Recording Progress in Soldier Training and Development with an Experience API Database Environment

Igor Franken MSc., Richard Kist MSc.
Netherlands Aerospace Centre (NLR)
Amsterdam, The Netherlands
Igor.Franken@nlr.nl, Richard.Kist@nlr.nl

Emilie A. Reitz
Alion S&T; Joint Staff DD C5I Joint Fires Division
Norfolk, VA
emilie.a.reitz.ctr@mail.mil

ABSTRACT

When training Soldiers, it is crucial to monitor the progress being made in Soldier performance to evaluate the effectiveness of a new training method. Whether an organization is considering the effectiveness of training or evaluating Soldiers' fitness for a task, the true performance cannot be judged without data. During two courses taken by military personnel in The Netherlands, and other data collections during the Bold Quest series, experimental use was made of an automated survey tool, QUEST, to record progress made by Soldiers and eventually to quantify course effectiveness. The data collection consisted of students' course-content test scores, and multiple choice survey questions asked to the Soldiers. Included were factual tests and questions seeking subjective feedback on certain aspects of the training, looking for areas to improve. These aspects and demographics of the Soldiers were collected in a single database, connecting all data. The Bold Quest series of exercises provides an operational test environment where new technologies and training methods are demonstrated, and the subjective data collected was paired post-event with the objective systems data, providing another use case.

When training feedback and effectiveness data is collected from several consecutive courses into a central database, insight into improvements in the training can be gained and trends clarified. It provides the means to correlate demographics, experience and skills, course scores, Soldiers' feedback and direct tests. Training methods can be improved using this analysis by tailoring training to candidates based on backgrounds and skills. Correlation between demographics, experience, skills and the shown aptitude to the training can also improve the effectiveness of recruiting. This paper focuses on the benefits of using automated surveys for these training improvements, as well as the long-term goal of integrating this collection technique with other technology-enhanced learning capabilities like Experience API (xAPI) (Mueller, Dikke, Dahrendorf, 2014).

ABOUT THE AUTHORS

Igor Franken received his Master of Science degree in Electrical Engineering at the Delft University of Technology. Shortly after his graduation Igor's professional career started in 2006 at the Netherlands Aerospace Centre (NLR) in Amsterdam in the Defense Systems department. From 2007 Igor was involved together with colleague Richard Kist in programs involving combat identification and Digitally aided Close Air Support (DaCAS). He was an active member of the Dutch team at several Bold Quest exercises. Igor and Richard developed the survey tool QUEST for use in this series, and played an important role in the successful subjective data collection during Bold Quest events.

Richard Kist received his Master of Science degree in Aerospace Engineering at the Delft University of Technology. He started his career in an IT company. In 1999 he joined the Netherlands Aerospace Centre (NLR). Starting in 2007 Richard was involved in Bold Quest. Richard and Igor developed the survey tool QUEST for use in this series, and played an important role in the successful subjective data collection during Bold Quest events.

Emilie A. Reitz, M.A., is a Research Analyst at Alion Science and Technology. She is currently supporting the Joint Fires Division of Joint Staff J6, Deputy Director for C4 and Cyber Integration. In this capacity, she is the data collection and analytical working group lead for Bold Quest. Her research focuses on integrating joint capabilities into modeling, simulation, and training, as performance enablers.

Recording Progress in Soldier Training and Development with an Experience API Database Environment

Igor Franken MSc., Richard Kist MSc.
Netherlands Aerospace Centre (NLR)
Amsterdam, The Netherlands
Igor.Franken@nlr.nl, Richard.Kist@nlr.nl

Emilie A. Reitz
Alion S&T; Joint Staff DD C5I Joint Fires Division
Norfolk, VA
emilie.a.reitz.ctr@mail.mil

INTRODUCTION

In military operations field training and experience are crucial in order to get the best equipped Soldiers for the battlefield. Monitoring the level of training or the amount of experience a Soldier has is not a straight forward task that can be evaluated with a simple test or exercise – it instead requires intense and long-term monitoring in a coordinated fashion, ideally centralized around each individual Soldier. It is clear that these processes are very different from Soldier to Soldier, task to task, and are not easily generalized among large groups.

The authors' 2015 I/ITSEC paper (Kist, & Franken, 2015) highlighted the way questionnaires can be made more effective lowering the burden to the respondent and at the same time improving data collection quality and quantity. It provided clear guidelines to efficiently constructing questionnaires and presenting them to respondents in a way that led to unbiased answers. Those guidelines result in a clean data set that is not affected by ambiguous questioning or hard to understand terms. This paper continues that theme, and describes how these methods and the resulting data set can be used to determine the effectiveness of training and the progress of individual Soldiers. Using an automated survey and collecting feedback from both the trainees and trainers provides the opportunity to look at a specific training event or exercise, or the longer term trends and the improvement of trainees over time.

During the Joint Staff J6 sponsored capability demonstration and assessment cycle, Bold Quest, an automated survey tool called QUEST was used. The tool collected respondent feedback on technologies demonstrated during the exercise as well as improvements in proficiency with the equipment over the course of the exercise. While combining these results with the respondent's demographics, past experiences and training, the participating NLR (Netherlands Aerospace Centre) researchers realized that the tool, with some minor adjustments, could easily collect experience and performance data – which, over longer periods or multiple exercises, could be used to indicate trends and improvement opportunities. In following Bold Quest events and other Dutch national events, these adjustments were partly incorporated and valuable experience-based data was collected on several occasions.

The goal of this paper is to show the potential to use a single survey tool for collection of trainee feedback on courses, of tutors' and teachers' ratings of trainee performance and tests to measure their understanding of the training, combined into a training evaluation and Experience API collection of the individual Soldiers. The next section describes what centralized experience data collection incorporates and which types of feedback are distinguished. The results section contains a description of past opportunities to use this method of experience data collection followed by a list of lessons learned and the conclusions.

EXPERIENCE API

The Experience API (xAPI) or Tin Can API provides a way of collecting events occurring in any learning experience. xAPI was released in 2013 as the outcome of an Advanced Distributed Learning CoLab project that aimed to both: (i) improve interoperability between E-learning systems that collect and exchange student learning data, and (ii) overcome the limitations of SCORM. The Tin Can API describes learning experiences in at least three standardized segments: actor (or noun), verb and object. In the record "Ellen completed a skydiving course" *Ellen* is the actor or noun, *completed* is the verb and *a skydiving course* is the object. When continuously storing learning experiences in this way a database will be formed that describes the level of experience of the individuals. Searches can be conducted for the person that is best trained for a certain skill like skydiving or another task. This basic set of

parameters can be complemented with a context and result field to get even more accurate data. "Ellen completed her skydiving exam in the rain with an average score of 9.1" already provides more data on the level of training and fitness of the subject. It is crucial that libraries are agreed upon to prevent that *completed*, *finished* and *succeeded* do not end up to become different data sets while they contain the same data. Once experiences are recorded over time the output opportunities of the dataset are endless (Tin Can API Homepage, 2016).

CENTRALIZED EXPERIENCE DATA COLLECTION

To monitor career progress and experience levels of individual Soldiers, it is important to collect data over an extended period of time. xAPI or Experience API is a method designed to add several types of experiences into a central data repository. During Bold Quest, a tool was developed using such a central database, collecting experience with equipment and methods over multiple exercises. From different requirements, the same flexible tool was used for collecting rating data, where tutors entered specifics on trainees' capabilities. In another application, the tool was used as a test tool to collect on trainee understanding of factual data. The latter was also used to indicate how new and experimental training affected their understanding by providing pre and post-testing. These types of collections can all be performed simultaneously in the same database. When this is done over time with a number of individuals and different training methods, new possibilities arise, both from monitoring the improvements of these individuals, and from measuring the effectiveness of an evolving training method.

Long-term data collection including demographics

The same trainee will logon on multiple occasions and enter data (answers) on the same questions while he/she is in a learning process, whether it is a one week exercise or a multiple-year career development. To monitor progress during training, results need to be logged and linked to the same individual user account. Adding demographics to this gives the opportunity to link certain behavior to the trainee's background and previous training, or to be able to select the right person for a task based on just his or her past experience. It provides a toolkit to a researcher that is not limited to the individual trainee, but also shows correlation between a general set of capabilities and background, as well as experience before being recruited, and the impact of different types of training. In the broader educational community, much work has been done on using demographics as a performance predictor in courses (Barber, & Sharkey, 2012), both to suggest courses and to pre-place support structures and additional tutoring for those trainees whose performance was expected to be negatively impacted by their demographics.

Types of feedback

Three main types of valuable feedback needed to create a high utility, centralized experience collection are:

- Feedback from the trainee on the training (survey)
- Feedback from trainers on trainee performance
- Test items indicating the trainees' understanding of content (learning curve)

Feedback from the trainee on the training (survey)

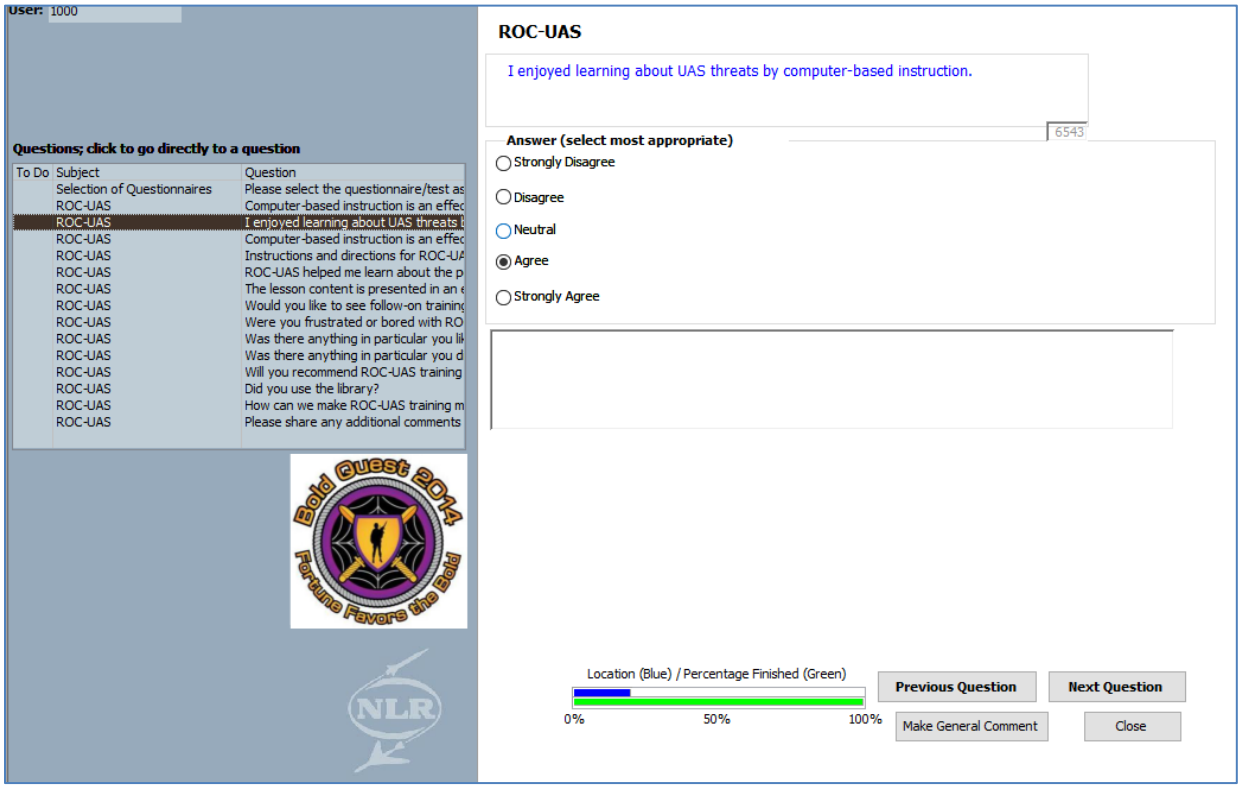

Figure 1: Sample of feedback questions on a new training method, used during Bold Quest

Feedback from the trainee on the training is useful for all training evaluations, especially when training is new and untested. The type and number of questions need to be optimized to lower the burden to the respondents (described in Kist, & Franken, 2015), but still manage to reveal details where improvements can be made to the training. Lines of questioning may seek feedback on how to the point the training is for operational use, determine where it lacks in clarity, whether the learning curve is too steep or shallow. A researcher would need to correlate these points of interest to the previous experiences of the respondents. One participant who is providing feedback may never have been operationally deployed, while another is a seasoned veteran. The level of education before they joined the training or their larger military career may also influence the comments made on the steepness of the learning curve. In Figure 1, a basic question can be seen regarding feedback on training. Usually the question is a statement on the training that can be answered by a Likert scale set of answers, strongly disagree to strongly agree in five or six steps.

Feedback from trainers on trainee performance

Feedback from trainers on trainee performance is aimed at establishing the rating of the trainer of each trainee in a training course or event. This can be a mark on certain types of mission performance, a check that one part of the training was successfully achieved, or contextualized, written feedback on specific skills or deficiencies in the trainee. The results of these data points may be correlated to the demographics of the trainee as well as the subjective feedback given by the trainee on both the course and their own perceived performance gaps. The combination of these may offer insights enabling a better and more tailor-made training method. In Figure 2 a sample of such a grading question can be seen. The instructor is indicated as “XXX” in the sample, and he is logging data for “YYY”, the student being graded. As can be seen on the left part of the form, the grading is done in several categories. By grading the same person over many days, and multiple training events and types of training, or grading many students per course, data is collected that may be used to see trends for each course, as well as trends for each individual.

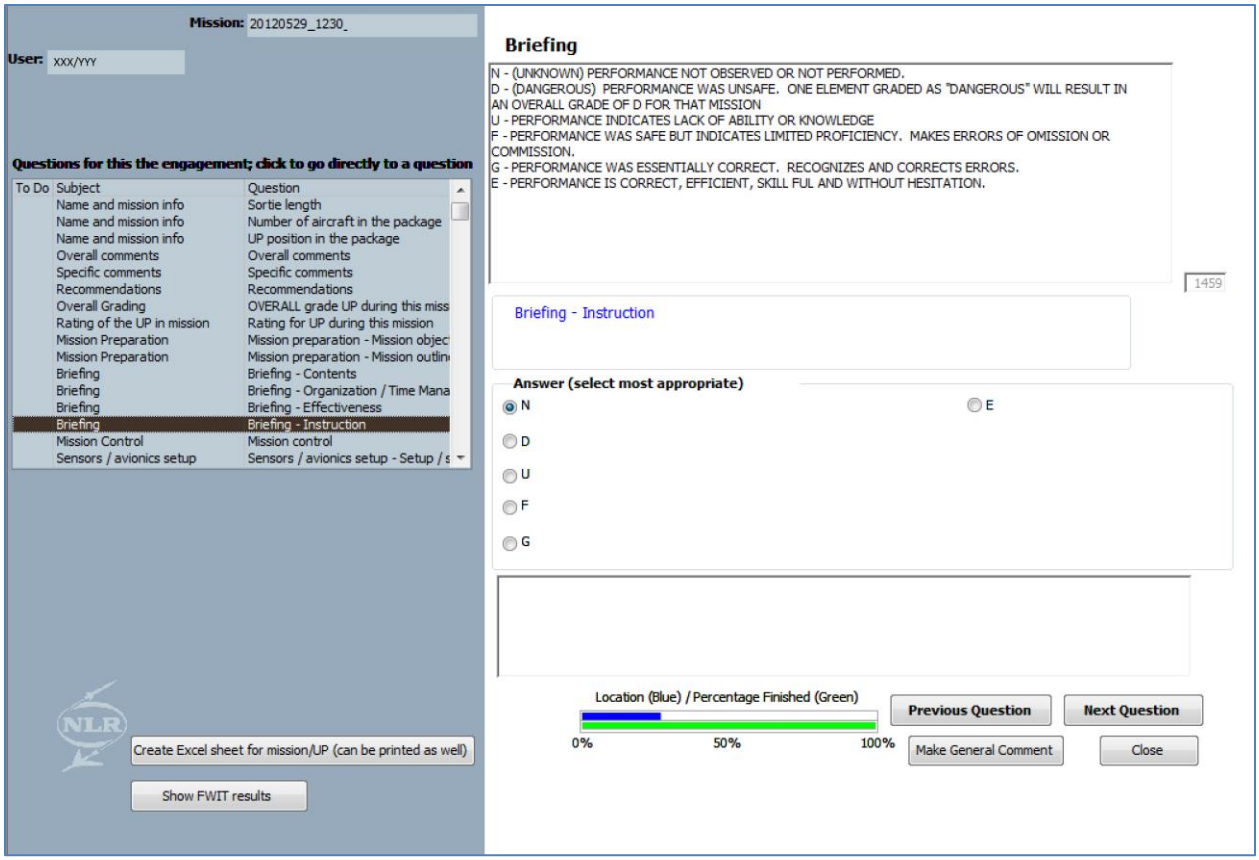

Figure 2: Sample of question on grading, of tutor on trainee, as used during the Fighter Weapon Instructor Training (FWIT)

Test items indicating the trainees' understanding of content (learning curve)

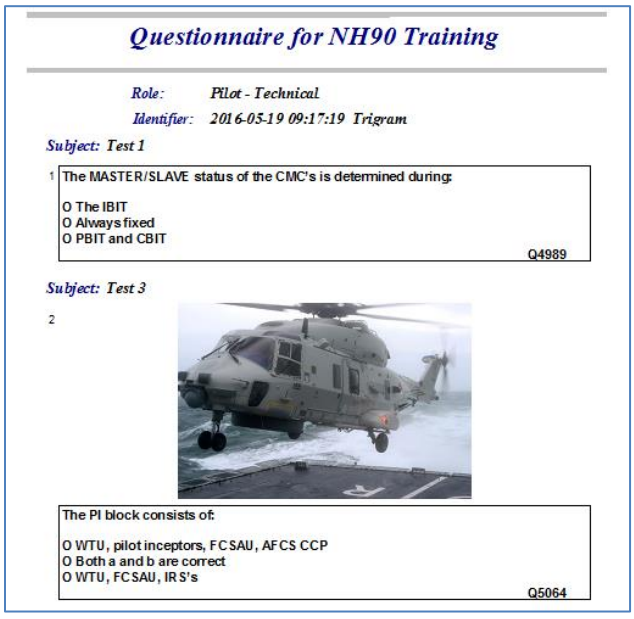

Figure 3: Sample of test of an NH90 related course; each question has only one correct answer

The third method is a conventional test of the trainee, where the trainee needs to answer questions with multiple options, where one (or more) choices are correct. These tests can be graded directly, but also be used to determine which specific concepts and knowledge fields are not fully understood by the trainee. A well-developed test may reveal that some fields are not understood by most trainees, or by most trainees with a certain characteristic. In Figure 3, a sample can be seen of an example test requiring the trainee to select the correct answer. The trainer can select any part of the test and a number of random questions and answers in a random order, preventing the trainee from seeing the answer elsewhere, or from copying over the shoulder of a nearby user.

Optimization of training is more efficient if performance data is collected with this level of detail. If only general grades are kept, no specific trends on details of the training can be retrieved. Many courses run for years, touching hundreds and thousands of students with little documentation about long-term

trends. Also, when full details of the tests are retained, it is possible to monitor the long-term progress of each individual trainee in specific areas where he or she previously showed deficiencies.

Other types of feedback and experiences that may be collected
Part of the additional collection will be done by determining demographics, experience and education as well as previous training. Another part needs to interface to other digital or digitized data, such as results from the firing range, and simulation data. In most cases, this will require a network connection to other digital systems, of a similar classification level.

In some circumstances, it is sufficient that the superiors of trainees (or the trainees themselves) enter their experiences and results of these manually into the database. Especially if it is a qualitative description, this may be a more practical method in the short term. The QUEST survey environment is capable of allowing the person entering data to use xAPI compliant terms to frame their input, as opposed to free text without any standardization. It may just be a variation on a multiple choice question that assists and actually forces the data to be xAPI compliant.

For instance, "John completed a course on tracking with a passing score of 95% in 4 hours" is a valid xAPI compliant statement. However, in the database just using free text, a person entering data may write this as "John did a tracking course pretty well in good time". This would not be standardized nor would it enable comparison and measurements.

To properly store xAPI compliant data in a searchable way, *John* as actor would need to be stored using the unique ID of the trainee, a different field in the data record would be used for the verb *completed*, another field for the object *course on tracking*, and fields for the resulting *passing score* and the *time to finish*. All standardized terms should be provided to the respondent with lists/combo boxes to help this process. If stored this way, assessment and finding trends becomes possible, as well as finding correlations between the abilities and experience and training outcome.

CORRELATION OF COLLECTED DATA

The U.S. Army is pursuing techniques to improve the human dimension (TRADOC, 2014) of their warfare capability – the Soldier him or herself. Trying to optimize the skillsets, background demographics and information of units to provide maximized training opportunities is not a new concept (Holtz, Hiller, & McFann, 1994); but with the technologies currently available, it is now opportune to gain ground in the field by using big data sets to quantify what those nuanced patterns might be. xAPI stands to potentially provide that big data set.

Using combinations of feedback data on related questions from not just the QUEST tool, but other similar databases as well, patterns can be recognized that will assist in improving the entire training environment and ultimately improve the quality of the warfighter by improving recruitment, producing more effective training and more awareness of the experiences that contribute to effective operations.

Ideally, in a training environment, each assessment environment (or module) should be based on evaluation of a different entity. One should focus on the training/experience evaluation. This entails measures for the level of understanding of trainees of the skills the training seeks to add or improve on. Sometimes this is factual (knowledge that can be tested). In that case, a test before and after is a simple solution. If it is a skill that is measured in a different environment, such as shooting performance, the data needs to be entered either manually or an automated interface is required that allows import of the data.

Another aspect required for evaluating any training is subjective feedback from trainees. This should in some cases be correlated to whether the trainee benefitted from the training, and if the trainee's experience and competence level before the training commenced matched the expected level needed to benefit from the training. If a number of trainees are not positive about the training, and did not benefit from it, it may mean the training itself needs adjusting; but if the unhappy group shares a limited level of skill, education or experience when compared to a group that was satisfied with the same training, the training may be unsuited for the group with less experience.

A capability that also needs to be provided is the ability for an individual Soldier to track which experiences he or she has encountered, as well as their own scoring on several skills. This can provide better insight into when this individual can best receive any type of training, how skilled and ready he is for a task, and what needs to be done to be prepared for a new task; but also allow the individual access to his own scores and skill levels over time and sortable by training content area.

The third necessary capability would focus on correlating performance to create better teams, leveraging previous skills, education and other capabilities with effectiveness in the current function. If correlations are found, recruitment and tailored training can be optimized, achieving lower drop-out rates and achieving better skilled Soldiers. The budgetary gains that can be achieved by this are significant.

RESULTS

During multiple Bold Quest events, several new training environments were evaluated through the use of knowledge tests before and after the training. The effectiveness of this training was measured by looking at the differences in the grading for the group, as well as leveraging audio recordings, subject matter expert collected performance measures, and instrumentation. For some training experiments the training was repeated several times, each time giving the trainees a knowledge or application-related test based on the course contents with each time a different order and selection of questions. By looking at the demographics, additional correlations could be shown between background and experience of the trainees and their individual performance during the training.

During a national exercise for fighter pilots in The Netherlands, data was collected on performance of 12 trainees in the exercise. This was intended to enable identification of trends in performance and to check whether all attending the course performed all their required manoeuvers and met all other performance goals in the course. From a technical point of view, this proved successful. However, to see trends, the data needs to be collected during more than one iteration of the training, which was not achieved yet. Practical issues were part of the problem, as the austerity of the training environment made it challenging to provide the database tooling directly during the course.

The tool is able to correlate responses to two separate questions. Figure 4 illustrates the correlation between the attendance of a course and the use of a skill (enhanced observation) trained during that course. The red bars indicate responders that did take the course and the blue bars the ones that did not. To clarify, the chart shows that 63% of the responders, that did not take the course on enhanced observation, never use the skill, where 31% of the responders that did take the course use it frequently. The correlation is evident and it illustrates that this is a helpful feature in finding more of these direct relations. It is obvious that in order to get relevant results a large data set is crucial and even that both groups in this example need to consist of a sufficient number of responders.

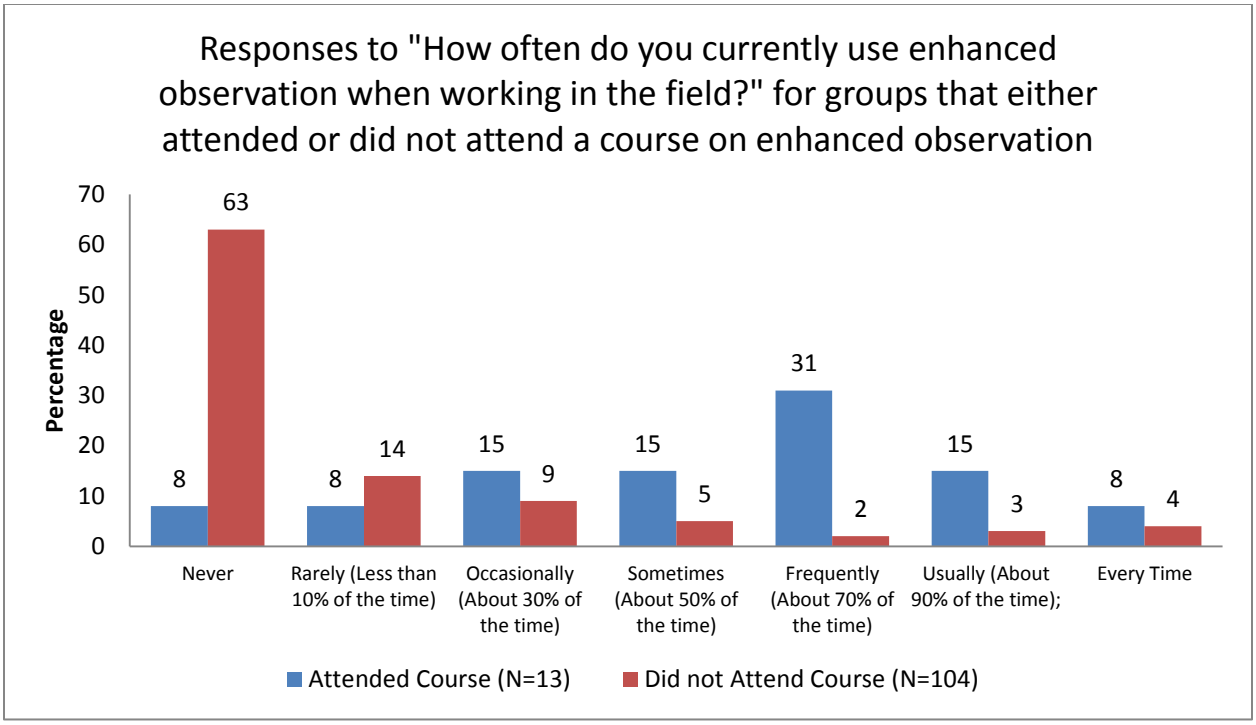

Figure 4: Correlation between two separate questions, one on attending a course on enhanced observation and one on the use of enhanced observation in the field

During the Bold Quest series of events, two training-focused technologies were assessed across a period of years, utilizing the demographics, trainee feedback and knowledge transfer data captured in the QUEST tool: dismounted simulation capability types for Joint Terminal Attack Controllers, and the Dismounted Soldier Training System (Reitz, & Seavey, 2014; Reitz, & Richards, 2013). In case of the Dismounted Soldier Training System, 348 Soldiers were trained in the capability and responded to surveys on it across five years. In both cases, the tool allowed for multi-year assessments to be robustly based in training transfer, demographics, and users' subjective ratings of the utility of the training and technology, as well as improvements needed. By controlling for the differences in user experience, we were able to assess whether poor ratings of technology were related to poor understanding of the situation and frustration at the challenge of a particular mission, or if the technology really was just performing poorly – and the data revealed a mix of both potential cases being true, with participants whose self-efficacy was higher than their actual mission performance having an overall more negative perception of the system.

But there were challenges creating these multi-year data sets. While the aspects of xAPI were included in the collection, the mechanisms have not been refined to the point where the community writ-large wants to go – a ubiquitous training environment where all past performance data is easily accessed and applied as necessary to tailor current training situations.

Without the application of xAPI to outlying, non-SCORM and non-learning management systems, performance data from instrumentation systems and other field-based performance data must be associated to surveys, demographics and knowledge-related data by hand or algorithmic queries to allow for the creation of a more robust dataset. This additional manipulation is also a space which allows for the injection of errors in the data. The final, analytically based product is not one which is available for the trainee and his or her leadership. It remains a research artifact. Until the multiple databases in the larger military learning community can be bridged, much of the data is not being correlated in a way which can impact the trainee in the longer term. In the current scenario, the trainee has no oversight of his or her performance in the mission, other than whether the squad, squadron or commander noticed that they performed poorly.

Although the full long-term data collection with a group and set of courses using the three types of data collection outlined above has yet to be commenced, all three elements to do this have been achieved already. It has been shown

these elements can indeed be combined in one database, as this happened during Bold Quest for two of these three and in The Netherlands for a different combination of two.

LESSONS LEARNED

During our testing and experimenting with the process for collecting and correlating data, we have experienced some important lessons to be learned, which pertains to the application of it to the Experience API methodology.

- It is critical to enable both the trainer and trainee easy access to the database through the interface (Kist & Franken, 2015). Going to another room or building after an exercise has proven to be a hindrance to collecting good data; many times trainees will skip the data input, or the added burden of moving to another location will lower their willingness to participate.
- Trainee burden for inputting data during one period of time should be kept as short as possible. Even for a one time training event, over 20 minutes is considered long. If data input is a daily event, 5 to 10 minutes is the time bracket within which the response to the questionnaire/xAPI input should be finished (Kist & Franken, 2015). Optimization of data collection is paramount to get the precise and accurate data that needs to be collected in the long-term.
- It is difficult to get a long-term data collection started and to keep it going. The long-term benefit, especially for identifying trends and making recruitment improvements, are seen only after months and even years of collection.
- High level 'ambassadors' for the survey/xAPI tool within the organization are required to be able to achieve this long-term collection. This was experienced at Bold Quest, as well as for exercises in The Netherlands. The leadership behind the course, training event, or collection must believe in the utility of it, convey that to participants as appropriate, and empower the data collectors and administrators.
- No data collection should be duplicated, as this is very frustrating for any respondent.
- Modules showing trends should be available from the very beginning of a collection. Although in reality these modules gain value only after months of collection, if they are not available in the beginning, respondents (and their leadership) may not recognize the long-term benefit of the collected data. If they see these modules or reports from the first day and get (at first limited) results through a data dashboard, the usefulness of the data and finding trends is more visible.

FUTURE WORK

Interest in the emerging xAPI standard and the way many existing data collection tools can be extended to be fully used in this concept resulted in a plan to upgrade the QUEST environment to this new standard. Part of what is needed already exists in the environment created for Bold Quest; grading from the outside, grading through a computer-based test and feedback (experience) data by the subject (trainee) are all possible in the same database.

Still required is a standardized set of high-utility experiences that can be entered into the database. Questions and multiple choice answers are present, but the presentation of entering experiences with a standardized set of actors, verbs and objects needs to be implemented. Interfaces need to be developed to communicate with other tools and databases, including simulations used for training and any appropriate E-learning tools. This way learning performance and skill as measured by the training aids is objectively and efficiently stored in the central database.

When these grades, experiences and demographics are all present in a single data set, these data points need to be searchable. Part of this is searching using the standard terms. If for a military mission persons are required to be able to climb a mountain and use a parachute, and it is desirable they speak a certain foreign language, a central xAPI-based database can be a tremendous assistance to find a group of forces that all share the required qualities, and some have the language skill that is desired. This matching capability can be used outside the military world in companies as well.

CONCLUSIONS

The automated survey environment was able to collect feedback on all three above mentioned types of feedback in the same database. The resulting database contains feedback from the trainee on the training as well as feedback from trainers on those trainees' performance and graded tests filled out by the trainees that quantify their understanding of the training content.

Using this data, it was possible to assess training both by measuring understanding by the trainees and by feedback from the trainees on the training. Using demographic data it was also possible to do rudimentary correlation between a trainee's opinion on the effectiveness of training and factors like operational experience or computer skills. It was also possible to distinguish whether a critical opinion on a technology or training was caused by shortcomings of these or rather related to, for instance, frustration at the challenge of a mission.

The gathered experience data can in the near future be used to evaluate and possibly improve the recruitment methods for military personnel for a specific function or task. Data may also be used to improve training by making separate modules customizable for individual trainees.

The work done so far has indicated that it is possible to incorporate xAPI in the current automated survey environment easily. With xAPI terminology and a long-term data collection it should be possible to search for specific personnel with certain experience or skills. Also this would allow for trends to be visualized when doing long-term data collection.

REFERENCES

Barber, R., & Sharkey, M. (2012). Course correction: using analytics to predict course success. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 259-262). ACM.

Del Blanco, Á., Serrano, Á., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2013, March). E-Learning standards and learning analytics. Can data collection be improved by using standard data models?. In *Global Engineering Education Conference (EDUCON), 2013 IEEE* (pp. 1255-1261). IEEE.

Holtz, R. F., Hiller, J. H., & McFann, H. H. (1994). Determinants of Effective Unit Performance: Research on Measuring and Managing Unit Training Readiness. ARMY RESEARCH INST FOR THE BEHAVIORAL AND SOCIAL SCIENCES ALEXANDRIA VA.

Kelly, D., & Thorn, K. (2013). Should Instructional Designers care about the Tin Can API? *eLearn, 2013*(3), 1.

Kist, R.A., Franken, G.E.A. (2015). Automated Surveys: Lowering the respondent's burden. In *Proceedings of the Interservice/Industry Training, Simulation and Education Conference 2015.*

Mueller, N., Dikke, D., Dahrendorf, D. (2014). Experience API vs SCORM – How xAPI benefits Technology-Enhanced Learning. In *EDULEARN14 Proceedings.*

Poltrack, J., Hruska, N., Johnson, A., & Haag, J. (2012). The next generation of scorm: Innovation for the global force. In *Proceedings of the Interservice/Industry Training, Simulation and Education Conference, 2012.*

Reitz, E. A., Richards, R. (2013). Optimum Dismounted Soldier Training Experience: Live or Virtual? In *Proceedings of the Interservice/Industry Training, Simulation and Education Conference 2013*.

Reitz, E. A., & Seavey, K. (2014) Distributed Live/Virtual Environments to Improve Joint Fires Performance. In *Proceedings of the Interservice/Industry Training, Simulation and Education Conference 2014.*

Tin Can API Homepage, (2016). *Programmable E-learning and Experience Tracking*. Retrieved June 14, 2016, from https://tincanapi.com/statements-101/

U.S. Army TRADOC [Training and Doctrine Command]. (2014) Human Dimension Concept (TRADOC Pam **525-3**-7. Department of the Army, Headquarters, United States Army, Training and Doctrine Command, Fort Eustis, Virginia.