

Leveraging Interoperable Data to Improve Training Effectiveness Using the Experience API (XAPI)

Jennifer Murphy^{1(✉)}, Francis Hannigan¹, Michael Hruska², Ashley Medford²,
and Gabriel Diaz²

¹ Quantum Improvements Consulting, Orlando, FL, USA
{jennifer.murphy, fhannigan}@quantumimprovements.net

² Problem Solutions, Johnstown, PA, USA
{mike, amedford, gdiaz}@problemsolutions.net

Abstract. This paper discusses our research efforts aimed at improving the training effectiveness and efficiency of the U.S. Army's gunnery and rifle marksmanship curriculum. Soldier assessments, typically in the form of final qualifications scores, are insufficient to conduct experimental comparisons required to evaluate training effectiveness. More importantly, this level of performance assessment does not speak to the root cause of the errors soldiers make during training. Using the Experience API (xAPI) specification, learning experiences are represented in terms of activity statements and can be used to track learning that happens both inside and outside of the classroom, enabling the development of robust, persistent student models. Importantly, xAPI data are interoperable across training systems, allowing a student's performance to be tracked across multiple platforms. Our research demonstrates the utility of xAPI to improve the effectiveness of Army simulation-based training through improved performance assessment capabilities.

Keywords: Adaptive learning systems · Experience API · Data interoperability · Training effectiveness evaluation

1 Introduction

Throughout military history, simulation has been used to train warfighting skills. The benefits of simulation include the ability to provide a controlled, safe environment to experiment at a relatively low cost. Simulation enables warfighters to experience high fidelity replications of operational settings prior to combat, which reduces the risk to lives and equipment. Unlike live operations, in simulated environments events can be replayed, paused, and customized, providing the opportunity for tailoring learning experiences to the individual or unit. Despite its clear benefits, quantifying the benefit of simulation-based training remains a challenge. Recently, the Government Accountability Office [1] critiqued the U.S. Army and the Marine Corps for insufficiently assessing the effectiveness of their simulation-based training systems. In response to this report as well as broader budgetary limitations, there has been a renewed interest in training effectiveness assessments (TEA) by the Department of Defense, and the Army in particular.

The value of training technology is typically conceptualized in one of two ways, depending on the rationale behind the evaluation. TEA focuses on the learner, and the extent to which he or she develops and applies new skills as a result of the training. When applied to new technologies, training effectiveness is usually evaluated through comparison to existing solutions or to live training. If learning outcomes are similar or improved relative to the standard, the training is considered effective. Another approach to evaluating training systems involves determining their cost effectiveness. As one of the primary benefits of simulation is reduced training costs, the purpose of these analyses is to demonstrate the extent to which a simulator can produce comparable learning outcomes to more expensive solutions. Variables of interest often include material costs, time to train, instructor hours, transportation costs, and increased safety. While TEA is centered on the trainee and often disregards cost considerations, cost effectiveness analysis (CEA) assumes similar learning outcomes and instead focuses on training resources. However, to fully address the value of training systems, both training and cost effectiveness should be considered. In other words, these systems should be evaluated in terms of training efficiency.

The concept of evaluating training efficiency is not new, but performing such an analysis has proven prohibitively challenging. Fletcher and Chatelier [2] described the goal of combining assessments of training and cost effectiveness for military training and identified barriers to conducting such an analysis. Training outcomes are difficult to quantify in financial terms. Additionally, modeling the cost element structure associated with developing, delivering and sustaining military training is in itself very complex [3]. Despite this, the Department of Defense does evaluate the cost of implementing any technology prior to a procurement. On the other hand, TEA is rarely conducted, and if so, it is typically not conducted well. Improving the TEA process, then, would go a long way toward improving our understanding of the true value of simulation-based training.

Why is TEA conducted so rarely? Evaluating military simulation as an effective means of instruction is not a new problem, and best practices for conducting TEA have been well-documented [4, 5]. Despite ample evidence to the contrary [6], an attitude that simply replicating the operational environment to the highest fidelity possible is sufficient to guarantee effective training persists in the military community. While the U.S. Defense laboratories are funded to conduct training research, acquiring troop support and equipment for data collection is consistently difficult. A more challenging issue, however, is a lack of objective, valid measures of warfighter performance during training events. Typically, performance is assessed either subjectively by an instructor or through a single qualification score at the end of a training event. These data are typically insufficient to conduct the experimental comparisons required to evaluate training effectiveness. More importantly, this level of performance assessment does not speak to the root cause of errors warfighters may make during training.

What is needed is a methodology to develop objective measures and metrics of warfighter performance within simulator systems. These metrics could be generated using the data simulators currently use to drive the training curriculum they provide. Currently, these metrics are not calculated, largely because training systems developers

are not required to do so. In addition, there is no guidance for these developers with regard to how to identify the appropriate metrics for use in these systems.

In this paper, we describe ongoing research efforts aimed at improving Army training effectiveness through the use of interoperable performance data. This work focuses on two critical warfighting domains: crew gunnery and basic rifle marksmanship. Using the Experience API (xAPI) as a means of standardizing performance data, we demonstrate the extent to which TEA can be improved. Importantly, our research also speaks to the larger challenge of assessing training efficiency through performance assessment.

1.1 Assessing the Effectiveness of Simulation-Based Training

Training effectiveness is usually thought of as the extent to which learners gain an understanding of a content domain as a result of a training intervention. As such, TEA has historically used assessments of learner knowledge as criteria. One enduring model for conducting these evaluations is Donald Kirkpatrick's Four Levels [7]. Using this approach, training effectiveness is evaluated based on four criteria, or levels. Level 1, "Reaction," focuses on the extent to which trainees enjoy the learning experience, which is usually assessed by a questionnaire at the end of the training. At Level 2, "Learning," effectiveness is conceptualized as a change in knowledge gained, skills acquired, or attitudes changed as a result of the intervention. Level 3, "Behavior," refers to the extent to which the training event influences subsequent actions. This is usually operationalized as an assessment of learner performance on the job by a third party, such as a supervisor or peer. Finally, the fourth level, "Results," describes the overall impact of a training intervention on the organization as a whole. In a corporate setting, Level 4 is often assessed in terms of a company's productivity or profitability. Defining the value of a training event on this level is challenging, and is rarely achieved. The benefits to an organization are not immediately evident, and determining them requires a long-term assessment strategy. Further, isolating the effects of a single intervention in the context of larger organizational shifts that naturally happen over time is difficult.

Although Kirkpatrick's Four Levels were designed to address the effectiveness of training in a civilian corporate context, this approach has been widely adopted to assess military training. Morrison and Hammon [4] and Simpson and Oser [5] advocate this framework as a basis for designing TEA for simulation-based training in particular. While this approach is certainly appropriate, its application to military training technology comes with unique challenges. A primary limitation is the feasibility of conducting Level 3 and 4 assessments. Typically, simulation is one component of a training continuum spanning introductory didactic instruction, hands-on exercises increasing in levels of complexity, and cumulating in a live exercise and qualification event. In this context, a Level 3 evaluation would involve assessing the extent to which simulator performance transfers to subsequent, higher fidelity training events. While the opportunity for this assessment exists, an individual warfighter's performance is not typically tracked across training events, and often final qualification scores are the only persistent record of the training experience. Level 4 evaluations are rare in any setting, but in a military context, "Results" translates to the effectiveness of a unit during combat operations. Opportunities to assess these events are rare, and the complexities of the

battlefield make definitive assessments of the impact of one training event nearly impossible. Examples of successful Level 4 evaluations of military training have involved air-to-air combat outcomes [8] and bombing accuracy [9]. However, the outcomes of ground operations are much more difficult to evaluate. As a result, most evaluations of simulation-based training are limited to Level 1 and 2. While impressions of training and retained knowledge are important, the true value of training is the ability to apply what is learned in an operational setting.

Ultimately, the success of achieving Level 3 and 4 evaluations is dependent upon the ability to assess the extent to which knowledge and skills gained during training transfer to higher fidelity, if not live, environments. Historically, collecting assessments of performance across a variety of training platforms has proven prohibitively difficult. However, recent developments in learning technology have supported the capture and analysis of more granular performance data. The rise of mobile technology and ubiquitous wireless data access have enabled both training and assessment anytime and anywhere. Improvements in low-cost wearable sensor technology have made unobtrusive assessments of a learner's location, physiological state, and activity level a possibility. Advances in machine learning have facilitated the interpretation of these data, and the ability to store massive amounts of data in a cost-effective way without reducing processing speed has made "big data" a reality. As a result of these recent advances, the data exist to inform real-time performance measurement in nearly any environment. A remaining challenge involves standardizing these data for use across multiple platforms. To address the need for performance data standardization, the Experience API (xAPI) was developed.

1.2 The Experience API and Data Interoperability

xAPI is a data specification developed by the Advanced Distributed Learning (ADL) Co-Lab as a means of tracking learning experiences across a wide variety of technology platforms. Although other data standards, such as High Level Architecture (HLA) and Distributed Interactive Simulation (DIS), are used in the context of training technology, xAPI is the only one designed specifically to capture and share human performance data. Using xAPI, learning experiences are represented in terms of statements in the format "Actor – Verb – Object" (e.g., "Chad read *Twilight*"), with the option of including additional contextual information and results. Performance data in xAPI format are stored in a Learning Record Store (LRS), which serves as a mechanism for multiple training and analysis systems to store and access these statements through a centralized point.

The primary benefit of using the xAPI specification is the ability it affords to store human performance data from multiple sources in a single, intuitive format. Because of its flexibility, xAPI enables the capture of a wide variety of learning experiences, both inside and outside the classroom. This data interoperability allows a much broader assessment capability than was previously possible. In terms of TEA, there are a number of implications. xAPI supports the development of robust, persistent learner models in training systems. As a result, tracking performance across multiple training events is possible. Importantly, all types of experiences can be represented in xAPI format, including events that occur completely outside of a training environment. Whereas Level

3 and Level 4 evaluations were previously limited in terms of reliable access to operational performance data, xAPI enables objective assessment of skill transfer to a higher fidelity or live scenario.

In addition to enabling more robust TEA, interoperable performance data support advanced training methodologies that have been shown to improve the efficiency of simulation-based training. Persistent models of learner performance enable the adapting of training content to the individual based on their performance in previous training events. These data allow for the predictive modeling of training outcomes, which can be used to prescribe a training curriculum based on existing knowledge, skills, and abilities.

Our research demonstrates the utility of xAPI to improve the effectiveness of Army simulation-based training through improved performance assessment capabilities. Below, we describe efforts to investigate the extent to which the effectiveness of an unstabilized crew gunnery simulator was improved by adapting training using interoperable data from an individual gunnery simulator. In addition, we address our research into the extent to which these data could be used to improve the overall efficiency of the Army training process by addressing the needs of multiple stakeholders in the marksmanship training community.

2 Adapting Training Using Interoperable Performance Data

In 2011, the Army's Training and Doctrine Command (TRADOC) published the Army Learning Concept for 2015 [10], a document outlining a vision for modernizing Army training. This new Army Learning Model (ALM) called for increasing the role of emerging technology as a way of improving the quality of soldier training while reducing costs. In particular, the ALM identified adaptive training technology as a means of efficiently tailoring learning experiences to the individual warfighter. Despite this guidance, adaptive training has still not been widely adopted by the Army. One reason for this is the expense associated with developing adaptive training systems, and research efforts such as the Army Research Laboratory's Generalized Intelligent Framework for Tutoring (GIFT) have focused on reducing these costs by improving the reusability of adaptive training content. However, a more significant barrier to the implementation of adaptive training is a lack of clear data showing the benefits of these systems. In their review of adaptive training technologies, Durlach and Ray [11] call for additional research to quantify the improvements in effectiveness expected from adaptive training.

A challenge in conducting TEA of adaptive training technologies is the need to provide a non-adaptive system as a standard for comparison. Typically, military training systems are developed to meet a specific requirement, and developing an additional adaptive or non-adaptive version for research purposes is prohibitively expensive. Under our current research effort, our team has been fortunate enough to have the ability to conduct such a comparison. Raydon Corporation's Unstabilized Gunnery Trainers (UGT) are simulators that support the training of Army gunnery crews. These trainers provide training on the individual level to gunners as well as gunnery crews in accordance with Army gunnery standards. Warfighters first learn gunnery basics, including how to maneuver the weapon, how to respond to commands, and how to quickly acquire

and destroy targets in the individual trainer. In this simulator, the gunner interacts with a virtual crew and engages targets in a number of scenarios under a variety of conditions (e.g. day/night, stationary/moving) per the relevant Army training manuals (TC 3-20.31). This trainer is unique in that instead of progressing the warfighter through the entire training tables, the curriculum is adapted in real time based on the performance of the gunner. As the gunner progresses through the tables, subsequent scenarios are automatically selected based on the score the gunner receives. (The specific details of how the training is adapted are documented in Long et al. [12, 13].)

After the gunner completes individual training, he or she progresses to a crew simulator in which a live crew composed of the gunner, a commander, and driver, trains together to complete the gunnery tables required for qualification. Similar to the individual trainer, the crew is required to master target engagement in a variety of positions and conditions prior to graduating to a simulation of the live qualification exercise. The training is facilitated by an instructor, who scores the crew's performance. (Again, specifics are described in [12, 13].)

For research purposes, an experimental crew curriculum was developed that adapted the crew's course of instruction based on the performance of the gunner during individual training. Specifically, the crew's training was accelerated based on the tasks and conditions in which the gunner demonstrated proficiency. This adaptation was made possible by leveraging the xAPI specification as a means of communicating performance data across simulators. Our task was to determine the extent to which using this adaptive training curriculum would increase the efficiency of the training process. To that end, we compared the performance of a group of participants who completed the crew training with no adaptation to a group who completed the experimental curriculum, which was tailored to their previous performance in the individual trainer. Our participants were a sample of Reserve Officers' Training Corps (ROTC) cadets from a local university. Performance was defined as the crew's final qualification score and the time required to complete the training.

The results of this experiment showed that while both groups performed exceptionally well, the adaptive group completed the training in nearly 40 % less time than the non-adaptive control. These findings speak directly to the current limitations of most TEA conducted with military simulations. If our evaluation had simply focused on comparing learning outcomes as most TEA do, the finding that both experimental and control groups performed well above standard would suggest no benefit to the adaptive curriculum. However, by assessing the time required to complete the training, our findings speak more to training efficiency. Further, because the manpower and material resources needed to conduct training using these simulators are knowable quantities, the resulting cost savings to the Army could easily be calculated.

An important caveat to this point is the finding that using these simulators, nearly all crews achieved a "distinguished" rating. While this speaks to the effectiveness of the training, it begs the question of whether the existing Army guidance on gunnery training could be further streamlined. While the goal of all Army training is to produce highly proficient soldiers, it would be worthwhile to investigate the extent to which training requirements could be reduced before a decrement in performance is noticed in order to maximize the efficiency of this training requirement.

3 Improving Training Efficiency with Interoperable Performance Data

Our research shows the utility of interoperable performance data to facilitate the evaluation of training technologies. While TEA is critical in demonstrating the value of training systems, what it does not capture is the broader context in which these technologies are used. Simulation-based training is rarely used alone, and is typically conducted as part of a curriculum involving introductory, didactic instruction followed by increasingly complex, hands-on practice. The Army refers to this process as “crawl-walk-run,” and simulation is often used as part of the “crawl” or “walk” phases of soldier training. In order to determine the true value of a training system, the extent to which it maximizes the efficiency of the entire training process should be considered. Our ongoing research efforts aim to determine a methodology and system for using human performance data to evaluate and improve the overall efficiency of the Army basic rifle marksmanship process.

Army rifle marksmanship is a skill every soldier must acquire during Basic Combat Training. Thousands of soldiers every year complete the basic rifle marksmanship curriculum, which involves classroom familiarization with marksmanship fundamentals, ballistics, and weapon care, simulation-based training on grouping and zeroing, and honing skills on live ranges. The management of this process requires extensive coordination between many groups of stakeholders, including drill sergeants, training developers, range control personnel, resource managers, and simulator operators. Each of these groups require specific data to accomplish their responsibilities. However, these data are often stove-piped in different databases, and coordination is difficult. As a result, precious training time is often wasted, drill instructors are overwhelmed, and soldiers are not receiving an optimal training experience. If access to performance data throughout the entire training process was improved, the marksmanship training process could be much more efficient.

A first step in conducting this research was the identification of the limitations of the current training process, which was carried out through a user needs analysis with representatives of the marksmanship training community at Fort Benning. Our team conducted interviews and focus groups with drill instructors, instructors from the Marksmanship Master Trainer Course and the 194th Armor Brigade, and training developers from the Individual and Systems Training Division of the Department of Training Development. Additionally, we consulted resource managers associated with marksmanship training, including managers from the Simulations Training Division, Range Control Operations personnel, and the Maneuver Center of Excellence’s Ammunitions Manager. Finally, we discussed the extent to which improved access to marksmanship data could support research goals with Research Psychologists from the Army Research Institute’s Fort Benning Research Unit. These discussions resulted in an understanding of the extent to which soldier marksmanship performance is currently being assessed, the challenges associated with delivering marksmanship training, and opportunities for improving the training process.

Our findings showed that while basic rifle marksmanship training takes place over the course of approximately two weeks, objective measures of soldier performance are

typically not captured nor maintained with the exception of a final qualification score. However, many opportunities for assessing performance over the course of training exist. If performance was assessed more frequently, the Army could realize many benefits. Tailoring training to the individual soldier would be possible, maximizing the potential of each soldier. More accurate estimates of soldier needs for simulation hours, ammunition, and range time could be produced. The effectiveness of the training curriculum could be evaluated, and experimentation with new training systems could easily be conducted.

Discussions with marksmanship instructors informed the most common and critical issues soldiers have when learning to shoot. The issues reported described various cognitive, psychomotor, and affective components of marksmanship performance. Based on these discussions, we developed prototype measures of these components for implementation. The cognitive measures include assessments of soldier knowledge and aptitude. Measures of the marksmanship fundamentals, vision, and handedness address the psychomotor components of the domain. Affective measures include perceived stress, grit, conscientiousness, and self-efficacy. Our future research will validate these measures with a sample of soldiers undergoing marksmanship training.

In order to provide access to these data across the marksmanship training community, our team designed a system for tracking trainee performance across multiple instances of Army basic rifle marksmanship training, across a variety of training technologies using the xAPI standard. The system enables (1) a historical view of trainee or unit proficiency, (2) a live view of performance, and (3) macro and micro adaptation.

4 Conclusions

Our research speaks to the extent to which human performance data can be used to improve the TEA process. At a broader lever, our aim is to not only demonstrate the extent to which training systems provide learning experiences to trainees, but to address the efficiency of the training process. To do this, the extent to which training systems provide benefit above and beyond existing solutions should be evaluated with an appreciation for the costs associated with delivering training. As our gunnery simulator research shows, systems can provide comparable learning outcomes with very different costs associated with implementing them. In addition, our research suggests performance data can be used to not only assess the efficiency of the training process, but to improve it through addressing the needs of the broader training community.

References

1. U.S. Government Accountability Office. Army and Marine Corps Training: Better Performance and Cost Data Needed to More Fully Assess Simulation-Based Efforts. U.S. Government Accountability Office, Washington (2013)
2. Fletcher, J.D., Chatelier, P.R.: An overview of military training. Institute of Defense Analysis Document D-2514. Institute of Defense Analysis, Alexandria, VA (2000)
3. Knapp, M. I., Orlansky, J.: A cost element structure for defense training. Institute of Defense Analysis Document P-1709. Institute of Defense Analysis, Alexandria, VA (1983)

4. Morrison, J. E., Hammon, C.: On measuring the effectiveness of large-scale training simulations. Institute of Defense Analysis Document P-3570. Institute of Defense Analysis, Alexandria, VA (2000)
5. Simpson, H., Oser, R.L.: Evaluating large-scale training simulations. *Mil. Psychol.* **15**(1), 25–40 (2003)
6. Salas, E., Bowers, C.A., Rhodenizer, L.: It's not how much you have but how you use it: toward a rational use of simulation to support aviation training. *Int. J. Aviat. Psychol.* **8**(3), 197–208 (1998)
7. Kirkpatrick, D.L.: Techniques for evaluating training programs. *Training Dev. J.* **33**(6), 78–92 (1979)
8. Hammond, C., Horowitz, S.A.: Flying hours and aircrew performance (IDA Paper P-2347). Institute for Defense Analyses, Alexandria, VA (1990)
9. Cavaluzzo, L.C.: OPTEMPO and training effectiveness (Professional Paper 427). Center for Naval Analyses, Alexandria, VA (1984)
10. U.S. Department of the Army. The U.S. Army Learning Concept for 2015. TRADOC Pam 525-8-2. Department of the Army, Washington, DC (2011)
11. Durlach, P.J., Ray, J.M.: Designing adaptive instructional environments: insights from empirical evidence (Technical report 1297) (2011). <http://www.adlnet.gov/wp-content/uploads/2011/11/TR-1297.pdf>
12. Long, R., Murphy, J., Newton, C., Hruska, M., Medford, A., Kilcullen, T., Harvey Jr., R.: Adapting gunnery training using the experience API. In: Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) (2015)
13. Long, R., Murphy, J., Ruprecht, C., Medford, A., Diaz, G., Kilcullen, T., Harvey Jr., R.: Evaluating adaptive training for teams using the experience API. In: Proceedings of the MODSIM World Conference (2016)