

## Total Learning Architecture Development: A Design-Based Research Approach

**P. Shane Gallagher, Ph.D.**

Institute for Defense Analyses  
Alexandria, VA  
pgallagh@ida.org

**J.T. Folsom-Kovarik, Ph.D.**

Soar Technologies, Inc.  
Orlando, FL  
jeremiah@soartech.com

**Sae Schatz, Ph.D.**

ADL Initiative  
Alexandria, VA  
sae.schatz@adlnet.gov

**Avron Barr**

Institute for Defense Analyses  
Alexandria, VA  
abarr@ida.org

**Sarah Turkaly**

Institute for Defense Analyses  
Alexandria, VA  
sturkaly@ida.org

### ABSTRACT

Organizations that use learning technology to educate and train are facing a new set of interoperability problems. Many new products—including adaptive systems, intelligent digital tutors, real-time data analytics, and interactive e-books—offer dramatic learning benefits. However, these products primarily “stand alone” and work outside of typical browser-based delivery environments controlled by traditional learning management systems. Furthermore, the effectiveness of these “intelligent systems” often depends on their access to data generated by and stored in other systems.

The U.S. DoD Advanced Distributed Learning (ADL) Initiative is designing a framework of specifications, called the Total Learning Architecture (TLA), to ultimately enable “plug-and-play” interoperability of learning technologies. That is, the TLA will allow these new products to interoperate with each other, with other existing learning systems, and even human capital management technologies.

Because of the rapid rate of innovation in such distributed technologies, we adopted a multiyear design-based research approach. During the project’s first year, an initial set of specifications have been developed and evaluated for technical and functional adequacy using a multi-round Delphi approach with a panel of international participants ( $n = 54$ ). Also, in partnership with the U.S. Army JFK Special Warfare Center and School’s Special Warfare Education Group, we conducted a live prototype test and demonstration with Special Operations Soldiers ( $n = 73$ ). This yielded data on the nascent system’s functionality, performance, user experience, and learning potential. Analysis of these data will lead to recommendations, which in turn will inform the second cycle of TLA development process.

This paper summarizes the TLA concept, development process, first-year analysis efforts and outcomes, and lessons learned leading to design improvements for the second year of TLA development.

### ABOUT THE AUTHORS

**P. Shane Gallagher, Ph.D.** is employed by the Institute for Defense Analyses and is supporting the ADL Initiative as a learning scientist and education specialist. He received his Ph.D. in Instructional Technology from George Mason University and his Master’s in Educational Technology from the University of New Mexico.

**J.T. Folsom-Kovarik, Ph.D.** is a Lead Scientist with Soar Technology, Inc. in the Intelligent Training research group. His research focuses on making intelligent automation capable and robust with advanced planning, user modeling, and contextual interpretation approaches to make technology meet individuals’ needs.

**Sae Schatz, Ph.D.** serves as the Director of the ADL Initiative, a research and development program under the Deputy Assistant Secretary of Defense for Force Education and Training.

**Avron Barr** is an Adjunct Research Staff Member at the Director of Institute for Defense Analyses and is supporting the ADL Initiative.

**Sarah Turkaly** is an Adjunct Research Associate at the Institute for Defense Analyses supporting the ADL Initiative in the evaluation and assessment of the TLA.

## Total Learning Architecture Development: A Design-Based Research Approach

**P. Shane Gallagher, Ph.D.**

Institute for Defense Analyses  
Alexandria, VA  
pgallagh@ida.org

**J.T. Folsom-Kovarik, Ph.D.**

Soar Technologies, Inc.  
Orlando, FL  
jeremiah@soartech.com

**Sae Schatz, Ph.D.**

ADL Initiative  
Alexandria, VA  
sae.schatz@adlnet.gov

**Avron Barr**

Institute for Defense Analyses  
Alexandria, VA  
abarr@ida.org

**Sarah Turkaly**

Institute for Defense Analyses  
Alexandria, VA  
sturkaly@ida.org

### INTRODUCTION

This paper summarizes the first spiral of research and development for the Total Learning Architecture (TLA), a set of specifications intended to support next-generation learning. More precisely, the TLA concept addresses the defense community's need for "a continuous, adaptive learning enterprise" that delivers the "right training at the right time in the right way" (U.S. Army Training and Doctrine Command, 2017, p. iii, U.S. Navy Personnel Command, 2017, p. 2). The paper begins by describing this demand signal in more detail, and then it proposes two possible acquisition strategies for addressing it. Next, we explain the TLA concept and features of the current prototype system. Following this, we summarize the spiral-1 testing methodology and preliminary results, and we close by highlighting key insights gleaned from this first stage of development.

### WHY AND HOW?

Why does DoD need a "future learning ecosystem" and how should we go about building one?

#### Demand Signal

The volatility and complexity of the evolving security environment place increasing demands on our military workforce. To thrive under these conditions, personnel require a broader set of competencies, higher levels of proficiency in them, and a greater ability to rapidly learn new material to confront novel challenges. In other words, personnel must develop an ever-expanding set of sophisticated, agile knowledge and skills—albeit without significantly increasing training and education time or costs (Schatz, Fautua, Stodd, & Reitz, 2015). Each of the U.S. Services, the U.S. Joint Staff, and many other federal programs has released publications highlighting this need and, correspondingly, calling for reforms to their learning and personnel development systems (see Raybourn, Schatz, Vogel-Walcutt, & Vierling, 2017 for a detailed summary). Remarkably, these numerous publications generally point towards a shared vision, which includes the concept of a modern "learning ecosystem" comprised of interconnected learning opportunities, supported by technology, driven by data, and integrated with other talent management capabilities.

This envisioned "continuum of learning" includes features such as the following:

- **Continuous:** Career-long, continuous learning replaces the status quo's stovepipe, episodic learning
- **Blended:** Formal training and education, just-in-time support, and informal learning are integrated
- **Enterprise-focused:** Training, education, and talent management are considered in concert, holistically
- **Diverse:** Disparate learning technologies and methods are stitched together into a cohesive ecosystem
- **Learner-centric:** Learning adapts to individual and team needs, contexts, and characteristics
- **Data-driven:** Learner data from across many sources are aggregated and analyzed to drive decisions
- **Competency-based:** Competency frameworks support assessment and guide developmental trajectories
- **On-Demand:** Modular training and education can be delivered at the point of need
- **Cloud-Based:** Software services and network-based repositories support flexibility and discoverability

To achieve this vision, organizational structures, policies, and cultures will need to adapt. Further, a novel technological capability—categorically different from current learning and development technologies—is needed. We can imagine two possible ways forward.

### **Option 1: Traditional Acquisition Approach**

One way to achieve this vision might be to embark upon a lengthy Joint Capabilities Integration and Development System (JCIDS) acquisition process. The Defense Department could spend millions (or perhaps billions) of dollars to develop a massive technology hub, which may be outpaced by the latest consumer technology by the time it reaches full operational capability.

To ensure compliance, top-down policy could direct defense training and education institutions to use the system. Assuming the acquisition process worked successfully, the technology would likely meet most, but certainly not all, of the different stakeholder communities' requirements. Those outlier groups would need to adapt their internal processes, or develop local modifications of the application, or perhaps petition for exemption (thereby limiting the system's effectiveness and likely spurring duplicative local initiatives). Mandating participation by fiat would do little to encourage commercial businesses (except for those specifically under defense contracts) to adopt the system or develop interoperable technologies to work with it. As a result, this application would most likely work poorly (or not at all) with private sector, college and university, and other government systems.

Upgrades to the technology may also prove troublesome, and given such an expansive system, the risk of complications would undoubtedly encourage more conservative modernization decisions. Further, for such a complex and customized technology, few vendors would be liable to possess the expertise necessary to manage it, potentially locking the Government into expensive operations and maintenance contracts with limited options for real, qualified competition.

If this scenario appears overly critical, the Defense Integrated Military Human Resources System (DIMHRS) program offers a cautionary tale. DIMHRS was intended to integrate over 90 legacy personnel management applications into a single authoritative system. However according to then Secretary Robert Gates, after nearly a billion dollars invested over 10-years, all that was produced was “unpronounceable acronym” (Philpott, 2010). This failure illustrates how the traditional, linear acquisition process no longer supports effective procurement of modern, software-heavy technologies—particularly for enterprise systems (Government Accounting Office, 2015).

### **Option 2: Community-Driven Specifications Development**

An alternative pathway involves agile development methods, open-architecture designs, and community-developed interoperability specifications. This approach essentially uses no enterprise acquisition processes. Instead, through collaboration with the wider public and private community, the Defense Department could encourage development of technical guidelines that define how any compliant technology can integrate with the “future learning ecosystem.” This approach achieves the vision—not through a massive and monolithic program—but rather through the federation of many smaller applications and micro-services, which combined create the emergent system.

The 802.11 specification offers an example. The 802.11 family of specifications provides the basis for wireless networking. Thanks to these detailed technical definitions, a diverse array of technologies can connect into a wireless local area network. Even yet-to-be-developed technologies can “plug and play” into a Wi-Fi environment, so long as they adhere to the 802.11 guidelines. By defining the interoperability specifications, 802.11 creators have enabled the diverse, adaptable, and pervasive wireless ecosystem the world now enjoys.

The specifications-based approach supports extensibility, encourages innovation within the system, and drives down costs through openness to competition. This approach also allows for local control of technology acquisition and operations, encourages use of the specification beyond defense systems, reduces the likelihood of version and vendor lock-in, and ultimately lowers costs and risks for the government. Clearly, for applicable modern software

systems, the specifications-based approach offers many advantages over the more traditional acquisition process. Consequently, the TLA is being created via this type of community-driven specifications development approach.

## TOTAL LEARNING ARCHITECTURE

The TLA is a set of internet and software specifications being developed to enable next-generation personalized, data-driven, and lifelong technology-enabled learning. Once complete, the TLA will include technical guidelines, such as Application Programming Interfaces (APIs) and data model descriptions, that define how training, education, and personnel management technologies “talk” to each other—both syntactically and semantically. The TLA will also define software services that perform whole-system processes using automation and artificial intelligence. The TLA is *not* a particular training device or educational tool; rather, it is the “glue” to connect all other learning technologies into an integrated, coherent system.

The Advanced Distributed Learning (ADL) Initiative, a program under the Deputy Assistant Secretary of Defense for Force Education and Training, is leading development of the TLA in close concert with many other industry and interagency partners, including the Office of Personnel Management and Army Research Laboratory. The development team has committed to using collaborative development methods, open-source licensing, and open-architecture design principles. The TLA research and development project began in late 2015, and (as of this writing), the current system meets the criteria for Technology Readiness Level 4 (“component validation in relevant environment”). The project uses an iterative design process, including iterative development and testing spirals, and system designers anticipate that scaled implementation of the TLA could begin as early as 2019.

### Current Prototype: TLA Version 0.1

The TLA project ultimately aims to deliver technical documentation—but defining software specifications without verifying their real-world instantiation is a recipe for disaster. Consequently, in addition to authoring the technical documents, the development team created a prototype TLA-enabled learning ecosystem (hereafter referred to as “TLA-1”). It includes various software services, technical components, and learning applications (also known as “learning activity providers”), all of which exchange data using the initial suite of TLA APIs. Importantly, the TLA-1 architecture represents just one example of a TLA ecosystem; many more arrangements of components are also possible, because a key design tenet underlying the TLA is generality and usefulness across learning use cases.

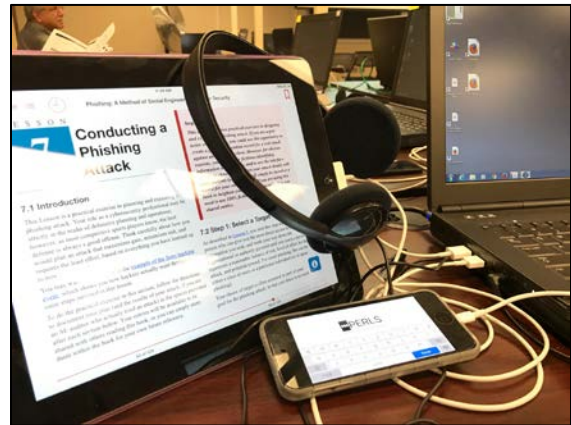


Figure 1. iPod Touch, iPad, and laptop from the TLA-1 April 2017 Ft. Bragg testing event

To ensure the spirit of the community-driven approach was maintained, different businesses, government agencies, and federally funded research and development centers were asked to participate. Overall, TLA-1 brought together a total of 22 applications and back-end services, created by 11 different providers. While this diversity of systems and vendors added complexity to the development process, it also helped reveal unique perspectives and insights much earlier, and more saliently, than could have occurred with a single-vendor development effort.

### User-Facing Components

As described in more detail below, TLA-1 was tested during a week-long event held at Ft. Bragg. This initial implementation included multiple learning activities, accessible via an iPod Touch, iPad, and/or laptop (Figure 1). Specifically, TLA-1 included the following user-facing applications:

- iPad-based Personal E-books for Learning (PEBL), developed by Eduworks

- iPod Touch–based PERvasive Learning System (PERLS), developed by SRI International
- Static content viewer for PDFs and static webpages
- YouTube video player
- Moodle, a traditional, open-source LMS loaded with cybersecurity learning materials
- Web-based tutor, built using the Army Research Lab’s Generalized Intelligent Framework for Tutoring
- Sero! a web-based concept map assessment tool, developed by Perigeon Technologies
- Project ARES web-based cybersecurity serious game, developed by Circadence
- CyberScorpion web-based cybersecurity simulation environment, developed by Sandia Labs

### **Back-End Components**

In addition to the learning applications, themselves, the TLA-1 prototype included back-end cloud-based services for providing cross-system learning recommendations, managing and asserting competencies, managing learning evidence, aggregating and visualization of learner data, and managing learner profiles. These back-end capabilities were designed to support and interact with the holistic system (i.e., across all of the learning applications). For example, the back-end learning record store collected data about learners’ performance from all of the applications; these aggregated data could then be visualized, revealing larger trends about learners’ performance.

The TLA-1 prototype included these back-end data processing components (i.e., services):

- Competency and Skill System (CASS) xAPI evidence-to-competency mapper, developed by Eduworks
- Learning context sensing service

...and these back-end data storage components (i.e., data core):

- Competency framework repository
- Index of the various user-facing learning activities and their metadata
- Learner record store to manage and exchange Experience API (xAPI) data

### **Application Programming Interfaces**

“API” is a generic software term. APIs define a standardized set of “rules” (i.e., code) that allow software systems to communicate with each other in prescribed ways, without giving access to all of an application’s internal code or capabilities. The current TLA specification (Version .01) encompasses 10 APIs. In addition to the mature xAPI specification (which is considered part of the TLA family of specifications), the TLA includes APIs for exchanging data about learning activities, learner profiles, competencies, learning evidence, and user authentication (Folsom-Kovarik & Raybourn, 2016). Both the front-end and back-end applications within a TLA-enabled ecosystem can use these APIs to access the TLA’s back-end services or data core, for instance, to pull information about the learning context to improve the learning experience or to tailor recommendations to a learners’ estimated mastery level.

### **Human Tutor**

In addition to its technical components, the TLA-1 prototype incorporated an expert human tutor. The tutor helped to create a more optimal and realistic blended learning experience for the learners. More than that, however, he encouraged the development team to consider how the TLA will support teachers, trainers, and other learning facilitators. For instance, based upon his feedback the development team created an instructor dashboard, which helped the tutor quickly identify which learners may be struggling with the content or feeling frustrated.

### **TLA-1 Learning Content**

The TLA specifications are content-agnostic; therefore, a TLA-enabled system can support any or all subject matter. For prototype testing, however, we selected a narrow subset of instructional topics. In this case, we asked all of the contributing organizations to design content around six cybersecurity terminal learning objectives (TLOs):

1. Describe network architecture, browser configuration, and hardware/software for secure Internet browsing
2. Understand social engineering concepts in cyber-attacks (e.g., phishing, waterhole attacks, lures)
3. Run penetration tests to locate potential security threats (Network enumeration using NMAP, Wireshark)

4. Know how to analyze a packet capture (pcap) file
5. Recognize, document, and analyze a successful attack from point of entry, pivoting, and systems controlled
6. Successfully execute the steps in creating an exploit

Instructional designers, working closely with subject-matter experts, subdivided each TLO into three subordinate learning objectives (ELOs), which were also defined at the novice, intermediate, and expert levels. This resulted in nine distinct objectives per TLO, for a total of 54 total objectives. For the purposes of this trial, each of ELOs was considered a discrete competency, and was recorded as such into the TLA-1 competency model. Each piece of learning content (e.g., quiz, article, scenario) was then mapped to one or more of these competencies, which informed system-wide content sequencing and the personalized recommendations given to learners.

For the Ft. Bragg event, the learning objectives were further clustered into two categories: “social engineer” and “cyber apprentice.” Learners attempted to demonstrate their proficiency in one, or both, of these areas by successfully completing final challenges (e.g., a series of simulation-based capture-the-packet scenarios) for each topic. Learners who completed these achievements earned a “badge,” in the form of a special stamp on their completion certificate. (As trivial as this may sound, these badging activities proved highly motivating!)

Rather than follow a prescribed, linear pathway through the learning content, learners could attempt any learning activity—even the final challenges—at any time. The larger TLA system, via its back-end competency-based applications and recommendation services, offered suggestions to guide learners towards the most appropriate activities that would foster the knowledge and skills each uniquely needed in order to successfully complete the final challenges. These back-end applications used learners’ pre-test data (to initially seed the learner profiles), as well as their performance and behavior in TLA-connected applications, to make these tailored recommendations.

## **RESEARCH AND DEVELOPMENT APPROACH**

The TLA project uses a human–systems integration approach for research and development, in accordance with Defense Department guidance (U.S. Department of Defense, 2007; U.S. Department of Defense, 2017; FY11 Department of Defense Human Systems Integration Management Plan, 2011). Among other things, this approach involves early and frequent coordination with stakeholders, iterative and incremental development, regular formative testing, an emphasis on human factors throughout the development process, and consideration of system lifecycle factors.

The development and testing procedures embedded within this approach use design-based research methods (National Science Foundation, 2007). Generically, these methods involve iterative phases of analysis, design, development, implementation, and evaluation. Systematic reflection and documentation are used continuously throughout to codify the process and facilitate understanding. The entire model repeats in spiral iterations until its goals have been achieved (Plomp & Nieveen, 2010; Wademan, 2005).

For the TLA project, the research methodology follows five phases, which repeat every 12–18 months:

- (1) **Scoping and understanding** – Process planning, such as determining objectives and their metrics
- (2) **Problem definition and context** – Building base understanding via expert interviews and published theory
- (3) **Design refinement** – Defining and refining design requirements as well as any existing prototypes
- (4) **Testing and evaluation** – Conducting formative testing of developed specifications and prototypes
- (5) **Analyses and outcomes** – Dissemination of testing results, particularly to inform the next research cycle

Beginning late in 2015, the TLA evaluation team collaborated with the TLA technical developers, potential end-users, and other government and industry stakeholders to inform the initial scoping and problem definition phases (i.e., phases 1 and 2). The outputs from these phases informed the system’s design (phase 3), which began around April 2016 and continued throughout September of that year. Finally, the testing and analysis phases (phases 4 and 5) began in October 2016. The results from these final two phases comprise the remainder of this paper.

## **METHODS**

The spiral-1 “testing and evaluation” phase included two major thrusts: (1) expert evaluation of the written specifications and (2) human-subjects testing of the prototype reference implementation model (i.e., a real-world instantiation of the draft specifications).

### **(1) Expert Evaluation of the Specifications: Delphi Technique**

The true “deliverable” of the TLA project is a set of specifications and associated software-service definitions (i.e., technical documentation). Consequently, these publications—both their concepts and specific content—must be usable, palatable, and actionable. To evaluate this, we used a Delphi technique to solicit feedback from stakeholders on the first draft of the materials.

The Delphi technique is a well-established method for converging expert opinions within specified domains. It uses a series of questionnaires to collect data from a select panel of reviewers via multiple rounds of anonymized data collection and analysis. Analysis from the preceding round informs the next, which allows participants to respond to one another’s feedback but with limited biases, due to the aggregation and anonymization of responses (Hsu & Sandford, 2007). Delphi studies have been used in all domains for planning, forecasting, issue identification, issue prioritization, and strategy development, and, although rarely applied to information architecture development, empirical studies have demonstrated their validity in this area (Day & Bobeva, 2005).

The TLA Delphi testing included three goals: (1) Gain feedback on the TLA concept, including its realistic viability for implementation and sustainment; (2) Gain feedback on the draft TLA technical specifications, including their clarity, comprehensiveness, and specificity; and (3) Encourage “diffusion of innovation” (i.e., buy-in and participation) by involving a broad range of key stakeholders in the development process.

### **Participants**

The success of Delphi studies relies, in part, on the quality of participating respondents. To ensure a rigorous selection process, we developed a technical advisory board comprised of TLA developers as well as other external stakeholders. The advisory board oversaw creation of the Delphi qualifying rubric, which researchers used to vet possible participants. Using it, the researchers identified more than 200 qualified stakeholders from across the U.S. Defense Department, other federal agencies, coalition defense enterprises, industry, academic institutions, and standards and specifications professional organizations. Researchers divided these stakeholders into two groups: Programmatic decision-makers and technical subject-matter experts.

### **Apparatus and Materials**

Potential participants were asked to complete an initial qualifying questionnaire, which the researchers reviewed to ensure respondents met all participation requirements. Qualifying participants then received the Round-1 package, including the 88-page TLA specifications documentation, corresponding questionnaire, and demographics survey. The programmatic decision-makers and technical experts received different questionnaires, consisting of either 22 or 17 questions. This process was repeated for Round-2 of the Delphi, with different questionnaires, based upon the responses from Round-1.

### **Procedure**

In November 2016, researchers sent the Delphi qualifying questionnaire and Round-1 package to the 200 identified participants, and 54 qualified participants completed the Round-1 package. These results were returned and analyzed using the qualitative research software Dedoose. Round-2 Delphi questions were then developed based on these responses as well as input from the technical advisory board. The Round-2 package was sent to participants in May 2017, and analysis of these results continues at the time of this writing (i.e., June 2017). To date, 35 participants have returned Round-2 results. The attrition rate between Round-1 and Round-2 is 35%, which falls within general expectations for Delphi studies (Adler & Ziglio, 1996). A third round of testing is under consideration; however, this may include a face-to-face workshop rather than the asynchronous written feedback—breaking the method’s typical anonymity but allowing for more fluid debate and open discussion.

## Preliminary Analyses

Preliminary analysis has been completed for Round-2 responses, but deeper inquiry will take place in the coming weeks. It is currently premature to present these results in this paper.

## (2) Human-Subjects Prototype Testing

Meanwhile, as experts across the community reviewed the TLA technical documentation as part of the Delphi study, the TLA development team used those same specifications to build, and then test, a prototype system. (This is the TLA prototype described above.) The testing event spanned four days (18–21 April 2017). Roughly six dozen ( $n = 73$ ) participants used this system, each accessing the TLA prototype for 2½ hours each day (for a total of 10 hours of spaced exposure). This testing event not only provided baseline data, against which subsequent TLA iterations can be compared, but it also helped the TLA developers better understand system functionality, how learners interact with it, and those learners' perceptions of it. Because the purpose of this testing was to inform subsequent design-based research iterations, a non-experimental within-subjects design was used (see Table 1, below).

**Table 1. Human-Subjects Research Design**

| Dates  | 10 April 2017  | 18–21 April 2017   | 21 April 2017   |
|--------|--|--|---|
| Events | O <sub>1</sub> – Pre-test<br>O <sub>2</sub> – Demographics | X – Treatment (TLA-based learning)<br>O <sub>3</sub> – System log data collected | O <sub>4</sub> – Post-test<br>O <sub>5</sub> – User-experience survey<br>O <sub>6</sub> – User-experience focus group |

## Location and Participants

Testing took place at the John F. Kennedy Special Warfare Center and School (JFKSWCS), specifically within the Special Warfare Education Group (Airborne) or SWEG(A) located at Fort Bragg, North Carolina. The population was drawn from the U.S. Army Special Forces Soldiers in active training, out of training waiting reassignment, and graduate students attending National Defense University through JFKSWCS. Ultimately, 73 personnel completed the four-day trial. A majority of participants (94%) were 18–30 years old. They consisted of E4s, E5s, E6s, and O3s from primarily 18X (Special Forces) (70%) and 38B (Civil Affairs) (14%) military occupation specialties. The majority of participants indicated they had strong prior experience with iPhones and traditional online e-learning, but they had less experience with iPads, e-books, and game-based and simulation-based training.

## Apparatus and Materials

Data were collected on system performance, system usability, user experience (Kirkpatrick's Level 1), and learning outcomes (Kirkpatrick's Level 2). Participants completed pre- and post-tests, surveys, and questionnaires via a web browser, and individuals were only identifiable through their user-ID numbers, which were randomly assigned.

- **O<sub>1</sub> – Pre-test and O<sub>4</sub> – Post-test:** To mitigate practice effects, two parallel forms of a 162-item multiple-choice test were developed. The tests included exactly three items per enabling learning objective, for each of the 54 ELOs. The pre-test used Form A, and the post-test used Form B.
- **O<sub>2</sub> – Demographics:** A short survey asked about participants' ages, ranks, occupational specialties, and levels of experience with the types of learning activities (e.g., e-books, online learning) used in this test.
- **O<sub>3</sub> – System log data:** Data were collected from the built-in TLA APIs (including xAPI), basic system component logs, and contextual time-stamped incident logs. In addition, the team developed a Java application, named the "Big Red Button." It consisted of a clickable red button, which sat in a window on each laptop's desktop. When clicked, it opened a text box where participants could record issues or observations, which were logged and time-stamped for later analysis.
- **O<sub>5</sub> – User-experience survey:** To capture usability and user experience data, we combined items from the *System Usability Scale* (Digital Equipment Corporation, 1986) and Turner's (2011) user experience questions. We also included five researcher-developed questions about participants' engagement and



specific TLA experience. For these items, participants rated each learning activity within the prototype on a scale of 1 to 5, based on their enjoyment and interest in it. The survey also included several free-form short-answer questions, which encouraged participants to elaborate on their specific TLA-1 experiences.

- **O<sub>6</sub> – User-experience focus group:** A handful of participants provided free-form feedback about their experiences in a small focus group, after their final day of system interaction. A set of 10 researcher-developed focus group questions were used.

## Procedure

A week before the testing event, researchers presented an initial walkthrough of the prototype system and planned schedule. At that time, volunteers also signed informed consent forms, obtained their system log-in credentials, and received instructions for taking the online pretest and demographic survey within the next two days (by 12 April).

On the first day of testing (April 18), participants chose their testing stations; each station included a laptop, iPad, and iPod Touch along with participants' system log-in credentials and a user guide. The research team also provided another, more in-depth, technical orientation. (This was a longer session to account for the extra time spent with orientation activities.) Following the orientation, participants were free to use any combination of the TLA learning activities to pass the final challenges and achieve the "social engineer" and/or "cyber apprentice" badges.

Roughly 30 participants interacted with the system during each scheduled block, which means three different blocks were required each day to accommodate all the participants. Following their 10-hours of interactions with the TLA prototype, the participants completed the post-test and user experience survey. Additionally, six participants, chosen at random, participated in a 90 minute focus group.

## RESULTS

### Learning Outcomes

Of the 73 total participants, only 67 completed both the pre and the post tests. Their scores were analyzed using a paired samples *t*-test. Results indicated a (rounded) positive change of 10% ( $p < .001$ ). Gains in scores increased from an average of 8% correct to an average of 18% correct resulting in a "very large" Cohen's *d* (effect size) of 1.14. (See Figure 2.) Also of interest were the individual differences in scores; 62 of the 67 participants showed individual gains between their pre- and post-tests. Table 2 includes the descriptive data and paired samples *t*-test results.

Given the design of this testing event, it is difficult to determine what portion of the "very large" effect size is attributable to the TLA versus the instructional design of the embedded content or on-site (human) tutor. However, it should be noted that pre- to post-test scores increased by more than one standard deviation after only 10 hours of interacting with the TLA-1 system.

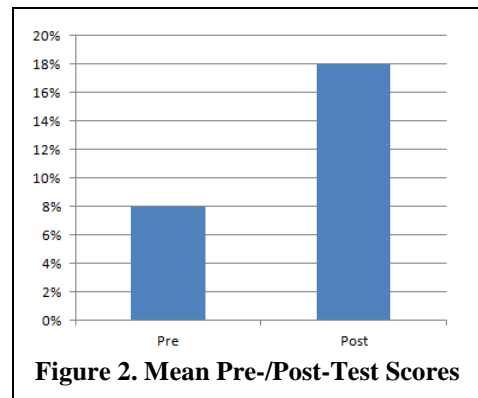


Figure 2. Mean Pre-/Post-Test Scores

Table 2. Paired Samples *t*-test Results

|  | Paired Differences |                |                 |   |        | t     | df | Sig. (2-tailed) |
|--|--------------------|----------------|-----------------|---|--------|-------|----|-----------------|
|  | Mean               | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference |        |       |    |                 |
|  |                    |                |                 | Lower                                     | Upper  |       |    |                 |
| Pair 1 Post-test overall score percent<br>- Pre-test overall score percent | .09904             | .08678         | .01060          | .07787                                    | .12021 | 9.342 | 66 | .000            |

## **Usability and User Experience**

Participants provided feedback on the TLA-1 prototype, changes they would recommend for the next iteration, and their overall experience. When asked to rate their TLA-1 experiences on a scale of 1–10 (lowest to highest), participants gave it an average of 6.1; however, responses varied widely from 1–9. Of those who responded to the question, “Would you recommend the TLA-1 to a friend?” 68% said they would recommend it, 20% said they would not, and the remaining 12% said “it depends.” Some changes requested from the “it depends” respondents included using different content, fixing technical bugs, adding introductory lessons on how to use the system as a whole, and incorporating introductions to the more difficult learning activities.

Most of the more specific user experience responses focused on learning activities or their embedded content, rather than the actual TLA capabilities (e.g., meta-adaptation across experiences, aggregation of cross-platform data). For instance, some participants remarked that the content was too information-dense or not suitable for beginners.

For those participants who did comment directly on the TLA-1’s capabilities, they had mixed experiences. On the negative side, they expressed frustration with technical glitches (such as activities not being registered as complete in the recommender system), loading and log-in system latency, not having the option to launch activities on the device the participant wanted to use, and communications failures between devices. Participants also wanted to be able to use any device to launch the learning activities. (In this trial, participants’ device options were constrained and intentionally required them to switch among the devices to access different learning activities.) Finally, other comments, from the focus group participants, indicated they wanted more feedback on their learning progression and performance, and more opportunities to practice what they had just learned.

Positive reactions to the system mostly concerned the efficiency it provided them as a learning platform. For instance, participants commented on how the content could be curated so only the most relevant information was offered. They also remarked on the system’s flexibility, such as its ability to deliver learning content through various platforms and activity types, and they commented on the advantages of self-directed learning.

Participants also indicated that the opportunity to complete badging activities alongside the system interaction kept them engaged throughout the week. In fact, the badging activities had the highest percentage of positive feedback, with 66% of participants rating their experience with them as “enjoyable” or “very enjoyable.”

## **IMPLICATIONS LEADING TO DESIGN IMPROVEMENTS**

Both the Delphi evaluation and human-subjects testing were intended to provide feedback into the larger TLA multi-year design-based research effort. Consequently, these results should not be viewed as a summative rating of the system’s ultimate utility. Rather, they should be viewed through the lens of an iterative TLA development process, where they provide valuable feedback to inform ongoing progress. This section highlights select lessons learned from the development and testing of TLA-1. In addition to guiding continued research on this project, these lessons have generalizable relevance for all researchers working towards the future learning ecosystem vision.

First, the Ft. Bragg testing event highlighted the value and feasibility of several TLA design tenets such as generalized data sharing across systems, ease of participation, and component interchangeability rather than component lock-in. For example, during the execution some supporting engineers organized a “hackathon” to visualize different aspects of data flow and learner experience. The dashboards they created could quickly draw on learner experience data from all the participating systems to build a more complete picture than with individual systems operating separately. The visualizations were also created in a matter of hours, because of the ease and low risk associated with joining new TLA applications into the existing prototype. The advantages of component interchangeability were also demonstrated during the lead-up to testing event. Load testing shortly before the main event revealed multiple points of inefficiency that could threaten the event. The affected components, such as one model of learning record store, were easily replaced within a short timeframe because the established TLA APIs, such as xAPI, make the components “plug and play.”

Second, in preparing for the Ft. Bragg testing event, manual authoring was needed to accomplish alignment between competencies and content. Hundreds of individual learning activities were created or imported from multiple sources. The learning activities all required markup within the activity index to describe competencies addressed, approximate difficulty level, expected duration, and other attributes. Because this task was manual, the markup was coarse-grained and sometimes inexact. For example, only 54 competencies were defined. Finer granularity of competencies, at the concept or sub-concept level, and more informative relations between competencies such as ordering and similarity relations, may let prototype components better interpret learner experiences in these activities to make more targeted recommendations for next activities.

The testing event revealed a need to identify components when they call into any API. Identifying information such as an API access key could support future improvements in trust and privacy, bandwidth or query limits, usage tracking, or even fees that could be charged for service requests. Because it was a distributed system with components built by many providers, the TLA-1 highlighted a need to monitor for possible inefficiencies caused by federated applications such as too frequent calls or too costly bandwidth usage. These service issues often arise from components other than the provider who is paying the cost of responding to the calls. As a result, tools could be built into TLA APIs to identify API callers and respond intelligently in situations that would otherwise be outside of local control.

Finally, an opportunity exists for TLA-enabled components to share more than learner experiences as defined in xAPI. TLA-1 provides draft APIs and data models that put those individual experiences in context and enable sophisticated processing of them such as inference, interpretation, and prediction. To realize this opportunity, there is a need to research and develop multiplicity of TLA components. The testing event demonstrated the potential for multiple learner-facing apps and future prototypes to provide multiple components in other roles, such as the evidence mapper or the activity index. As TLA gains formality and it becomes harder to make large changes to the APIs and data models, the research carried out now will pay off in design insights and accordingly in a more effective final product with reduced disruption for all TLA stakeholders.

## **CONCLUSION**

This paper described the research and development methodology and current state of the TLA. The initial spiral of the design-based research process has helped the system's developers better understand the technical challenges, consider how learners will interact with it, and ways to enhance learning through the TLA. Already, the "alpha" TLA prototype shows a strong potential for increasing the effectiveness, efficiency, and flexibility of learning.

Technical achievements for the integration experiment were threefold. First, all components participated in data sharing by communicating to another component through the TLA. Examples of increased data sharing included writing xAPI to a single common LRS, adding metadata to enable interpretation into the activity index, and enabling user identity management OpenID, which played the role of authentication service in TLA-1. Second, some components output interpretations and inferences such as learner mastery estimates or adaptation recommendations, in a manner that other systems were able to use. In this way, TLA enabled improved understanding of learners compared to each system working in isolation. This objective was met by components like the evidence mapper and instructor-facing analytics tools. Third, some components input learning record store data, activity data, or learner mastery estimates from other systems and then used the data to adapt their own internal behavior or change their outputs. The systems that changed their behavior leveraged the potential of TLA to improve distributed learning.

Although room for improvement remains, the prototype implementation and integration of the TLA for this year's study represented a successful team effort that resulted in a usable prototype and supported a week of interaction with dozens of real users. The findings reported here support the realized value and promise of the distributed learning vision. In addition, specific lessons learned that we report will immediately inform design improvements in the next iteration of the TLA.

## ACKNOWLEDGEMENTS

This work is supported, in part, by the ADL Initiative (Contracts BE-02-4039 and W911QY-16-C-0019-P00005). The views and conclusions contained in this document are the authors' and should not be interpreted as representing the official policies, either expressed or implied, of the ADL Initiative or U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes. We would also like to thank Mr. Curtis Price, Ms. Michelle Worsley and the administration and staff of JFKSWCS/SWEG(A) for their generous support.

## REFERENCES

- Adler, M., & Ziglio, E. (1996). *Gazing into the oracle: The Delphi method and its application to social policy and public health*. Jessica Kingsley Publishers.
- Day, J., & Bobeva, M. (2005). A generic toolkit for the successful management of Delphi studies. *The Electronic Journal of Business Research Methodology*, 3(2), 103–116.
- Digital Equipment Corporation. (1986). *System Usability Scale*. Retrieved from [hell.meiert.org/core/pdf/sus.pdf](http://hell.meiert.org/core/pdf/sus.pdf).
- Folsom-Kovarik, J.T., & Raybourn, E.M. (2016). Total Learning Architecture (TLA) enables next-generation learning via meta-adaptation. In *Proceedings of the IITSEC*. Arlington, VA: NTSA.
- FY11 Department of Defense Human Systems Integration Management Plan. (2011). Washington, DC: DDRE, Directors, Mission Assurance and Human Performance, Training & Biosystems.
- Government Accountability Office. (2015). *DoD business systems modernization: Additional action needed to achieve intended outcomes*. Washington, DC: Government Accountability Office.
- Hsu, C., & Sandford, B. (2007). The Delphi technique: making sense of consensus. *Practical Assessment, Research & Evaluation*, 12(10), 1–8.
- Kelly, A. E. (2010). When is Design Research Appropriate? In T. Plomp & N. Nieveen (Eds.), *An Introduction to Educational Design Research* (3rd Print, pp. 73–88). Enschede: SLO.
- National Science Foundation. (2007). *Cyberinfrastructure Vision for 21st Century Discovery*. Retrieved from <http://www.nsf.gov>.
- Phipott, T. (2010). (February 20, 2010). Military Update: DIMHRS program dumped as ‘a disaster. Stars and Stripes. Retrieved from <https://www.stripes.com/opinion>.
- Plomp, T. (2010). Educational design research: an introduction. In T. Plomp & N. Nieveen (Eds.), *Educational design research: An introduction* (3rd Print, pp. 9–35). Enschede: SLO.
- Plomp, T., & Nieveen, N. (2010). *An introduction to educational design research*. In T. Plomp & N. Nieveen (Eds.), *Educational design research: An introduction* (3rd Print). Enschede: SLO.
- Raybourn, E., Schatz, S., Vogel-Walcutt, J.J., & Vierling, K. (2017). At the Tipping Point: Learning Science and Technology as Key Strategic Enablers for the Future of Defense and Security. In *Proceedings of the IITSEC*. Arlington, VA: NTSA.
- Schatz, S., Fautua, D., Stodd, J. & Reitz, E. A. (2015). The Changing Face of Military Learning. In Doughty, R., Wells II, L., and Hailes, H. (Ed.), *Innovative Learning: A Key to National Security* (pp. 35-50). Fort Leavenworth, Kansas: The Army Press.
- Turner, N. (2011, April, 7). 15 useful user feedback questions for online surveys. Retrieved from <http://www.uxforthemasses.com/online-survey-questions/>.
- U. S. Department of Defense. (2007). *The Defense Acquisition System*. (DoD Directive 5000.1). Retrieved from <http://www.dtic.mil/whs/directives/corres/pdf/500001p.pdf>.
- U. S. Department of Defense. (2017) *Operation of the Defense Acquisition System* (DoD Instruction 5000.2). Retrieved from [http://www.dtic.mil/whs/directives/corres/pdf/500002\\_dodi\\_2015.pdf](http://www.dtic.mil/whs/directives/corres/pdf/500002_dodi_2015.pdf).
- U.S. Army Training and Doctrine Command (2017, April). *Army Learning Concept for Training and Education for 2020–2040* (TRADOC Pamphlet 525-8-2). Ft. Eustis.
- U.S. Navy Personnel Command. (2016). *Sailor 2025*. Retrieved from [www.public.navy.mil](http://www.public.navy.mil).
- Wademan, M. R. (2005). *Utilizing development research to guide people capability maturity model adoption considerations*. Unpublished doctoral dissertation, Syracuse University, Syracuse.