INSTITUTE FOR DEFENSE ANALYSES

Bridging the Archipelago: An Assessment of the Advanced Distributed Learning Initiative's Total Learning Architecture (TLA)

P. Shane Gallagher
Avron B. Barr
Sarah J. Turkaly

October 2018

Approved for public release; distribution is unlimited.

IDA Document D-9077

Log: H 18-000182

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

About This Publication

This work was conducted by the Institute for Defense Analyses (IDA) under contract HQ0034-14-D-0001, Project BE-2-4039, "Analyses and Review for the Advanced Distributed Learning Initiative," for the Under Secretary of Defense Personnel and Readiness, USD(P&R). 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.

For More Information

Poornima Madhavan, Project Leader
pmadhava@ida.org, 703-578-2822

Leonard J. Buckley, Director, Science and Technology Division
lbuckley@ida.org, 703-578-2800

INSTITUTE FOR DEFENSE ANALYSES

IDA Document D-9077

Bridging the Archipelago: An Assessment of the Advanced Distributed Learning Initiative's Total Learning Architecture (TLA)

P. Shane Gallagher
Avron B. Barr
Sarah J. Turkaly

Executive Summary

Background

The advantage of bridging time and distance through Web access has motivated the leaning industry to develop and vigorously apply distributed learning (DL) in education and training. DL is described in DoDI 1322.26 as “learning content and systems, mediated through technology that are accessed through a network or experienced via portable media.” DL courses of learning are now common and accompanied by learning management systems that allow learners to authenticate themselves, register for courses, complete courses, and take assessments. These systems are increasingly available today, albeit with more advanced components and functionality.

This report describes the first cycle of a significant multi-year research and development project undertaken by the Advanced Distributed Learning (ADL) Initiative. This project, Total Learning Architecture (TLA), is being developed to support and modernize the government’s DL technology, strategy, infrastructure, and management systems. The TLA project is intended to provide viable interoperability specifications and standards for interconnected learning opportunities, supported by technology, driven by data, and integrated with other talent-management capabilities.

Methodology

Design and development of the TLA is based on a Human Systems Integration approach using 12- to 18-month development and testing cycles. A design-based research approach was used to assess evolution of TLA education and development and to maximize use of its findings in subsequent design and development cycles. All assessment is formative using this approach. Three primary areas of assessment were examined. First, the technical viability of the design was assessed using expert opinion. Second, expert opinion was also used to assess the viability of usage and adoption throughout the ADL constituency. Third, a fully operational implementation of the TLA specification suite was tested with human subjects and learning content. This approach provided assessment of the system’s functionality based on its specifications, learning potential, usability, and user experience.

Initial design and development of the TLA produced a suite of 10 application programming interfaces (APIs). An API is a set of “rules” (code) that allows software systems to communicate with each other using standards to facilitate their integration. The communication between services and components cover issues such as learning activity, learner profiles, competencies, learning evidence, and user authentication. The APIs were documented in the initial technical documentation set and later applied in a working test implementation known as TLA-1.

TLA-1 consisted of a learning system that was designed and built with the cooperation of ADL Initiative Broad Area Announcement providers, industry partners, and Federally Funded Research and Development Centers. The system provided infrastructure and services to manage and assess learners' competencies, recommend learning content based on their evolving competencies track their learning interactions, and provide a full range of learning activities. Content for TLA-1 consisted of cyber-security learning objectives and content relevant to the testing population.

Assessment

Three formative assessments were performed during the first design and development cycle of the TLA. Their findings provided recommendations for the subsequent design and development cycles. They captured a baseline of activity and performance for the first year of TLA multiyear research, development, and assessment activities. Two assessments were performed using the Delphi method, with experts commenting on the initial design and implementation documentation. The experts assessed: (1) technical viability of the design and (2) the adoption and diffusion potential of the TLA. Assessment (3) concerned the functionality of the TLA specifications' usability, user experience, and potential for providing relevant learning. The assessment was based on the experiences of authentic learners interacting with an operating TLA-based system.

The Delphi panel consisted of 54 panelists who were selected from a pool of 200 from relevant organizations, including the Institute of Electrical and Electronics Engineers Learning Technology Standards Committee and the National Science Foundation CyberLearning Infrastructure program. The panelists were further divided into two subpanels focused on either technical implementation or adaptation issues. There were three rounds of input and analysis. Each was based on TLA documentation concerning technology, implementation, and adoption. The analysis targeted the technical viability and the adoption/diffusion potential of the TLA. It identified gaps in technologies and implementation and adoption considerations.

Operational and Human Subjects Testing consisted of 73 total participants drawn from U.S. Army Special Forces soldiers and military personnel attending National Defense University at Ft. Bragg, North Carolina. They interacted with TLA-1 for 10 hours over a span of 4 days. An additional day was reserved for pretesting and orientation. The purpose of the pretest was to determine baseline knowledge about cybersecurity. Those were followed by the TLA-1 interaction and a posttest to compare with pretest cybersecurity proficiency. All 73 participated in applying user experience and gathering usability data; however, only 67 were able to complete both pre- and post-testing due to duty assignments.

Additional data were gathered using data collected from TLA-1. Analysis for this report was limited to xAPI data where xAPI or "Experience application programming interfaces" incorporate experience and context into software systems in prescribed ways facilitating their standardized integration.

The findings from each assessment are found in the tables below. The first table summarizes findings concerning the technical viability of the TLA design, followed by the adoption potential of the TLA. The next table summarizes findings concerning the TLA learning potential, usability, user experience, and functionality of the specifications within the TLA-1 implementation.

Summary of Technical Viability and Adoption Potential Findings

Delphi Assessment	Finding
Technical Viability of the Design	• The design is viable for supporting further iterative development. Technological decisions are sound. • Incorporation of the current xAPI specification was essential and a positive design choice. • The system may be overly complex. Simplicity is necessary for implementation. • Security, privacy, and governance must be addressed. International standards need to be incorporated. • Participation in the greater standards community is essential.
Adoption and Diffusion Potential	• More concrete use cases are needed. • Methods for trial use and observation of use are critical. • Evidence of support for learning processes is needed. • Cost-benefit models need to be developed. • The organizational and cultural barriers that may exist with TLA affordances need to be addressed. • Building community support and capacity is essential.

Summary of Findings from Operational and Human Subjects Testing

Operational and Human Subjects Testing	Finding		
Learning Gains	Direction	Significance	Effect Size
	Post > Pre	$p < .0001$	1.14
Usability and User Experience (UX)			
Usability (System Usability Score)	56.95	OK	
UX Overall Rating (1–10)	6.1	OK	
UX Recommendation to Others	Yes = 68%		
UX Comments	• Mostly positive throughout. Provides recommendations and direct access to topics needed, facilitating navigation and utilization easy. Positive about the efficiency provided as a training platform. • Criticism was constructive. More feedback is needed on learning progression and feedback to wrong answers. • More practice opportunities are needed. • Learners should be allowed to use any device they personally own.		
xAPI Data	• Learner tracking is limited value due to implementation constraints on tracking many possible learner interactions. • 169 learning activities were accessed, but only 71% of them were completed. • Six user learning paths were captured and analyzed; logical flow of activities with a variety of static webpages, e-books, YouTube videos, and assessment concept maps were generally followed.		
Overall Functionality of the TLA-1	• The functionality of the initial TLA suite of specifications worked as intended during operational testing of TLA-1, data were run through the system to allow components to interact with each other as intended. • The capture of large amounts of data concerning user interaction were enabled. They are relevant and valuable for analyzing and improving learning experiences.		

Recommendations

Findings from all three formative assessments produced recommendations and potential next steps for design improvements for the next design and development cycle. The following table presents a summary of the high-level recommendations.

Summary of High-Level Recommendations

Recommendation	
Technical	• Security should be a prominent consideration in the architecture and in implementation practices as the TLA evolves. • International practices of privacy and security should be considered when going forward. • Incorporate or help develop new international standards for TLA communication elements where needed. • Consider mapping standards that may be used for more than one purpose. • Consider consolidating or reducing the number of service definitions to simplify design. • Actively participate in the greater standards community.
Adoption/Diffusion	• Focus on one use case that might incentivize an organization to adopt. • Develop an implementation/transition plan for early adopters. • Provide guidance to existing and potential developers incorporating machine intelligence to understand how it may best integrate within the TLA ecosystem. • Develop a set of learning activities sufficient to demonstrate the chosen use case(s) to mitigate a deficit in the diffusion enabler of observability by others. • Consider implement a plan for community and capacity building.

All data gathered and analyzed represented an attempt to capture a baseline of activity and performance for the first year of a multiyear research, development, and assessment endeavor. They are conclusive only for determining the development of year one and do not necessarily predict future performance of the system or human subjects. The first year of the design-based research process for assessing the TLA development was successful. It produced baseline findings concerning the architecture as currently defined and an operational view of the architecture.

From an initial formative assessment point of view the ADL Initiative also was successful. Feedback from the panelists concerning the need for solutions the TLA is targeting was strongly positive. Additional assessment addressing all levels of ADL objectives will be undertaken in further assessments.

Contents

1. Background

The mission of the Advanced Distributed Learning (ADL) Initiative is to provide policy oversight and support associated innovation and coordination across the Department of Defense (DoD), DoD coalition partners, and other Federal agencies for distributed learning (DL) activities and technologies. DL is used in education and training to refer to learning experiences that are distributed across a variety of geographic settings, time, and various interactive media (Dede 2004). As described in the DoDI 1322.26, DL is "learning content and systems, mediated through technology, which are accessed through a network or experienced via portable media" (Department of Defense 2017). In supporting interoperability of DL technologies, the mission of the ADL Initiative promotes personnel readiness by ensuring that the right people receive the right training at the right cost. DL interoperability relies on the development and use of common specifications and standards that are under the stewardship of the ADL Initiative. For effective stewardship, the initiative conducts ongoing research and development, tests new technologies, and reviews software specifications, standards, and best practices for their implementation. These activities occur primarily under the ADL Initiative's DL Modernization line of effort.

With the advent of the Internet and especially the World Wide Web, not only browsing the web but intentionally leveraging the advantages of bridging time and distance through Web access was attractive to the learning industry. Online learning or DL courses (essentially a collection of Web pages with embedded graphics, videos, and various learning interactions and assessments) became common, along with methods to allow learners to authenticate themselves, register for courses, complete courses, and take assessments (Gallagher 2010). These systems, called learning-management systems, are still in place today performing the same functions, albeit with more advanced components and functionality.

This report provides an assessment of the first cycle of a significant multiyear research and development project to support the modernization of the government's distributed learning technology, strategy, and infrastructure. The project, called the Total Learning Architecture (TLA), intends to produce viable interoperability specifications and standards to implement the concept of a future "learning ecosystem" comprising interconnected learning opportunities, supported by technology, driven by data, and integrated with other talent management capabilities (Gallagher et al. 2017).

A. The Need for the TLA

Learning materials initially were developed to work with, and communicate or send data back and forth with, only one type or brand of learning management system (LMS). As an early adopter of DL, the DoD became aware that the tight coupling of learning materials to a particular LMS was dramatically increasing costs due to the need to redevelop the same materials repeatedly. In 1998, the ADL Initiative, working closely with industry and with other government and civilian organizations, developed a set of packaging and reporting guidelines that allowed any distributed learning materials to be run in any LMS. This solution, based on a collection of data specifications and standards is called the Sharable Content Object Reference Model (SCORM). It was developed by an industry-government collaboration for the ADL Initiative. SCORM specified how to correctly display learning material in a browser and how that activity was to exchange a small amount of data with the LMS (student name, start time, pause/resume times, finish time, quiz scores, and other data) regardless of the LMS in place (Gallagher 2010). The DoD, as a perennial market leader, required all procured learning materials and LMS products to conform to the SCORM model. The result transformed the distributed learning industry—innovative new LMS products were introduced without redeveloping the materials, and distributed learning publishers created materials that were sold and deployed across all the dozens of different LMSs on the market. As a result, SCORM is in broad use internationally and is a de facto standard.

1. The Effect of Changing Technology

With advances in technology and learning science, the traditional LMS model is transitioning to the use of multiple distributed systems (Web services and stand-alone immersive trainers). The change causes a need for learning content access and the tracking and reporting of learners' progress across a variety of systems. These DL technology advancements combined with innovative implementers have, over the last 20 years, significantly evolved the way in which teaching and learning occurs online. For example:

Individualization—the fixed-curriculum sequence is dropped and learners' experience is tailored to their performance, background, learning progress, and other contextual information.

Competency-based training and training—materials and assessments are presented or made available that are most relevant to learners' current state and learning objectives.

Immersive learning and practice environments—simulations, multiplayer games, virtual reality, and augmented reality allow realistic low-cost distributed rehearsal activities.

- Personal learning environments—learners (especially autodidacts) compile and use multiple resources to reach learning goals.

- Smart system and artificial intelligence—these can teach, coach, answer questions, grade papers, offer guidance, recommend activities, alert instructors, and provide continual novel methods to assist.

As these developments emerge and adopt new technological and pedagogical directions, there are at least two primary limitations on the LMS/SCORM model:

1. The range of interactive environments on current devices can no longer be managed by an enterprise LMS running a SCORM "player." Distributed learning applications may exist on varied apps, stand-alone systems, or even cyber-physical crosscutting applications. SCORM, however, has limited ability to take advantage of how, when, and where learning applications exist and communicate data.
2. The amount of data that needs to flow across current systems is more extensive and varied than SCORM permits. More data are needed by new developments and especially by applications applying intelligent tutoring approaches, which use vast amounts of data to make better decisions and recommendations for guiding learning.

2. The Experience API

SCORM's limitations have become increasingly apparent over the last several years. Issues include a need to exchange data about the learner's activity across several systems in real time and a need to track the learning experiences of individuals beyond frame-based computer-based training. These issues led to launching a capability called the Experience API, or xAPI, by the ADL Initiative. The "API" in xAPI refers to application programming interfaces (APIs), a standardized set of "rules" (code) that allows software systems to communicate with each other in prescribed ways, facilitating their standardized integration. The xAPI approach assumes that data about learning experiences and outcomes are distributed across a wide variety of sources, platforms, learning approaches, and technologies. They must be accessed and analyzed to support the growing domain of learning analytics (ADL Initiative 2017)—the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and learning environments (Learning Analytics and Knowledge 2011). The xAPI has also evolved as learner data are shared by providing:

- A new, flexible data format that allows a learning activity to post many types of information about the learner's activity, progress, and context.
- A new cloud-based database product, the Learning Record Store (LRS), that accepts input in the form of xAPI statements and allows other processes to query and analyze the data in various ways.

The xAPI is still maturing and is being continually refined by the existing xAPI community led by the ADL Initiative. There are many commercial LRS products on the market today. Innovative organizations are using this technology in several ways, among them providing early warnings about students who may drop out; correlating job performance with training; and tracking how readers are using or misusing electronic publications. The xAPI addresses one aspect of modern distributed learning, the communications layer, and provides a foundation for the evolution of current DL technologies.

3. The Total Learning Architecture

The xAPI also functions as a primary component under the ADL Initiative's strategy for modernization, culminating with the TLA project. The TLA project explores and tests a new way to organize distributed learning systems for a learning ecosystem and addresses the limitations of the LMS/SCORM and other existing DL models. The goal of the TLA project is to understand and standardize data communications that support distributed learning advances and facilitate creation of a learning ecosystem consisting of many small applications and micro-services. Together these applications and micro-services create an emergent federation based on common standards and specifications.

B. Purpose and Scope of the Study

The primary purpose of this study is to assess whether design goals for the TLA are being met and to determine what improvements may need to be incorporated to consistently reach those goals. The focus is on the first cycle of a multiyear iterative research and development approach. The work provides recommendations for research and development as the project enters its next cycle. This assessment has three longitudinal goals:

1. Iteratively assess the state of the TLA design in accord with TLA goals following the design and development cycles.
2. Evaluate design changes as the TLA design evolves in response to TLA goals and its operating context.
3. Evaluate the evolving potential of the TLA to support learning and add value to TLA end users and stakeholders.

Keeping in mind the evolving nature of the longitudinal goals, the scope of this assessment report is limited primarily to goal (1) and partly to goal (3). The first goal is addressed by all the following assessment questions for year one, and the third goal is addressed partly through the third question:

1. Is the TLA design technically viable as described in its technical documentation?

2. What is the diffusion potential of the TLA based on its technical documentation and operational testing of an initial or reference implementation[1] involving human subjects (users)?
3. How well do the TLA specifications perform, what is the user experience with them, and can quality learning occur?

[1] In software development, a reference implementation is the standard from which all other implementations and corresponding customizations are derived. A reference implementation is, in general, is an implementation of a specification to be used as a definitive interpretation for that specification (NIST, 2003).

2. Current State of the TLA Research and Development

The ADL Initiative using a human-systems integration approach for research and development of the TLA uses cyclical iterative development activities intended to span at least 3 years. This approach, which involves coordination with stakeholders, agile development processes, formative testing, and emphasis on human and life-cycle factors, has currently finished the first development cycle spanning the fourth quarter of CY 2016 through the second quarter of CY 2017. This first and initial development cycle is referred to as Year-One.

A. Initial TLA Architecture

The initial TLA specification included 10 APIs allowing communication between services and components concerning issues such as learning activity, learner profiles, competencies, learning evidence, and user authentication. Such services and components exist as either front-end or back-end applications. Front-end applications support a specific functionality and may have direct interaction with a user or learner. Back-end applications, however, are cloud-based services that support broad functionality across one or more front-end applications. (Gallagher et al. 2017).

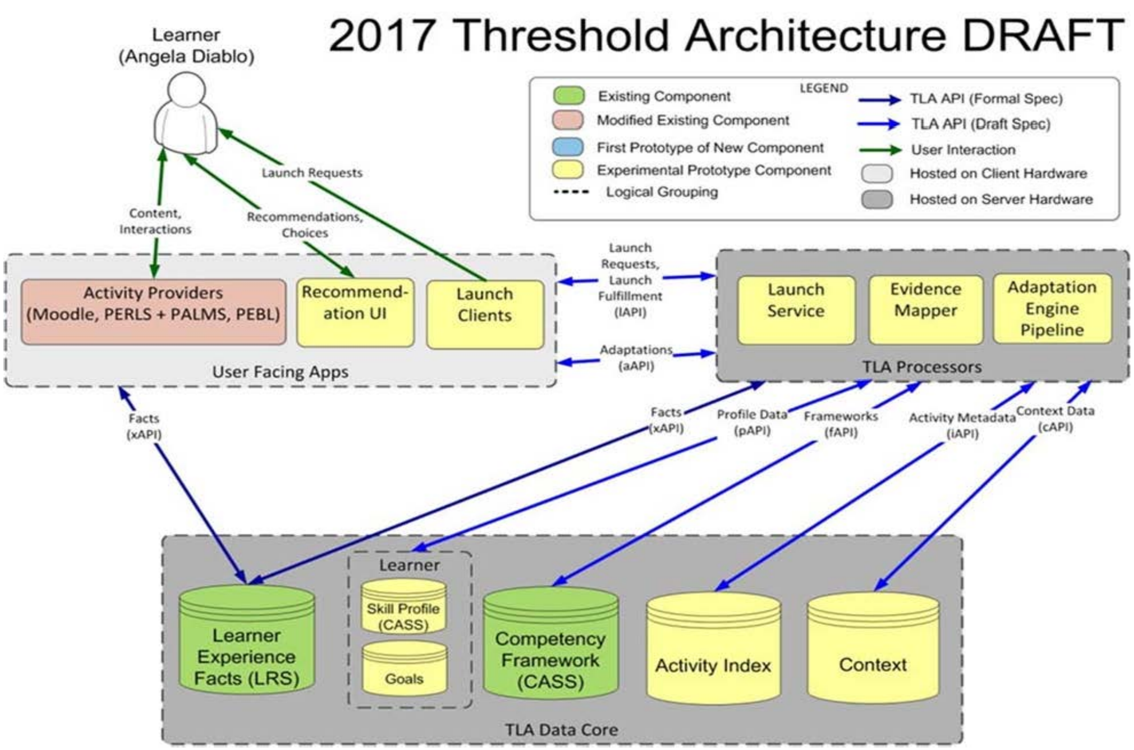
The TLA comprises APIs and can be best described in terms of their functions, and what types of components would most likely use them. The following section describes the initial, or Year-One, architecture by focusing on API descriptions and types of components that could typically comprise a TLA-enabled learning system.

1. Year-One Architectural Description

The Year-One architecture consists of a suite of 10 initial APIs designed to communicate data about learners, competencies, learning activities, circumstances or context, adaptation, and system support. With the exception of the xAPI, the other nine have been developed specifically for integration within the TLA suite of APIs or were otherwise existing open specifications (i.e., OpenID); they have also been integrated or developed tangentially within the ADL Initiative's research program.

For implementation purposes, note that a given instance of the TLA is not forced to use every API. The intention is to provide the APIs that meet the needs of a specific context or application. These APIs when considered together are referred to as an application profile of the TLA. For example, it may be that all a specific implementation should do is to communicate competency data for updating a future credentialing component. In this case, the APIs for that application would be used for that specific implementation. To show the entire range of functionality intended by the initial architecture, Figure 1 depicts archetypal components and the communication architecture with the labeled APIs.

To move from archetypal components to an actual implementation, an operational instance, earlier referred to as a reference implementation, was developed. Figure 1 shows movement in that direction with the labeling of various components such as the Learner Experience Facts database or the Competency and Skill System (CASS) competency framework database and to determine what improvements may need to be incorporated to consistently reach those goals.



Source: Soar Technology, Inc. (2017).

Figure 1. TLA Initial Architectural Diagram

B. TLA Reference Implementation (TLA-1)

Specifications and data models are abstract concepts in technical documentation. To determine if the design reflected in the documentation enabled a system to meet the design goals, these concepts had to be made concrete. A reference implementation was developed to instantiate the design, facilitate understanding of technical challenges, and identify specific points of reference for assessment labeled TLA-1. Technical documentation accompanied the reference implementation to reflect the current design state of the TLA. This documentation included guidance for implementing the communication protocols and data models needed to enable data flow between nodes, component systems, or services in an interoperable manner. The guidance also aided discovery or subscription to a service, allowing service sets to be assembled or disassembled statically or dynamically for adaptable learning experiences. Development activities then provided both a reference implementation for a direct assessment and technical documentation for an indirect assessment of the architecture.

3. Research Design and Methods

A. Design-Based Research

To accommodate the cyclical human-systems integration approach, understanding whether design aligns with design goals can be approached through research. Design of any modern software architecture is a complex problem (National Science Foundation 2007). This complexity is especially evident when coupled with the aim to advance the understanding of what constitutes the TLA and the most desirable characteristics and processes necessary for its implementation, while facilitating efficient and effective learning. Achieving these aims is an open problem within the TLA development community and congruent with a design-based research (DBR) approach (Plomp and Nieveen 2009; Kelly 2009). Therefore, in this report, the assessment procedures apply a multiyear DBR and associated methods (Gallagher et al. 2017), with emphasis on the First-Year transitioning into the next.

Methodologically, DBR uses iterative formative evaluation incorporating generalizable design approaches (approaches that contribute to the greater body of DBR) common to both instructional design and systems engineering. It is loosely defined as a cyclical set of five activities: (1) analysis, (2) design, (3) development, (4) implementation, and (5) evaluation. Each activity is itself cyclical, creating the potential for nested designs. These activities translate into a generic DBR model with the following phases: contextual problem definition, identification of potential products and design principles, tentative products and theories, prototyping and assessment of preliminary products and theories, and problem resolution and theory advancement (Wademan 2005). Systematic reflection and documentation are used continually to codify the process and facilitate understanding. At the end or final cycle, the entire process is portrayed through a retrospective analysis that includes lessons learned leading to the explication of design principles and their connection to the conceptual framework (Nieveen, McHenney, Van den Akker 2006). The second and fourth phases generally iterate within a cycle, and the entire DBR model cycles as necessary until it is agreed that the initial and goal states have achieved a desired balance. Reaching this balance requires substantial collaboration between end users and researchers. In the TLA development and assessment, this collaboration extended to system developers and stakeholders.

Adapting Wademan's (2005) model required a minor variation. IDA and the ADL Initiative collaborated with system developers and stakeholders through the first two phases and then advanced to the third phase of design-based research—design and technical refinement. The fourth and fifth phases, which encompass this report, began in October 2016 and have culminated in the phases of data analysis and preliminary outcomes. Figure 2 illustrates the adapted model with general timeline information. Assessing each cycle and providing design input toward the next, the DBR is anticipated to iterate for at least three complete cycles, with a final set of specifications for the TLA. For this report, however, the scope is for Year-One, or the first complete cycle of TLA research, including development and assessment activities.

In design-based research, data-gathering methods vary depending on the phase and the questions being answered. Analysis activities generally focus on dimensions ranging from user experience and usability to learning achievement. Because the TLA is an enabler for effective digital learning interventions, user-facing affordances must also be addressed, as well as assessment and evaluation of specifications and data models.

1. Design-Based Research and Diffusion Models

The goal of producing a learning ecosystem supported by standardized data communication protocols for use by many different communities and constituencies creates additional design questions. These questions are based on the need for the outcomes to be broadly adopted and diffused among all constituent communities, thereby establishing a need to align the design-based research model with that of the diffusion of innovations.

One method of this alignment is found in the design process. The diffusion literature identifies five attributes of an innovation that are necessary for successful adoption. These attributes can be considered and manipulated for the best chances of success. First is the degree to which an innovation is perceived by users as being better than the idea it supersedes. It is measured in terms that matter to these users, such as economic advantage, social prestige, convenience, or satisfaction. Second is compatibility with existing values and practices. Third is complexity, the degree to which an innovation is perceived as difficult to understand or use. Fourth is trialability, the degree that an innovation can be experimented with low levels of risk. Fifth, is observability, or ease with which individuals can see the results of an innovation. The more observable an innovation is, the more likely it will be adopted (Rogers 2003).

Another method of aligning with diffusion is found in the overall research and development process. A common shortcoming among many design and development initiatives funded by the Federal Government is premature termination, ending with the publications that are produced by the researchers with the intervention never being used or adopted by the intended recipients. The Integrative Learning Design Framework (ILDF) was developed to address this issue. It articulates specific steps within Wademan's (2005) phases of design-based research, practically aligning it with diffusion theory (Rogers 2003). The ILDF overlays an overarching iterative phase called "web-enabled proto-diffusion” across the steps of design, formative testing, refinement, implementation, and evaluation. This allows the process itself to support diffusion through the methods and populations chosen during these phases. The model also recognizes adoption, adaptation, and diffusion together as an integral step in research and development (Bannan 2009). The ILDF indicates that diffusion should be included within the detailed design for the TLA to have its intended impact.

Under Bannan’s (2009) model, the TLA underwent detailed design and advanced to formative testing, followed by refinement, end-user testing, and evaluation and dissemination of results after Year-One of TLA development. To bridge the gap from research into practice, diffusion concerns should be addressed during the detailed design phase to enable evaluation in formative testing. Because the model is iterative, however, knowledge and theory building for diffusion may emerge after formative testing during theory and system refinement or even later in preparation for the next iteration. The ILDF framework assumes the current state of the TLA design and development within. By considering both the five design attributes and the ILDF model, the diffusion potential of the TLA can be fully assessed at this stage.

B. Method

To focus the assessment and facilitate analysis, the study was guided by specific assessment-oriented research questions, presented again here for emphasis:

1. Is the TLA design technically viable as described in its technical documentation?
2. What is the diffusion potential of the TLA based on its technical documentation and operational testing of an initial or reference-implementation involving human subjects (users)?
3. How well do the TLA specifications perform, what is the user experience with them, and can quality learning occur?

Answering the questions outlined above required methods that could assess the architectural specifications’ potential for implementation and for adoption or diffusion throughout the considerable international ADL Initiative constituency. As there are only abstractions and concepts embedded within such documentation, more traditional methods of assessment alone would be insufficient. The assessment questions also required an instantiation of the architecture in a system for actual observation and testing of data, data types, data flows and user interactions. The reference implementation described previously was used for implementing, observing, and testing the architecture and user affordances.

Answering the questions required the use of multiple methods. The Delphi method was used to understand the viability of the architecture and its potential for diffusion.

Concurrently, Operational and Human Subjects Testing was used to assess the implementation of the architecture and suite of APIs. This relied on mixed methods to understand the behavior of a TLA-enabled system with human subjects.

1. Delphi

TLA viability and diffusibility were assessed using the Delphi method. This method allowed researchers to identify and explicate implicit expert knowledge concerning the technical documentation and how it might be actualized. This method works by using a series of questionnaires to collect data from a select panel of experts (Hsu and Sandford 2007). Current information and communication technology tools allow new means and modes of collecting data beyond traditional questionnaires. Delphi uses multiple rounds of data collection and analysis to reach consensus. Each round may and most likely will be shaped by the data and analysis produced by the preceding round. Delphi approaches typically incorporate two or possibly three rounds; however, the number of rounds is determined as the study unfolds.

The validity of Delphi studies in addressing information systems issues has been empirically affirmed by design and application of a framework for information architecture for business networks (Day and Bobeva 2005). The method and procedures in the current study build on this prior work. Delphi methods are "particularly suitable for situations where the gathering of subjective judgements moderated through group consensus is the only approach possible in the absence of precise analytical techniques..." (Linstone 1978 as cited in Day and Bobeva 2005). Various RAND studies have found that Delphi methods produce more reliability than individual judgment (Helmer 1963; Brown and Helmer 1964; Dalkey 1969). Ziglio (1996) affirmed that Delphi methods should be used when seeking clarity and informed judgment on the nature of the problem at hand.

Two primary validity risks present within Delphi methods are accuracy in qualifying experts and maintaining anonymity among the panelists. To reduce the qualifications risk and to guide the Delphi, a Technical Advisory Board (TAB), an initial qualifying rubric, and initial qualifying questionnaire were created. The board guided creation of the initial qualifying rubric for panelists and assisted in piloting instruments to provide a broader analytical lens and guidance for the next round of questions.

A primary characteristic of a Delphi study, and one of its historic advantages, is maintaining participant anonymity. This can reduce the effects of dominant personalities in group dynamics, such as manipulation or coercion to conform to a particular viewpoint (Hsu and Sandford 2007). Anonymity was maintained by strict adherence to an anonymity protocol for all email correspondence and survey administration. Anonymity was broken after the last round of questions and analysis, leading on August 2, 2017, to public acknowledgment of the panel at an open TLA Working Group meeting sponsored by IDA.

2. Operational and Human Subjects Testing

The operational characteristics of the technical specifications, user experience, and learning potential were assessed through human subjects testing. This testing, which occurred from April 10, 2017, to April 21, 2017, provided operational and functional testing of TLA-enabled components to feed back into the design-based research process.

The first goal was to understand if the initial specifications and components using the TLA specifications could function together. This trial with 73 users for 4 days was the first functional test of TLA-enabled components working together for a specific learning purpose. A secondary goal was to test the potential for knowledge gains after sustained interactions by learners with the TLA-instrumented components assembled in this first reference implementation (hereafter known as TLA-1). A third goal was to understand the end users' interactions with and reactions to the TLA-1, as well as the stakeholders' perceptions.

To meet these goals, users interacted with the TLA-1 in a quasi-experimental, within-subjects research design. This step facilitated data collection to support all three goals—users interacting with the TLA-1 to produce system interaction data, learning data captured through pre/post methods, and user experience data to be captured throughout and after the interaction ended. Using a mixed-methods approach, data were collected from TLA APIs (including xAPI), basic system component logs, contextual time-stamped incident logs, learning gain data, and quantitative and qualitative user experience data. These data were analyzed for understanding end-user interaction experiences, reactions, learning gains, and relative performance of the TLA-1 both functionally and technically.

a. Research Design

Table 1 represents the data-collection points, treatment, and dates in the research design. The pre-and post-tests are denoted by O1 and O4, respectively. Other data points indicated in the design were for the purposes of collecting demographic, user-experience, and system-level data. The timeframe for data collection is also represented in Table 1.

Table 1. Research Design

Dates	10 April 2017	18–21 April 2017	21 April 2017
Events	O1 – Pre-test O2 – Demographics	X – Treatment (TLA-based learning) O3 – System log, xAPI data collected	O4 – Post-test O5 – User-experience survey O6 – User-experience focus group

The following are descriptions of the data collected as indicated by Table 1:

- O1 – Pre-test and O4 – Post-test—162-item multiple-choice test.
- O2 – 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).
- O3 – System log data—Data collected from the built-in TLA APIs (including xAPI), basic system component logs, and contextual time-stamped incident logs, including the "Big Red Button" described later.
- O5 – User-experience survey—Combined System Usability Scale (SUS) user-experience questions plus five researcher-developed questions about participants' engagement and specific TLA experience.[2] According to a paper from the *Journal of Usability Studies,* "The System Usability Scale (SUS) is an inexpensive, yet effective tool for assessing usability … [that can] aid in explaining results to non-human factors professionals" (Bangor, Kortum, and Miller 2009). SUS is calculated through a 10-statement survey where users designate to what extent they agree or disagree with each statement as it pertains to their experience with the system in question, in this case the TLA reference implementation. The form has been demonstrated to be fundamentally sound, and its statements are worded both positively and negatively.
- O6 – User-experience focus group—Seven participants provided freeform 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 was used.

a. Treatment Design

The treatment consisted of participants interacting with a variety of learning activities delivered by TLA-enabled learning activity providers. In addition, two overarching badging activities provided end goals that were supported through interaction with the TLA learning activities. Participants interacted for 10 hours over five days. During this time, providing learner scaffolding was an on-site human coach and subject matter expert who also assessed that attainment of either of the two badges. The learning activities centered on six terminal learning objectives (TLOs) within the cybersecurity domain. These TLOs functioned as competencies for functional testing purposes and pedagogically to advance a learner toward attaining a badge as a social engineer or as a cyber apprentice:

[2] 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 freeform short-answer questions, which encouraged participants to elaborate on their specific TLA-1 experiences.

- Describe network architecture, browser configuration, and hardware/software for secure Internet browsing.
- Understand social engineering concepts in cyberattacks (e.g., phishing, waterhole attacks, lures).
- Run penetration tests to locate potential security threats (network enumeration using NMAP, Wireshark).
- Know how to analyze a packet capture (pcap) file.
- Recognize, document, and analyze a successful attack from point of entry, pivoting, and systems controlled.
- Successfully execute the steps in creating an exploit.

For each TLO three enabling learning objectives (ELOs) were also defined at the novice, intermediate, and expert levels. This design yielded 9 distinct objectives per TLO, for a total of 54 total objectives. For Operational and Human Subjects Testing purposes, each ELO was considered a discrete competency and recorded as such in the TLA-1 competency model. Each learning activity (e.g., quiz, article, and scenario) was then mapped to one or more of these competencies to inform competency-attainment tracking, system-wide content sequencing, and more individualized learning recommendations.

The learning objectives and mapped learning activities were also aligned toward one of two overarching learning goals: cyber apprentice and social engineer. By successfully completing a final challenge the learner (user) could attain a badge for attaining one or both goals. Those successfully completing the challenges earned a special stamp or badge on their certificates of completion. These challenges were used to situate the learning activities within the domain and motivate the learners effectively, thereby enabling the TLA-1 to guide learners through activities that would support their successfully completing the challenge. Throughout the learning activities, a human tutor well versed in cyber operations and individual and small group tutoring was present. The tutor, with little to no prior knowledge about functioning as a coach or guide, helped learners progress through complicated tasks. This in effect converted the Operational and Human Subjects Testing into a blended 10-hour learning experience. The tutor also graded the badge challenges and publicly stamped the participation certificates with the appropriate badge.

C. Data-Collection Procedures

1. Delphi

The Delphi study began in November 2016 and ended in July 2017. The purpose and aims indicated the need for two thrusts of inquiry, and the study was therefore executed by two branches. The Technical Branch accessed a targeted pool of technical software developers or implementers who might potentially build products that work with in the TLA. The Diffusion Branch accessed a target pool of decision-makers from organizations that might adopt the TLA architecture.

An initial pool of possible panelists was compiled from a list of past ADL Initiative Technical Working Group members, the Institute of Electrical and Electronics Engineers (IEEE) Learning Technology Standards Committee (LTSC), IMS Global Learning Consortium members, the National Science Foundation CyberLearning Infrastructure program, and various international industry and academic contacts with known experience in enterprise-learning technology projects.

An initial qualifying questionnaire was used to identify panelists. For Round 1, TLA technical documentation, including use cases, was assembled along with an initial 17–22 question survey that included demographic questions. Two packages were developed, targeting either the Technical Branch or the Diffusion Branch. Some panelists qualified to complete both technical and diffusion questionnaires. The process was repeated for Round 2 with different questions based on responses from Round 1 and input from the TAB on gaps in knowledge from the first round. Round 3 used a simpler package and the opportunity to provide virtual or face-to-face input at the TLA Working Group Meeting. During Round 2, a few short Likert-scale surveys were used to rank responses from Round 1 and to gather information concerning diffusion potential. Instrumentation for Rounds 2 and 3 are found in Appendix A.

In November 2016 a cover letter and link to the qualifying survey were sent anonymously to over 200 potential panelists. After qualifying, 54 panelists were sent the Round 1 package according to their branch (technical or diffusion) in December 2016. Results were returned and analyzed in January through March 2017, and new questions were prepared. Because Reference Implementation Testing activities continued through March and April, the Round 2 package was not sent until May 2017, with initial analysis performed through the first week of June 2017.

For Round 3, the panelists were provided a preliminary paper on end-user testing. They were only asked to provide final input for an upcoming TLA Working Group meeting that they were all invited to attend virtually or in person. A total of 8 Delphi panelists attended the working group meeting in person and about 10 other panelists attended remotely online, demonstrating their continuing desire to engage in the research and development process.

2. Operational and Human Subjects Testing

a. Setting and Test Population

The TLA-1 or reference implementation consisted of a TLA-enabled learning system, designed and built with the cooperation of several ADL Initiative sponsored performers, with front-end and back-end functionality in the form of critical services, components, and learning applications (also known as learning activity providers). These cloud-based and mobile services systems were developed or modified to exchange learning data, including competencies and activity streams, using the initial suite of TLA APIs as defined in the baseline version of the architectural documentation. Users accessed the applications through an iPod Touch,[3] iPad, or laptop (depending on the application), with seamless communication of learning tracking and learner data among the devices (Figure 2).

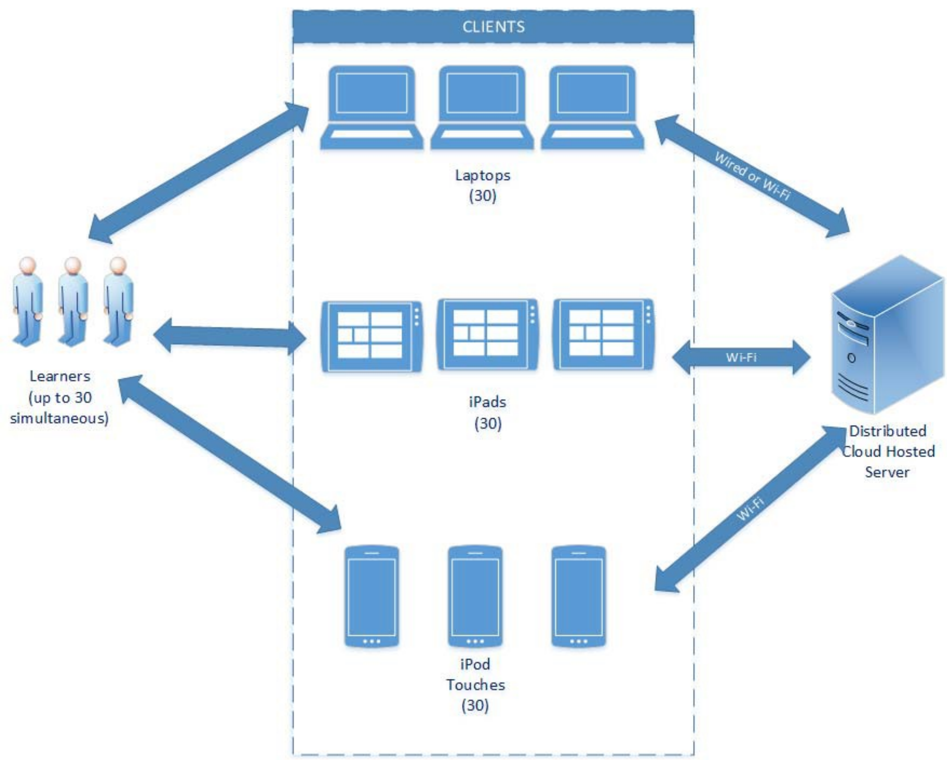


Figure 2. High-Level Testing Topology

The primary application for user interaction was a learner recommender system called PERvasive Learning System (PERLS). Accessed through the iPod Touch and based on a set of cybersecurity competencies and results from an initial pretest, PERLS recommended a variety of different types of learning activities dealing with cybersecurity issues.

---

[3] The security environment precluded use of a mobile phone so the IPod Touch was used as a proxy for an iPhone.

Recommendations were modified to accommodate competencies developed through system interactions that were tracked and communicated to PERLS through implementation of TLA communication specifications.

In addition to PERLS, users were able to access a dashboard providing another entry point to each learning activity without artificial-intelligence-based recommendations. Using the dashboard, any learning activity could be launched at any time. All learner interactions were tracked using the Experience API (xAPI). The dashboard also functioned as a backup for any stress load issues that cropped up with users accessing PERLS simultaneously.

Learning content recommended or accessed through PERLS or the dashboard was provided by learning activity providers. One application (PERLS) acted as a primary access point/recommender as well a learning activity provider. The following is a complete list of the learning activity providers with developer and access modality:

- PERLS, developed by SRI International accessed through an iPod Touch.
- Personal E-books for Learning (PEBL), developed by Eduworks, accessed through an iPad.
- Static content viewer for PDFs and other documents by the ADL Initiative, accessed through a laptop.
- YouTube video player by the ADL Initiative, accessed through a laptop.
- Moodle, a traditional, open-source LMS loaded with cybersecurity learning materials, accessed through a laptop.
- Intelligent tutor built using the Army Research Lab's Generalized Intelligent Framework for Tutoring (GIFT), accessed through a laptop.
- Sero!, a concept-acquisition assessment tool developed by Perigean Technologies, accessed through a laptop.
- Project ARES cybersecurity serious game by Circadence, accessed through a laptop.
- CyberScorpion cybersecurity simulation environment developed by Sandia National Laboratories, accessed through a laptop.

Other services not providing learning content but enabling the communication, data storage, and retrieval are referred to as the back-end, cloud-based services and the data core. They provided competency management, learning-evidence management, learning-profile management, intelligent learning content recommendations, content metadata management, and xAPI data storage and retrieval.

The primary back-end cloud-based services and the data core consisted of:

- Learning recommendations (PERLS), developed by SRI International.
- Competency management, learning evidence management, and learning profile management (CASS), developed by Eduworks.
- Metadata repository and index, developed by the ADL Initiative.
- LRS for xAPI data storage and retrieval (open-source managed by the ADL Initiative).

The physical setting for Operational and Human Subjects Testing was 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. This location housed the array of laptops, iPod Touches, and iPads within a large multipurpose room used for meetings, classes, etc. Several tables and chairs arranged in six pods accommodating groups of four to six subjects were placed within the room. Laptops were labeled using an alphanumeric scheme to facilitate initial registration of all devices within the TLA-1 before any testing activities (Figure 3).

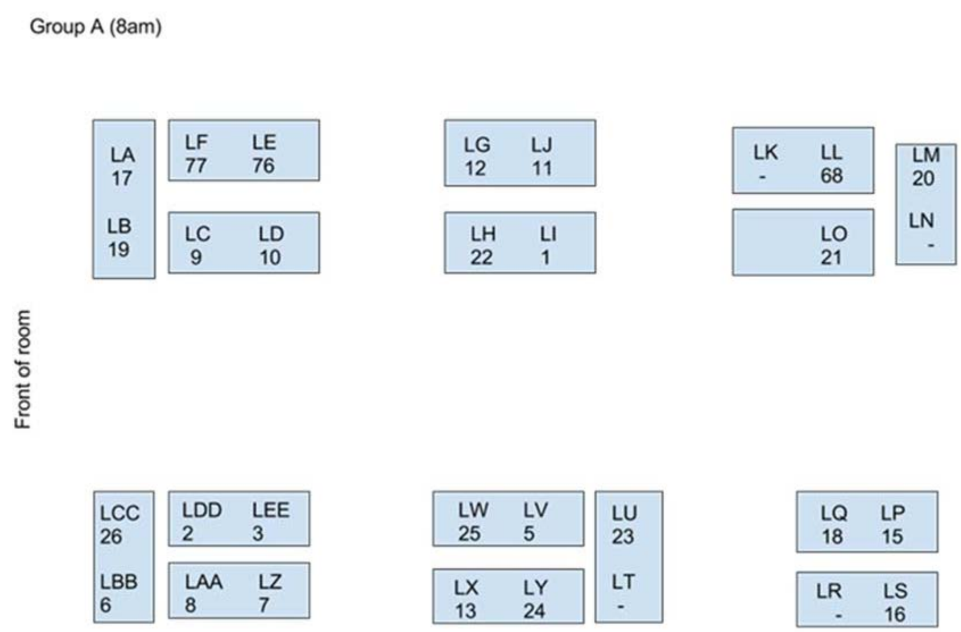


Figure 3. Testing Layout

The human subjects population was drawn from the U.S. Army Special Forces soldiers in active training or out of training and waiting for reassignment, and graduate students attending National Defense University through JFKSWCS. The sample frame consisted of 73 E4s, E5s, E6s, and O3s from primarily 18X (Special Forces), 37F (Psychological Operations), and 38B (Civil Affairs) military occupation specialty codes

(MOSCs). Research using this population was deemed exempt (i.e., judged not to involve more than minimal risk) by the Chesapeake Institutional Review Board.

b. Instrumentation

For pre- and post-testing, a 324-item pool of relevant cybersecurity questions was developed, with 6 items addressing each enabling learning objective. The purpose of the pretest was to determine baseline knowledge about cybersecurity. It was followed by the TLA-1 interaction and a posttest to compare with pretest results for changes in knowledge about cybersecurity. Subsequently, parallel forms (Form A and Form B) of a 162-item test were developed. Form A was used for the pre-test, and Form B was used for the post-test. The pretest was administered 1 week before the treatment or interaction with the TLA-1 and, combined with the use of parallel forms, was used to mitigate practice effects. Both the pre- and post-tests were accessible online.

The System-Usability Scale (Brooke, 1996), Turner's (2011) user experience (UX) survey questions, and a set of 10 researcher-developed UX focus group questions were used to gather usability and UX data. The usability and UX survey questions were accessible online and identifiable only through a user identification number randomly assigned to each individual. A five-question user-engagement survey captured user engagement and final comments at the end of the overall testing period. This engagement survey was developed ad hoc and was not initially considered in the instrumentation. These instruments can be found in Appendix B.

A Java application named the Big Red Button was developed to capture user technical, functional, or other concerns with reference-implementation interactions. The application featured a clickable icon (red button) that when activated opened a text box for free-text capture. When submitted, the text was logged and time stamped for later analysis. The Big Red Button was available to users from the desktops of their assigned laptops.

System-level behavior throughout was obtained using the following data sets: component system logs, the xAPI data stored in the learning record store, and the initial state data of back-end systems (i.e., CASS and PERLS). These data sets were then available for system behavior analyses.

c. Data Collection

On April 10, 2017, participants were given an initial orientation to the TLA and End User Testing activities and schedule. They received an informed consent form, which was signed and returned by those volunteering. Participants were also given log-on credentials and the URL to the online pretest and demographic survey, with instructions to complete them by close of business on April 12, 2017.

On the first day of testing, April 18, 2017, participants chose their testing station, which had a laptop, iPad, and iPod Touch available along with system log-on credentials and a user guide to the TLA components and learning activities. They were then given a more comprehensive orientation on how to interact with the TLA-1, including log-on and launching of learning activities. The first day was longer than later sessions to provide extra time for orientation activities. For the following 3 days users repeated their interaction session with TLA-1. On the fourth day, April 21, 2017, users took the post-test, UX survey, and engagement survey. In addition, six users and one volunteer were chosen at random to participate in a 90-minute focus group.

D. Analysis

1. Delphi

Responses from Round 1 diffusion and technical participants were uploaded into the collaborative qualitative research platform Dedoose. To determine areas of potential design improvement and adoption issues for future iterations of the TLA, the responses were analyzed thematically by question, coded by emergent codes, or coded according to Rogers' (2003) five steps of adopting an innovation (i.e., relative advantage, compatibility with existing values and practices, complexity, trialability, and observability). Round 2 Delphi questions were then developed with input from the TAB as well as insights and questions from Round 1 responses.

2. Operational and Human Subjects Testing

Data associated with human subjects were anonymized using only randomly assigned identification numbers. All human subject associative demographic data or personally identifiable information was kept with SWEG(A) for additional security. Demographic data (O2) were then analyzed for general descriptive statistics.

Pre- and post-test scores (O1, O4) were analyzed for individual differences and group means. Mean scores of the pre-and post-test scores were compared using a paired-samples *t*-test in SPSS. A sample of individual score pairs was used to compare learning paths derived from tracking data collected from the xAPI data set.

xAPI data were analyzed as much as possible for descriptive data about matters such as counts of launching a learning activity, duration spent with a learning activity, and completions of a learning activity. Due to the limited implementation of xAPI in TLA-1, more meaningful tracking data could not be obtained. xAPI data were analyzed descriptively using Kibana open-source analytics and visualization platform. In addition to xAPI data analysis, all qualitative data obtained were input and analyzed thematically through an emergent coding scheme in the Dedoose cross-platform application for analyzing qualitative and mixed-methods data. Due to the limitations of this project, other data sources from log files of CASS, PERLS, CyberScorpion, and Moodle (as well as all the Big Red Button data) have yet to be analyzed fully.

SUS was calculated for each survey response and then was averaged to determine the overall score from all participants. Scoring the SUS required that for each of the odd-numbered questions, 1 was subtracted from the score. Next, for each of the even- numbered questions, the value of the score of the even-numbered questions was subtracted from 5. The new values were then added for the total score, which was subsequently multiplied by 2.5. This value was the score out of 100 points.

4. Findings

A. Delphi

A total of 51 surveys were completed in Round one, where 31 were submitted from the Diffusion Population, and 20 were submitted from the Technical Population. A few participants qualified to participate in both the diffusion and technical streams, so some individuals completed both surveys. In Round 2, a total of 35 surveys were submitted, where 18 surveys were from the Diffusion Population and 17 were from the Technical Population. Again, a few participants completed both surveys. Round three combined technical and diffusion respondents into a single group, and a total of 12 surveys were completed. Ziglio (1996) established the attrition rate for a Delphi study to often be 40% or more per round. The attrition rate for this Delphi study was 31% between Rounds 1 and 2, and 66% between Rounds 2 and 3. Based on the large number of participants who attended the working group meeting (which took place after the close of Round 3) either virtually or in person, this group of participants was a relatively engaged population.

The Delphi findings across both branches coalesced into five main categories: perceived value of the TLA, standards, compatibility with existing practices, security, and documentation. Many findings overlap between categories, and it is difficult but potentially important to separate them. For example, the technical branch participants and diffusion branch participants both agreed that the TLA has potential to bring stakeholders together to create standards and foster collaboration; legacy systems and Defense Information Systems Agency (DISA) rules are barriers; more refined competencies and educational learning objectives across institutions for TLA success are required; and evidence of benefits and return on investment (ROI) for TLA adoption are needed for TLA to become widespread. Findings will be initially presented in accordance with these categories, followed by some specific findings per branch and finally specific combined findings

1. Perceived Value of the TLA

The panel agreed that the TLA addressed serious and pressing market issues and has the potential to open new opportunities such as smart data, performance support, and competency-based technology. Other agreements noted were the value of TLA in producing or leveraging common standards and the potential to share data across applications in ways that haven't been possible before. The panel commented, "All of these would be significant wins," adding that the "TLA project is seen as highly relevant and addressed the problem of disparate systems enabling the integration of multiple systems for richer solutions and presents similarity to other current work."

2. Standards

The panel agreed that standards are useful in making implementation easier or cost effective, but not everyone agreed that standards were always workable. There is implied trust in standards bodies and a desire to avoid reinventing the wheel. The stability that standards bodies provide is seen as positive, and the issue of interoperability is also important; however, using multiple standards or specifications for the same type of functionality creates confusion and frustration. This concern is particularly salient in an international context, as illustrated by one comment: "However, an issue is the fragmented approach to standardization across countries and continents, and the ongoing 'community-based' standardization, in Europe and outside Europe, also resulting from the funding approach to standardization activities."

3. Compatibility with Existing Practices.

Many points the panel made suggested the need for the TLA to be compatible with existing practices. These points span technological, organizational, and cultural matters. Primarily, backward compatibility with SCORM is essential. This issue spans all three areas and must be considered in the TLA development. Other points of emphasis included the following:

- Learning measurement practices, specifically in managing grain size of content.
- Controlling versioning and version management.
- Determining how technology choices are limited by current practice and how that should be accounted for.
- Dealing with organizational and cultural change in implementing the TLA.
- Developing and applying international security practices.
- Locating and including the expertise.
- Accounting for time-switching costs using multiple devices.
- Developing ways to understand and manage costs.
- Privacy.

The panel also noted incompatibilities with existing practices:

- Lack of homogenous infrastructure.
- Scale.
- Poor understanding of data use in meeting learning goals.

- No market demand.
- Forced use of new technologies.
- International privacy laws.

4. Security

Managing security was a primary area of agreement. It is considered paramount to the success of the TLA and should be of the highest priority; even the property of "plug and play" should be subordinate to security. The TLA must meet government security policies, which may be the biggest technical challenge. There should be a model to tie security to classification for government agencies that require or use classified environments. Some panel members specifically noted that security is more of a TLA implementation issue, not necessarily an inherent TLA concern. The need for best practices in this area is important for setting up a TLA environment; the perception of lax security measures with cloud- based hosting remains. One interesting suggestion was to make the security pluggable in a plug-and-play environment.

Some participants reported that the approach was too narrow—one person described it as a "specific vendor's approach." This can be prevented by better outlining TLA terminology to participants in future iterations. These responses indicated that some Delphi participants might not have understood that part of the TLA documentation they received for this study describes only one reference implementation and not all of the TLA.

5. Documentation

The documentation findings spanned both Technical and Diffusion Branches, but some were specific to one or the other. The Technical Branch was more concerned with preparing documentation sets for audiences concerned with technical specifications and implementation details. Standards produced also need to be included within the documentation.

Diffusion panelists had different ideas. Anecdotally, they noted that the documentation needs to provide more use cases and ROI evidence to convince developers to adopt recommended practices, but it wasn't rated as highly as expected when rating relative importance. They also saw a need for specifications and standards incorporated within the documentation. A final point was that software tools and tutorials are needed to encourage and allow for running trial applications. Figure 4 shows the types of documentation and their relative importance.

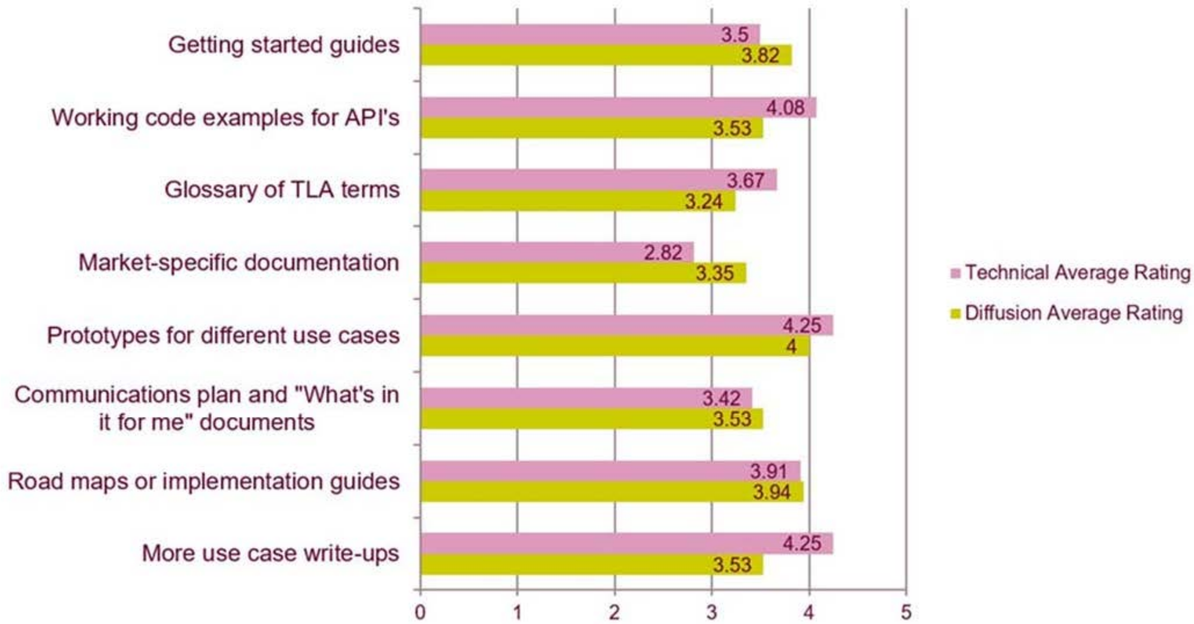


Figure 4. Documentation Types and Ratings

6. Technical Branch

Technical panelists specifically recommended more focus on developing standards, in addition to the work of the xAPI effort, and working toward the development and incorporation of specifications with standards as a goal. They also agreed that marketing "what is truly novel about the TLA" would drive adoption. Many agreed that visibility, market demand, and even a mandate for adoption will be required for widespread use to occur.

One of the most controversial discussions among the panelists was the type of communications protocols that should be used. The consensus was that the Representational State Transfer (ReST) protocol in use with the xAPI is acceptable, but the panelists indicated the need for both synchronous and asynchronous protocols.

Several UX considerations arose. The need to keep UX in mind while designing underlying communications protocols was recognized as being difficult, but it was found necessary for an overall design philosophy.

The panel was asked to provide input concerning what should be included in future designs to help the ADL Initiative make strategic and sound technological choices. Figure 5 shows the technologies that were rated and the technical participants' responses to which technologies should be considered for the next iteration of TLA design and development. Participants were asked to rate the priority of each on a scale of 1–5, along with a blank box to fill in anything that was not included on the list. The top-five responses were learner profiles including competency models, Credential Engine, CMI5, IMS Learning Tool, and publishing to a repository.

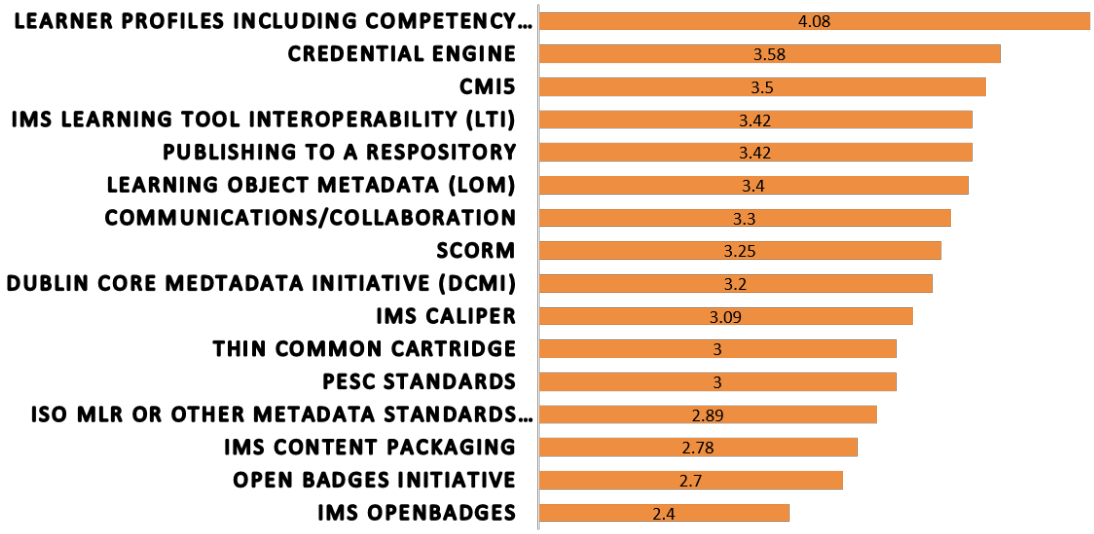


Figure 5. Technologies for Consideration. Acronyms used in the above chart are SCORM (Shareable Content Object Reference Model, PESC (Postseconday Electronic Standards Council), and ISO MLR (International Standards Organization Meta Standard for Learning Resources)

7. Diffusion Branch

A few primary agreements emerged across all rounds of the diffusion Delphi study. One common concern was the difference between a functioning architecture and the quality of the learning experience. This concern arose from questions about instructional design and the current state of DoD Distributed Learning (DL) environments. A few participants agreed that instructional design aspects should be left to adopting organizations and that instructional components should be intuitive enough for faculty, educators, and others to develop or individualize. Others agreed that DoD DL activities will require numerous quality improvements before the TLA becomes advantageous to adopt. One participant suggested that the TLA (or, more generally, the ADL Initiative) could bridge the gap between adopters and existing DL services by establishing a "marketplace" to make the search for complementary products easier or acting as a "matchmaker" for vendors.

Diffusion participants also observed a lack of detail in the existing documentation for data management, storage, and privacy. Specifically, participants mentioned barriers around classified content, legacy systems as a barrier to integrating new technologies, and DISA rules and regulations. This issue was to be expected while the documentation is in its alpha stage. Life-cycle management concerning the responsibility for maintaining and storing data over the course of a student's lifetime was also emphasized.

The panel members further noted the importance of focusing on data-management plans, development of standards, and establishing solid use cases to help with both understanding and marketing TLA for adoption. Adopting it themselves or convincing others to adopt would require good communications plans, cost/benefit data, proof of positive educational effects, and tutorials.

In Round 2, diffusion panelists were asked to consider Rogers' attributes for successful diffusion (described previously in Chapter 2) and determine to what extent the TLA in its current form displays those attributes. Figure 6 shows the average ratings for each. The TLA was perceived as an advantage over what currently exists, but it is not always compatible with current systems, organizational culture, and values. This aligns with previous data provided and indicates that organizational and cultural change present the principal barriers. It is also apparent that the initial architecture and documentation may be too complex to easily implement (lower value for complexity is desirable in the wording of the rating description). This result suggests a need for more simplicity in design.

8. Combined Diffusion and Technical Branches

Figure 6 and Figure 7 contrast diffusion and technical responses to other types of documentation and project priorities for the next 12 months. Prototypes for different use cases were ranked as the highest priority by both groups for types of documentation they would like to see. Market-specific documentation was ranked as the lowest priority by technical participants; a glossary of TLA terms was ranked as the lowest priority by diffusion participants.

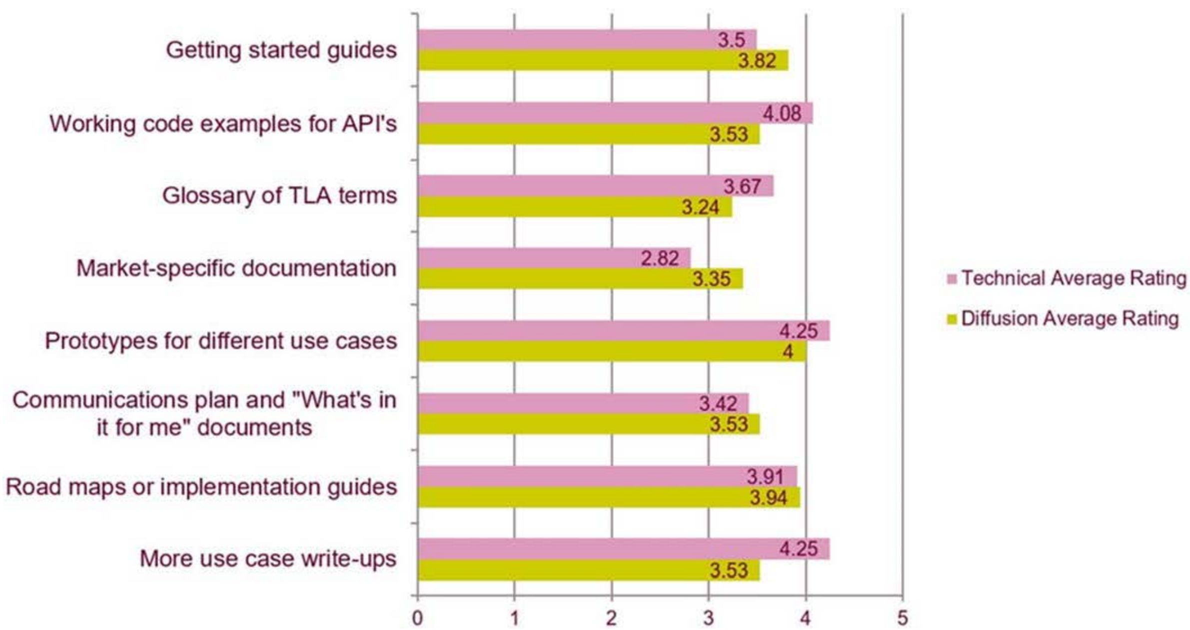


Figure 6. Diffusion and Technical Responses

As far as project priorities for the next 12 months, both diffusion and technical participants ranked security issues as the most critical, and both rated academic workshops as the least critical. Other high priorities included building a community, developing TLA-specific tools, working with commercial vendors, and addressing user-interface accessibility issues. At the third level of ranking (UI/accessibility issues), the two branches diverged somewhat, but overall it was apparent that the priorities were close.

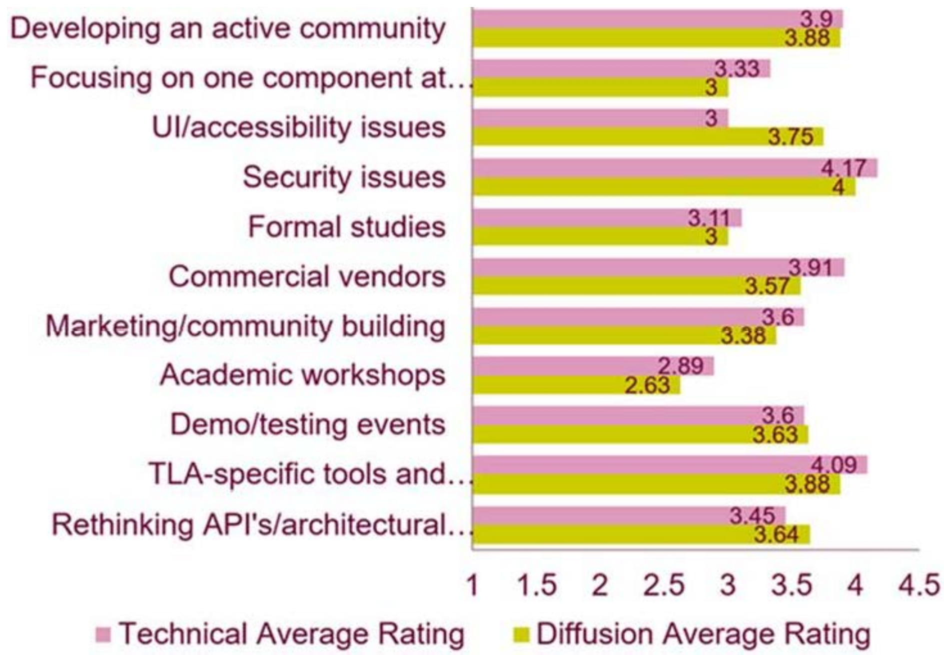


Figure 7. Top Project Priorities

B. Operational and Human Subjects Testing

1. Demographics

Demographic analysis indicated that 70% of the participants were Special Forces trainees; 14% were civil affairs; 3% were psychological operations; and 3% were infantry. Participants also mentioned several other occupational specialties. The age ranges of the participants varied: 58% were 18–25; 36% were 26–30, and 6% were older than 30. After analyzing device experience, the majority indicated strong experience with the iPhone, less experience with the iPad, and even less experience with e-books. Most indicated moderately strong experience with traditional online distributed learning, but less experience with game-based or simulation-based training. Surprisingly, experience with blended Web-based and instructor-led training was also less common.

2. Learning Gains

Of the 73 total participants, only 67 completed both the pre-and the post-tests, which the following analyses were based on. The smaller number of the testing sample was the result of military command and assignment decisions and not based on self-selection. Results from the paired samples *t*-test indicated a (rounded) positive change of 10% ($p <$ 0001). 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. Also of interest were the individual differences in scores. Out of 67 participants, 62 (93%) showed individual gains between their pre- and post-test scores.

3. User Experience and Usability

In addition to the pre- and post-testing for learning potential, other UX and usability attributes were examined. Therefore, capturing and understanding these types of data are critical. During this iteration, we addressed these data via the usability and UX surveys, the engagement survey, and a focus group.

To capture participant experience with the different activity providers and before the usability and UX survey was administered, the short ad hoc exit or engagement survey was given on the last day of system interaction. Participants ranked each activity on a scale of 1 to 5 to indicate how enjoyable or interesting they perceived each. Each ranking question was followed by several free-form short-answer questions that allowed participants to elaborate on their experience with the activity types offered in the TLA-1 instance.

Results indicated that participants’ frustrations with specific activities and the TLA system as a whole largely stemmed from loading and log-in speeds; content that was too difficult for their experience level; and operational issues, including system delays and bugs. Participants also indicated that the opportunity to complete badging activities alongside system interaction kept them engaged throughout the week and, although badging activities may have not run perfectly parallel to the TLA content, they had the highest percentage of positive feedback, with 66% of testers ranking their badging activity as “enjoyable” or “very enjoyable.”

Participants also completed a UX survey on their last day of testing. Their responses provided additional insight into aspects of the TLA-1 that testers liked or disliked, changes they would recommend, and their overall experience with this initial product. When asked to rate the TLA on a scale of 1–10, where 1 was the lowest and 10 was the highest, testers responded with an average of 6.1, but the distribution of answers ranged from 1 to 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 recommend it, and the remaining 12% said they might or “it depends,” based on various conditions. Some changes requested from the “it depends” responders included different content, bug fixes, and introductory lessons on how to use the system as a whole and for the more difficult learning activities.

The SUS was included in the UX survey, calculated for each survey response, and then averaged to determine the overall score from all participants.

The average score was 56.95. According to the SUS score interpretation chart found in Appendix B, that score puts the reference implementation in the “ok” range. For an alpha implementation of the TLA, this is impressive. Commercial systems and products scored with SUS average a rating of 68, and anything above 80.3 is in the top 10% of scores documented (Sauro 2011).

Another UX survey question asked participants to describe the TLA-1 in one or more words. As expected, these responses focused on the learning activities rather than the TLA-enabled capabilities, because their experience was almost exclusively with content. Those who had negative reactions were likely to comment that the information was dense, confusing, or not suited for beginners with little prior knowledge. For participants who did comment directly on the TLA-1’s capabilities, frustrations included glitches such as activities not being registered as complete in PERLS, not having the option to launch activities on the device the participant wanted them on, and communications failures between devices. Participants also responded that for the TLA-1 to reach its full potential, learners should be able to use any device they have. (In the trial run, learners were asked to switch among devices to access different learning activities.)

Positive feedback about the system mostly pertained to the efficiency it provided participants as a learning platform. Examples of this feedback included observations that the content could be limited only to the most relevant and up-to-date information offered to the user, that the system’s flexibility allowed for various learning platforms and activity types, and that there were advantages of self-directed learning. One user commented, “I like how user friendly it is. It is very simple and gets right to the lessons and topics that I need to address without a lot of fluff. This makes navigation and utilization very easy for the end user.” Other comments coming from the focus group suggested that participants wanted more feedback on their learning progress and guidance when they got something wrong; they needed opportunities to practice what they had just learned; they would prefer that the learning activities relate more directly to their badging path; and they enjoyed having a person in the room to help when they got stuck.

4. System Data

The xAPI data were used to produce basic metrics only. With more design attention to the xAPI implementation, more robust data could be collected for meaningful and useful real-time or post hoc metrics. Nevertheless, by analyzing the existing xAPI data using the verbs “accessed” and “completed,” it was found that 169 learning activities were accessed, but only 71% of them were completed. This result does not have any bearing on the learning or success of the completions. It only acts as a gross measurement that could be refined with a more robust implementation.

Using xAPI data, the researchers analyzed a random sample of end users' learning paths (learning activities experienced by recommendation or choice) over the course of their interactions, as indicated by the statements generated from their interactions with the system. Analyzing statements gave insight into improvements needed for future iterations in addition to information on how individual users were interacting with individual learning activities. Based on the sample of six end users who's interactions provided the most robust xAPI data, it was determined that a logical flow of activities with a variety of static webpages, e-books, YouTube videos, and assessment concept maps was generally being followed. In cases where users failed, assessment statements indicated they would often retake the assessment instantly and continue to do so until receiving a passing grade. It was also possible to aggregate data sets for a better idea of the background and attitudes toward TLA-1 in terms of their test scores, goal choice, SUS scores, and in-system behavior. Figure 8 shows how those data can be presented in comparing User 16 with User 38.

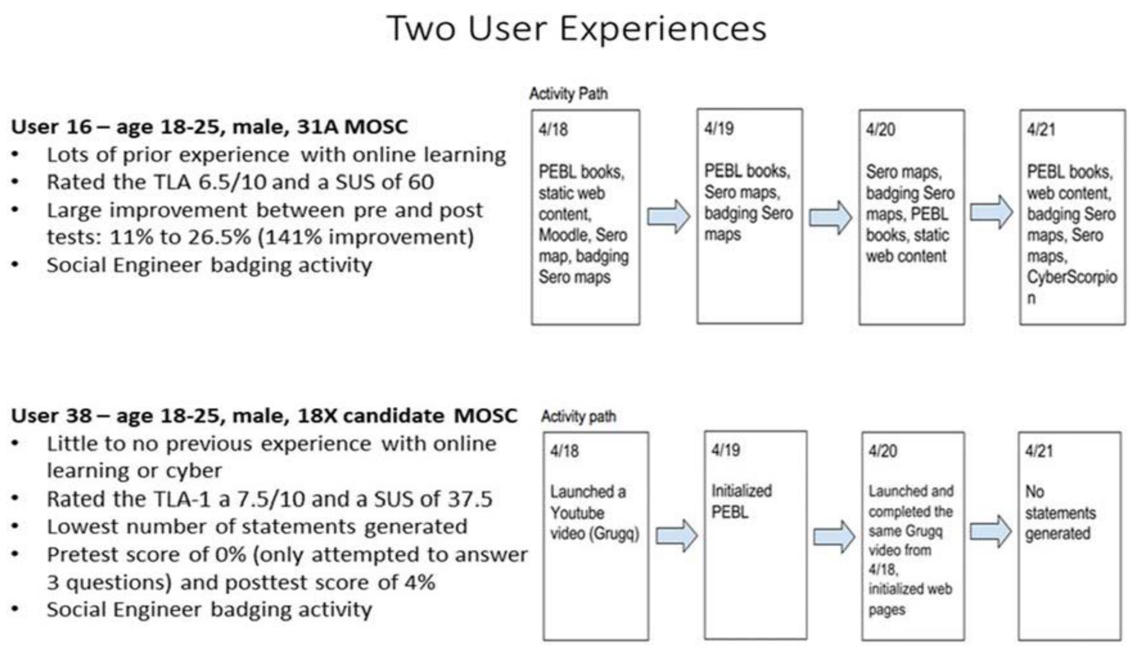


Figure 8. Two User Experiences

The automatically generated xAPI statements tracked only some of the activities. When an end user worked on a badging activity, statements were not generated because of constraints in development time and resources. In further xAPI statement analysis, when the indication was that no statements were being generated for a given amount of time, it was assumed that the user was simply offline and not interacting with the system. But the users could have been completing a badging activity or doing their own research outside the TLA. This result indicated a need for future iterations to include more comprehensive tracking of activities and more streamlined and complete statement generation.

5. Discussion, Limitations, and Conclusion

A. Discussion

The goal of the ADL Initiative's TLA project is to design standardized data communications that support distributed learning and facilitate the creation of a learning ecosystem accommodating a wide variety of applications and micro-services. The primary purpose of this study is to assess whether the goal of the TLA is indeed the right one and whether the resulting research design and activities are successfully meeting the first cycle goal for TLA research and development—to produce an initial set of viable interoperability specifications and standards. The study also provides potential recommendations for research and development as the project enters its next cycle. This discussion centers on the current strategic and tactical success of the design goals, considering the state of the design, the potential for adoption and diffusion of the TLA throughout the ADL constituencies, and recommendations that could be made for redesign before the next development cycle.

Overall, the first iteration of research and development for the TLA was successful in that it incorporated external opinions, insight, and collaboration; generated a functioning reference implementation; and led to collection of baseline data that can be used to assess the TLA in future iterations. However, as with any initial research and development endeavor and especially one as robust and, in some ways, abstract as the TLA, there is a need to reflect on what was learned and what can be improved.

Therefore, this discussion addresses the following:

1. What is the state of the TLA design specifications as embodied within the technical documentation?
2. What is the diffusion potential of the TLA as embodied within both the technical documentation and the operational testing of an initial or reference implementation incorporating human subjects (users)?
3. By interacting with a TLA-enabled reference implementation, how well do the TLA specifications perform, what is the user experience, and can quality learning occur?

1. State of the TLA Design

In its initial state, the TLA design was found to be valuable and necessary for addressing emerging and common problems with today's learning technologies. It may be useful, however, to reduce the number of service definitions in use. For implementing the TLA, incorporating simplicity as a design tenet is desirable, maybe even essential.

The technological decisions are perceived as sound, although some are considered controversial. The controversy centered on types and methods of common and emerging communication protocols (e.g., synchronous vs. asynchronous). However, by Round 3 these issues came to a stable level of agreement, with the consensus on the soundness of the choices made.

Note, however, that several areas of improvement need to be considered in the design. Of primary concern to the Delphi panel was how to address security and governance models. Although security was not within the scope of this first year, it should be a prominent consideration in the architecture or in implementation practices as the TLA evolves. From an international perspective, security and privacy are paramount. They require consideration of international practices that may not be included within the initial architectural design documentation. It was unclear what practices should be considered at this point, but they deserve further exploration.

Another consideration was the incorporation or creation of international standards for the TLA communication elements. With the incorporation of the xAPI specification and other capabilities developed for competency management, there is some movement toward establishing standards. Effort should be made, however, to identify which existing specifications and standards should be incorporated and which should be fostered for development by a standards body. There are possibilities for both incorporating standards and fostering them in the following areas: competencies (frameworks, learner model, and evidence); metadata for learning activities (vs. learning resources or materials); learner context (location, devices available, and schedule); PEBL e-reader; student data governance; and xAPI profiles.

Also to reduce complexity, consideration should be given to facilitating or mapping more than one standard that could be used for the same purpose. Methods should be used to allow standards from different standards bodies to be used interchangeably. This ability could help the international community concerned with improving the consistency of implementations.

The need for standards versus specifications was demonstrated in the Delphi study by their concern about implementation and adoption. Creating or using standards will be essential for a TLA-type innovation to become widely implemented. Continued awareness of and participation in the greater learning technology standards community is essential for that to occur.

2. The Diffusion Potential and Recommendations

As evident by placement of the TLA development within Bannan's (2009) ILDF and the feedback from the Delphi panel, there are many proto-diffusion concerns that could be beneficially considered as the TLA development approaches diffusion. In examining the results of Rogers' (2003) attributes, these concerns were manifested in terms of how concrete or abstract the current state of the TLA is. Understanding what the TLA is intended to support (i.e., use cases) and how it can be tested and observed are critical, along with recognizing, and mitigating the effects of, organizational and cultural hurdles that are in place. What is known is that the idea of a TLA presents advantages over what is currently available within the DL marketplace.

One explicit panel concern was the need for the architecture to support good learning processes, high-quality learning experience, and proof of positive educational effects. Testing in this area should be continued in further development activities. Other concerns are data management and governance, cost/benefit data, and community building. These all fall into an overarching concern about organizational and cultural barriers. Explicitly addressing how the TLA is overcoming or addressing these concerns will help ensure a positive inflection toward diffusion.

To support the process, demonstrations and tests of the TLA in the next design cycle should focus on the issues of importance to potential early adopters and their constituencies. The benefits of TLA adoption should be clearly articulated along with estimates of costs and risks. Several different potential benefits of the TLA were indicated by the first TLA reference implementation.

One suggestion is to focus the next and subsequent versions of the implementation and operational testing events on only one use case or one education/training issue that might encourage, or offer incentives to, an organization to adopt the TLA. For example, the next assessment could focus on demonstrating the TLA's support of a competency-based training program, especially one with ambitions to issue certifications or credentials that require capabilities, components or usages that might drive early adopters toward a TLA-type system. These could include:

- Cross-institution competency sharing.
- Commercial credentials.
- Personal learning assistants or recommendation systems vs. self-directed learning.
- Individualized adaptive learning and other competency-aware systems.
- Learning analytics (dashboard, feedback to learner, intervention, product improvement, research).

- Plug-and-play integration of activities, apps, or tools. For example, there could be a comparison of Personal Assistant for Learning (PAL) algorithms.
- Advanced activity providers such as intelligent tutoring systems (ITSs), Actionable
- Data Book (ADBook)/PEBL, or concept maps.
- Tracking learning across institutions, jurisdictions, and commercial (online) offerings.

Beyond demonstrations and operational testing, the ADL initiative needs to outline an implementation/transition plan for early adopters—possibly one for each of several identified use cases. It may be helpful to include the following in these plans: what components need to be acquired (if any), how are they assembled, and what new jobs and workflows are required. The plans should also address the possibility that some of the TLA back-end components and utilities may emerge as new product categories (e.g., learner competency profile, personal data locker, content repository/index, competency management, or TLA instance manager).

It was evident from the Delphi panel that many TLA use cases involved the deployment of some type of machine-intelligence component such as a personal assistant, intelligent tutoring system, or analytics engine. Beginning in iteration 2, attention should be given to guide the developers of those applications, perhaps through use of the GIFT framework (Sottilare et al. 2012). Paying attention to how machine-intelligence components integrate or are incorporated could help address the question of how to help commercial product developers build systems that take advantage of the services built into the TLA.

A related adoption issue emerged during operational testing concerning the corpus of learning and assessment activities itself. A set of activities sufficient to demonstrate the use cases (e.g., individualized or competency-based learning paths through a corpus) should be chosen or developed. This set of activities will help to concretely demonstrate how the TLA addresses deficits in observability. Adopters will also need to know how to develop a curriculum in the domain they chose and how to determine its cost and return on investment compared with traditional learning materials.

3. Community Building

An outcome that was identified through the Delphi study and was essential for successful diffusion is that of building community and capacity. Because the TLA's intention is to facilitate the development of distributed learning ecosystems, it is important to foster the development of communities throughout which these ecosystems are woven.

Adoption and diffusion are essentially social phenomena, and it is through this lens that barriers and enablers of adoption of the TLA should be understood, along with how the TLA can provide relative advantage and compatibility to members of the adoption communities.

Community building has been occurring in two ways. The first is through the engagement of the Delphi panelists not only in using their knowledge and input to help shape the design of the TLA but also in providing input toward the essential design issues enabling or hindering adoption. The panelists also provided a forum and sense of community for those who have participated in the multiple rounds of the Delphi study.

Second, building community has been and should continue to be fostered through design and testing partnerships such as the one with the JFKSWCS/SWEG(A). This partnership has been invaluable to fielding and testing the first reference implementation (TLA-1) and for expanding the design and testing community to include multiple services, including the Joint and Special Forces communities. This work should continue within the Department of Defense but expand to other domains, Federal agencies, and civilian economic sectors. In addition, the outcome of the second testing cycle should be the formation of a self-sustaining community of implementers, designers, publishers, researchers, and early adopters.

There are other ways in which the necessary community development could occur. It could be accomplished through the ADL Initiative's existing constituencies across the Federal Government, industry, and academia; through an existing or new industry association; or through a standards-development organization. Although there are many open questions about how to support the launch of the community, community involvement is an essential component of the ILDF (Bannan 2009) for full adoption and diffusion and is imperative to evolve the TLA beyond a research project into common use and acceptance.

4. Functionality of TLA Specifications and User Interactions

The functionality of the initial TLA suite of specifications worked as intended during operational testing of TLA-1, with data in the system allowing components to interact with each other as intended. It was also demonstrated during testing that a large amount of data were captured for each user interacting with the system. These data are valuable for analyzing and improving learning experiences both in real time and in an ad hoc fashion. However, in agreement with the Delphi panel, using these data comes with challenges for security, privacy, and ownership. Although these issues were not within the scope of the first iteration of the reference implementation development, they should be addressed next iteration.

A need for consensus on TLA terminology, agreement on what comprises the core system, and how elements should be implemented for future testing were all found in the testing. Addressing these issues should aid in testing, marketing, and explaining the TLA to both adopters and end users. The discussion of generated xAPI statements suggests that consistently applied rules for statement construction and their application are necessary. This need was demonstrated when different activity providers created different statements, making analysis more difficult and less meaningful. For example, an activity was reported as "initialized" when it was opened, but PEBL generated an "initialized" statement every time a page was turned in a book. This led to a disproportionate number of statements generated for the users of PEBL. Refining xAPI "verbs" and associated rules for each type of statement should resolve this issue in the future. In addition, the back-end components incorporated within the initial implementation should be carefully considered for clarity, using precisely what is needed for each specific use case.

The reference implementation and subsequent Operational and Human Subjects Testing primarily incorporated an overabundance of implicit use cases with no explicit use cases defined. This created challenges for answering research questions and attributing success. At least five primary implicit use cases were examined, identified, and tested:

- Competency-based training in cybersecurity, including credentialing.
- Sharing competency/mastery information across apps from multiple vendors.
- Self-directed learning using a recommender system.
- Self-directed learning without a recommender system...
- Real-time learning analytics dashboard.

These and other use cases included the integration of a recommender engine, a human tutor working alongside the TLA system, and badging activities to incentivize completion of learning activities. Using fewer use cases in future iterations should allow for simplified end-user testing and better data generation.

Operational and Human Subjects Testing included learning activities, primarily from research organizations, some with pre-developed cyber content that could quickly include xAPI code within their existing applications to provide data to the TLA-1 competency system and other components. Developing xAPI code occurred mostly independently by each of the developers Granularity of statements - i.e., the level of interactivity reporting, the verbs used, and how those verbs were defined produced inconsistent xAPI enabled reporting and tracking. Consequently, learning activities had the effect of not producing consistent xAPI statements or a coherent learning experience. Because iteration 2 may expand the development of the next reference implementation (i.e., TLA-2), it will be important to guide learning activity developers and instructional designers to take advantage of the TLA affordances such as competency tracking and reporting matched to appropriate recommendations. Considerations could include learning models and applied theories in the context of the TLA, as well as implications for the field of instructional design.

An unexpected finding from the first design cycle concerned the description of learning activities (metadata). It was informally hypothesized that the TLA would require extra metadata to support its vision of multiple delivery devices and accommodation of the learner's current context and preferences (e.g., delivery platform, activity duration, text, or video). It was discovered, however, that introducing a recommender system or personal learning assistant to recommend next steps for the learner may require a fundamental reconceptualization of the description of learning activities. It is likely that each assistant, intelligent tutoring system, or adaptive system will have its own pedagogy and unique characterization of the relevance, difficulty, and prerequisites for learning activities. In cycle one, this issue emerged and was addressed manually with much effort. Going forward, development of metadata remains a key research issue.

As demonstrated through human subjects testing, the considerable learning potential of a TLA-enabled system depends on how it is implemented. The TLA-1 in the context of a blended learning experience enabled impressive learning gains over the course of 10 hours of interaction or instructional time. The specific objectives considered were chosen because of their importance to the target population and were therefore highly motivating, thereby contributing to success.

TLA-1 learning activities combined with motivating challenges or badging activities and a very skilled and knowledgeable coach were all components of the complete system that produced the learning gains. Unfortunately, during this baseline testing event, it was not possible to determine which component produced which gain, so learning gains can only be attributed to the full system plus the learning design in place.

In terms of usability and user experience, the TLA-1 can be considered successful for this stage of development. Positive user feedback centered on the targeted learning resources for the objectives that were already part of the overall schoolhouse curriculum. Having open Internet resources was a helpful component. Most participants recognized the state of the development and could see past it to the future potential.

Positive attitudes toward the learning objectives and learning activities were the result of much work on the design of the learning experience. As with the metadata development, leveraging open resources for learning activities required much manual effort in researching resources and mapping them to an extensive set of enabling learning objectives. Note, however, that the alignment produced through the mapping effort most likely contributed to the relative success of the user experience.

5. Potential Next Steps for the Next DBR Cycle

Based on the range of open issues, lessons learned, and design implications from the first design cycle for Cycle 1 of development, launching the second DBR cycle should include the following steps:

- Define the research questions, outreach goals, and success metrics for Cycle 2.
- Design the evaluation and data-collection plan, including human subjects and site considerations.
- With implementation site partners, choose the use cases for pilot tests, demonstrations, and operational testing, considering subject matter, student population, learning goals and objectives, and user experience.
- Select participating activity providers and other applications.
- Create the learning design that can leverage what components will be included and guide the design of learning activities, including metadata and xAPI implementation needs for learning analytics.
- Cooperatively design the corpus of activities, motivation (badges), competency structure, activity metadata, UX, and xAPI profiles.

B. Limitations

Many factors should be considered in assessing a project of this complexity and magnitude. One factor is ROI. Discussions about ROI have been present throughout the assessment process. Considerations such as how to define ROI overall, how to design costing models for this project, how these models should be applied and to whom, and what measures to use have all been considered. It was determined that a valid approach was to hold focus groups and require each participating to clarify what processes were undertaken in development and how that translated into rough levels of effort and cost. Essentially, the intention was to use this first year as a baseline for understanding what metrics and models might be developed and applied for determining the ROI. This endeavor would have required several site visits and was too resource intensive for the scope of the first year's assessment.

Another limitation due to resource constraints was the ability to analyze all available data. As discussed previously, data were captured from many system log sources, including initial state data on at least two components. These data, although captured and available for analyses, were beyond the resources that could be applied in the first year of TLA assessment and remain to be analyzed in the future.

It should also be recognized that all data gathered and analyzed represented an attempt to capture a baseline of activity and performance for the first year of a multiyear research, development, and assessment endeavor. Therefore, they are conclusive in only determining "what was going on" for the development of year one and not predictive of future performance for the system or human subjects; they are only an indicator of potential. These results should be compared longitudinally to performances in subsequent cycles for any deeper meaning.

C. Conclusion

The first year of the DBR process for assessing the TLA development was successful inasmuch as it produced baseline findings concerning the architecture as currently defined and an operational view of it. Drucker (1993) emphasizes the balance between the strategic and the tactical by doing things right and doing the right things. By using multiple methods, the process could provide assessment of specifications needed to understand if the ADL Initiative is doing the right things and doing those things right.

The feedback received through multiple rounds of the Delphi process, overwhelmingly suggests that the ADL Initiative is doing things right. The positive feedback on the need for the types of solutions the TLA offers was resounding. Determining what right things are being done is more complex, however, because it depends on all levels of objectives and addressing many still open questions. In terms of architecture, open questions remain about technology choices and processes, as well as complexity. However, all indications are that these questions may be the result of particular experience and development practices, not necessarily choices made by the ADL Initiative.

The tested reference implementation (TLA-1) communicated data that resulted in expected actions. Components functioned adequately, considering they were in various stages of development and most were, in fact, individual research and development projects. The APIs used for inter-component communication and the focus of the TLA also seemed to function correctly. Functionally, component behaviors produced results as expected, such as providing a recommended learning activity or resource based on an updated competency.

As a final note, users learned material leveraging the TLA-1 in a very short period as indicated by score gains on the extensive pre- and post-tests. They could also be tracked according to their in-system behavior. These results can be considered as successful, especially for an alpha-stage prototype. The user experiences were on par with other developing systems as measured by the SUS and qualitative UX data gathered. This measure also should be considered as indicating success.

As noted throughout this report, many matters that still need to be addressed. As the research evolves into the next cycle, the knowledge gained from the first round of DBR and the baseline operational performance will allow the next cycle to provide more answers (and more questions).

One question in the user experience survey was, “If the TLA-1 was a car, what type of car would it be?” Two of the answers suggest potential: “Audi or BMW. It’s nice, has a lot of features, but you're not going to fix it yourself and depending on the model, would not be useful in every situation,” and “Apple Car—could be the future but no one is really sure how well it works yet.” But what might be the most definitive answer points out what one might expect in the initial year of this type of research and development: “1973 Volkswagen Beetle … This car has to have a push start. As in the passenger has to push the car until the driver can get it on and in gear, then the passenger runs up to his side and hops in. Annoying to start using, but workable.”

In other words, the first year of the TLA has had a promising start with much potential, but much remains to be done.

Appendix A.
Delphi Data Collection Instruments

Technical Qualifying Rubric

Qualification	Degree Qualified				
	Least				Most
	1	2	3	4	5
Experience implementing enterprise learning technology products and architectures	☐	☐	☐	☐	☐
Experience using various learning technology products and architectures	☐	☐	☐	☐	☐
Experience evaluating strengths and weakness of product functionality and product integration	☐	☐	☐	☐	☐
Experience evaluating architectures using systems wiring diagrams, block diagrams, and technical documentation	☐	☐	☐	☐	☐
Familiarity with past and current learning technology platforms	☐	☐	☐	☐	☐
Familiarity with interoperability specifications and standards efforts	☐	☐	☐	☐	☐
Understanding of current trends in learning technology and underlying related technologies	☐	☐	☐	☐	☐
Degree of application of current learning technology trends and underlying related technologies applied.	☐	☐	☐	☐	☐

Technical Qualifying Questionnaire

1. Are you a professional:
    a. Software developer? ______
    b. Systems engineer? ______
    c. Systems architect? ______
    d. Other? ______

2. How many years have you worked with learning technology software?
    a. 1 to 3 years?
    b. 3 to 5 years?
    c. 6 to 9 years?
    d. 10 to 15 years?
    e. 15 + years?

3. Where are you currently employed? ____
    a. What is your job title? ____
    b. How long? ________________________________

4. Are you familiar with modern systems architecture and the inner workings of learning technology products (LMSs, LRSs, learning technology content, learning technology specifications and standards)? Yes/No ____
    a. How would you rate your familiarity on a scale of 1-5 (5 most familiar)?

       1 2 3 4 5

5. Are you able to read and comment deeply on technical software documentation? Yes/No

Technical Delphi Round One Questions

After analyzing the provided technical documentation of the TLA, please answer and/or comment to the following:

1. As a product developer or systems integrator, does the design of the TLA and the technology choices we have made appeal to you? Why or why not?
2. Would you be comfortable working with the TLA spec, in terms of the technologies used and the complexity of implementation? Why or why not?
3. If you are familiar with xAPI, how do you feel about the TLA's use of and extensions to xAPI?
4. Will the TLA support the full range of new elearning products? If not, what should be supported?
5. What real problem do you think it addresses for you and your customers?
6. Will the TLA support products developed to take advantage of modern distributed systems architecture? If not, why?
7. What better way exist to do this? Could you describe alternative technologies or approaches?
8. What's missing? Are there any use cases or needed functionality we haven't thought of?
9. Did you find the documentation understandable, clear, and complete?
10. Did you notice any ambiguities or missing elements? If so, what are they?
11. How can we make the documentation better?
12. Will TLA instances work? Explain why or why not.
13. Will the TLA adequately support the deployment of data intensive products like interactive ebooks, adaptive practice environments, personal learning assistants, analytics engines, multiplayer games, or intelligent tutoring systems?
14. Do you feel that the effort required to move your software towards TLA conformance is a serious problem? If so, how can we address it?
15. Will the TLA get us close to a plug-and-play architecture for new products? If not, what else is needed?
16. Does the TLA address all the kinds of data that anticipated new products might need to share? If not, what should be addressed?
17. What are the barriers to adoption of the TLA?

18. What problems would you foresee moving from your current product or system towards the TLA?
19. How can we overcome those barriers?

Diffusion Qualifying Rubric

Qualification	Degree Qualified Least 1	2	3	4	Most 5
Current or prior responsibility for applying educational technology in operational settings (e.g. real students, real budgets)	☐	☐	☐	☐	☐
Familiarization with educational and instructional technology products and how they are used	☐	☐	☐	☐	☐
Awareness of trends in educational and instructional technology and the potential of emerging technologies (e.g. ebooks, augmented reality, analytics).	☐	☐	☐	☐	☐

Diffusion Qualifying Questionnaire

1. Who is your current employer?
   a. What is your job title?
   b. How long?

2. Are you responsible for the procurement and deployment of learning technology products and systems?

3. Where do you find out about new learning technology products and trends?

4. In your opinion, what new learning technology, if any, has the most potential for improving training?
   a. Why?

Diffusion Delphi Round One Questions

After analyzing the provided documentation of the TLA, please answer and/or comment to the following:

1. Do you believe that the problems the TLA is designed to address are real and properly understood?
2. As described in the documents you've seen, does the TLA approach make sense?
3. Are you aware of any other solutions for distributed elearning systems being put forward?
4. The following table lists 20 existing and emerging elearning product categories. For each one, please indicate whether:
    a. You are currently using a product in this category as part of your operation
    b. You have plans to purchase or build a system of this type
    c. You think this type of product has potential in your organization, but aren't yet trying it out
    d. You don't feel that this product category would be of interest

Elearning Product Category	Currently Using	Plan to Use	Has Potential	Not of Interest
Learning Management Systems (LMS)	○	○	○	○
Mobile learning	○	○	○	○
Ebooks	○	○	○	○
Personal Learning Assistants (AIs that advise students)	○	○	○	○
Adaptive learning activities	○	○	○	○
Intelligent tutoring systems	○	○	○	○
Analytics Engines	○	○	○	○
Stand-alone learning activities not launched by the LMS	○	○	○	○

Apps for instructors (outside of the LMS) to manage schedule, assignments, status, lesson plans, rosters, grades,	◯	◯	◯	◯
Classroom dashboards or leaderboards	◯	◯	◯	◯
Student learning lockers (personal data store) or e-portfolios	◯	◯	◯	◯
Cloud-based apps or learning activities	◯	◯	◯	◯
Badging or some form digital certification of competence	◯	◯	◯	◯
Content repositories, registries, or aggregators	◯	◯	◯	◯
Competency framework and tools to consistently describe learning objectives, job requirements, etc.	◯	◯	◯	◯
Content authoring tools	◯	◯	◯	◯
MOOCs	◯	◯	◯	◯
Content curation and management tools	◯	◯	◯	◯
Chat, collaboration, communication technology	◯	◯	◯	◯
Augmented reality	◯	◯	◯	◯
Simulations or virtual reality	◯	◯	◯	◯
Games (PC, mobile, multi-player, …)	◯	◯	◯	◯
Other: (list)	◯	◯	◯	◯
Other:	◯	◯	◯	◯
Other:	◯	◯	◯	◯

Other:	◯	◯	◯	◯

5. Would the TLA accelerate your adoption of intelligent products like personal learning assistants, analytics engines, or intelligent tutoring systems?
6. Do you feel that informal learning, student-initiated learning from outside resources and other institutions, and on-the-job learning are important components of enterprise training that need to be managed? Do you manage them?
7. Are you aware of elearning infrastructure issues that the TLA should address, but doesn't currently?
8. Did you feel the TLA description you've received explained the purpose and design of the TLA clearly and completely?
9. Did you have questions about any sections or features?
10. How can we make the documentation better?
11. Have you experienced costs or frustrations associated with the kinds of problems the TLA is intended to address: integrating new kinds of products; incorporating web-based or tablet-based learning activities; etc. Please help us understand the issue(s) you were dealing with.
12. Would TLA-conformant product have addressed these issues, if such products had been available? What else would make it easier for you to evaluate and deploy innovative new products: tools like analytics engines and content like intelligent tutoring systems?
13. If you see the value of the TLA's approach, would you be ready to move toward a TLA instance in your organization?
14. Do you feel that student data security is a major issue in the management of your elearning installation?
15. What are the barriers to adoption of the TLA?
16. Do you foresee any problems in moving from your current elearning systems installation towards a TLA instance?
17. How can we overcome those barriers?

Appendix B.
Operational and Human Subjects Testing Data Collection Instruments

Operational and Human Subject Testing Demographics

Last Name ____

First Name ____

Gender ______

Age _____

Service ____

Duty Station ____

Office/Directorate ____

Duty Assignment ____

If Military, indicate E, W or O and level (e.g. E5) ________

MOSC with a short description of what you actually do

____

____

____

If Civilian, indicate GS level (e.g. GS 11) _______________

Last Education level attained: HS, Some College, BA/BS (or equivalent), MA/MS (or equivalent), Doctorate (e.g. PhD, EDd, DSc, etc.)

Education field of study ____

IT/IA Certificates (e.g. SEC+) ____

____

IT/IA Work Experience (Years) ________________________

Cyber Operations Experience (Years) ___________________

Operational and Human Subject Testing Sample Pre- and Post -Test Questions

(2 items used for the pre-test and 2 items used for the post-test per ELO)

Part I: TLO 1 - Describe basic network architecture, browser configuration, and hardware/software considerations for secure Internet browsing.

ELO 1

Novice - Identify the correct definition of secure Internet browsing including VPN, firewall, plugins

1. Which are the two main types of VPN's
   - **a. Remote Access / Site-To-Site**
   - b. Secure / Unsecure
   - c. Remote Access / Local Access
   - d. Unencrypted / Encrypted
2. Which of the following is an open-source software application that implements virtual private networks techniques for creating point-to-point or site-to-site connections in routed or bridged configurations.
   - a. . IPSec
   - b. PPTP
   - **c. OpenVPN**
   - d. SSH
3. When securing your browser plugin settings, enabling Click-to-Play will allow you to:
   - a. . Run all plugin content
   - **b. Let the user choose when to run the plugin content**
   - c. Disable all plugin content
   - d. Detect and run plugin content for you
4. Which ports will you need to allow for internet browsing (Choose Two)
   - **a. 80**

   b. 22
   c. 25
   d. 443
5. Which of the following is the most popular browser plugin that will kill third party scripts and widgets, pop-up ads and banners?
   a. . Ghostery
   b. AdBlock Plus
   c. HTTPS Everywhere
   d. Ads4You
6. A VPN Tunnel is used for:
   a. . Secure tunnel between two sites via the internet
   b. Unencrypted traffic
   c. Bypassing the firewall
   d. Connecting to the internet
7. The most widely sold solution to combat the challenges of network security is the
   a. Antivirus
   b. Managed service
   c. YARA
   d. Firewall
8. A type of firewall which acts as a proxy for one or more services. Examples include mail filters that weed out spam and web proxies that block or remove undesirable content.
   a. Circuit Gateways
   b. Packet Filtering
   c. Application Relays
   d. Velocity Gates
9. This is the simplest function of a firewall, available as standard on most machines, and is used for mundane firewalling tasks.
   a. Host based security
   b. Basic packet filtering

c. Disk defragmenting

d. CleanFeed mechanisms

Operational and Human Subject Testing UX Questionnaire

ID ______________________________

Part 1 System Usability

	Strongly disagree 1	2	3	4	Strongly agree 5
1. I think that I would like to use this system frequently					
2. I found the system unnecessarily complex					
3. I thought the system was easy to use					
4. I think that I would need the support of a technical person to be able to use this system					
5. I found the various functions in this system were well integrated					
6. I thought there was too much inconsistency in this system					
7. I would imagine that most people would learn to use this system very quickly					
8. I found the system very cumbersome to use					
9. I felt very confident using the system					
10. I needed to learn a lot of things before I could get going with this system					

User Experience

These questions are based on Turner (2011):

1. Would you recommend TLA-1 to a friend?
2. How would you describe TLA-1 in one or more words?
3. If TLA-1 were a car, what car would it be? For example, Ferrari, Cadillac, Toyota Corolla? User your own car type.
4. How does TLA-1 compare to Blackboard or other learning technologies system you know about?
5. If you were to review TLA-1 what score would you give it out of 10?
6. What do you find most frustrating about TLA-1?
7. Overall, how easy to use do you find TLA-1?
8. If you could change one thing about the TLA-1 what would it be and why?
9. What features could you not live without?
10. What do you like best about the TLA-1?
11. What do you like least about the TLA-1?
12. Which features could you live without?
13. How can we improve the TLA-1?
14. Anything else you care to share or get off your chest?

Operational and Human Subject Testing UX Focus Group Questions

1. As you began to interact with it, what were your expectations about your learning using TLA-1?
2. How well did TLA-1 meet your learning expectations?
3. What happened as you interacted with TLA-1?
4. What did you expect to happen when you interacted with TLA-1?
5. How does TLA-1 compare to Blackboard or other learning technologies system you know about?
6. What do you find most frustrating about TLA-1?
7. If you could change one thing about TLA-1 what would it be and why?
8. What do you like best about TLA-1?
9. What do you like least about TLA-1?

10. How can we improve TLA-1?

11. Anything else you care to share?

References

ADL Initiative. 2017. “Part One: About the Experience API.” April 18, 2017. Retrieved from https://github.com/adlnet/xAPI-Spec/blob/master/xAPI-About.md#partone.

Bangor, A., P. Kortum, and J. Miller. 2009. “Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale.” *Journal of Usability Studies* 4 (3): 114–23.

Brenda Bannan. 2009. “The Integrative Learning Design Framework: An Illustrated Example from the Domain of Instructional Technology.” In *An Introduction to Educational Design Research*, edited by T. Plomp and N. Nieveen, 53–72.

Enschede: SLO, 3rd print.

Brooke, J. (1996). SUS: a "quick and dirty" usability scale. In *Usability Evaluation in Industry*, edited by P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland, 189-194. London: Taylor and Francis.

Brown, B., and O. Helmer, O. 1964. “Improving the Reliability of Estimates Obtained from Consensus of Experts.” P-2986. Santa Monica, CA: The RAND Corporation.

Dalkey, N. C. 1969. “The Delphi Method: An Experimental Study of Group Opinion.” RM-5888-PR. Santa Monica, CA: The RAND Corporation.

Day, J., and M. Bobeva. 2005. “A Generic Toolkit for the Successful Management of Delphi Studies.” *The Electronic Journal of Business Research Methods* 3 (2): 103–

16. Retrieved from http://www.ejbrm.com/issue/download.html?idArticle=154.

Dede, C. 2004. Distributed-Learning Communities as a Model for Educating Teachers. In *Proceedings of SITE 2004: Society for Information Technology & Teacher Education International Conference*, edited by R. Ferdig, C. Crawford, R. Carlsen,

N. Davis, J. Price, R. Weber & D. Willis, 3–12. Atlanta, GA: Association for the Advancement of Computing in Education (AACE). Retrieved January 19, 2018, from https://www.learntechlib.org/p/14523/.

Department of Defense. 2017. “Distributed Learning.” DoDI 1322.26. Drucker, P. F. 1993. *Managing for Results.* New York: HarperCollins.

FY11 Department of Defense Human Systems Integration Management Plan. 2011.

Washington, DC: DDRE, Directors, Mission Assurance and Human Performance, Training & Biosystems.

Gallagher, P. Shane, J. T. Folsom-Kovarik, S. Schatz, A. Barr, and S. Turkaly. 2017. “Total Learning Architecture Development: A Design-Based Research Approach.” Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2017. Paper number 17117.

Gallagher, Patrick. 2010. "The Development of SCORM Sections I and II". In *Learning on Demand – ADL and the Future of e-Learning*, edited by R. Wisher and B. Khan, 49–70. Washington, DC: Department of Defense.

Helmer, O. 1963. "The Systematic Use of Expert Judgement in Operations Research." P-2795. Santa Monica, CA: The RAND Corporation.

Hsu, C., and B. Sandford. 2007. "The Delphi Technique: Making Sense of Consensus."

Practical Assessment, Research & Evaluation 12 (10): 1–8.

Kelly, A. E. 2009. "When Is Design Research Appropriate?" In *An Introduction to Educational Design Research*, edited by T. Plomp and N. Nieveen, 73–88.

Enschede: SLO, 3rd print.

Learning Analytics and Knowledge. 2011. Call for Papers. 1st International Conference on Learning Analytics and Knowledge 2011. Retrieved from https://tekri.athabascau.ca/analytics/call-papers.

Linstone, H. A. 1978. The Delphi technique. In *Handbook of Future Research*, edited by

J. Fowles, 271–300. London: Greenwood Place.

National Science Foundation. 2007. "Cyberinfrastructure Vision for 21st Century Discovery." Retrieved from http://www.nsf.gov.

Nieveen, N., S. McKenney, and J. Van den Akker. 2006. "Educational Design Research: The Value of Variety." In *Educational Design Research*, edited by J. Van den Akker, K. Gravemeijer, S. McKenney,. and N. Nieveen, 151–58. London: Routledge.

Orlando, FL: Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2017.

Plomp, T., and N. Nieveen. 2009. "An Introduction to Educational Design Research." In *An Introduction to Educational Design Research*, edited by T. Plomp and N. Nieveen, 9–35. Enschede: SLO, 3rd print.

Rogers, E. M. 2003. *Diffusion of Innovations.* New York: Free Press.

Sauro, J. 2011. "Measuring Usability with the System Usability Scale (SUS)." Retrieved from https://measuringu.com/sus/.

Soar Technology, Inc. 2017. "TLA Design Document." Orlando, FL: Soar Technology, Inc.

Sottilare, R. A., K. W. Brawner, B. S. Goldberg, and H. K. Holden. 2012. "The Generalized Intelligent Framework for Tutoring (GIFT)." Orlando, FL: U.S. Army Research Laboratory – Human Research & Engineering Directorate (ARL-HRED).

Turner, N. 2011. "15 Useful User Feedback Questions for Online Surveys." *UX for the Masses.* http://www.uxforthemasses.com/online-survey-questions/.

Wademan, M. R. 2005. "Utilizing Development Research to Guide People Capability Maturity Model Adoption Considerations." PhD diss., Syracuse University, Syracuse, NY, unpublished.

Ziglio, Erio. 1996. “The Delphi Method and Its Contribution to Decision-Making.” In *Gazing into the Oracle: The Delphi Method and Its Application to Social Policy and Public Health*, edited by M. Adler and E. Ziglio, 3–33. London: Jessica Kingsley Publishers.

Abbreviations

aAPI	Adaptation Engine API
ADL	Advanced Distributed Learning
AI	artificial intelligence
API	application programming interface
CASS	Competency and Skill System
dAPI	Discovery API
DBR	design-based research
DL	distributed learning
DoD	Department of Defense
eAPI	Evidence Mapper API
ELO	enabling learning objective
GIFT	Generalized Intelligent Framework for Tutoring
iAPI	Activity Index API
IEEE	Institute of Electrical and Electronics Engineers
ILDF	Integrative Learning Design Framework
JFKSWCS	John F. Kennedy Special Warfare Center and School
LMS	learning management system
LRS	learning record store
LTSC	Learning Technology Standards Committee
oAPI	Authorization and Authentication APIs
pAPI	Learner Profile API
PEBL	Personal E-books for Learning
PERLS	PERvasive Learning System
ReST	Representational State Transfer

ROI	return on investment
SCORM	Sharable Content Object Reference Model
SWEG(A)	Special Warfare Education Group (Airborne)
TLA	Total Learning Architecture
TLO	terminal learning objective
UX	user experience
xAPI	Experience API

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED *(From–To)*
October 2018	Final	JAN 2018 – OCT 2018

4. TITLE AND SUBTITLE

Bridging the Archipelago: An Assessment of the Advanced Distributed Learning Initiative’s Total Learning Architecture (TLA)

5a. CONTRACT NUMBER
HQ0034-14-D-0001

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Gallagher, P. Shane
Barr, Avron B.
Turkaly, Sarah J.

5d. PROJECT NUMBER
BE-2-4039

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Institute for Defense Analyses
4850 Mark Center Drive
Alexandria, VA 22311-1882

8. PERFORMING ORGANIZATION REPORT NUMBER

IDA Document D-9077

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of the Assistant Secretary of Defense for
Readiness Force Education and Training Directorate
Advanced Distributed Learning Initiative
1901 North Beauregard St., Suite 600
Alexandria, VA 22311

10. SPONSOR/MONITOR’S ACRONYM(S)

USD(P&R)

11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited (19 November 2018).

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This report provides an assessment of the first cycle of a significant multiyear research and development project undertaken by the Advanced Distributed Learning (ADL) Initiative to support the modernization of the government’s distributed learning technology, strategy, and infrastructure and to mitigate the limitations of learning management systems as a total solution. The project, called the Total Learning Architecture (TLA), intends to produce viable interoperability specifications and standards to implement the concept of a future “learning ecosystem” comprising interconnected learning opportunities, supported by technology, driven by data, and integrated with other talent-management capabilities.

15. SUBJECT TERMS
ALD; Advanced Distributed Learning; learning ecosystem; TLA’ Total Learning Architecture

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Schatz, Sarah
Uncl.	Uncl.	Uncl.	SAR	71	19b. TELEPHONE NUMBER *(include area code)* (407) 920-4981

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18