

# Alternate Rubric for Performance Assessment of Infantry Soldier Skills Training

Douglas Maxwell<sup>1</sup>(✉), Jonathan Stevens<sup>2</sup>, and Crystal Maraj<sup>2</sup>

<sup>1</sup> U.S. Army Research Laboratory, Orlando, FL, USA  
douglas.maxwell13.civ@mail.mil

<sup>2</sup> University of Central Florida, Orlando, FL, USA  
jonathan.stevens@knights.ucf.edu, cmaraj@ist.ucf.edu

**Abstract.** Gauging the impact of simulation-based training (SBT) technology has been straightforward in the past when applied to domains such as pilot training and ground vehicle operator training. In the dismounted infantry soldier skills domain, the low hanging fruit for effective use of (SBT) are weapons and equipment operations training. However, the complexities of the operational environment are often too difficult to replicate in current virtual environments to represent an accurate or effective training for the skills requiring identification of enemy activity or reacting to enemy contact. This paper discusses the need for an alternate method of performance assessment when comparing traditional training means to SBT.

**Keywords:** Simulation based training · Infantry soldier training · Rubric · Return on investment

## 1 Introduction

In early 2013, a research team composed of personnel from the U.S. Army Research Laboratory and the University of Central Florida performed a data collection activity with the Florida Army National Guard's 2/124<sup>th</sup> Regiment Apache Company. The data collection activity focused on individual and team performance of soldiers performing room clearing exercises. In this experiment, a company of soldiers were divided into two groups, where the control group only was provided training using a traditional classroom method and the experimental group was provided training that included a game-based simulator. Subject matter experts (SMEs) were tasked with assigning a pass/fail rating to each soldier and fire team. The research team did not wish to interfere with the method of assessment used by the SMEs, as the guidance for this is given in training support packages provided by the U.S. Army Training and Doctrine Command. Analysis of this categorical data proved to be problematic and yielded results that could only indicate differences in soldier performance, but not by how much. In essence, it was only possible to determine if the virtual training treatment had an effect.

Since mid-2014, the research team has been conducting studies using large numbers of soldiers to determine how effective virtual training methods are in comparison to traditional training methods for dismounted infantry soldier skills in the Warrior Leader Course (WLC) at the Florida National Guard's 211<sup>th</sup> Regional Training

Institute. The ARL/UCF team worked with the WLC course managers to incorporate lessons learned from the previous experiment to create a new assessment methodology for collecting more meaningful data. A new rubric was jointly developed and used during data collection activities. This paper discusses the rationale for incorporating new ideas that enhance the development of the performance rubric. The rubric allows for a more in-depth understanding of the required tasks. Instead of using the traditional assessment “GO or NO-GO” ratings for the task evaluation, the revised rubric incorporates each major task and list all related subtask activities. Modifying the rubric not only allows for clarification of each major task but also assess whether the subtasks were successfully accomplished. In addition to discussing the performance rubric, this paper also integrates sample performance data collected from 20 squads of soldiers and demonstrates the value of the revised rubric to the data analysis.

## 2 Background

The United States Army has invested significant funding dedicated to the use of virtual environments (VEs) for training infantry soldier skills. There is a pervasive attitude in the acquisition community that a simulation-based training (SBT) system’s graphics quality are the strongest indicators of utility and training quality. Very little data exists to quantify the return on investment (ROI) provided by these training systems (Bell et al. 2008). There is also a lack of formal methodologies for the identification of where in the training cycle these technologies belong as well as which training tasks they should be applied (Kincaid et al. 2003; Salas et al. 2003). The United States Government Accountability Office issued a report in August of 2013 which calls for better assessment of performance and accounting of costs to accurately assess SBT systems throughout the United States Army and Marine Corps (Pickup 2013). As a result of limited empirical data supporting training effectiveness using VEs (Haque and Srinivasan 2006), there is little guidance for the program manager’s to follow in the decision-making process. This leaves the requirements generation team and the acquisition process to attempt to simulate the training provided by traditional means using VEs. There is too much leeway in the interpretation of this replication and lack of empirically driven data to make informed decisions.

To further complicate matters, the lack of formal requirements and performance measurement methodologies has led to a fracturing of the training space within the United States military that utilizes game-based virtual environments (GBVE). Although, there is a GBVE training system listed as the program of record called Virtual Battlespace, it is limited for specialized training needs of some organizations. Pockets of innovation and product development in recent years have resulted in numerous training systems specializing in different utilization such as education (McLennan 2012) and military applications (Buede et al. 2013).

The United States Army requires a mechanism for properly assessing the performance of infantry soldiers who have been trained using virtual simulations in order to establish statistically significant differences (if any) for comparisons to traditional training methods. This research initiative was conducted through Cooperative Agreements (CA) #W911NF-14-0012 and #W911NF-15-0004 between the United States

Army Research Laboratory and the University of Central Florida. These CAs were created to facilitate the investigation of training effectiveness of operationally relevant tasks in a VE as compared to traditional classroom and live training. The desired outcome of this work is to establish a methodology for quantitatively defining the training effectiveness differences between traditional and virtual methods, and acquiring data through field experimentation to apply the methodology.

A literature review has revealed a lack of knowledge surrounding the efficacy of the practical application of virtual world technology for infantry soldier training, specifically ground combat skills training such as room clearing and reaction to contact (Lackey et al. 2014). Due to the current subjective nature of gauging training effectiveness of VEs, it is difficult to calculate a ROI. Lastly, it is difficult to determine comparisons of knowledge transfer between traditional and virtual training activities for ground combat skills.

Whether it is labeled *virtual world technology*, *GBVEs*, or *VEs*, the technology is becoming ubiquitous in the lives of infantry soldiers. However, it is unclear as to where in the ground combat skills training cycle this technology is applied most effectively. The literature is terse in identifying the appropriate tasks the technology is most suitable for training. Further, the assessment methodology is not standardized across the combat skills training cycle and is often assessed through subjective means.

### 3 Infantry Soldiers Skills Assessment

Minimal empirical evidence exists regarding the effectiveness of game-based and virtual world SBT (Whitney et al. 2014; Sotomayor and Proctor 2009), especially at the collective echelon. While some virtual training has been empirically proven to be effective in the transfer of skills to the live environment (Blow 2012; Hays et al. 1992), this has been primarily demonstrated for platform-centric training, such as aviation and vehicle-type training. However, platform-based training is restricted to low-density specialties in the United States Army; the vast majority of soldiers do not require this type of training and are not tethered to a platform. In contrast, all soldiers are required to be proficient in basic infantry skills, yet minimal SBT capabilities exist to support this training need and are rarely examined for efficacy. Therefore, this study's primary objective was to examine the training efficacy of SBT for infantry skills.

Training effectiveness evaluations are generally subjective in nature, making it difficult to ascertain whether or not training technology, methods and/or approaches are effective. The literature indicates that some of the primary challenges to effective, collective training in simulation are the lack of clear performance measures (Seibert et al. 2011) as well as a lack of comprehension of the simulation's capabilities by the unit trainers (Seibert et al. 2012). Thus, technology is only part of the analysis; the selection of the proper instructional strategy is equally as critical to whether or not training is effective (Salas et al. 1999).

Current approaches to measuring training effectiveness remain primarily subjective in nature (Wong et al. 2012; Sotomayor and Proctor 2009; Beal and Christ 2004; Kunche et al. 2011), employing techniques such as questionnaires, knowledge review and evaluation of training by trainees or SME observers. These means of assessment

offer insight into trainees and trainers' perceptions of training environments, but reporting might be influenced by factors other than those directly attributed to the training itself. Therefore, a secondary goal of this study was the creation of a rubric that minimizes subjectivity in performance assessment, while not increasing raters' overhead, in order to determine whether game-based simulation and virtual world simulation-based training is truly effective or not. Objective evaluation of performance is critical in order for the Army to design, create and implement the next generation of simulation-based trainers for infantry-centric skills training.

### 3.1 Pilot: Assessment of Room Clearing Task Performance

The ARL research team worked with the 2/124<sup>th</sup> Florida Army National Guard to design a training event that coincided with data collection activities. The data collection event represent the presentation of a single ground training task to the unit. The training condition chosen for this study is a room clearing task that requires a fire team composed of four soldiers to enter and search a room. The participants are assessed both at the individual performance level as well as group performance level.

For this data collection event, two training conditions were selected for the soldiers. The first condition represented the control group and was composed of traditional classroom and slide presentation of the procedures described by FM 3-21.8, Field Manual for the Infantry Rifle Platoon and Squad (U.S. Army Training and Doctrine Command 2007) for room clearing (Fig. 1). The experimental condition comprised of training materials presented to the soldiers using a prototype virtual training simulator,



Fig. 1. 2/124th classroom training site

called the Military Open Simulator Enterprise Strategy (MOSES) (Ortiz and Maxwell 2016). MOSES was used to provide a virtual training arena utilizing practice task scenarios (Fig. 2) (Maraj et al. 2015).

Room clearing exercises represent one of the most common tasks performed by an infantry soldier, and is considered to be of the most dangerous tasks to complete. Although this is a collective task, each of the individual positions in the task is assessed independently. This allows for both an individual performance assessment and a collective team assessment.

On the day of the experiment, 64 soldiers were divided into two groups, 32 soldiers each and placed into groups of four to compose 8 fire teams. All of the soldiers were assembled and provided with a briefing to explain the intent of the experimentation. Each soldier signed a consent form agreeing to participate in the study. Due to the nature of the experimental design and utilizing a targeted population, the ARL/UCF team required two formal review processes. One process occurred through the UCF Institutional Review Board (IRB) and another through ARL IRB before data collection efforts began for the two groups.

One group received virtual training while the second group received the traditional training. After training, all soldiers were asked to perform a live room clearing exercise



Fig. 2. 2/124th virtual training site

during which time the SMEs assessed their performance according to the rubric. Although the task of clearing a room is performed as a collective effort, each position in the team is unique and can be assessed individually. The assessment is provided using a “GO or NO-GO” rating, which indicated whether the soldier completed their task to the SME’s satisfaction.

The SMEs rated the performance of 64 individuals and 16 fire teams by following a 4 step rubric (Fig. 3). Table 1 shows the tasks and assessments for collective performance. Step one of the rubric (entry phase) was to assess the speed of entry, removal of self from the entry area, follow the path of least resistance and flow of movement. Step two (eliminate threat phase) was to maintain correct sector of fire throughout the flow. Step three (position of dominance) was to assess the soldier’s ability to move to the correct position of dominance for their position in the entry team and for the team leader to announce “CLEAR.” Step four (Consolidation and Reorganization) is to assess the team’s ability to report ammunition, casualty, and equipment status (ACE report).

The data collected from this event provided enough information to determine the performance effect of different training conditions of individual soldiers. In this case, the independent variable is the training condition and the performance assessment is the

**Table 1.** Rubric for collective performance assessment for 2/124<sup>th</sup> FLANG Leesburg trial

Task	Assessment
Step 1. Entry phase	Go/No-Go
Step 2. Eliminate threat phase	Go/No-Go
Step 3. Position of dominance phase	Go/No-Go
Step 4. Consolidation and reorganization	Go/No-Go



**Fig. 3.** 2/124th live assessment activity

dependent variable. The use of a “GO/NO-GO” performance metric limited the data analysis to simply determining dependence of the variables to each other. This categorical data lent itself to Chi-Square analysis and could indicate whether differences between the two training conditions were significant. However, this data could not be used to determine by how much the performance differences between the two training conditions. The ROI of the virtual treatment could not be established using this method.

### 3.2 Warrior Leader Course: Assessment of Dismounted Infantry Soldier Skills

The Florida Army National Guard’s 211<sup>th</sup> Regional Training Institute, located at Camp Blanding, incorporates the Warrior Leader Course (WLC) as part of its training curriculum. This leadership course is designed to teach squad leadership skills to the infantry soldier (Association of the United States Army 2010), specifically the squad leader position.

The course managers, or Small Group Leaders (SGLs), worked closely with the ARL/UCF research team to examine the WLC to determine how to create a comparison study similar to the one described in Sect. 3.1. An examination of the course revealed that it would be possible to use a between-treatments experimental design to compare a traditional training treatment to a virtual training treatment.

Table 2 shows the original evaluation rubric used in the course for team performance evaluations. In order to gather the data from the WLC, it was necessary to make adjustments to the rubric so that a more meaningful comparison could be made between the control and virtual training treatment. As with the room clearing tasks from the 2/124<sup>th</sup>, the WLC training also relied on a “Go/No-Go” performance evaluation metric.

**Table 2.** Original rubric for collective performance assessment for 211<sup>th</sup> FLANG RTI pilot

Task	Assessment
1. React to indirect fire while dismounted	Go/No-Go
2. React to improvised explosive device	Go/No-Go
3. React to near ambush	Go/No-Go
4. React to far ambush	Go/No-Go

This research focus seeks to determine applicability of specific infantry soldier skills against different training treatments. The current rubric indicates differences between the training treatments, but a new rubric is required to provide a comparison of how much one treatment differed from another. Each major training task was divided into subtasks and the assessment performed utilized a four point Likert scale (Garland 1991). A four point Likert scale provides an opportunity to make a choice by eliminating the midpoint responses. This research expands the rating categories from two (i.e., GO/NO-GO) to four (i.e., needs improvement, adequate, successful, and excels) which enables the research team, as well as the course cadre, to gain greater insight into whether or not the preceding training condition had an effect on trainee performance.

Coded categorical data can be treated as numerical and lends itself to deeper analysis if the optimal number of categories are employed. For this study, four categories of rated performance were created through a questionnaire and used in order to not overload cadre rating requirements (i.e. performing the actual evaluations); while providing the research team with quantifiable data for analysis. Further, subjectivity in evaluation was reduced by decomposing the training tasks to the subtask level, thereby allowing the cadre to increase their objectivity ratings of the performance evaluation at each atomic step. Table 3 shows the adjusted rubric the ARL/UCF team provided to the SGLs for use in their final squad performance evaluations.

**Table 3.** Revised Rubric for Task 1

Task	Assessment			
	Needs improvement	Adequate	Successful	Excels
1. React to indirect fire while dismounted				
1.1 Shout “incoming” in a loud, recognizable voice				
1.2 React to the Instruction of the leader and look for guidance				
1.3 Seek nearest cover				
1.4 Assess situation				
1.5 Report situation to leader				
1.6 Continue mission				

The period of instruction (POI) for this course is 20 days. For days 1–17, the soldiers receive the same classroom-based training including PowerPoint slides and SME instruction. Typically, day 18 is reserved for practical exercises with a four hour block allocated for scenario-based training. The practical exercises consists of “walk-throughs,” where the four major tasks are posed to the soldiers and they are given the opportunity to practice responding to the task. The 211<sup>th</sup> RTI uses the United States Army’s Virtual Battlespace 3 (VBS3) simulation platform for training. On days 19 and 20, a formal assessment of the squad’s performance is evaluated during an on-site situational training exercise (STX).

For this experiment, an adjustment was made to the POI such that the class was separated into two groups and provided with different walkthrough training treatments. This control group represented a traditional method of providing practical instruction. The practical instruction comprised of sending the control group into the near-by wooded areas where an instructor provides guided instruction during the practical exercises (Fig. 4.). Alternatively, the experimental group received the training treatment in a computer lab using the VBS3 suite (Fig. 5.) The VBS3 scenarios were developed by onsite contractors who replicated the STX lanes the soldiers would encounter the following day during their performance evaluations.





**Fig. 4.** Warrior leader course control group in traditional “walkthrough” treatment

On day 18 of the POI, all soldiers enrolled in the WLC were assembled and provided a brief describing the experiment. The soldiers were given the opportunity to ask questions pertaining to the experiment before signing consent forms. The UCF/ARL IRB reviewed the consent form to ensure the study had minimal risk to the soldiers. After the soldiers signed the consent forms, they completed a series of pre-experimental questionnaires.



**Fig. 5.** Warrior leader class using virtual battlespaces 3 (VBS3)

Following the walkthroughs on day 18, the two treatment groups assembled at the STX lanes (on day 19) to participate in the live simulation that evaluated the performance of the training tasks (Fig. 6.). The SGLs assessed the squads (or groups) according to a new rubric in real-time. This evaluation process was repeated for nine WLC class rotations, starting in April 2015 and ending in December of 2015. A total of 23 squads were evaluated and over 250 soldiers provided questionnaire input. Twenty squads followed the experimental protocols explicitly, yielding data for 10 squads exposed to the control treatment and 10 squads with the virtual treatment. By applying the new rubric, a large number of data points were generated allowing the team to apply parametric statistics to the survey data and calculate analysis of variables on the squad performance data. Implementation of the new rubric was instrumental for gathering expanded data points to calculate comparisons between the two treatment groups.



**Fig. 6.** Live performance assessment in the situation training exercise (STX) lanes

## 4 Discussion

The purpose of this study was to initiate an examination and analysis of the training effectiveness of game-based and virtual world simulation as it applies to infantry-centric skills training. The selection of infantry-centric skills for analysis represents a novel approach, as most SBT for the United States Army is affiliated with a particular platform, such as a helicopter or tank. However, the large majority of soldiers

have no requirement to be proficient with this type of equipment; in contrast *all* soldiers must be proficient in basic infantry skills. Therefore, it is authors hope that this longitudinal study provides the Army with meaningful data to create effective, next-generation simulation-based trainers that are applicable to more than just niche specialties.

Removing the implicit subjectivity of training effectiveness evaluations is an ambitious goal, which may or may not be achieved. The creation of a rubric that reduces subjectivity in performance assessment, through the use of multiple categories on a four-point Likert scale with training tasks decomposed to the subtask level, will hopefully contribute to and improve the quality and accuracy of SBT efficacy evaluations. This is a critical first step to determine whether game-based simulation and virtual world SBT are truly effective. Future papers will expand on SBT as an effectiveness tool for training soldiers on basic infantry skills.

**Acknowledgements.** The U. S. Army Research Laboratory and the University of Central Florida research team would like to express our profound gratitude to the dedicated men and women of the 2/124<sup>th</sup> and 211<sup>th</sup> without whom this work could not have been performed. We thank you for your service to our country.

## References

- Association of the United States Army: Preparing for Warrior Leader course (WLC). Arlington, Virginia, United States of America (2010). [http://www.ausa.org/RESOURCES/NCO/TRAINING/CAREERADVANCEMENT/PREPARINGFORSCHOOL/Pages/PreparingforWarriorLeadercourse\(WLC\).aspx](http://www.ausa.org/RESOURCES/NCO/TRAINING/CAREERADVANCEMENT/PREPARINGFORSCHOOL/Pages/PreparingforWarriorLeadercourse(WLC).aspx). Accessed 1 Feb 2016
- Beal, S., Christ, R.: Training effectiveness evaluation of the full spectrum command game (No. ARI-TR-1140). Army Research Institute for the Behavioral and Social Sciences, Fort Benning, GA (2004)
- Bell, B.S., Kanar, A.M., Kozlowski, S.W.: Current issues and future directions in simulation-based training. CAHRS Working Paper Series, 492 (2008)
- Blow, C.: Flight School in the Virtual Environment. United States Army Command and General Staff College Fort Leavenworth, Kansas (2012)
- Buede, D., Maxwell, D., McCarter, B.: Filling the need for intelligent, adaptive non-player characters. In: Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (IITSEC), Orlando, FL (2013)
- Garland, R.: The mid-point on a rating scale: is it desirable. *Mark. Bull.* 2(1), 66–70 (1991)
- Haque, S., Srinivasan, S.: A meta-analysis of the training effectiveness of virtual reality surgical simulators. *IEEE Trans. Inf. Technol. Biomed.* 10, 51–58 (2006)
- Hays, R., Jacobs, J., Carolyn, P., Salas, E.: Flight simulator training effectiveness: a meta-analysis. *Mil. Psychol.* 4, 63–74 (1992)
- Kincaid, J.P., Hamilton, R., Tarr, R., Sangani, H.: *Simulation in Education and Training*, pp. 437–456. Springer, US (2003)
- Kunche, A., Puli, R.K., Guniganti, S., Puli, D.: Analysis and evaluation of training effectiveness. *Hum. Resour. Manag. Res.* 1(1), 1–7 (2011)

- Lackey, S., Salcedo, J., Matthews, G., Maxwell, D.: Virtual world room clearing: a study in training effectiveness. In: Interservice/Industry Training, Simulation and Education Conference (IITSEC), Orlando, Florida (2014)
- Maraj, C., Lackey, S., Badillo-Urquiola, K., Ogreten, S., Maxwell, D.: Empirically derived recommendations for training novices using virtual worlds. In: Interservice/Industry Training, Simulation and Education Conference (IITSEC), Orlando, pp. 1–8 (2015)
- McLennan, K.: Interactivity, engagement, and increased learning outcomes in 3D virtual world educational simulations. In: Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications, p. 843. AACE, Chesapeake, VA (2012)
- Ortiz, E., Maxwell, D.: Military Open Simulator Enterprise Strategy (MOSES) (2016). Retrieved from Military Metaverse: <http://militarymetaverse.org/>
- Pickup, S.: ARMY AND MARINE Better Performance and Cost Data Needed to More Fully Assess Efforts. United States Government Accountability Office, Washington, D.C. (2013). <http://www.gao.gov/assets/660/657115.pdf>
- Salas, E., Fowlkes, J., Stout, R., Milanovich, D., Prince, C.: Does CRM training improve teamwork skills in the cockpit?: two evaluation studies. *Hum. Fact. J. Hum. Fact. Ergon. Soc.* **41**, 326–343 (1999)
- Salas, E., Milham, L.M., Bowers, C.A.: Training evaluation in the military: misconception, opportunities, and challenges. *Mil. Psychol.* **15**(1), 3–16 (2003)
- Seibert, M., Diedrich, F., Ayers, J., Dean, C., Zeidman, T., Bink, M., Stewart, J.: Addressing Army Aviation Collective Training Challenges with Simulators and Simulations Capabilities. Aptima, Woburn (2012)
- Seibert, M., Diedrich, F., Stewart, J., Bink, M., Zeidman, T.: Developing Performance Measures for Army Aviation Collective Training. Army Research Institute, Fort Benning (2011)
- Sotomayor, T., Proctor, M.: Assessing combat medic knowledge and transfer effects resulting from alternative training treatments. *J. Defense Model. Simul. Appl. Methodol. Technol.* **6**, 121–134 (2009)
- U.S. Army Training and Doctrine Command: FM 3–21.8 The Infantry Rifle Platoon and Squad. Washington, D.C. (2007). [http://armypubs.army.mil/doctrine/DR\\_pubs/dr\\_a/pdf/fm3\\_21x8.pdf](http://armypubs.army.mil/doctrine/DR_pubs/dr_a/pdf/fm3_21x8.pdf)
- Whitney, S., Tempby, P., Stephens, A.: A review of the effectiveness of game-based training for dismounted soldiers. *J. Defense Model. Simul.* 319–328 (2014)
- Wong, J., Nguyen A, Ogren, L.: Serious Game and Virtual World Training: Instrumentation and Assessment. Naval Undersea Warfare Center, Newport, RI (2012)