Held, & Weissmuller, 2008). While this is true, at some point the focus of requirements determination, definition, and solution development for military training systems shifted focus away from human performance and skill acquisition towards advanced technology. Operational and system requirements documents (ORD/SRD) have driven the increasing focus on the technological aspects of possible training solutions while marginalizing the importance of the front-end human analysis. This situation has resulted in the common practice of providing technical requirements specifications for training systems to defense contractors and then requiring the defense contractor to provide the government with a detailed explanation of how the training system will support the user (Klein, Johns, Perez, & Mirabella, 1985).

Between conflicts, the Armed Forces rely heavily on integrated training environments (ITE) to maintain warfighting skills. ITE are comprised of various live, virtual, constructive, and game supported training aids, devices, simulators, and simulations, which allow Soldiers, Sailors, Airmen, and Marines to practice the skills and engrain the knowledge necessary to execute their combat missions successfully on the battlefield (Hodges, Darken, & McCauley, 2014). ITE are extremely resource intensive and are rarely described as lightweight or turnkey. They require verification, validation, and accreditation (VVA) just as their analytical counterparts that support budgetary and force structure decisions. A major difference between ITE that support training and other types of simulation is how they are evaluated.

The most common method of determining ITE effectiveness is through the use of empirical transfer of training (TOT) studies that are expensive and often provide limited or misleading insight into ITE utility. Some researchers have attempted to use non-empirical means to evaluate ITE in an effort to reduce costs and accurately capture positive system attributes (Tufano & Evans, 1982; Keesling et al., 1999; Sticha, Campbell, & Knerr, 2002, Gilligan, Elder, & Sticha, 1990). Despite their best efforts only a handful of researchers have had their techniques successfully implemented outside of the research arena and of those few have been used more than a handful of times (Johnston, Nolan, & Caldwell, 2015). Most of the techniques developed have not been extensible, user friendly or well documented to facilitate reuse. Additionally, many have used mathematical equations that have not been validated with empirical data. Many have been automated due to their extreme complexity without concern for program documentation making them nearly impossible to implement by others and their focus has been similar to that of empirical attempts.

Until 2012, the United States Army (USA) had a system to provide analysis of training programs called the Training Effectiveness Analysis (TEA) system. The TEA system, established in 1975, was a Training and Doctrine Command (TRADOC) program focused on the impacts associated with training and hardware costs, hardware development cycles and complexity, training resources, and the overall effectiveness of Army programs to prepare Soldiers for battlefield conditions (Neal, 1982). Prior to 2012, the TRADOC Analysis Center at White Sands Missile Range (TRAC-WSMR) was the Army's lead agency for providing technical assistance and conducting TEA for training systems. TRADOC Regulation 350-32 governed the TEA program. At the time, Simpson (1995) offered that the Army TEA system was the most robustly defined training analysis system that existed. Several system analysts have described the use of TEA studies for the benefit of their respective programs and offered examples of how they conducted TEA studies (Carter, 1982; Maitland, 1982). Despite this, in the summer of 2012, the USA officially concluded its last TEA study, eliminating both the office responsible for the conduct and oversight of TEA, and the regulation that governed the TEA system (Drillings, 2013).

HUMAN ABILITIES AND AFFORDANCES

Human ability (HA) research has been ongoing since the 1960's and has been used as a tool for empirical work investigating training system design and fidelity (Hays & Singer, 1988; Napoletano, 2013). Initially, the Defense Advanced Research Projects Agency (DARPA) sponsored research into HA to assist the military with job placement and training (Cockayne, 1998). The HA body of research has been developed as part of an umbrella taxonomic effort attempting to standardize the way human performance is described. The objective of the ability requirements approach was to identify and define the fewest number of independent ability categories that would be useful and meaningful for describing performance in the widest variety of tasks (Fleishman & Quaintance, 1984). HA development is an iterative process intended to produce a list of verified abilities that are empirically derived from patterns of responses to different tasks. The assumption is that specific tasks require certain abilities and that tasks requiring the same types of abilities can be categorized similarly. This assumption allows researchers to discuss task performance in relative terms. The HA project, through experimentation and collaboration with multiple subject matter experts, derived 52 HA with the possibility of adding more. Examples of HA are oral comprehension, deductive reasoning, dynamic strength, peripheral vision, and sound localization. HA are grouped into one of four categories (i.e., physical, sensory, psychomotor and cognitive). The United States Department of Labor uses HA as the basis for their O*NET (<http://www.onetonline.org>) program that provides information about jobs based on the HA needed to execute them.

Through years of research, Fleishman and his colleagues analyzed various jobs and tasks to ascertain and develop the list of 52 human abilities that can be found throughout various human activities. During this process, they executed numerous task analyses (TA). Through their process of defining ability requirements, they linked information dealing with task characteristics to HA (Fleishman & Mumford, 1991; Fleishman & Quaintance, 1984; Fleishman & Bartlett, 1969). The results of their efforts led to a means of description, understanding and categorization of human activity (i.e., work) based on HA instead of through the use of TA. HA are viewed as enduring attributes of the human being (i.e., they are the same in the real or virtual world) and they play an important role in the methodology discussed here.

Affordance theory comes from ecological psychology and James J. Gibson. Gibson (1986) coined the term "affordance" to capture the essence of what an environment offers or provides an animal in either a positive or negative fashion. Affordance theory provides a context for discussing the qualities of the human-environment relationship within an ITE. Precedent exists for the use of affordance theory in supporting computer science and human factors research (Bærentsen & Trettvik, 2002; Chemero & Turvey, 2007; Lintern, 2000; Rome, Paletta, Şahin, Dorffner, Hertzberg, Breithaupt, Fritz et al., 2008). Affordance theory is naturally associated with HA, most notably with human perception. Gibson's theory of affordances has been met with varying degrees of enthusiasm and criticism over the years (Jones, 2003). As initially described, the concept of affordances was simple, clear and appealing (Michaels, 2003). However, Gibson's later attempts to describe affordances in more detail, resulted in a situation that "makes them seem like impossible, ghostly entities, entities that no respectable scientist (or science worshipping analytic philosopher) could have as part of their ontology" (Chemero, 2003, p. 182). Attempts at providing clarity and concrete definitions for affordances have been offered and debated (Stoffregen, 2003; Turvey, 1992).

Affordances are used in this research as a means of identifying the qualities and characteristics of the ITE that are absent or present in relation to the HA associated with specific tasks. We have elected to use affordances as part of our methodology because they provide context and allow us the opportunity to view an ITE unlike any other approach. Using affordances we are not only able to identify the characteristics of an ITE that support deliberate practice; but also why those identified characteristics are important to the trainee's execution of the tasks. Through the use of affordances, we are able to determine specific task elements with the highest likelihood of positive TOT.

INTEGRATED TRAINING ENVIRONMENT ASSESSMENT METHODOLOGY (ITEAM)

Figure 1 depicts the integrated training environment assessment methodology (ITEAM), a human-centered systems engineering approach to ITE analysis. ITEAM was developed based on the lessons learned from the literature and based on the recognition that front-end human analysis is critically important to training system development. Of the pieces of a training program (e.g. technology, requirements, humans) it has been established that computer technology evolves the fastest (i.e. Moore's Law). Requirements determination occurs more slowly. Human beings

evolve the slowest yet their evolutionary stability is ignored in ITE development in favor of an emphasis on advanced technology. ITEAM takes this into account by focusing on the support provided by an ITE to the deliberate practice of specific tasks and not on any specific technology. ITEAM was developed as a set of three main processes each containing multiple sub-processes. All of the sub-processes are iterative in nature and steps may be abbreviated or skipped depending on the time available and level of detail required. Requirements definition occurs first and proceeds from left to right beginning with determining the need and ending with determining the real world (RW) affordance requirements. Verification follows and builds on requirements definition by determining the ITE HA and ITE affordances. Assessment of ITE support to training happens last and only after the RW and ITE affordances have been identified for comparison.

Requirements Definition

Proper problem description and analysis are critical to the ITE development process. ITEAM groups the activities of determining the need for the ITE, how it will be used, which functions will be performed by the ITE and the human, description of the tasks to be executed during training and the desired learning outcomes, within the boundary of requirements definition. Also included is a list of

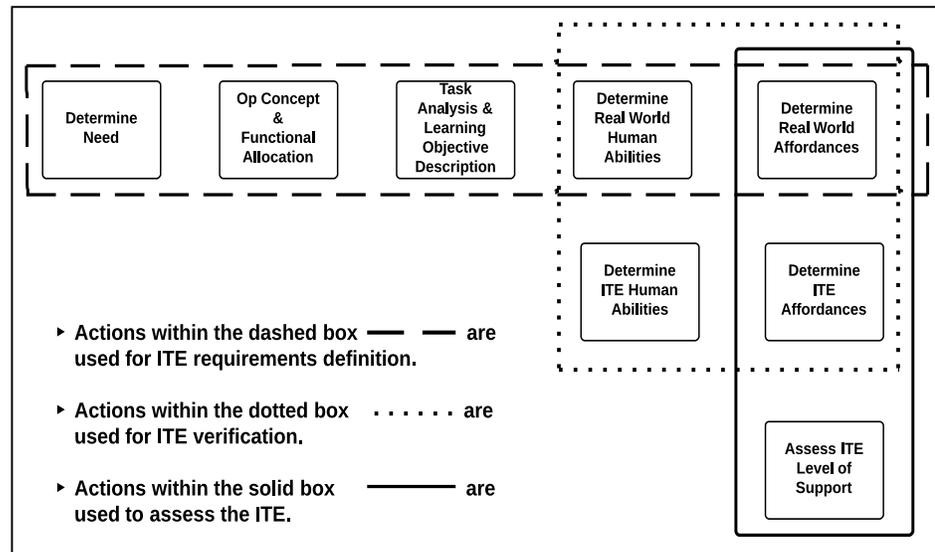


Figure 1. Integrated Training Environment Assessment Methodology

RW HA and RW affordance requirements

that are necessary to accomplish the training tasks. HA are used to help illuminate the critical aspects (i.e. environmental affordances) required of the ITE. Affordances are used to describe the attributes of the ITE that are necessary to support the execution of the desired training.

Verification

Verification is defined as “the process of determining that a model or simulation implementation and its associated data accurately represent the developer’s conceptual description and specifications” (Under Secretary of Defense, 2009, p. 10). The sub-processes of ITEAM considered to be useful for verification consist of compiling the identified RW and system-supported HA as well as the RW and system-provided affordances. During this process, the evaluator uses the TA to determine the RW HA and affordance requirements associated with the tasks to be trained. Then, the ITE is investigated to determine what HA it supports and what affordances are available. Comparison of these items provides the basis for an initial judgment on whether or not the ITE will support the execution of the desired training.

Assessment

The final process of ITEAM assesses ITE ability to support desired training by quantifying ITE affordance resources based on ITE affordance requirements. The quantification of resources provides the customer/stakeholder/user with an estimate of the level of support that the ITE provides. ITE scoring is based on a subject matter expert (SME) evaluator’s judgment on the absence or presence of specified affordances using the scale seen in Figure 2. The scale was set up so that the first two scoring levels (Poor and Fair) each contain 25 percent. The next two scoring levels (Good and Very Good) each contain an additional 20 percent and the final scoring level (Excellent) contains only 10

percent. Constructing the scale in this manner provides a progressive level of difficulty in reaching a rating of excellent, which requires that an ITE contain 90 percent or better of the affordances identified as being required to support the deliberate practice of specific tasks.

Subtask Affordance Scoring in Detail

Subtask Affordances are Unique. A unique affordance is one that has not been previously evaluated or accounted for as part of another subtask evaluation. If a subtask's affordances are unique, then a simple average of the number of affordances present divided by the total number required provides the percentage of affordances available for the subtask. This percentage is compared to the rating scale (Figure 2) and results in a rating of 1–5 Poor to Excellent.

Scale Definition	
5–Excellent	– the ITE contains all but a few (90–100%) of the affordances determined during the analysis
4–Very Good	– the ITE contains a significant portion (70–89%) of the affordances determined during the analysis
3–Good	– the ITE contains a good portion (50–69%) of the affordances determined during the analysis
2–Fair	– the ITE contains some (25–49%) of the affordances determined during the analysis
1–Poor	– the ITE contains very few (0–24%) of the affordances determined during the analysis

Figure 2. ITEAM scoring scale definition

Subtask Affordances are Previously Accounted for. If a subtask's affordances are completely accounted for in other analyses (referred to as affordance rollups), then those analyses are consulted and the ratings for their subtask affordances are obtained and averaged together to compile a numerical score for the current subtask under assessment. If more than one rollup is listed, then this process is executed for each of those rollups. Once all of the subtask affordance scores are collected they are summed and then divided by the total number of subtasks involved to obtain an average score. The average score now represents a number on the rating scale of 1–5 (see Figure 2). Raw scores containing 0.50 or less are rounded down to the nearest whole number for scoring purposes. Scores containing 0.51 or greater are rounded up.

Subtask Affordances are Partially Unique. If the affordances for a subtask are partially unique and partially accounted for in other analyses then the calculation is conducted in three steps. **Step one**—Treat each affordance rollup as an individual affordance that is present and unique. **Step two**—Evaluate and account for the presence of any unique affordances associated with the subtask. Once every affordance is accounted for, the calculation for determining the percentage present is conducted as described in (subtask affordances are unique). The result (rating of 1–5) is temporarily assigned as the subtask score. **Step three**—Obtain the values (scores) for the subtask affordances from the previous analyses (see subtask affordances previously accounted for) and sum them. Add the temporary value for the subtask currently under evaluation. Average this value by the total number of subtasks (including the current one). The derived number represents a number on the scale between 1 and 5 (see Figure 2) that when rounded appropriately (0.50 and lower round down) provides the qualitative rating for this subtask.

Subtask Affordances Contained in Multiple Analyses. In the case where a task's affordances are accounted for in multiple nested layers of sub-analyses, we have elected to stop the decomposition at the top of the second nested level. In such a case the top-level raw score of the high-level task at the second nested level is used in the value calculation for the current subtask. By our estimation, conducting further decomposition during the analysis leads to inflated results.

High-level Task Scoring

High-level tasks are also scored using the scale seen in Figure 2. The procedure to score a high-level task consists of summing all of the subtask scores and dividing them by the total number of subtasks. The result is a numerical value

that is associated with a level of support provided (Poor to Excellent) by the ITE to the deliberate practice of the task.

APPLYING ITEAM TO ASSESS THE GAME VIRTUAL BATTLESPACE 2

The Assessment in a Nutshell

This study re-examined the training effectiveness analysis (TEA) study of game supported training using Virtual Battlespace 2 (VBS2): U.S. Army (USA). Version 1.23 of VBS2 was used in the original evaluation. Unfortunately, that version of the software was not available for this analysis so version 1.40, resident on the U.S. Marine Corps (USMC) Deployable Virtual Training Environment (DVTE), was used. While this version of the software was technically VBS2: USMC and newer, it contained all of the same base models and behaviors as VBS2: USA. Personnel working in the TRADOC capability manager for gaming (TCM-Gaming) office confirmed this fact. To avoid any confusion in the discussion below, the general acronym VBS2 is used. Readers interested in viewing the full analysis of VBS2 are encouraged to visit Hodges (2014).

Method

Brief Description of Empirical TEA Study

In 2009, TCM-Gaming in conjunction with the Army Research Institute (ARI) and Aptima Inc., conducted research designed to empirically shed light on the issue of game supported training effectiveness (Ratwani, Orvis, & Knerr, 2010). The study employed observational methods to several small unit events at the USA installations of Fort Hood Texas and Fort Lewis Washington to collect data in support of six hypotheses. The purpose of the TEA was to study the overall effectiveness of VBS2 and the impact of situational variables on training outcomes. Situational variables were used to build metrics that supported measuring skill acquisition during training. The measures applied as part of the evaluation protocol were skill preparedness, training motivation, task performance, unit process, unit cohesion, unit efficacy, and unit effectiveness. Surveys designed to gather data for each of these measures were used.

Application of ITEAM to VBS2 Training Environment

During this study, the full range of ITEAM processes and sub-processes were employed as depicted in Figure 1. The introduction of the game supported training effectiveness TEA stated the following as the USA's need. "The Army needs methods for providing soldiers and leaders with effective training and opportunities to practice tasks effectively and efficiently" (Ratwani et al., 2010, p. 1). Supporting this statement of need was additional language indicating that USA personnel were already exploiting low-cost, technology-based solutions and innovative training methods in order to increase the impact and effectiveness of training. TRADOC, a significant stakeholder in the USA training and educational domain, "recognized that games have the potential to *augment and improve* [emphasis added] military training for both individuals and collectives" (Ratwani et al., 2010, p. 1). This recognition was based on the USA use of the games DARWARS Ambush! and Tactical Iraqi for convoy and language training support. The need stated in the TEA served as a point of departure for this analysis but provided little assistance to our efforts. According to Ratwani et al. (2010, p. 1) the USA needs "methods for providing soldiers and leaders with effective training" and "opportunities to practice tasks." The former statement is extremely vague and the latter has to do with time not ITEs. The meanings of "effectively" and "efficiently" all depend on how the USA defines the terms.

The operational concept for employing game support to the deliberate practice of skills in this specific instance was a controlled classroom setting. A stated assumption of the TEA was that this training was the "crawl" part of the USA crawl, walk, run tiered training approach. The tactical scenario on which the evaluation was based and most of the training that occurred revolved around combat convoy operations. Trainees utilized both desktop and laptop computers with standard keyboards, mice, and headsets that allowed for their control and communication during mission execution. Commercial off-the-shelf (COTS) vehicle controls were available for use by trainees designated as vehicle drivers. Synthesis of this operational concept, tactical scenario, and individual feedback from Fort Hood and Fort Lewis, resulted in the general recognition that VBS2 support to the deliberate practice of tasks was mainly cognitive and sensory in nature. The functional allocation derived from the available information supported this

conclusion. VBS2 provided the entire stimulus and scenario environment where soldiers would control avatars to execute specific activities. Classroom space at the local mission command training center (MCTC) provided the physical environment and surroundings. The scenario development capability was part of VBS2. The unit leader and/or the civilian controller at the MCTC executed all scenario manipulations. Soldiers were represented in the game environment via avatars. Trainees accomplished avatar control through the use of the keyboard and mouse. Trainees were responsible for vehicle control using a steering wheel/pedal interface. Enemy actions were controlled by VBS2 artificial intelligence or by a MCTC employee. All equipment and vehicles necessary to conduct mission practice were provided by VBS2 within the game scenario.

Table 1 depicts the partial analysis sheet for two of the tasks investigated as part of the ITEAM VBS2 analysis. The remainder of this section describes how the analysis was conducted for VBS2. The task analysis (TA) conducted in support of this effort began using the 13 items listed as skill preparedness items in the TEA. Additional tasks were listed under task performance, but those items were interpreted as the tasks necessary to use VBS2 (i.e., avatar control “buttonology”) and were not included in the TA. Each skill was used to search for relevant doctrine in the Central Army Registry (CAR). The CAR subsumed what was previously known as the Reimer Digital Library (RDL). For each skill, multiple doctrinal references were consulted and used to develop the TA. For those readers interested in viewing the full TA consult Appendix D of Hodges (2014). The training objectives used to guide the conduct of unit training during the original TEA were not available for our ITEAM assessment of VBS2.

The real world human ability (HA) inventory associated with the TA was conducted by reviewing the definition of each of the 52 HA in Fleishman & Quaintance (1984) and determining their applicability to the tasks. Those determined to be applicable were listed next to the tasks and subtasks in the spreadsheet. The rule of thumb applied to this process was to err on the side of commission versus omission. Only a small amount of iteration was applied to refine the HA list for this study due to time constraints. Familiarity of the HA’s through iterative review ultimately sped up the analysis process of this and other case studies.

The real world affordance list flowed from the description of the tasks as well as the real world HA associated with the tasks. These items were also listed next to the tasks and HA in the spreadsheet. Environmental HA supported by the game, were identified by conducting the same process as described above for the real world HA. The ITE was interrogated looking at each task/subtask to determine which HA were related to game play or manipulation of the game. Example questions that guided this sub-process may be viewed at appendix B of Hodges (2014).

Table 1. Example ITEAM analysis

The affordance inventory for VBS2 was conducted in the following manner. First a facsimile of the TEA study ITE was developed using one DVTE suite consisting of four laptop computers, headsets, mice, and keyboards. The laptops were connected in a closed loop network so that four individuals could participate together as a crew. One station was equipped with a steering wheel and pedals and was designated as the driver station. Next we played the three built-in training scenarios that VBS2 provided. Following this exposure, we investigated the

Real World (RW) Preconditions: Sector of responsibility for visual/optical/weapons scan is provided to the SM.			Environmental Evaluation		
RW Tasks	RW Human Abilities	RW Affordance Requirements	Environment HA	Environmental Affordances	Level of Support to Training
Confirm sector of responsibility	<p>COGNITIVE: Oral Expression; Oral Comprehension; Spatial Orientation; Flexibility of Closure;</p> <p>SENSORY: Night Vision; Peripheral Vision; Glare Sensitivity; Depth Perception; Far Vision; Near Vision; Visual Color Discrimination; Speech Clarity; Speech Recognition;</p> <p>PSYCHOMOTOR: None</p> <p>PHYSICAL: None</p>	<ul style="list-style-type: none"> - Weather/Climate - Urban/Suburban environment with sprawl - Identifiable terrain or other features - Optics (Weapon, Vehicle, Personal) - A Means of communication with leader (FM/Digital communication device) 	<p>COGNITIVE: Oral Expression; Oral Comprehension; Spatial Orientation; Flexibility of Closure; Memorization; Time Sharing;</p> <p>SENSORY: Near Vision; Visual Color Discrimination; Auditory Attention; Speech Clarity; Speech Recognition;</p> <p>PSYCHOMOTOR: Finger Dexterity; Rate Control; Multitimb Coordination; Wrist-Finger Speed;</p> <p>PHYSICAL: None</p>	<ul style="list-style-type: none"> - Scenario editor that allows for the building of scenarios containing weapons, optics, replication of weather/climate, terrain, enemy units, friendly units, vehicles; pedestrians etc. - Avatar with 1st person or 3rd person view wearing appropriate personal clothing and equipment - Headset to communicate with others - Keyboard control cheat sheet - Mouse 	<p>*5 – Excellent 4 – Very Good 3 – Good 2 – Fair 1 – Poor</p> <p>All present</p>
Using optics/eyes: observe an area of assigned responsibility for threats using an appropriate search pattern	<p>COGNITIVE: Spatial Orientation; Memorization; Selective Attention; Speed of closure; Flexibility of Closure</p> <p>SENSORY: Night Vision; Peripheral Vision;</p>	<ul style="list-style-type: none"> - Area to observe - Potential threat activity 	<p>COGNITIVE: Spatial Orientation; Flexibility of Closure; Speed of Closure; Memorization; Time Sharing;</p> <p>SENSORY: Near Vision; Visual Color Discrimination;</p>	<ul style="list-style-type: none"> - Keyboard control cheat sheet - Mouse - Keyboard - Scenario editor that allows for the building of scenarios containing weapons, optics, replication of weather/climate, 	<p>*5 – Excellent 4 – Very Good 3 – Good 2 – Fair 1 – Poor</p> <p>All present</p>

model library available to all users and developed several small-scale vignettes to better understand the capabilities of the scenario editor. During this process we contacted personnel at the Fort Hood MCTC and the TCM-Gaming

office with questions about game capability. Finally, we enlisted the aid of one Army officer who had experience as a trainer and developer of VBS2 scenarios to help us develop and work through focused vignettes. Scoring of affordances, subtasks and high-level tasks was conducted as described in appendix D of Hodges (2014). Table 1 provides an example of the analysis conducted.

DISCUSSION

Comparison of VBS2 TEA and ITEAM Analysis Results

Readers interested in the full results of the TEA should view Ratwani et al., (2010). The results presented here for comparison are those items investigated to shed light on the skill preparedness of the trainees both before and after using VBS2. Subjective questionnaires were used as the data collection method for the original TEA. The questionnaires asked subjects to rate their preparedness both prior to and post training with VBS2. Table 2 depicts the pre and post-training results of the seven items reinvestigated as part of this case study. 141 participants in two locations were asked their opinions of their preparedness to conduct the activities listed after conducting convoy training in VBS2. TEA scoring used an ordinal scale ranging from 1–5 with 1 = Unprepared; 2 = Slightly Prepared; 3 = Neither Unprepared nor Prepared; 4 = Slightly Prepared; and 5 = Prepared.

Table 2. Skill preparedness results from TEA

Skill Preparedness Item	Mean Pre/Post Tng	Stand Dev. Pre/Post Tng
Scan my sector of responsibility	4.39 / 4.35	.83 / .92
Comply with rules of engagement	4.12 / 4.29	1.06 / .91
Communicate with members of your unit	4.20 / 4.13	.82 / .92
React to IED	3.94 / 4.04	1.02 / .99
Coordinate activities with your chain of command	3.84 / 3.90	.96 / .98
React to an attack	3.96 / 3.85	.91 / 1.04
Conduct CASEVAC/Recovery Operations	3.68 / 3.75	1.05 / 1.02
Scale		
5 - Prepared		
4 - Slightly Prepared		
3 - Neither Unprepared nor Prepared		
2 - Slightly Prepared		
1 - Unprepared		

The scores were averaged and the mean results and standard deviations are provided below. Table 3 depicts the results of the ITEAM assessment of VBS2 1.40. During the assessment, only seven items were reevaluated due to a time constraint on the study. One of the seven (Conduct CASEVAC/Recovery Operations) was broken into two tasks to simplify the analysis. The results of the scoring of the two subsequent tasks were combined and averaged for purposes of comparison. At the outset of the evaluation we recognized that the myriad of physical tasks associated with the skills under investigation could not be supported using VBS2. Our evaluation discovered that VBS2 contains many if not all of the affordances we listed as necessary for the deliberate practice of the skills identified. This result led to one interesting question and one interesting finding. Since VBS2 was initially designed as a first-person shooter game where players control an avatar that conducts actions directed by the player, is it possible that a soldier can be trained in an activity through the control of an avatar and to what extent? Discussion of this question with other researchers, soldiers, and civilians who support the development of the ITE reached similar conclusions. No, it is not acceptable to assume that soldiers are trained in a task by controlling the actions of an avatar through a process or action.

Table 3. ITEAM Analysis of VBS2

Skill Preparedness Item	ITEAM Score
Scan my sector of responsibility	5.00
Comply with rules of engagement	5.00
Communicate with members of your unit	3.66
React to IED	5.00
Coordinate activities with your chain of command	5.00
React to an attack	5.00
Conduct CASEVAC/Recovery Operations	3.42
Scale	
5 - Excellent: ITE contains all but a few (90–100%) of the affordances determined necessary	
4 - Very Good: ITE contains a significant portion (70–89%) of the affordances determined necessary	
3 - Good: ITE contains a good portion (50–69%) of the affordances determined necessary	
2 - Fair: ITE contains some (25–49%) of the affordances determined necessary	
1 - Poor: ITE contains very few (0–24%) of the affordances determined necessary	

It is reasonable to assume that there is an amount of learning involved but investigating the extent to which that is true was beyond the scope of our investigation. We believe that this highlights an intuitive finding that game supported ITE are best suited for supporting the practice of cognitively dominant tasks. In VBS2 it is possible to extract a wounded soldier from a damaged vehicle and drag him to a non-standard medical vehicle where he can be evacuated. VBS2 goes so far as to automatically place the casualty in the evacuation vehicle.

Practice of the cognitive tasks of assessing a situation and taking the appropriate actions are definitely supported. However, the physical tasks of lifting and dragging the casualty, opening doors, and walking are not supported even though the trainee controls the avatar that does those physical actions. The finding involved bias and its effect on assessment. Our bias against a game's ability to usefully support the deliberate practice of tasks involving physical skills almost derailed our ability to objectively assess the capabilities of VBS2. This dilemma occurred even though we were following a logical process to conduct our assessment. If evaluator bias is capable of derailing assessment when a logically unbiased process is used, what can we expect if and when no logical process is used? This situation strengthens our belief that a methodical process must be enforced and used during ITE assessment.

CONCLUSIONS

Several conclusions were drawn from this basic study into the use of ITEAM as methodology for evaluating the utility of an ITE. First, while it is acceptable to apply survey mechanisms to a training audience to establish the likeability and surface usability of an ITE, this approach does little to provide substantive feedback on the true capabilities of the ITE. ITE capability is best determined by interrogating the ITE itself to determine all of the features and capabilities that the ITE affords a user. Relying on users who may not understand or use the full capabilities of an ITE as a basis to judge ITE utility is only marginally useful. Through the use of ITEAM to investigate the capabilities of VBS2, we identified several system capabilities that were never introduced to the training audience during the original TEA study that resulted in our scores of ITE capability being higher than those in the TEA study. The context in which an ITE is employed and the scenarios used play a critical role in determining ITE utility. In this study and others using ITEAM, we often found that users surveyed about the effectiveness of an ITE would rate it poorly because the scenarios used to test the ITE did not fully leverage or highlight the features and capabilities of the ITE. Using ITEAM to investigate VBS2, resulted in a deep understanding of the system's capabilities as well solid insights into where and how the game could be best used.

The use of human abilities as a layer of analysis coupled with the task analysis provided useful information into the gaps and capabilities of the game. Additionally, analyzing the human abilities required to accomplish the tasks in the real world highlighted the true nature of the tasks (i.e. whether they were mainly cognitive, sensory, physical or psychomotor or a combination). This is extremely useful in two ways. First, it is useful in assisting the designers and developers of an ITE to recognize where resources should be focused (i.e. what aspects of the system should be high fidelity or not) based on the tasks to be practiced. Secondly, it is useful to help the user community understand the best ways to employ the ITE to support training (i.e. inform the user what the system is good for).

ITEAM does not attempt to determine appropriate levels of fidelity for system affordances. It cannot be relied upon to answer questions about the quality of the affordances in the ITE. ITEAM only provides insight as to whether or not specific affordances are available. With that information, a generalized rating of game support for training was provided. This research approached all affordances as being equal. Reality demonstrates that some things are more important than others, so a way of factoring that into the evaluation and scoring process of the methodology is necessary.

This research investigated the use of an analytical assessment methodology anchored in systems engineering principles, affordance theory, and human abilities, to measure the potential of an integrated training environment (ITE) to effectively support training. Empirical investigation of ITE is costly, lacks formal guidance, and is therefore often unreliable. Ad hoc studies, commissioned by individual organizations, constitute the current state of Army ITE evaluation. These assessments are often entirely based on subject matter expert judgment through surveys, which produce results that are linked indirectly and loosely to the ITE. What is required is a repeatable, inexpensive, analytical approach to ITE assessment that bounds the potential of a given system to the support it provides to the deliberate practice of specific tasks. ITEAM is offered as one possible solution to this identified gap.

REFERENCES

- Bærentsen, K., & Trettvik, J. (2002). An activity theory approach to affordance. In *Proceedings of the second Nordic conference of Human Computer Interaction* (pp. 51–60). New York: ACM.
- Carter, R. (1982). Methodologies for Evaluating Training Products and Processes. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 26(3), 258–262.

- Chemero, A. (2003). An Outline of a Theory of Affordances. *Ecological Psychology*, 15(2), 181–195.
- Chemero, A., & Turvey, M. (2007). Gibsonian Affordances for Robotocists. *Adaptive Behavior*, 15(4), 473–480.
- Cockayne, W. (1998). *Two-handed, whole-hand interaction*. Naval Postgraduate School.
- Fleishman, E., & Mumford, M. (1991). Evaluating Classifications of Job Behavior: a Construct Validation of the Ability Requirement Scales. *Personnel Psychology*, 44(3), 523–575.
- Fleishman, E., & Quaintance, M. (1984). *Taxonomies of Human Performance: The Description of Human Tasks* (1st ed.). Orlando: Academic Press.
- Fleishman, E., & Bartlett, C. (1969). Human abilities. *Annual Review of Psychology*, 20(1), 349 to eoa.
- Gibson, J. (1986). *The Ecological Approach to Visual Perception*. Hillsdale: Lawrence Erlbaum Associates, Inc.
- Gilligan, E., Elder, B., & Sticha, P. (1990). *Optimization of Simulation-Based Training Systems: User's Guide*. Alexandria.
- Hays, R., & Singer, M. (1988). *Simulation Fidelity in Training System Design: Bridging the Gap Between Reality and Training*. New York: Springer-Verlag.
- Hodges, G. (2014). *Identifying the Limits of an Integrated Training Environment using Human Abilities and Affordance Theory*. Naval Postgraduate School.
- Hodges, G., Darken, R., & McCauley, M. (2014). An Analytical Method for Assessing the Effectiveness of Human in the Loop Simulation Environments : A work in progress. In *Proceedings of the 2014 Spring Simulation Multi-conference*. Tampa: The Society for Modeling and Simulation International.
- Johnston, J., Nolan, M., & Caldwell, J. (2015). Capability Assessment of Test and Live Training Systems for Real-Time Casualty Assessment. In *Inter-Industry Training, Simulation and Education Conference*. Orlando: NTSA.
- Jones, K. (2003). What is an affordance? *Ecological Psychology*, 15(2), 107–114.
- Keesling, J., King, J., & Mullen, W. (1999). *Simulation Training Strategies for Force XXI Final Technical Report*. Alexandria.
- Klein, G., Johns, P., Perez, R., & Mirabella, A. (1985). *Comparison-Based Prediction of Cost and Effectiveness of Training Devices: A Guidebook*. Alexandria.
- Lintern, G. (2000). An Affordance-Based Perspective on Human-Machine Interface Design. *Ecological Psychology*, 12(1), 65–69.
- Maitland, A. (1982). Training Effectiveness Analysis: Where the Operator Meets the Equipment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 26(3), 255–257.
- Michaels, C. (2003). Affordances: Four Points of Debate. *Ecological Psychology*, 15(2), 135–148.
- Napoletano, N. (2013). The Eyes Have It: Simulated Sound Visualization for Testing. In *The Interservice/Industry Training, Simulation and Education Conference (IITSEC)* (p. 9). Orlando: Interservice/Industry Training, Simulation and Education Conference.
- Neal, G. (1982). Overview of Training Effectiveness Analysis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 26(3), 244–248.
- Ratwani, K., Orvis, K., & Knerr, B. (2010). *Game-Based Training Effectiveness Evaluation in an Operational Setting*. Arlington.
- Şahin, E. (2008). Formalization for Robotics The Concept of Affordances, (December 2007), 1–28.
- Salas, E., Rosen, M., Held, J., & Weissmuller, J. (2008). Performance Measurement in Simulation-Based Training: A Review and Best Practices. *Simulation & Gaming*, 40(3), 328–376.
- Simpson, H. (1995). *Cost-Effectiveness Analysis of Training in the Department of Defense*. Monterey.
- Sticha, P., Campbell, R., & Knerr, M. (2002). *Individual and Collective Training in Live, Virtual and Constructive Environments. Training Concepts for Virtual Environments*. Alexandria.
- Stoffregen, T. (2003). Affordances as Properties of the Animal-Environment System. *Ecological Psychology*, 15(2), 115–134.
- Tufano, D., & Evans, R. (1982). *The Prediction of Training Device Effectiveness : A Review of Army Models*. Alexandria.
- Turvey, M. (1992). Affordances and prospective control: An outline of the ontology. *Ecological Psychology*, 4(3), 173–187.
- Under Secretary of Defense. (2009). DoDI 5000-61: Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A) Instruction. Washington, DC: Author.