Using Digital Twins in Maintenance Operations and Training

Deepak Haste, M.S.[1], Sudipto Ghosal, Ph.D.[1], Valarie Yerdon, M.S.[2], Maartje Hidalgo, Ph.D.[2], Jeffrey M. Beaubien, Ph.D.[2], Jason H. Wong, Ph.D.[3]

Qualtech Systems, Inc.[1], Aptima, Inc.[2], Naval Information Warfare Center Pacific[3]

Rocky Hill, CT[1], Woburn, MA[2], San Diego, CA[3]

deepak@teamqsi.com, sudipto@teamqsi.com, vyerdon@aptima.com, mhidalgo@aptima.com jbeaubien@aptima.com, jason.h.wong.civ@us.navy.mil

ABSTRACT

Maintenance Operations and Training is an integral part of equipment lifecycle management. The DoD has been using interactive electronic technical manuals (IETMs) for troubleshooting and training purposes since the 1980s. However, IETMs are poorly suited to the complexities of modern equipment due to the fact that their linear troubleshooting packages are unable to capture complex cross-subsystem interrelationships. Moreover, every time the equipment is upgraded, the IETMs must also be upgraded to reflect these changes. By comparison, model-based intelligent reasoners (i.e., digital twins) go a step further than IETMs by using a model as the "single source of truth." With updated underlying models, the training content and troubleshooting capabilities remain relevant. Coupled with COTS head-mounted Augmented Reality (AR) devices, model-based intelligent reasoners can support hands-free, just-in-time (JIT) training and troubleshooting support. Previous research suggests that such tools can help novices perform like experts by reducing the number of troubleshooting steps by half (Schlueter, 2018). Along with lowering the amount of time-in-training, the cost-savings of virtualized and immersive training have been found to be 20% of their hardware counterparts to build and update (Orlinski & String, 1981).

In this paper, we present a model-driven, agile *DevOps* (Wikipedia, 2022)-like "*TrainOps*" process that leverages digital twins, COTS AR technologies, and learning sciences concepts. We provide an overview of how digital twins can be used to quickly and efficiently model downstream technical faults for use in maintenance training and operations. We also describe an automated process that captures standardized learning records to yield quantifiable metrics of learner performance, providing on-time feedback for both the instructor and the learner. Finally, we conclude with a series of best practices when adapting a model-based decision support tools for classroom-based training purposes through the use of scaffolding (Wood et al., 1976) and progressive hints (Shute, 2008).

ABOUT THE AUTHORS

Mr. Deepak Haste is the Director of Engineering at QSI, with focus on customer-driven enhancements, productization and commercialization of QSI's products through several key SBIR projects with NASA and DoD. His recent productization efforts include extensibility through plugin frameworks and prognostic capabilities, and embedding and interfacing diagnostic software within onboard platforms. He is currently leading efforts to add Augmented Reality / Virtual Reality (AR/VR) and Maintenance Training capabilities into QSI's product suite. Mr. Haste holds an M.S. in Electrical Engineering from Clemson University and a B. Tech. in Electrical Engineering from Indian Institute of Technology (IIT), Bombay, India.

Dr. Sudipto Ghoshal is the Vice President of Engineering at QSI, with experience in system diagnostics and prognostics, and design-time assessment and mitigation of operational risk and improvement of system availability. His research involves development of failure-space modeling methodology leveraging general purpose system design languages such as SysML and adapting it for QSI's products. He has been a committee member of IEEE SCC20 Diagnostics and Maintenance subcommittee since 1991. Dr. Ghoshal holds an MBA from Kelley School of Business, Bloomington, both Ph.D. and M.S. degrees in Biomedical Engineering from University of Connecticut, and a B. Tech. in Electrical Engineering from Indian Institute of Technology (IIT), Kharagpur, India.

Ms. Valerie Yerdon is a Scientist at Aptima, Inc. with focus on human-centered engineering initiatives. Ms. Yerdon has worked on the Sailor 2025 Ready Relevant Learning (RRL) initiative at the Naval Air Warfare Center Training Systems Division (NAWC-TSD) as a Rating Coordinator, Training Analyst, and Instructional Systems Designer. She has been developing competency models and knowledge architectures for learner-centric systems, team dynamics, and human-AI teaming. She is a Ph.D. candidate in Human Factors at Capitol Technology University, and is completing her Instructional Systems Designer Graduate Certificate at the University of Central Florida (UCF). Ms. Yerdon also holds a M.S. in Modeling and Simulation, a Graduate Certificate in Cognitive Science, and a B.S. in Psychology from UCF.

Dr. Maartje Hidalgo is a Research Engineer at Aptima, Inc., who has industry, academia, and government experience in human-centered engineering, front-end analyses, course analyses/design, and adaptive learning systems. Her research on human-agent teaming, standardized assessments, and training evaluations has been published and presented at international conferences. Dr. Hidalgo holds a Ph.D. in Industrial Engineering and Management Systems with a concentration in Human Factors and Ergonomics from the University of Central Florida. She also holds a B.S. in Psychology and a M.S. in Neuropsychology from Leiden University, The Netherlands.

Dr. Jeffrey Beaubien is the Chief Behavioral Scientist at Aptima, Inc. For the past 20 years, his work has focused on training and assessing leadership, teamwork, and decision-making skills. His research has been sponsored by the U.S. Navy, the U.S. Army, the U.S. Air Force, and the Telemedicine and Advanced Technologies Research Center, among others. Dr. Beaubien holds a Ph.D. in Industrial and Organizational Psychology from George Mason University, a M.A. in Industrial and Organizational Psychology from the University of New Haven, and a B.A. in Psychology from the University of Rhode Island.

Dr. Jason H. Wong is a Cognitive Scientist with the Naval Information Warfare Center Pacific, where he supports the Office of Naval Research in improving warfighter decision making and human-technology teams. His government career has included measuring cognitive performance for training and simulation at the Naval Undersea Warfare Center and fostering international scientific collaboration at the Office of Naval Research Global in Tokyo, Japan. Dr. Wong holds a Ph.D. in Human Factors and Applied Cognition from George Mason University and a B.S. in Psychology from the University of Illinois Urbana-Champaign.

Using Digital Twins in Maintenance Operations and Training

Deepak Haste, M.S.[1], Sudipto Ghosal, Ph.D.[1], Valarie Yerdon, M.S.[2], Maartje Hidalgo, Ph.D.[2], Jeffrey M. Beaubien, Ph.D.[2], Jason H. Wong, Ph.D.[3]

Qualtech Systems, Inc.[1], Aptima, Inc.[2], Naval Information Warfare Center Pacific[3]

Rocky Hill, CT[1], Woburn, MA[2], San Diego, CA[3]

deepak@teamqsi.com, sudipto@teamqsi.com, vyerdon@aptima.com, mhidalgo@aptima.com jbeaubien@aptima.com, jason.h.wong.civ@us.navy.mil

INTRODUCTION

Maintainers frequently lack access to modern training capabilities that provide up-to-date knowledge and expertise to successfully maintain systems. This leads to system failures, inefficient maintenance and repair, and reduced operational availability. This in turn leads in increased costs to deliver and sustain readiness. There is a need to centralize management of maintenance training devices and simulations, so training organizations do not have to seek information from various program offices to obtain updated technical information and training equipment. This would ensure that maintainers will always be taught up-to-date technical information to better prepare them for field maintenance operations.

A model-driven dynamic reasoner can guide maintenance technicians in isolating and correcting the root cause(s) of anomalies, and Augmented Reality/Virtual Reality (AR/VR) can support on the job training, virtual co-location, and just-in-time and just-in-place technical assistance. An agile *DevOps*-like "*TrainOps*" capability where the maintenance performed on the field with the aid of AR is an integral part of system life-cycle, and where the performance and experience gained on the field is channeled back into classroom-based training, will improve DoD's maintenance training effectiveness. In this arrangement, maintenance is driven by intelligent model-based diagnostic/prognostic reasoners, and performance and experience gained on the field is channeled into the learning experience to improve Maintenance Operations and Training. Thus, a complete life-cycle view incorporates maintenance inputs of field data and model-based approaches to predict asset readiness (Martin et al., 2018).

Background and Current State of Maintenance Training

Typically, technicians undergo maintenance training prior to deploying for field maintenance operations, and this transition can take several months. During this gap, the knowledge and skills gained during training often decay. Furthermore, the technology underlying their occupational specialties may also have changed, thereby further diminishing the knowledge and skills that the technicians have at their disposal in the field. The problem is further compounded by the lack of availability of skilled technicians due to personnel attrition. Hence, there is an acute need to provide the right learning, in the right format, at the right time. In order to achieve this, an agile process needs to be in place that takes the field operational data/experiences and rapidly transforms them into instructional content that can easily be re-purposed for maintenance training accessible via multiple devices such as laptops, desktops, mobile devices, and AR/VR goggles.

The *TrainOps* process is powered by the "digital twin" – a functional failure model system that captures the causes, effects, and dependencies among various subsystems and their components – and provides maintenance technicians with fault simulation scenarios using immersive AR/VR capabilities so that they can practice fault diagnosis, isolation, and repair/replace procedures on their devices, thereby turning downtime into learning time. As part of the digital thread, the digital twin employs efficient reasoners to run real-time diagnostics and prognostics for remote health monitoring, and generate interactive guided troubleshooting trees that maintainers can use to access uncommonly-accessed reference aid material while in the field. Since this process is model-driven, it enables an efficient and continuous improvement process encompassing field deployment, efficient troubleshooting sessions driven by intelligent reasoners, and model refinements through maintenance data capture and technician feedback. The training content is presented in a sharable on-demand web-based format, thereby supporting anytime/anywhere distance learning. A single-source-of-truth model used across multiple life-cycle development of the system - Concept; Design; Development; Testing; Deployment; Maintenance and Operations – significantly reduces the cost of technology preparation and deployment and also leads to seamless transition between training and actual usage of the technology in the field. In the remaining sections of this paper, we demonstrate how to leverage digital twins, combined with learning sciences and human performance measurement, to support training and performance assessment for learners across the continuum of expertise.

THE *TRAINOPS* PROCESS

Concept of Operations (CONOPS)

Digital twins are powerful tools for accurately capturing the functional behavior of a system for the purpose of field maintenance and troubleshooting. However, leveraging these same digital twins for learning requires the inclusion of one or more capabilities, such as:

- Ability to organically generate instructional material for distributing this training content over time to maximize retention (Donovan & Radosevich, 1999)
- Ability to interlace instructional content to promote deep understanding (Schmidt & Bjork, 1992)
- Ability to highlight subtle background clues that novice learners should pay attention to while performing the task (Hagemann et al., 2006)
- Ability to provide process-based feedback so that learners can benefit from their mistakes (Lefroy et al., 2015)
- Ability to adapt the instructional content based on the learner's unique profile of individual strengths and weaknesses (Ericsson et al., 1993)
- Ability to track the learners' performance across multiple learning events, thereby supporting long-term learning (Raybourn et al., 2017)

The *TrainOps* process aims to capture these key tenets into an integrated workflow for digital twin-enabled training. The key steps of this process are shown below (see Figure 1):

1. The process begins with building a functional failure cause-effect dependency digital-twin (model) of the interconnected components, for maintenance troubleshooting and sustainment.
2. The next step involves generating the AR content to render the guided troubleshooting steps by the digital-twin reasoner. We used 3D modeling frameworks such as Web-XR (WebXR, 2022) to facilitate web-based authoring. Additional multimedia content, or external links to a multimedia repository, can be rendered in troubleshooting procedures.
3. The digital twin is then deployed inside a centralized "repository" that can be accessed by the training facility for

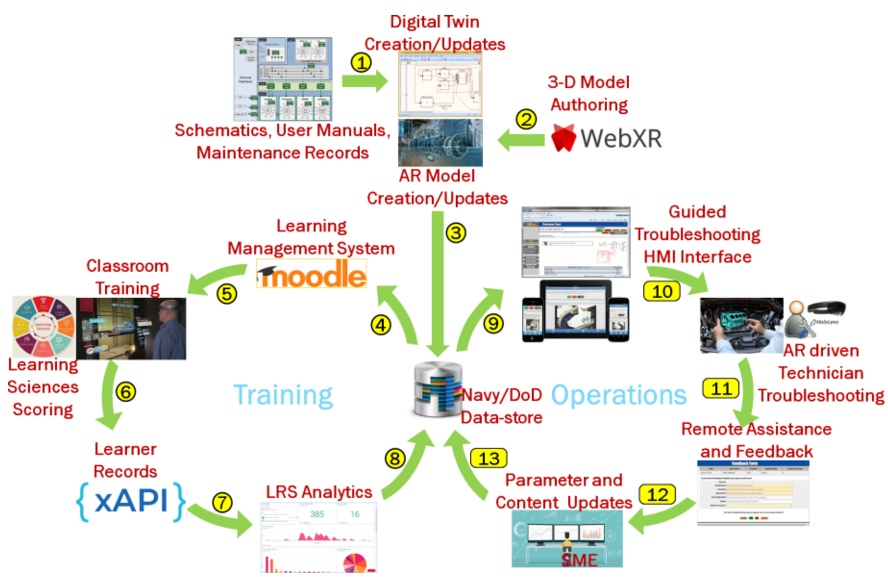

Figure 1: ***Train-Ops (Training + Operations)*** **Solution Architecture**

in-house technician training, as well as for generating guided troubleshooting strategies for field operations.

4. **Classroom Training Cycle**: E-learning systems such as Moodle (Moodle, 2022) are used for knowledge management and testing.
5. Instructors use Moodle to create training content. Based on the learner's experience level, Moodle will assign appropriate guided training sessions. The scores from those sessions are logged in Moodle.
6. The digital twin expert reasoners are used for maintenance training in the classroom by seeding of simulated failures, which will generate concomitant condition-indicators (alarms/symptoms) that can be troubleshot. The training session utilizes learning sciences capabilities such as hints to assess the cognitive grasp of the learners, and tailor the training content according to the learner's skills.
7. The expert reasoners have a built-in audit log capability that records all the steps taken by the learner. These recorded actions are transformed into xAPI (Rustici Software LLC, 2022) formatted learner records and logged to a Learning Record Store (LRS).
8. Analytics from the LRS store are mined to assess the problem areas in the training regimens, assess the training return on investment (ROI), etc. This helps the training instructors assess the learners' cognitive skills, and revise the training regimen to maximize the impact of the training.
9. **Field Operations Cycle**: The exported digital twin is also deployed for field troubleshooting.
10. Field technicians are provided with an IETM-like guided troubleshooting interface.
11. During troubleshooting, the technician may experience various "nuances" of the guided instructions that need to be organically incorporated into the digital twin. These can be captured in the Feedback form, with input-fields to enter comments, and upload pictures/videos, etc.
12. The technician responses and media uploads are relayed to the Subject Matter Expert (SME), who is notified of the new feedback and "pending" uploads. The SME then reviews the changes within the context of the existing model.
13. **Classroom Training Cycle (Iterative)**: Once the SME is satisfied with the results, the model is updated and re-deployed within the training facility for in-house training, and subsequent field deployment. This cycle continues iteratively. Thus, updates from the field continuously improve the maintenance training within the schoolhouse.

Digital Twins for Maintenance Training

The digital twin, which is the underlying driver of the *TrainOps* process, captures the functional failure cause-effect dependencies between various subsystems. First-order cause-effect dependencies are modeled between system faults (functional or physical), and their resulting manifestations and test points (monitoring mechanisms) that detect functional changes arising from these faults. Higher-order dependencies can then be inferred from the first-order dependencies (Mathur et al., 1998). This multi-signal modeling methodology (Deb et al., 1995) can be conceptualized as a dependency graph, where the structural components are modeled as nodes and the directional links denote dependencies based on structural adjacency. Figure 2 shows the digital twin of the interconnected subsystems and lower-level multi-functional dependency model (structure and functional behavior).

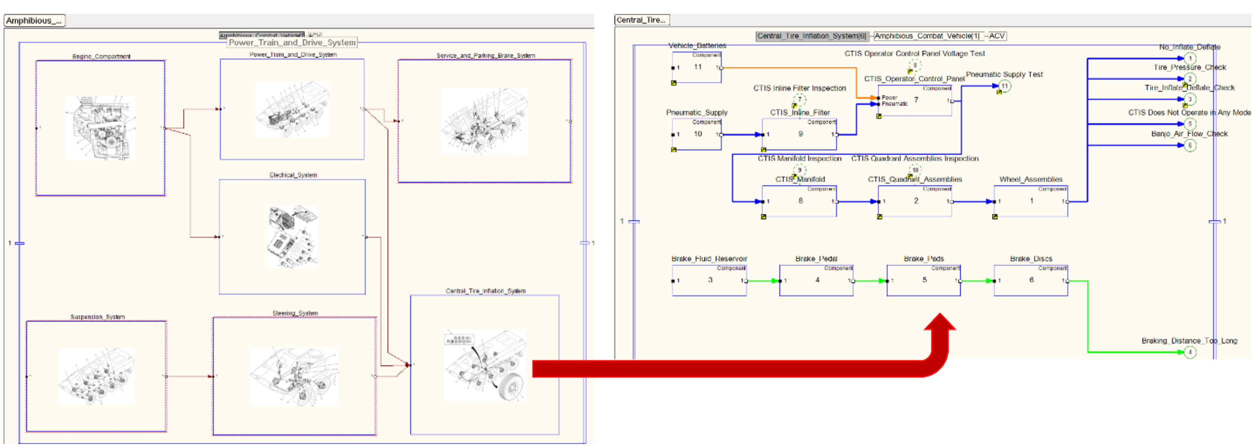

Figure 2: Example Cause-Effect Digital Twin

System functions are layered onto the nodes as signals. The interconnections of tests (on-board error codes, automated, visual indictors, manual troubleshooting) and faults (at different levels of system hierarchy) gleaned from the graphical model, coupled with repair-procedures, enable the modeler to generate a diagnostic strategy for the system. The underlying expert diagnostic reasoner executing the digital twin optimizes the test sequencing for diagnostic resolution of failure causes. Thus the “guided troubleshooting” strategy is the most optimal path to root cause isolation. The cause-effect dependencies can also be employed to generate simulations that target specific types of faults/failures for the maintainer to practice. This is key to assessing technicians on their guided troubleshooting knowledge and skills. This form of maintenance training is a step further than pure symptom-based troubleshooting procedures typically found in IETMs. The single-source-of-truth digital twin makes this possible by virtue of knowledge of system’s inherent cause-effect relationships. It captures the interconnected nature of the system which often results in a root-cause manifesting as multiple alarms, which need to be expertly resolved. Sometimes the digital twin diagnostic strategy can prioritize a test that is common across those symptoms since it might be cheaper to perform. Ultimately, the reasoner picks the most efficient tests that cover one or more component(s) across multiple symptoms. This leads to an improved and more efficient troubleshooting strategy. This key ability of the digital twin enables the instructor to create challenging training scenarios where the learners are tasked to disambiguate between multiple error-codes/alarms and select an optimal test sequence for root cause isolation.

Moodle Training Content

The digital twin’s expert reasoner exports the diagnostic tree and guided troubleshooting strategies in an XML format, which can then be imported into the Moodle LMS using a plugin that is freely available from the first author. Figure 3 shows the imported troubleshooting trees for a system that the maintainers can browse in Moodle.

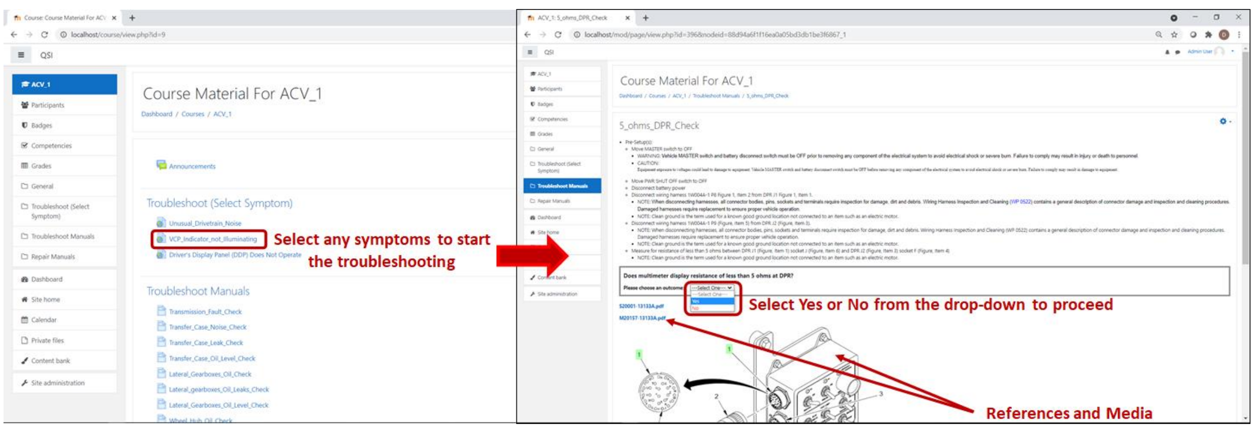

Figure 3: Example Diagnostic Tree Imported into Moodle

Adaptive Training In Moodle using Scaffolding Instructional Approach

The scaffolding instructional approach (Wood et al., 1976) provides learners with support that is tailored to their level of proficiency, displaying tiered hints that are dynamically computed through a scoring system based on the learner’s performance. This approach utilizes the expert reasoner-based Guided Troubleshooting to *proceduralize* the troubleshooting task for maintenance training (learning-by-doing). The key hint tiers of the scaffolding approach are:

- *Metacognitive* hints (Fiorella et al., 2012) help the learners critically examine their thought processes (such as clarifying their goals) **(Expert level)**. The digital twin expert reasoner shows the history of test-PASS/test-FAIL sequence and the history of likely causes in the prior steps to help the learner understand the consequence of those tests on the current diagnostic state.
- *Teaching* hints (Vanlehn, 2006) provide information to the learner and describe the implications of that data, but don’t explicitly tell them the correct answer **(Experienced level)**. The digital twin expert reasoner displays additional hints that show the current diagnostic status and most likely causes based on probabilities computed through Bayesian updates by the expert reasoner, the test coverage for each of the suspects, and the test costs and times. This knowledge will help the learner to learn the effectiveness of the tests given the current diagnostic status of the subsystem.

- *Pointing* hints (Vanlehn, 2006) direct the learners to informational cues such as Theory of Operations, associated multimedia and detailed documentation etc. associated with the correct test that might help them to answer the question or resolve the problem **(Intermediate level)**. The expert reasoner progressively provides more visual cues in its troubleshooting screen;
- *Bottom-out* hints tell the learners the correct answer if they are unable to use the cues, and make several mistakes in a row **(Novice level)**. The expert reasoner will display the default screen with all clues that is used for operational maintenance.

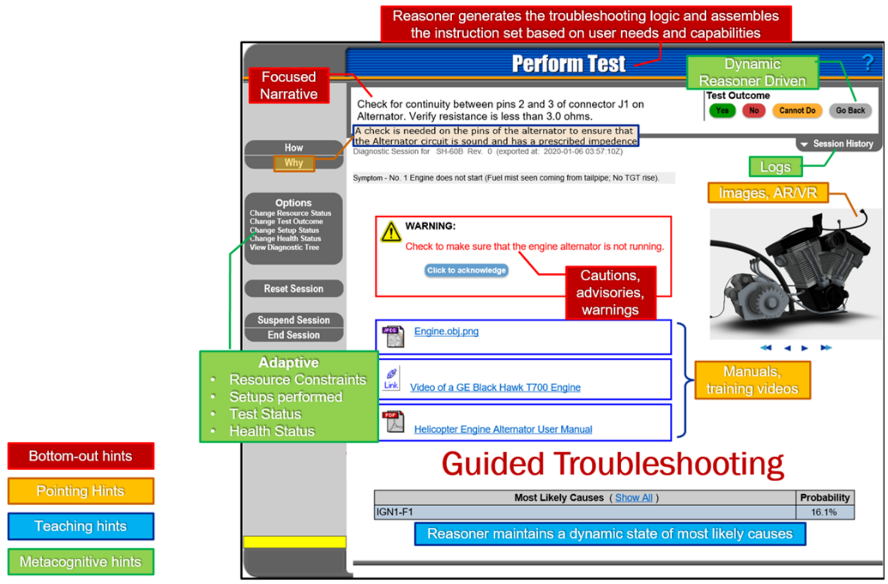

Figure 4: Scaffolding method in a Guided Troubleshooting environment

The Scaffolding approach is applied for the training loop of the *TrainOps* process. For the simulated faults(s), the digital-twin expert reasoner creates an optimal sequence of tests that leads to the root cause isolation of the fault. During Operations, the technician is presented with the optimal sequence in the guided troubleshooting tool.

During Training, on the other hand, the test sequence is hidden, and the learner is tasked to select the correct sequence. The guided training module asks the learner, from a list of possible options, the best test to perform. After the user selects an option, the training module scores the performance of the learner for each troubleshooting step and computes a cumulative total score based on: (a) The deviation of the maintainer's response from the optimal response (determined by the digital twin's expert reasoner and calculated based on the Information Gain of the tests i.e. how useful this test is for the most optimal root-cause-isolation); and (b) the number of consecutive incorrect answers leading to scaffolding hints. The scores can be averaged across multiple sessions. Based on the score, the proficiency of the learner is assessed, and appropriate hints are presented. The guided training interface shows appropriate hints of the next best test based on the skill levels and masks the remainder of the hints. Figure 4 shows different tiers of information extracted from the Guided Troubleshooting interface, based on expertise level. Figure 5 shows a training session for one of the simulated faults generated by the digital twin and

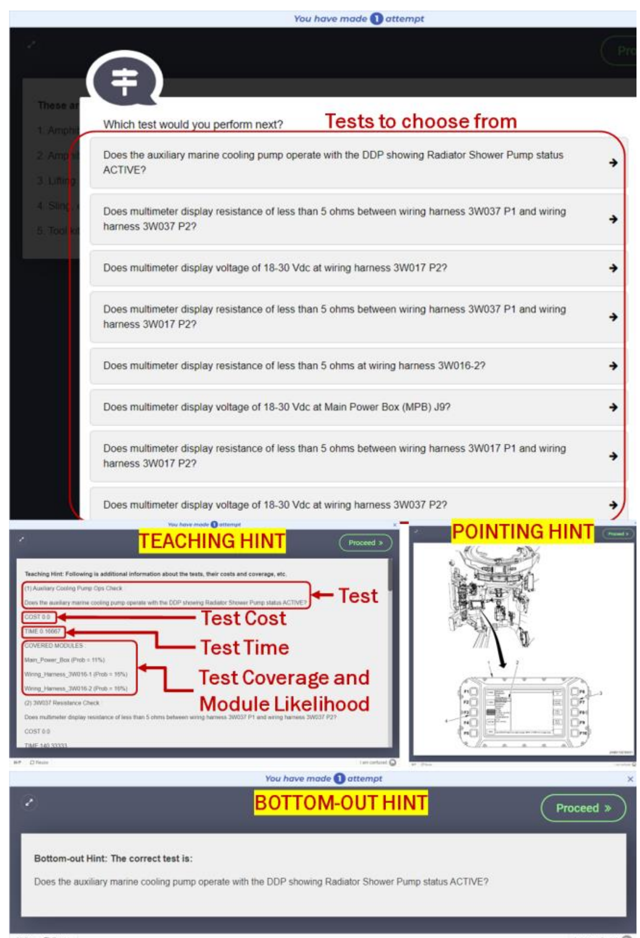

Figure 5: Scaffolding in Moodle Branching Scenario

converted to Moodle Branching Scenario.

In the figure above, the learner is asked to choose from different answers. The "Teaching Hints" provide clues about most likely causes, their likelihoods, and test coverage for those likely causes. "Pointing Hints" provide media, references and visual clues associated with the correct test, while "Bottom-out Hints" reveal the correct answer.

The Moodle Branching Scenario is adaptive with respect to learner's choice of correct or incorrect responses by providing instantaneous hints. An initial incorrect response leads to displaying of the "Teaching" hints. Two incorrect responses lead to "Pointing", while three-in-a-row incorrect responses lead to "Bottom-out" hints. Each instance the learners provide an incorrect response, their scores are penalized, thereby dissuading them from making too many mistakes and providing them with adaptive hints. The learner scores are relayed to the Moodle dashboard, and can be analyzed by the instructor or supervisor to tailor subsequent training sessions so that they address gaps in the learner's understanding.

Use of Experience API (xAPI) and Learning Locker to Find Training Gaps

After each guided training session, the expert reasoner logs the learner's performance (see Figure 6) such as time to complete each step, number of UNDOs and CANNOT-DOs, etc. using Experience API (xAPI) learner records (Gallagher et al., 2017) which are archived in a centralized Learner Record Store (LRS).

xAPI is lightweight communications protocol that records learning experiences and actions in a uniform manner and which can be used by any learning platform. An LRS is a data store for receiving, storing and returning xAPI statements. The objective is to apply the science of adaptive tutoring by collecting xAPI learner records across a population of learners, storing those records in the centralized LRS, and then mining the data to customize the training curriculum. If certain parts of the instruction are very difficult (as determined by a large percentage of failed tasks e.g., 70% of learners were not able to complete Step A in a learning course, etc.), then there may be a need to provide the learners with more explicit (rather than indirect) prompts. SMEs and instructors can centrally perform these type of digital twin updates to improve learning. Figure 6 shows the mapping of the maintainer task report fields generated by a training session to relevant xAPI statements.

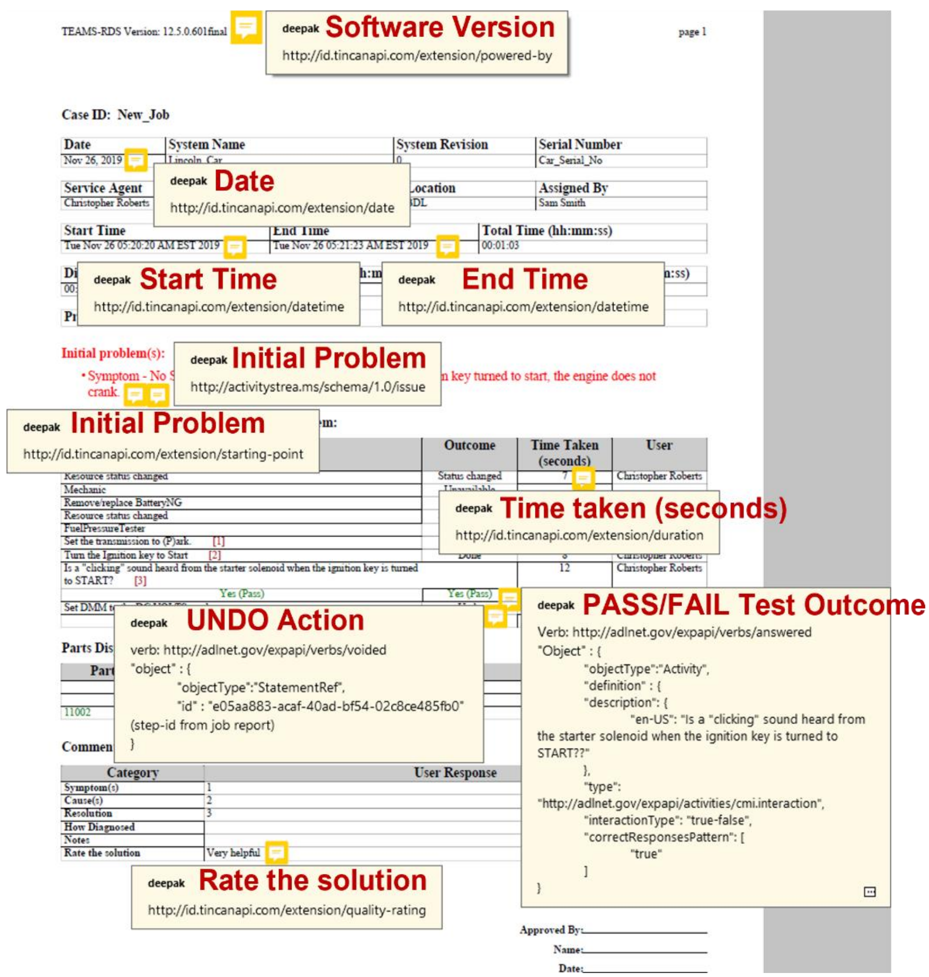

Figure 6: Mapping of Task Report and xAPI Statements

The mapping from the guided training session logs to xAPI statements is user-configurable. This enables dynamic mapping from maintainer session logs to various xAPI profiles. The example above shows the mapping to the government's ADL xAPI profile. xAPI analytics from the LRS such as learning activity, role readiness, training compliance, etc. are used for customizing training tutorials in Moodle. After reviewing the LRS analytics page for steps (xAPI "Verbs") that take the most time (or Top Activities) to complete, the instructor can customize the training accordingly.

Figure 7 shows how aggregate xAPI learner records (stored in the LRS) can be visualized to identify trends. The graphs provide insights about learning activities, role readiness, training compliance, etc. and help instructors to locate problems that may benefit from curriculum re-design. For instance, Figure 7 shows the most time-consuming activities across all learners while Figure 8 shows top engaged Actors and Individual Performance

Figure 7: LRS Analytics for Task Duration across all Trainees

Figure 8 here shows metrics such as number of "do-overs" (UNDOs) and "skipped-steps" (CANNOT-DOs). Other LRS analytics include: (1) training hot spots; (2) problem areas that are experienced by multiple learners; (3) individual learner performance data such as diagnostic time, time at individual step, etc. Using these metrics, the overall organizational metrics can also be generated, such as:

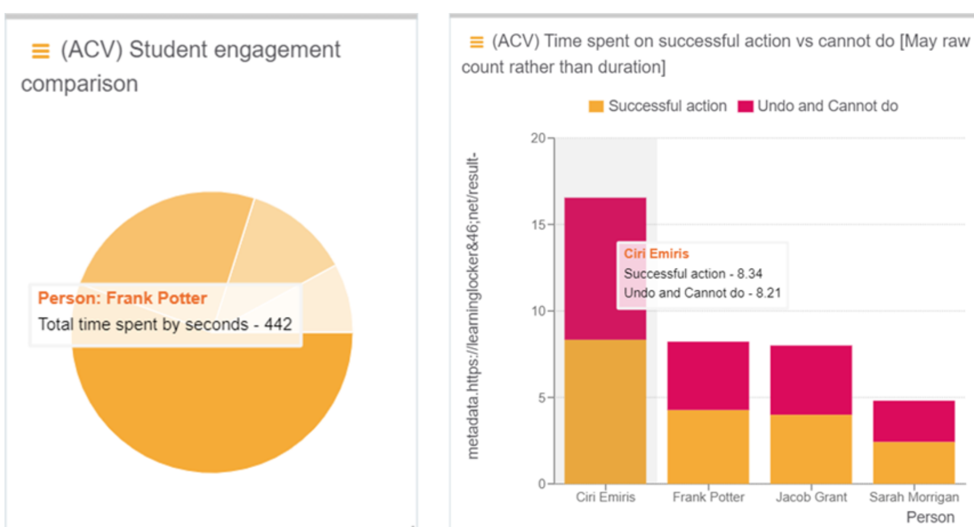

Figure 8: LRS Analytics for Individual Trainee Performance

- Cost-benefit-analysis of training investment;
- Troubleshooting time vs. diagnostic accuracy;
- Field vs classroom performance;
- Performance based on KSAs (Knowledge, Skills and Abilities) – experts vs novices.

METHOD AND RESULTS (FRONT END ANALYSIS)

The Office of Naval Research (ONR) and MARCORSYSCOM (Marine Corps Systems Command) Manpower, Personnel, Training, and Education (MPTE) sponsored a Front-End Analysis (FEA) survey with various U.S. Marine Corps (USMC) schoolhouses to assess the availability of different kinds of legacy training data (e.g., quizzes, instructor handouts, checklists, animations, etc.) which could potentially be repurposed for use in digital-twin driven immersive training environments. For example, some legacy data (e.g., multiple choice quizzes and tests) could potentially be hosted in Moodle as a digital quiz. By comparison, other legacy data (e.g., maintenance checklists, technical flowcharts) could be potentially integrated into AR- or VR-based training scenarios, while other data (e.g., technical manuals) could be incorporated into using an expert reasoner-based adaptive training tool to help learners choose the best answer. Prior to conducting the FEA, however, MARCORSYSCOM didn't have an accurate assessment of what types of legacy content currently exist, much of what would be required to convert this content into digital form.

The FEA survey – which was designed to identify the most common types of legacy training data at USMC schoolhouses – consisted of four parts. Part I identified courses and equipment in the schoolhouse. Part II surveyed their training requirements. Part III identified metrics of interest for student learning and successful maintenance activities. Part IV sought out legacy data types that can and should be converted to immersive maintenance training content. The FEA survey was sent out to 3 separate schools: (1) Assault Amphibian School (AAS) with the PM AAA (Advanced Amphibious Assault), (2) Marine Corps Combat Service Support School (MCCSSS) and (3) Marine Corps Engineering School (MCES). These schoolhouses identified a total of six courses as eligible for conversion into immersive maintenance training content.

According to the survey respondents, the most commonly available materials are common task lists, technical/OEM manuals, maintenance checklists, instructional videos, quizzes, lesson plans, and student handouts/outlines. The majority of these materials are saved in Microsoft Office or Adobe Acrobat formats. A host of student metrics are currently in widespread use; however, the most prevalent is learner/class throughput. The respondents indicated a desire to reuse these materials in various ways: static learning exercises within Moodle; as part of interactive AR/VR-based training scenarios, and; to help provide learner with teaching and/or pointing hints as part of a web-based adaptive troubleshooting tool.

METHOD AND RESULTS (FEASIBILITY TEST)

At the conclusion of the FEA survey, the research team developed a "maintenance training content creation pipeline/process" to show MARCORSYSCOM personnel how digitized legacy training materials could be quickly and cost-effectively captured in a digital twin which in turn would drive Moodle-based AR/VR environments, and web-based troubleshooting tools. To this end, we designed and conducted a feasibility exercise for the *TrainOps* process with the Program Of Instruction (POI) material supplied by PM AAA on their New Equipment Trainer (NET) course for the Amphibious Combat Vehicle (ACV). The training material used for capturing into digital twin included:

- Training Presentation (TP) for each lesson in the form of PowerPoint slides;
- Lesson Plans (LP) in the form of Word documents which are instructor's aids for the Training Presentation;
- Student Handout (SH) in the form of Word documents which are Student lessons for the Training Presentation;
- Operator Manuals and Maintenance Manuals.

Finally, we demonstrated this maintenance training creation process using open-source and low-cost software tools such as Moodle and Learning Locker during the course of a 2.5 day workshop. During the workshop, the USMC visitors logged into a hosted Moodle instance and performed the following activities:

- Moodle Lesson on theory of operations of the selected subsystem
- Practicing Diagnostic Trees in Moodle for set of symptoms in the subsystem
- Performing Guided Training with Scaffolding and Scoring for one or two symptoms. Scores were published in Moodle and user job report was transmitted to Learning Locker

At the end of the workshop, we conducted a brief usability test with a sample of USMC instructors and training program managers to solicit their feedback about the maintenance training process. In the following sections, we describe the usability test procedure, results, and findings.

Participants

The sample included six representatives from the USMC's AAS at Camp Pendleton, CA, three AAS training Program Managers (civilian contractors), and three ACV maintenance Subject Matter Experts (SMEs)/Instructors. Two SMEs were active-duty senior enlisted Marines and the third was a civilian contractor with recent active-duty USMC experience.

Measures

The primary criterion measure was the Post Study System Usability Questionnaire (PSSUQ; Lewis, 1992). The PSSUQ is a 16-item usability scale that measures perceived satisfaction with a process or system with anchors that range from "Strongly Agree" (1) to "Strongly Disagree" (7), with a high internal consistency reliability (.97) (Lewis, 1992). The scale is frequently used in Human-Computer Interaction (HCI) research. The PSSUQ has been translated into numerous languages, including European Portuguese (Rosa et al., 2015), Arabic (Al-Tahat, 2021), and Greek (Katsanos et al, 2021). The PSSUQ questions are organized into three categories. Questions 1-6 focus on the system/process quality (in our case, the content creation process). Example questions include "It was simple to use this system" and "I was able to complete the tasks and scenarios quickly using this system." Questions 7-12 focus on the quality of information provided by the system or process. Example questions include "Whenever I made a mistake using the system, I could recover easily and quickly" and "It was easy to find the information I needed." Finally, Questions 13-16 focus on the interface quality of the system or process. Example questions include "I liked using the interface of this system" and "This system has all the functions and capabilities I expect it to have."

The research team included five additional questions that were custom developed for the current study. The questions compared the new content creation pipeline to the USMC's current ways of developing electronic instructional content. The five questions were as follows:

- "The amount of time to complete was worth the product"
- "The number of steps to complete made sense for the product"
- "The number of people required to complete was not over expectations for the product"
- "The level of technical knowledge to complete was appropriate"
- "This system and process makes sense for automation"

These were worded in the same direction as the PSSUQ, with smaller numbers indicating more favorable responses.

Procedure

The usability test occurred during the course of a 2.5 day workshop, where the research team described five different sections of the Maintenance Training process, including the content creation. Two of the sections - Powerplant Theory of Operations (Course Content), Maintenance Training Exercise (Troubleshooting Content) – involved the participants listening to the process being described, observing the procedure being performed by a member of the research team, and then practice performing the procedure on their own. Three other sections - Course Content Creation, Troubleshooting Content Creation, Student Metrics Visualization in Learning Locker – involved the participants only listening to the process being described and observing it being performed by the research team. This was due to time and other resource constraints. After completing each section of the workshop, the participants immediately completed the PSSUQ to provide feedback on that section. This process continued iteratively throughout the multi-day workshop.

Results

Survey means and standard deviations (SDs) are summarized in Table 1. Columns 2-4 contain the PSSUQ scale scores of System Quality (mean and SDs of PSSUQ items 1-6), Information Quality (mean and SDs of items 7-12), and Interface Quality (mean and SDs of PSSUQ items 13-16), respectively. Finally, Columns 5-9 contain the five survey items that were custom developed for the current study: amount of time required, number of steps required, number of people required, amount of technical knowledge required, and the extent to which the process was suitable for automation. Because these were single-item ratings, no means or standard deviations were computed. As a reminder to the reader, all of the items were scored such that smaller values indicate more favorable responses.

Table 1. Feasibility Test Results

	PSSUQ Means and Standard Deviations			Custom-Developed Survey Questions				
Source	**System Quality**	**Information Quality**	**Interface Quality**	**Time Required**	**Steps Required**	**People Required**	**Technical Knowledge**	**Suitable for Automation**
Course Content Creation	1.81 (0.47)	2.25 (0.73)	2.17 (0.99)	3.67	2.33	2.17	1.67	1.50
Maintenance Training Exercise	1.72 (0.57)	2.28 (1.09)	2.28 (1.18)	2.83	2.17	1.67	2.00	1.83
Powerplant Theory	1.70 (0.79)	1.63 (0.67)	1.92 (0.67)	2.20	2.00	2.00	1.75	1.50
Student Metrics (Learning Locker)	2.44 (0.91)	2.64 (1.17)	2.17 (0.71)	2.67	2.33	2.33	2.50	1.67
Troubleshooting	2.13 (0.82)	2.23 (0.50)	2.40 (1.06)	3.40	2.40	2.40	3.20	1.80

All of the scores were below the scale midpoint of 4.0, which indicates the participants perceived the new maintenance training process favorably. An analysis of the PSSUQ scale scores indicated that participants thought the system/process was well designed, that it provided useful information, and that the system interfaces made sense. Finally, the custom-developed questions showed that the new maintenance training content creation process was perceived as favorable regarding the amount of time required, number of steps required, the number of people required, the amount of technical knowledge required, and that it is suitable for automation. It is important to note that this workshop did not include an opportunity for student to practice, due to time and other resource constraints. Furthermore, majority of the content creation steps mentioned above will be semi-automated in the future.

During the course of this training workshop, while interacting with the USMC instructors, a notable pattern was observed with respect to the use of traditional classroom instructional techniques with Word Lesson Plans, PowerPoint Presentation materials, etc. These traditional classroom training techniques, with student handouts and presentations, coupled with stove-piped legacy material, result in training being out-of-date with system operations. As more complex systems are fielded, not only will maintaining them be more challenging, but keeping training up to date will be even more difficult. Thereby, a key takeaway coming out of this exercise was to transition the classroom instructional format towards an enterprise-level distance learning platform, powered by the single-source-of-truth digital-twin, and immersive maintenance training techniques. This entails transitioning towards Moodle or equivalent learning management platforms with adaptive training capabilities such as the Branching-Scenario, that is auto-generated from the digital twin. Furthermore, learner standards such as xAPI (eXperience Application Programming Interface) would lead towards a more objective and automated metrics measurement. Learner activities are logged as xAPI statements into a Learning Record Store (LRS) which has built-in dashboards and analytics to display metrics related to the learner engagement and performance during their maintenance training exercises. These dashboards will provide instructors with insights into common problem areas and learning gaps which can then be used for more personalized course generation suited to individual KSAs (Knowledge, Skills and Abilities). The ultimate goal is to empower the schoolhouse to take ownership of their maintenance training activities so that they don't have to rely on outside agencies or contractors to develop and administer the training content.

CONCLUSIONS

In this paper, we highlighted the *TrainOps* process using a model-based intelligent guided-training and guided-troubleshooting framework that combines interactive media, references, modern web-based AR/VR technologies, etc.

from the operational context, and seamlessly bridges the gap between training environments and field maintenance. The training approaches uses a model (digital twin) that is the “single source of truth” for operational as well as training scenarios. There is a methodical way of incorporating content, whether from operational or training context, and is able to systematically categorize and incorporate into the digital model. The Moodle E-learning system serves as a central repository of distant learning where the instructor can semi-automatically generate training content from the digital twin, view the learners’ progress, gaps in their learning and general training effectiveness metrics, and issue course curriculum remediations. Using a model-based approach for training enables creation of comprehensive simulations/scenarios since they are based on inherent system cause-effect relationships and hence more accurately represent actual system behavior and resemble real-world troubleshooting experiences. Subsequently, Scaffolding hints during the training motivate maintainers with their cognitive thinking. Ultimately, a reasoner-guided immersive training can help achieve the DoD’s strategic goal of training technicians to be efficient maintainers.

The *TrainOps* process was demonstrated to a sample of USMC training managers and domain SMEs/instructors from the AAS at Camp Pendleton, CA during a 2.5 day workshop. During the workshop, five different components of the *TrainOps* process were demonstrated. Two of the components involved the research team describing the process, demonstrating the process, and the participants practicing the task on their own. The remaining three components did not include the practice component because of time and other resource constraints.

All five components were assessed using a validated usability scale. Five additional survey questions were specifically developed for this study that compared the *TrainOps* process to current USMC maintenance training processes.

A key finding emerging from the analysis of the questionnaire responses was that all mean scores were below the scale midpoint of 4.0, indicating that the participants viewed the new *TrainOps* process favorably. Overall, the results suggest that the sample of USMC instructors and training managers perceived the *TrainOps* process favorably, both in an absolute sense and compared to current USMC maintenance training creation and dissemination.

ACKNOWLEDGEMENTS

We would like to thank Dr. Peter Squire from ONR for supporting and funding the research, and Dr. Jason Wong for his guidance. We also would like to thank Mr. James Aurilio, PM AAA Manpower, Personnel, and Training Lead for providing the ACV training material, and for heading the USMC contingent to participate in the *TrainOps* workshop. And finally, we would like to thank CWO Micah Soboleski at MCCSSS and GySgt Patrick C. Wantuch at the MCES for providing access to their training material and facilities, and for providing inputs to the FEAs.

REFERENCES

Al-Tahat, K. S. (2021). Arabic translation, cultural adaptation and psychometric validation of the Post-Study System Usability Questionnaire (PSSUQ). *International Journal of Human–Computer Interaction*, 37(19), 1815-1822.

Deb, S., Pattipati, K. R., Raghavan, V., Shakeri, M., & Shrestha, R. (1995). Multi-signal flow graphs: A novel approach for system testability analysis and fault diagnosis. *IEEE Aerospace and Electronic Systems Magazine*, 361–373.

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology, 84*(5), 795-805

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406.

Fiorella, L. & Vogel-Walcutt, J. J., & Fiore, S. (2012), Differential impact of two types of metacognitive prompting provided during simulation-based training. *Computers in Human Behavior*, *28* (2). 696-702.

Gallagher, P., Folsom-Kovarik, J., Schatz, S., Barr, A., & Turkaly, S. (2017). *Total Learning Architecture development: A design-based research approach*. *17*, 1 - 12.

Hagemann, N., Strauss, B., & Cañal-Bruland, R. (2006). Training perceptual skill by orienting visual attention. *Journal of sport and exercise psychology, 28*(2), 143-158.

Katsanos, C., Tselios, N., & Liapis, A. (2021, November). PSSUQ-GR: A first step towards standardization of the

Post-Study System Usability Questionnaire in Greek. In *CHI Greece 2021: 1st International Conference of the ACM Greek SIGCHI Chapter* (pp. 1-6).

Lefroy, J., Watling, C., Teunissen, P. W., & Brand, P. (2015). Guidelines: The do's, don'ts and don't knows of feedback for clinical education. *Perspectives on Medical Education, 4*(6), 284-299.

Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7(1), 57-78.

Lewis, J. R. (1992, October). Psychometric evaluation of the Post-Study System Usability Questionnaire: The PSSUQ. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 36, No. 16, pp. 1259-1260)*. Los Angeles, CA: Sage.

Martin, B., Yardley, R. J., Pardue, P., Tannehill, B., Westerman, E., & Duke, J. (2018). *An approach to life-cycle management of shipboard equipment*. RAND Corporation. | https://www.rand.org/pubs/research_reports/RR2510.html

Mathur, A., Deb, S., & Pattipati, K. R. (1998). Modeling and real-time diagnostics in TEAMS-RT. *Proceedings of the 1998 American Control Conference*, *3*, 1610–1614.

Moodle. (2022). *Moodle - Open-source learning platform | Moodle.org*. https://moodle.org/

Orlansky, J., String, J. (1981). Cost-Effectiveness of Maintenance Simulators For Military Training. *Prepared for Office of the Under Secretary of Defense for Research and Engineering.* Institute For Defense Analysis.

Raybourn, E. M., Schatz, S., Vogel-Walcutt, J., & Vierling, K. (2017). At the tipping point: Learning science and technology as key strategic enablers for the future of defense and security. *Proceedings of the 2017 Interservice/Industry Training, Simulation, and Education Conference, 1 - 12.*

Rosa, A. F., Martins, A. I., Costa, V., Queirós, A., Silva, A., & Rocha, N. P. (2015, June). European Portuguese validation of the Post-Study System Usability Questionnaire (PSSUQ). In The *2015 10th Iberian Conference on Information Systems and Technologies (CISTI)* (pp. 1-5). IEEE.

Rustici Software LLC. (2022). *What is an LRS? Learn more about Learning Record Stores*. XAPI.Com. https://xapi.com/learning-record-store/

Rustici Software LLC. (2022). *What is xAPI aka the Experience API or Tin Can API*. XAPI.Com. https://xapi.com/overview/

Schlueter, Caroline & Fraenz, Christoph & Pinnow, Marlies & Friedrich, Patrick & Güntürkün, Onur & Genç, Erhan. (2018). The Structural and Functional Signature of Action Control. *Psychological Science*.

Schmidt, Richard & Bjork, Robert. (1992). New Conceptualizations of Practice: Common Principles in Three Paradigms Suggest New Concepts for Training. *Psychological Science*, *3*, 207-217.

Shute, V. (2008). Focus on formative feedback. *Review of Educational Research*, *78*, 153-189.

Vanlehn, Kurt. (2006). The Behavior of Tutoring Systems. *International Journal of Artificial Intelligence in Education, 16*. 227-265.

WebXR. (2022). *Immersive Web Developer Home*. https://immersiveweb.dev/

Wikipedia. (2022). DevOps. *In Wikipedia*. https://en.wikipedia.org/w/index.php?title=DevOps&oldid=1084496188

Wood, D., Bruner, J.S. & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*, 89-100.