Support for Training Effectiveness Assessment and Data Interoperability

Gregory A. Goodwin
U.S. Army Research Laboratory
Orlando, FL
gregory.a.goodwin6.civ @mail.mil

Jennifer S. Murphy
Quantum Improvements Consulting, LLC
Orlando, FL
jennifer.murphy @quantumimprovements.net

Ashley L. Medford
Problem Solutions, LLC
Johnstown, PA
amedford @problemsolutions.net

ABSTRACT

The U.S. military faces increasing pressure to train and maintain readiness with fewer resources. New training technologies offer the promise of lower-cost, effective training, but assessing the effectiveness of those technologies is expensive. To reduce this cost, technologies are needed to automate performance assessment. We present a research project investigating the design of a system to automatically assess the learner performance during Army rifle marksmanship training. Recent efforts from the Army Research Laboratory (ARL) on interoperable performance assessment (IPA) for individuals and teams have made advances on defining and persisting human performance data. The effort has concentrated on leveraging the work on the Experience API (xAPI) of the Advanced Distributed Learning (ADL) Initiative in an effort to produce human performance data that has intersystem data value. The intent of the effort is to enable the collection of data across the training continuum that 1) enables a historical view of proficiency, 2) demonstrates a live view of performance, 3) enables macro and micro adaptation, and 4) collects data for trends analysis for efficiency and effectiveness studies. In this paper, we present the first phase of work in this effort which included a user needs analysis of trainers, training developers, resource managers, and researchers to determine the ways in which they would need to access and the tools they would need to use marksmanship performance data. From this, we developed a library of measures and an xAPI registry to encode this library of measures. This registry also acts as a stand-alone reference for training simulator vendors to easily create compatible, marksmanship-specific, xAPI measures. Finally, we developed an architecture for IPA data capture and usage that maps key use cases for data capture, identity management, storage, secure access, and data analysis. The use cases were employed to create the system architecture designs.

ABOUT THE AUTHORS

Dr. Gregory A. Goodwin is a Senior Research Scientist at the Army Research Laboratory – Human Research and Engineering Directorate in the Advanced Training and Simulation Division in Orlando, FL. For the last decade, he has worked for the Army researching ways to improve training methods and technologies. He holds a Ph.D. in Psychology from Binghamton University and an M.A. in Psychology from Wake Forest University.

Dr. Jennifer Murphy is the founder and Chief Executive Officer of QIC. She is passionate about using her expertise in research psychology to improve performance, especially for those who serve our country. She has over 15 years of research experience, with a focus on the use of emerging technologies for training and assessment. She holds a Ph.D. and M.S. in Cognitive and Experimental Psychology from the University of Georgia. She is an active member of the Orlando simulation training community, and is a member of the National Simulation Center and National Defense Industrial Association. She supports I/ITSEC on the Emerging Concepts and Innovative Technologies subcommittee.

Ashley L. Medford, PMP is the Director of Government Services at Problem Solutions. She is an experienced program manager, specializing in the development and use of immersive technologies for Commercial and Defense education, training, analysis, and collaboration. Ms. Medford has led various immersive learning technology programs and projects for major Government and Defense clients, including the Air Force Research Laboratory, U.S. Naval Air Systems Command, Army Research Laboratory, NASA, Office of Naval Research, U.S. Special Operations Command, and U.S. Central Command. She is also a member of the e-Learning Guild, the American Society of Training and Development (ASTD), the Project Management Institute (PMI), and the National Defense and Industrial Association (NDIA).

Support for Training Effectiveness Assessment and Data Interoperability

Gregory A. Goodwin
U.S. Army Research Laboratory
Orlando, FL
gregory.a.goodwin6.civ @mail.mil

Jennifer S. Murphy
Quantum Improvements Consulting, LLC
Orlando, FL
jennifer.murphy @quantumimprovements.net

Ashley L. Medford
Problem Solutions, LLC
Johnstown, PA
amedford @problemsolutions.net

INTRODUCTION

The use of simulators and computer technologies for training purposes by the Armed services continues to grow substantially, especially for the training of ground systems and personnel (U.S. Government Accounting Office, 2013). Rapid improvement in hardware, software, and network capabilities coupled with decreasing costs for these technologies have contributed to this trend. Investment in these technologies is also fueled by the promise that simulation-based training can save money without sacrificing readiness. However, the Government Accounting Office (GAO) has repeatedly found that the services do not have adequate measures of either the total cost or the overall effectiveness of using training technologies (GAO, 1991, 1993, 2006, 2012, 2013). Lacking such data hampers sound acquisition decision making and limits the services' ability to effectively manage their simulation training capabilities. One must ask, why do the services fail to collect information that is so clearly important to guiding both the acquisition and employment of these training systems?

One reason is that collecting training data is challenging. Cohn et al. (2009) identifies three major challenges to collecting training data for the purpose of evaluating training systems. First, valid measures of performance are time-consuming and expensive to collect. This is why most studies focus on trainee reactions or opinions about the training. Second, most training effectiveness evaluations (TEA's) are done after a new system is fielded at which point it is too late to make changes to the system design if it is found to be ineffective. Finally, many logistical constraints impede sound analysis. Because unit training takes precedence, researchers rarely have experimental control over conditions of training, participant numbers and selection, and measurement options.

We propose that many of these challenges could be overcome if the services implemented systems to automatically collect and analyze human performance data during training events. Furthermore, linking performance data from multiple training events as well as from operational and/or garrison environments, would make it possible to understand not only the immediate impact of a training event, but also the organizational and operational impact.

A framework for understanding these relationships is described by Alvarez et al. (2004). In that paper, the authors analyzed research studies between 1993 and 2002 to produce an "Integrated Model of Training Evaluation and Effectiveness" (IMTEE) presented in Figure 1. Their main objective was to identify the individual, training and organizational variables that are empirically related to the targets of evaluation in training effectiveness studies. As can be seen, the causal relationships are mapped to Kirkpatrick's well known rubric for evaluating training programs (Kirkpatrick, 1979). His rubric examines reactions of learners, learning, behavioral change, and organizational impact. Arrows in Figure 1 denote significant relationships among variable sets. Because the authors found no published data examining organizational outcomes, there are no verified causal links to this outcome, though in theory they should exist.

In this paper, we present work to date on an effort to develop an integrated performance measurement system to support this IMTEE model in a marksmanship use case. This effort is called Support for Training Effectiveness Analysis with Data Interoperability (STEADI).

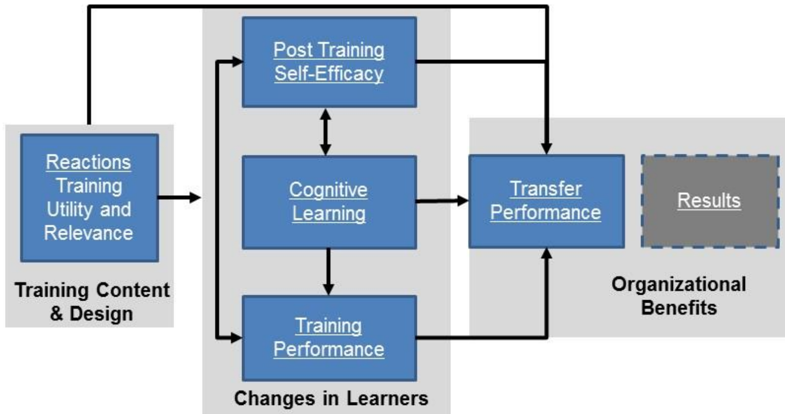

Figure 1. Model of Training Evaluation and Effectiveness (adapted from Alvarez et al., 2004)

Measurement Standards

To develop STEADI, a standard for recording human performance data was needed. Fortunately, there are industry standards for collecting human performance data. One of these is called the Human Performance Markup Language (HPML), an eXtensible Markup Language (XML) activity structure. This schema was designed to capture and assess performance across distributed simulations by a language that identifies critical fields and stores them within an XML activity structure. The major goal in designing HPML was to bridge the gap between simulation generated data that reflects human performance and assessments that are relevant to training professionals, instructors, operators, and researchers (Stacy, Ayers, Freeman, & Haimson, 2006). The current effort used HPML as a framework for describing current performance data that is being collected by various Army simulators, as well as to understand what type of system-based and observer-based data is being tracked across environments.

Another, more recent, standard is one developed by the Advanced Distributed Learning (ADL) Initiative called the Experience API (xAPI) (Advanced Distributed Learning, 2014). The xAPI is a non-proprietary specification for tracking and storing experiences across learning platforms (e.g., simulators, virtual worlds, web content, mobile devices, games, and observer-based measures). The learner’s activity stream (a series of xAPI statements) is stored in a JavaScript Object Notation (JSON) database called a Learning Record Store (LRS). Each statement includes the actor, verb, object, and optional information about results, context, etc. Within STEADI, the xAPI provides a means to encode performance data across different training systems (i.e., how to collect data), while the HPML offers a general structure to the specific data we collect (i.e., what data we collect).

Goals of the Present Study

The primary goals of STEADI were to conduct the foundational research to guide development of effective measures and metrics for TEA in the marksmanship domain, to represent those measures and metrics in xAPI format, to design and develop an architecture to support the use of TEA measures for key user groups, and to support the use of performance data for assessing other organizational benefits.

Specifically, the phases of this effort include:

1. Measure identification
2. Instantiation of measures in xAPI format
3. Design of an architecture to manage the data to support key user groups
4. Validation of measures and architecture
5. Analysis and reporting on data collected for the evaluation effort

This paper reports on work to date performed on the first three of these phases of work.

PHASE 1: MEASURE IDENTIFICATION

Methods

The identification of measures for the marksmanship use case entailed two related tasks. First, we conducted a literature review of research on measures of marksmanship training. Second, we conducted a user needs analysis with several communities of users of the STEADI architecture being developed.

Literature Review. To familiarize ourselves with the issues involved in Army basic rifle marksmanship, and potential measures of performance, we examined journal articles, technical reports, and conference proceedings from databases including: ESBCO (Academic Search Complete, Military and Government Collection, PsycArticles), Defense Technical Information Center (DTIC), and Google Scholar. In addition, we reviewed proceedings from conferences including the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) and Human Factors and Engineering Symposium (HFES), among others. This literature review focused on three areas: (1) previous assessments of training effectiveness of marksmanship training technology; (2) the knowledge, skills, and abilities (KSAs) relevant to marksmanship performance, and; (3) the ways shooter skills can be directly assessed.

Importantly, we identified a model of skill development to serve as the basis for measuring marksmanship skill going forward. Using Bloom's (1956) framework for learning domains which proposes that complex skills have cognitive, psychomotor, and affective components, and Fitts and Posner's (1967) skill development framework, we constructed a model for marksmanship skill progression. This model was used to identify prototype measures of marksmanship performance that would be measured and used in the STEADI system to assess overall marksmanship skill.

Focus Groups. Prospective user groups for the STEADI system were identified. These included instructional developers, trainers, training managers, and resource managers. Semi-structured interview and focus group protocols were developed to guide the discussions. The questions centered around designing, delivering and managing marksmanship training data in ways that were useful to these user groups. For individuals directly involved in marksmanship training, including instructors, the questions dealt with how performance was diagnosed, the typical errors shooters make, and remediation strategies. Interview protocols with resource management personnel focused on the processes underlying supplying trainers with the necessary ammunition, range and simulation time. Training developer interviews included questions about how programs of instruction (POIs) are generated, how marksmanship training is evaluated, and how emerging training strategies are being developed and deployed.

Table 1. User Needs Analysis Participants

BRM Training Support Areas	***Participants***
Marksmanship Research	ARI Research Psychologists (3)
Marksmanship Instruction	MMTC Instructors (2) 194th AR BDE Instructors (11) Drill Sergeants (5)
Training Development	Task Developers (1) Training Developers (4)
Resource Management	Ammunition Management (1) Simulation Management (2) Range Management (4)
Acquisition Support	EST Proponent Instructors (4)

Targeted user communities were identified and interviews and focus groups were coordinated through Maneuver Center of Excellence (MCoE) leadership. The data collection was conducted from September 28, 2015 to October 2, 2015 at Fort Benning, GA. Interviews and focus groups were conducted with Research psychologists from the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI), Engagement Skills Trainer (EST) proponent instructors, 194th Armor Brigade (AR BDE) instructors, Marksmanship Master Trainer Course (MMTC) trainers, Training and task developers from the MCoE Department of Training Development (DOTD), Simulation center managers, Range control operations personnel, Ammunition resource managers, and Drill Sergeants (see Table 1).

Results

Literature Review. Extensive literature exists that examines how well various measures of shooters predict performance on standard marksmanship tests (e.g., see Chung et al., 2011). What is clear from this literature is that whether you consider prior shooting experience (e.g., Tierney, Carter, & Johnson, 1979), performance in a marksmanship simulator (e.g., Hagman, 1998), trigger pull (e.g., Chung et al., 2008), aptitude (Carey, 1990), previous record fire performance (Chung et al., 2004), or measures of learner state (Chung et al., 2004), it is only possible to predict at best about 50% of the variance in performance on a standard marksmanship test.

Literature examining the acquisition of marksmanship skills is less extensive. For this, we have to turn to research on skill acquisition of complex tasks which supports a perceptual-motor learning model where learning occurs in three general phases: cognitive, associative, and autonomous (Fitts, 1962; Fitts & Posner, 1967; Anderson, 1983; Dreyfus & Dreyfus, 2005). Initially, students have an understanding of the task they are to perform, but they have little practice. Their efforts at this stage are very deliberate and require a lot of mental effort. As learners practice the skill, they are able to execute it with less mental effort and are learning to execute under a variety of different conditions. This is the associative phase. In the final autonomous phase, the skill has become automatic and can be executed with minimal mental effort in a variety of conditions. (Fitts & Posner, 1967; Gagne, 1989; Kraiger, Ford, & Salas, 1993). The framework from this skill acquisition model has also been successfully investigated as a theoretical structure to evaluate training effectiveness based on cognitive, psychomotor, and affective training outcomes (Kraiger, Ford, & Salas, 1993).

In our literature review, we identified measures in these three domains (cognitive, psychomotor, and affective) that could be used to track and predict skill acquisition. These measures were based on the theoretical support of their predictive effectiveness, predictive validity, and ease of implementation. While pilot data on these measures has yet to be collected, our findings suggest the inclusion of the measures described in Table 2.

Table 2. Rifle Marksmanship Assessment Constructs and Measures

Domain	*Construct/Measure*
Cognitive	General Cognitive Ability, Marksmanship Domain Knowledge, Openness to Experience, Hunting Experience, Videogame Experience
Psychomotor/Physical	Visual Acuity, Handedness, Eye Dominance, Height, Physical Fitness, Sports Experience, Musical Ability, and Performance Measures from Simulation and Live-Fire Training Events
Affective	Perceived Stress, Resiliency/Hardiness, Grit, Self-Efficacy, Initiative

User Needs Analysis. The data collected during the User Needs Analysis were compiled and overall themes were extracted. The three main themes that emerged from the user needs analysis were training issues, resource issues, and data management issues. A brief synopsis of these themes follows.

One of the biggest training issues had to do with the high student to instructor ratio (1 to 30+) and the sometimes limited experience of instructors for diagnosing and remediating shooting problems. Diagnosing student errors requires a fair amount of experience and skill by the instructor. When the student pulls the trigger, a lot happens in a very short period of time. Accurately and quickly diagnosing shooter errors is essential in the high throughput environment of basic training. Fortunately, efforts are underway to improve the training given to the instructors; however, many of the instructors we spoke to saw value in devices that could help them to make faster and more accurate identification of shooter problems as well as provide the ability to review a record of the student's performance from previous training periods.

The final theme from the user group analysis had to do with data management. The effectiveness with which training is conducted depends on much more than just the actions of the instructors at the various training sites. Trainees have to be delivered to the training site in a state of readiness to learn, resources have to be available for training, and the training site has to be available and capable of supporting the training. This requires coordination across training, facilities, and resource agencies. Currently, these groups use several different data management systems to do their jobs. While the Army's long-term vision is to develop a unified Army Training Information System (ATIS) to support scheduling, developing, and managing training, such a system does not currently exist. So, for example, shooter performance data may reveal that a specific target on a range is not being hit. This is often due to a malfunctioning target or to some factor that affects target visibility. Range operations does not have access to this data, so they are dependent on a training unit bringing this to their attention. Another problem is that the only way to get data from one system to another is to manually enter it. This results in errors and additional workload for instructors and other personnel. For instance, automated ranges will score student performance, but the instructors must either bring a paper printout or a disk back from the range to then enter that data into a spreadsheet and/or upload it to the Army Learning Management System (ALMS).

Interim Conclusion

In this first phase, we determined that while there is a significant body of research examining ways to identify shooter problems and predict future performance, Army marksmanship instructors still rely primarily on direct observation of trainees to identify problems and provide remediation. Given the high instructor to student ratio, this means that instructors use most of their time and resources to train the worst performers at the expense of the better trainees. Additionally, stovepiped data systems prevent different user groups from working together to solve problems and generally increase workload and data entry errors.

In this context, assessing the training effectiveness of new systems or training methods would face all the obstacles identified by Cohn et al. (2009). This first phase of work has also enabled us to identify the important assessment constructs and measures to be collected in the STEADI system and to identify how key user groups would need to access and visualize that data.

PHASE II: DEVELOPMENT OF xAPI PERFORMANCE MEASURES

Method

In order to understand human performance data, contextual information is essential. For example, if a student hits a particular target, simply recording that the student hit the target is not a particularly useful indicator of the shooter's skill. More specifically, we don't know all the facts, such as the firing position of the student, whether the student was using a scope or iron sights, whether this was the student's first attempt to hit the target or the 100^{th} attempt, whether it was a timed target, or whether the event took place during the day or at night, etc. As noted earlier, the xAPI specification allows for the inclusion of any needed contextual information (e.g., weapon, time of day, weather, shooting position, etc.) in addition to information about the actor, action, and object (e.g., Major Medford hit the target).

Although the xAPI specification allows for the inclusion of all of this information, it does not provide any recommendations for specific performance data. In a training context like marksmanship, different training devices are recording similar behaviors. For example, a trainee might do an exercise in a simulator and then repeat that exercise on a live range. In order to be able to track and analyze that trainee's performance across those two training periods, it is necessary for each system to produce the same xAPI statements for the same measures. To solve this problem, we developed the STEADI xAPI registry, which acts as a stand-alone reference for training simulator vendors to easily incorporate marksmanship-specific, xAPI measures. In addition, the online registry provides a central means to quickly update the core measures and metrics associated with STEADI during future research efforts.

Prior to developing the online registry, the constructs and related measures identified in Phase 1 were documented in a matrix. All applicable information was included, such as name, description, source, or xAPI property type. This matrix facilitated rapid development of the online registry. Table 1 provides a sample matrix with marksmanship measures.

Table 3. Sample Marksmanship Measures for Pipeline Update

Measure	Description	Data Type	Data Range	Source	xAPI Property
Team Sport Experience	Team sport participation frequency	Integer	1-5; not at all to very frequently	Demographics Questionnaire	Context
Eye Side Dominant	Eye dominance: left or right	String	(a) left eye, or (b) right eye	Demographics Questionnaire	Context
BRMT	Basic Rifle Marksmanship Test score	Integer	0 to 10	Basic rifle marksmanship testing score	Context

Results: xAPI Registry for Marksmanship

An online, xAPI code registry incorporating the identified marksmanship measures was designed and created based on the measure matrix. A web interface was chosen for implementation. Figure 2 shows a sample view from the website. The performance measure is listed, and an example JSON statement can be generated for that measure by pressing the example button. This can be used to easily standardize the implementation of xAPI in other related systems.

Figure 2. STEADI xAPI Registry: Main Table

Figure 3 displays an example of the xAPI formatted statement for the Team Sport Experience measure.

Figure 3: STEADI xAPI Registry: Example xAPI Statements

Interim Conclusions

In this second phase, we developed a tool to facilitate standardization of xAPI statements across different training systems. Standardizing these statements is extremely difficult but absolutely essential to having truly interoperable performance measurements. Although this tool was developed to standardize xAPI statements during Army marksmanship training, it can easily be extended to other training domains. Almost any training domain is apt to have multiple live, virtual, and/or constructive training environments with different training systems assessing learner proficiencies. As with marksmanship, there is a need to standardize xAPI statements across those systems to enable trainers, managers, researchers, and others to automatically track skill acquisition, examine training transfer, and measure training effectiveness. Although this tool is a means to an end, additional work in other training domains will help to develop methods for identifying critical measures and the necessary contextual information in any given domain.

PHASE III: ARCHITECTURE DESIGN

Method

Leveraging input from the user needs analysis and literature review, we developed an architecture for xAPI data capture and usage. Key use cases for data input, identity management, storage, secure access, and data analysis were identified and the use cases were employed to create system architecture design using the Unified Modeling Language (UML). The UML provides a means to describe requirements and design intent, guides development, and facilitates reverse engineering and documentation (International Organization for Standardization, 2005). The UML use case model and diagram can be seen in Figure 4. The delineation of key use case requirements served as building blocks for the final architecture design.

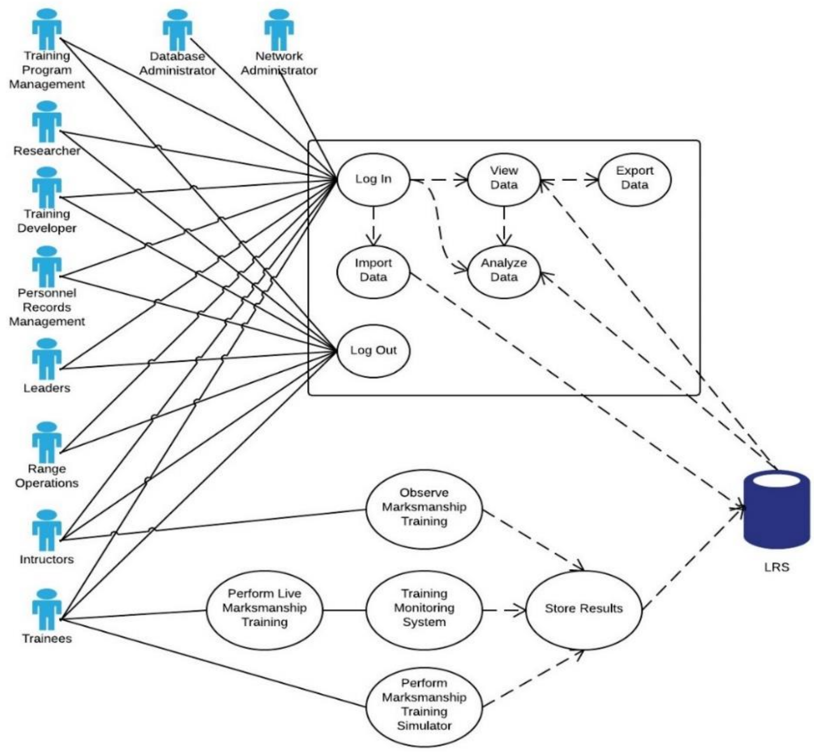

Figure 4: STEADI Use Case Diagram

The reason for including so many user groups in the design of a system meant to examine training effectiveness is to support the IMTEE training effectiveness model proposed by Alvarez et al. (2004). In that model, training effectiveness impacts the whole organization. To identify these effects, it is necessary to look beyond changes in trainee performance. For example, if training time is reduced (because training has become more effective), range capacity goes up. This, in turn, can alleviate scheduling conflicts or may reduce wear and tear on ranges, equipment, etc. If researchers identify best practices of high performing units or instructors those best practices can be used to change training outcomes across the institution. Finally if performance scores are consistently lower on certain ranges, range operations may need to inspect the range to verify that all targets are functioning properly and are visible from firing points.

Figure 4 illustrates how key user groups would access and/or generate data in this architecture. More specifically, user group requirements were translated into actions (i.e., what and how would users access data within the architecture). All users would need to log into the system and would use tools tailored to their user roles to view and analyze the data. This use case diagram focused on the human actors who would use the system and what they could do with it. It does not show interactions with other systems like learning management systems or intelligent tutoring systems; however, those interactions are described in more detail in the final architecture design described below.

Enterprise Architect, a UML architecture tool developed by Sparx Systems Pty Ltd., was used to develop both the use case and architecture diagrams. The STEADI architecture was developed using the use case analysis and diagrams. The requirements visualized in the use case diagrams provided the basis to develop the overarching system diagram.

Results

The final STEADI architecture design is shown in Figure 5. The architecture design meets all use case requirements and depicts the system boundaries, data flows, scenarios, and other attributes. The diagram provides a system-level overview of the STEADI architecture, which will provide guidance for developing a reference architecture in following efforts.

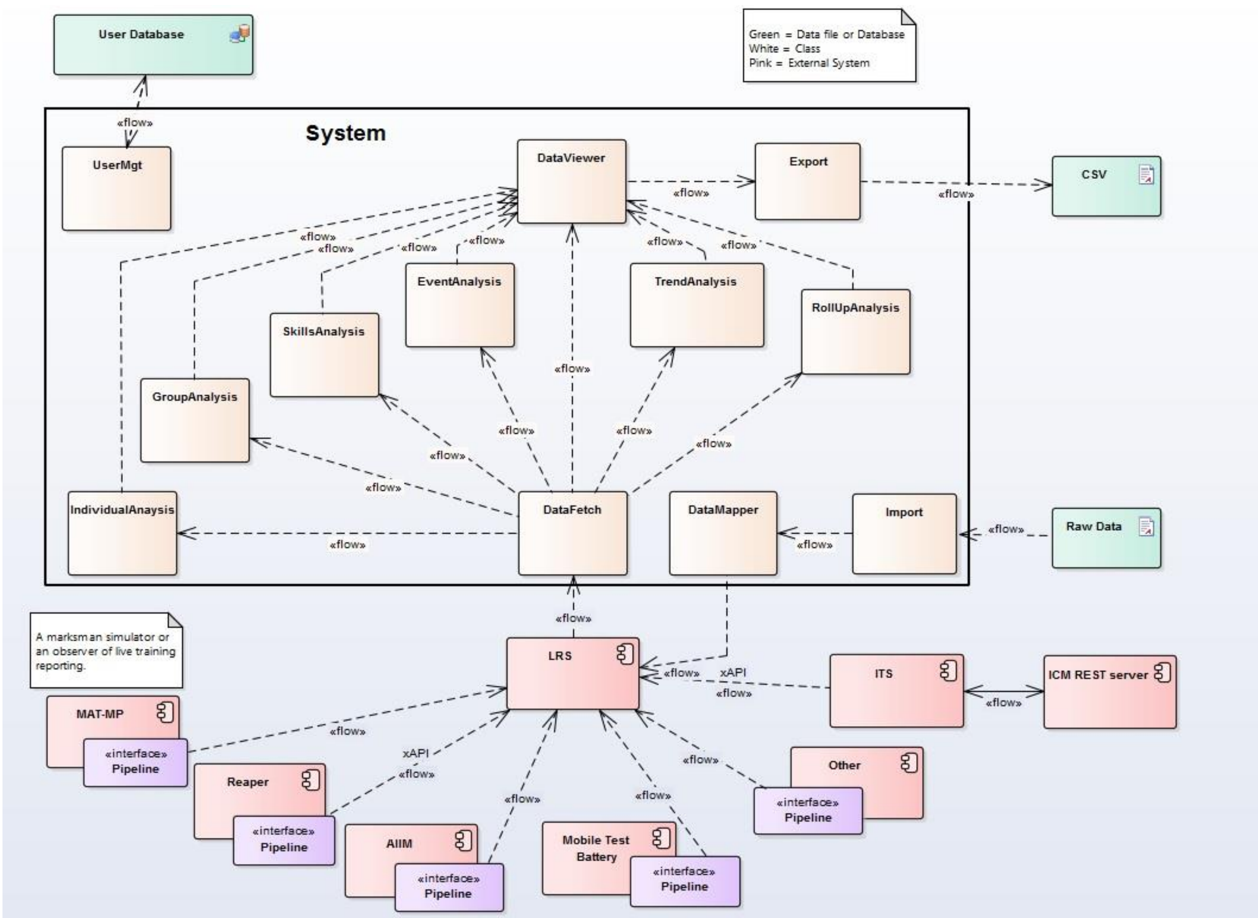

Figure 5: STEADI Architecture Diagram

The core STEADI system is contained within the black frame in Figure 5. There are six analysis classes inside this frame (individual, group, skills, event, trend, and roll up) which enable visualization of data through the DataViewer class. A user database external to the system (the LRS) stores information about different users. The UserMgt class accesses that database to determine the best interface for that user. Data summaries can be exported from the DataViewer class to comma separated value (CSV) files for further analysis or reporting.

The LRS, which contains the trainee performance database (in xAPI statements), is the primary data repository for the STEADI system. Raw data, not in xAPI format, can be transformed into the xAPI format by Import and DataMapper classes. Other training systems (e.g., simulators, automated ranges) external to STEADI provide trainee data via the LRS. Pipeline, shown in purple on the architecture diagram, is a Microsoft.NET dynamic link library (.dll) written in C#, which provides the ability for simulators to rapidly produce performance data using the xAPI specification. Finally, intelligent tutoring systems (ITS) can access trainee performance data and map to an interoperable competency model (ICM) to provide individualized training independently or even through interfaces with the various training systems (links not shown).

Interim Summary

The architecture illustrated in Figure 5 is a system that embodies the IMTEE model proposed by Alvarez et al., (2004). That model suggests that training impacts the entire organization, not just the learner. Although the use cases we identified are for a training unit, they could easily be adapted for an operational unit. Understanding the relationships between the training delivered to unit members, performance changes in those members, and the efficiency with which the unit to executes its mission is just as important for training as for operational units.

SUMMARY AND CONCLUSIONS

Information technologies offer a means for the Armed services to employ the IMTEE model to understand and manage their large-scale training enterprises. Our research identified five key phases to accomplishing this and we have completed three of those phases. We identified necessary performance measures based on Fitts and Posner's (1967) skill development framework and on a thorough user needs analysis. Using an xAPI registry that we developed, we were able to create a method for standardizing those performance measures across multiple training systems for storage in an LRS. And finally, we developed an architecture that will provide data access and visualization to support analysis and decision making. The next step of this work will involve developing an implementation of critical components of this architecture for use in marksmanship training so that we can validate performance measures and predictive models, evaluate user interfaces, and demonstrate its utility for conducting training effectiveness analyses.

The complete implementation of the STEADI system would entail establishing network connectivity to all live-fire ranges at an installation like Fort Benning. These ranges are currently not connected to any networks and the cost of adding those network connections is beyond what we have available. Because this is a research effort, the first thing we need to do is to quantify the potential benefit of connecting all ranges to an Army network so that we can determine if the cost would be justified. This means that some of the connections described in Figure 5 above will be established by manually moving data from the training systems into the LRS. Fortunately, we do have one range with a network capability, so we can compare the costs and benefits of moving data into the LRS automatically vs. manually.

Developing the STEADI architecture has led to the identification of challenges associated with the use of xAPI to store trainee performance measures. For example, something that is both the biggest strength and the biggest weakness of xAPI is that it does not prescribe how particular performance measures are to be described. It is a strength because it creates flexibility for systems that would use it to save user activities. On the other hand, it is a weakness because different systems can end up encoding the same behavior in several different ways. This makes it challenging to compare performance across systems. ADL has relied on communities of practice to solve this problem. Communities of practice are expected to develop their own standards for encoding specific behaviors to include specific naming conventions, etc. The creation of an xAPI registry, as described above, provides a means to catalogue those conventions in a tool that facilitates the generation of standard xAPI statements.

Although the xAPI registry establishes common ways of describing trainee behaviors using xAPI statements, it does not solve the problem of two different systems measuring the same behavior in different ways. For example, the Engagement Skills Trainer, a simulator, measures trigger pull by looking at movement of the trigger. On the other hand, a device used to record trigger pull during live-fire training measures trigger pressure. While both of these systems measure trigger pull, the sensors used would not be expected to produce equivalent output. One solution to this problem would be to have individuals of different ability levels produce comparable trigger pulls on both systems and then attempt to develop a transformation function to produce comparable output. Another would be to treat these as different but related measures that each feed into diagnostic and/or predictive models, each one being validated independently. These solutions will have to be evaluated in future research.

This leads to the final challenge of using xAPI to record data about the learner: raw performance measures are frequently not informative in and of themselves. In addition to the need for supplemental contextual data, as discussed above, models are often needed to provide metrics that an instructor can interpret. Using the trigger pull example just described, a graph of pressure change over time or even trigger movement over time would not make sense to even an experienced instructor. Models are needed help the instructor interpret whether the data represents a proper trigger pull for a student at that point in the training program. As we proceed with our research, we will be investigating the development of these models to facilitate interpretation of data and to determine the best way to present that data to trainers and other user groups.

ACKNOWLEDGEMENTS

The authors would like to thank the leaders, trainers, training developers, and resource managers at Fort Benning, GA for their support and participation in our user needs analysis.

REFERENCES

Advanced Distributed Learning. (2014). *Training and Learning Architecture (TLA)*. Retrieved from Advanced Distributed Learning: http://www.adlnet.gov/capabilities/tla

Alvarez, K., Salas, E., & Garofano, C. M. (2004). An integrated model of training evaluation and effectiveness. Human Resource Development Review, 3(4), 385-416.

Cohn, J., Stanney, K., Milham, L., Carroll, M.B., Jones, D., Sullivan, J, & Darken, R. (2006). in J. Cohn, D. Nicholson, & D. Schmorrow (Eds.), *The PSI Handbook of Virtual Environments for Training and Education: Developments for the Military and Beyond. Vol 3. Integrated Systems, Training Evaluations, and Future Directions* (pp 158-172), Westport, CT: Praeger Security Intl.

Fitts, P.M., & Posner, M.I. (1967). Human performance. Belmont (CA): Brooks/Cole.

Gagné RM. Studies of learning. Tallahassee (FL): Learning System Institute; 1989.

International Organization for Standardization. (2005, January). Unified Modeling Language Specification: Version 1.4.2. Retrieved November 2015, from Object Management Group: www.omg.org/spec/UML/

Kirkpatrick, D.L. (1979). Techniques for evaluating training programs. *Training and Development Journal, 33,* 78-92.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of applied Psychology, 78*, 311-328

United States General Accounting Office (1991). Army Training: Computer simulations can improve command training in large-scale exercises. Report to the Chairman, Subcommittee Readiness, Committee on Armed Services, House of Representatives. (Report GAO/NSIAD-91-67). United States Government Accountability Office, Washington, D.C.

United States General Accounting Office (1993). Simulation Training: Management framework improved but challenges remain. Report to the Chairman, Committee on Armed Services, House of Representatives. (Report GAO/NSIAD-93-122). United States Government Accountability Office, Washington, D.C.

United States General Accounting Office (2006). Contract Management: Service contract approach to aircraft simulator training has room for improvement. Report to the Chairman, Subcommittee Readiness, Committee on Armed Services, House of Representatives. (Report GAO-06-830) United States Government Accountability Office, Washington, D.C.

United States General Accounting Office (2012). Air Force Training: Actions needed to better manage and determine costs of virtual training efforts. Report to Congressional Committees. (Report GAO-12-727). United States Government Accountability Office, Washington, D.C.

United States General Accounting Office (2013). Army and Marine Corps Training: Better performance and cost data needed to more fully assess simulation-based efforts. Report to Congressional Committees. (Report GAO-13-698). United States Government Accountability Office, Washington, D.C.