W6QK ACC-APG NATICK
CONTRACTING DIVISION

Contract No. W911QY-16-C-0081
Purchase Request Number: QL6H5R66F004MPPERIGEAN

Maturation, Evaluation, Alignment and Testing of Sero!

CDRL A007

Technical Report- Study/Services;
Ad Hoc Reports and Works (Quarterly)

Base Period Of Performance (POP) Final Report

Submitted to:
Dr. Sae Schatz
Contracting Officer's Representative
via sarah.l.schatz.civ@mail.mil

Submitted electronically by:
Brian Moon
Chief Technology Officer
Perigean Technologies LLC (Perigean)
8108 Lee Jackson Circle
Spotsylvania, VA 22553

Submitted on:
September 30, 2017

UNCLASSIFIED

Executive Summary

This Technical Report-Study/Services; Ad Hoc Reports and Works is intended to deliver 3.2.4a Base POP Final Report by summarizing the work performed during the Base POP, assess progress, identify solved and unsolved technical problems, and set forth specific work details for the Optional POP. An itemized assessment of progress is included in Appendix A. Such sections follow the format specified in DI-MGMT-80227. This report assumes familiarity with the acronyms and terminology in use under the Total Learning Architecture (TLA) program and by ADL.

It includes in Appendix B the following Technical Deliverable:

4.3.10	Report of findings from the ITD event	Analysis of the findings captured during ITD event.

This report also initiates, in Appendix C, the development of technical documentation.

TABLE OF CONTENTS

SECTION 1. SUMMARY OF PROGRESS 5

USER CENTERED-DESIGN 5
CONTENT DEVELOPMENT 5
DEVELOPMENT 5
ITD PLANNING, ENGAGEMENT AND ANALYSIS 5
REPORTING AND AWARD 5

SECTION 2. SOLVED AND UNSOLVED TECHNICAL PROBLEMS 6

INTEGRATION WITH TLA 6
MODULE DEVELOPMENT 6
ASSESSOR 6
LEARNER 6
SCORING 6
TRIPLEPARSER 7
ENABLE ASSESSORS TO ACCESS EMBEDDED ASSESSMENT DESIGN INSTRUCTION 7
ENABLE ASSESSORS TO PROVIDE ASSESSOR FEEDBACK 7
ENABLE ASSESSORS TO ACCESS REPORTING OF LEARNERS' PERFORMANCES 7
ENABLE LEARNERS TO ACCESS ASSESSMENT FEEDBACK REPORTING 7
ENABLE MORE COMPLEX, SEMI-AUTOMATED ANALYSIS, SEMI-AUTOMATED ASSESSMENT. 7

SECTION 3. SPECIFIC WORK DETAILS FOR THE OPTIONAL POP 8

(OPTIONAL) 2.2.5 THE CONTRACTOR SHALL USE THE RESULTS FROM TASK 2.5.3 (ITD EVENT) TO REVISE THE DESIGN OF RELEVANT SERO! FEATURES AND FUNCTIONS 8
FURTHER INTEGRATION WITH XAPI 8
FEEDBACK, SCORING AND REPORTING MECHANISMS 8
ADDITIONAL ASSESSMENT TYPES 8
GUI ENHANCEMENTS 9
(OPTIONAL) 2.2.6 THE CONTRACTOR SHALL CONDUCT A DESIGN CHECKOUT / USABILITY REVIEW OF SERO! 9
(OPTIONAL) 2.3.11 THE CONTRACTOR SHALL EXPLORE INFORMATION ASSURANCE (IA) REQUIREMENTS FOR TRANSITION OF SERO! INTO DOD LEARNING ENVIRONMENTS 9
(OPTIONAL) 2.3.12 THE CONTRACTOR SHALL EQUIP SERO! TO ENABLE ASSESSORS TO ACCESS ADVANCED REPORTING OF LEARNERS' PERFORMANCES 9
ADDITIONAL WORK DETAILS 9
ITD PLANNING, ENGAGEMENT AND ANALYSIS 9
REPORTING 9

APPENDIX A. ITEMIZED ASSESSMENT OF PROGRESS 10

APPENDIX B. REPORT OF FINDINGS FROM THE ITD EVENT 13

INTRODUCTION 13
DATASET 15
COMPLETION RATES 15
ANALYSIS 15
SCORES ACROSS TLOS AND TIME 16
ANALYSIS 18

FOCUS GROUP COMMENTS **18**
ANALYSIS 19

APPENDIX C. TECHNICAL DOCUMENTATION **20**

ARCHITECTURE DESCRIPTION **20**
SERO! GUI 20
CLOUDANT 20
AWS 20
LOGIN 20

FIGURES

FIGURE 1: SERO! SCORES, BY TLO AND DAY OF THE ITD EVENT 17
FIGURE 2: PRE-TEST, SERO!, AND POST-TEST SCORES FOR PARTICIPANTS WHO TOOK ALL THREE 17
FIGURE 3: COMPARISON OF AVERAGE SCORES ACROSS ALL PARTICIPANTS 18

TABLES

TABLE 1: SUMMARY PERFORMANCE DATA 14
TABLE 2: COMPLETION RATES BY ITEM TYPE 15

Section 1. Summary of progress

Based on feedback received from the Government throughout the Base POP, Perigean's performance can be characterized as successful in both the technical and managerial aspects. An itemized assessment of progress in included in Appendix A. To summarize, during the Base year, Perigean achieved the following successes.

User Centered-Design. Perigean executed a UCD process through several activities. Perigean designed wireframes describing functionality, workflows, and Graphical User Interface (GUI) elements of the envisioned system. Perigean published Use Cases describing how Sero! was envisioned for use within the Total Learning Architecture environment for two user classes: assessor and learner. And Perigean conducted two formal and one informal (i.e., outside the scope of work) usability tests, with a total of 5 assessors and 23 learners. Each round of usability testing employed maturing versions of Sero! and increasingly complex tasks, and findings from each session informed ongoing development. System Usability Scores of 61/60/71 (across the three learner usability tests) and 87 (assessors) provide strong support for the usability of Sero!.

Content Development. Perigean executed three content development activities, all focused on developing Concept Maps for use in the ITD event. The first activity created maps from content of Perigean's finding; the second on content provided by ADL. The third activity, which resulted in maps used for the ITD event, focused on a set of multiple choice questions developed by IDA and its cybersecurity expert that were intended to be used for pre- and post-testing at the ITD event. The questions were aligned with learning objectives, and were re-imagined by Perigean into 53 Concept Maps, then Assessment Maps. This innovation represented the first known attempt to convert existing assessment items into Assessment Maps.

Development. Perigean advanced the architecture and functionality of Sero!. Key development efforts included:

- importing of triples and spatial arrangement of graphs to initialize Assessment Maps,
- GUI and database development to enable the authoring and taking of three new assessment item types (Drag-and-Drop-On, ConnectTo, Errors),
- algorithm development to enable robust scoring and semi-automated authoring,
- generation of xAPI statements,
- hosting on ADL's AWS instance.

ITD planning, engagement and analysis. Perigean participated in several activities focused on planning for the ITD event, and played an active role in ideating on overall the user experience. Perigean recommended an approach to generating and managing competencies. Perigean provided on-site support during the event, and conducted a robust analysis of user performances with Sero! (provided in Appendix C).

Reporting and award. Perigean provided an IMP and IMS for the Base year, monthly reports, and two ad hoc reports covering the usability events. Perigean provided a poster for iFest and published the following:

Moon, B. & Rizvi, S. (2017). Sero!: A Learning Assessment Platform for Adult Learning Environments**.** Proceedings of the 8th International Conference on Applied Human Factors and Ergonomics. Springer.

Sero! was awarded Highly Commended for Innovation by the e-Assessment Association.

Section 2. Solved and unsolved technical problems

Perigean made significant progress toward solving numerous technical problems during the Base POP.

Integration with TLA. Sero!'s xAPI statement capability now includes the following statements, which were used during the ITD event:

- "start": statement is sent when a user loads a map;
- "complete": statement is sent when a user submits a map;
- "pass": statement is sent when a user answered 60% or more items correctly on a map;
- "fail": statement is sent when a user answered 59% or less items correctly on a map.

With this development effort achieved, Sero! is capable of sending and receiving additional and more complex xAPI statements.

Module development. Sero! provides two interface modules; one for the assessor role and one for the learner role.

Assessor. The assessor page has a navigation pane on the left side that manages creating new assessments and displaying existing assessments. The assessor page allows the user to create a new assessment map in three ways: from user entered triples in a text box, from importing a text document of triples, or from choosing an existing assessment map available to the user. Once the map is created, the assessor can update elements of the graph into assessment items and assign the newly created assessment.

Learner. When the learner page is loaded, the open and completed assessments are available in the left navigation pane. Upon choosing an assessment, the graph pane on the right side of the screen renders that graph. If it is an open assessment, the user can update the graph, when complete they may submit the graph, where it is sent to a database (Cloudant) and updates the assessment display in the interface.

With this development effort achieved, Sero! is capable of implementing additional and more refined GUI functionality.

Scoring. The scoring algorithm used at the ITD event tested each assessment item in the map to check, first, if the user entered an answer, and then whether the answer matches one-to-one with the correct answer for that assessment item. There were four assessment types tested at Bragg: Drag-and-Drop-On, ConnectTo, Fill-In and Multiple Choice. Both Multiple Choice and Fill-In assessment items were matched one-to-one with the user's input. A point (1) was rewarded if correct or 0 if incorrect. For both the Drag-and-Drop-On and ConnectTo items, Sero! checks every link that the user made with the assessment item. During the ITD event, these items only checked if the correct link was made and disregarded incorrect links. After getting the percentage correct, Sero! checks if the result is over the pass/fail threshold, which for the ITD event was 60%. Learners received only a pass/fail indication.

With this development effort achieved, Sero! is capable of implementing additional and more mature scoring schemes.

TripleParser. A framework of string parsing algorithms that takes text from .txt and/or .csv files and generates semantic triples. It also parses raw triples into the semantic graph and manages updates to the graph's triples and triggers any restructuring if necessary.

Perigean made progress toward solving numerous other technical problems during the Base POP; additional work remains to solve the following technical problems.

Enable assessors to access embedded assessment design instruction. Given that all content for the ITD event was to be generated by Perigean, significant effort was not expended to develop embedded assessment design instruction. Sero!'s assessor GUI is, at the end of the Base POP, mature enough to support the design, generation and integration of assessment design instruction. Perigean expects this technical problem to be solved early in the Option POP.

Enable assessors to provide assessor feedback
Enable assessors to access reporting of learners' performances
Enable learners to access assessment feedback reporting
Collectively, these development technical problems comprise the reporting and feedback mechanisms that will enable more robust learning assessment. Significant development was undertaken for the graph structure, GUI, and user account architecture, which provided a framework to implement complex feedback on and reporting of learner performance. Perigean expects this technical problem to be solved mid-way through the Option POP.

Enable more complex, semi-automated analysis, semi-automated assessment. This exploratory task targeted semantic matching and Natural Language Processing (NLP) capabilities to enhance Sero!'s capability to automate assessment. Basic engineering effort was expended to develop a semantic matching mechanism that employs WordNet comparisons to generate similarity ratings for Fill-In answers. In addition, deployment of Sero! to the AWS environment opens numerous opportunities for NLP technology integration. Perigean expects to make progress toward solutions to this technical problem by the end of the Option POP.

Section 3. Specific work details for the Optional POP

The Performance Work Statement outlines five Optional tasks. Perigean will perform the following work detailed under these tasks, as well as three additional work tasks.

(Optional) 2.2.5 The Contractor shall use the results from Task 2.5.3 (ITD event) to revise the design of relevant Sero! features and functions. Under this task, Perigean will pursue enhancement and re-engineering efforts to further mature Sero!, to include the following.

Further integration with xAPI. Perigean envisions extending the current set of xAPI statements to incorporate:

- "interact" – statement sent whenever a user uses the click or drag functionality on a tracked item in the app. Interact statements can also be sent for each question, with markup information that includes the question format and the correct answer. The object of an interact statement could be the activity itself or a topic in the activity or an assessment item in the activity or any user interface item.
- "answer" – statement will be sent along with each attempted assessment item. It will contain context data that includes the question, possible answers, and the user answer. This data can be used downstream by other activities or consumed by Sero! to generate assessment items, possibly including ad hoc assembly of Assessment Maps.

Feedback, scoring and reporting mechanisms. Perigean envisions significant effort in the Optional POP toward maturing the Base POP capabilities to provide the following:

- Enable assessors to provide assessor feedback,
- Enable assessors to access reporting of learners' performances,
- Enable learners to access assessment feedback reporting, and
- Enable more complex, semi-automated analysis.

In addition, Sero!'s scoring algorithms will be matured to include capabilities for taking completion rates into consideration and providing varied levels of correctness – e.g., Right/Acceptable/Wrong/Egregiously wrong. (The latter capability was suggested by an Assessor participant during the usability testing conducted at USUHS.)

Additional assessment types. Base POP engineering efforts on the graph structure introduce a range of opportunities to introduce new assessment types without significant effort. Perigean envisions extending the assessment types to incorporate some of the following:

- Drag-and-Drop-On + ConnectTo = Assemble the entire map
- Multiple Choice – "All but the following"
- Error, comprising: Designate, Correct, Provide Rationale
- Arrowhead direction
- Proposition level responses
- Map level responses
- Distractor Drag-and-Drop-On
- Formative learning feedback, e.g., as ConnectTo and Drag-and-Drop-Ons are attempted.

GUI enhancements. Base POP engineering efforts on the HTML framework and graph structure introduce a range of opportunities to embed instructional guidance and other GUI enhancements. Perigean envisions implementing GUI enhancements following from findings from the ITD event, as well as the Base POP usability tests, the Optional POP usability test, and other User-Centered Design activities.

(Optional) 2.2.6 The Contractor shall conduct a design checkout / usability review of Sero!. Perigean will design and execute a design checkout / usability review of Sero! during the Optional POP, following the protocols developed during the Base POP, with appropriate revisions. The task will no less than 10 Participants. The previously identified candidate organization is USMA at West Point. Perigean will also pursue other organizations to replace or augment USMA.

(Optional) 2.3.11 The Contractor shall explore information assurance (IA) requirements for transition of Sero! into DoD learning environments. Base POP discussions concerning security revealed requirements and challenges for Sero!. Perigean will pursue Government guidance regarding security protections and the necessity for an IA audit by the end of the Optional POP.

(Optional) 2.3.12 The Contractor shall equip Sero! to enable assessors to access advanced reporting of learners' performances. / (Optional) 2.3.13 The Contractor shall extend Sero! to enable learners to access enhanced assessment feedback reporting. Base POP engineering efforts and envisioned Optional POP efforts toward feedback, scoring and reporting mechanisms, will inform the design and development of advanced reporting. Perigean will pursue reporting mechanisms, to include aggregated and cumulative visualizations of learner responses, longitudinal and trend reports, and comparison with descriptions of Knowledge, Skills, and Abilities (KSAs), competencies, and learner objectives provided by the TLA.

Additional Work Details.

ITD planning, engagement and analysis. Perigean will engage in all planning for the Spring 2018 ITD event, provide on-site support and analyze results from the event.

Reporting. Perigean will provide monthly reports, an Optional POP Final Report, and an ad hoc report covering the usability review. Perigean will also provide complete Technical Documentation by the end of the Optional POP, to include:

- a listing and detailed explanation of algorithms, coding schemes, software, and empirical results;
- a user manual explaining how to run the delivered software; and
- a programmer's manual which includes documentation (READ ME files) for the delivered software.

Appendix A. Itemized assessment of progress

Sero! is a learning assessment platform that uses Concept Maps. For the purposes of

2.1 Integrated Management Plan (IMP) and Integrated Master Schedule (IMS)

ACHIEVED.

2.2 The Contractor shall conduct a User-Centered Design (UCD) process to facilitate Task 2.3. UCD is an iterative design approach that represents explicit understanding of users, tasks and environments, and conducts user-centered evaluation throughout design and development.

OVERALL TASK ACHIEVED.

2.2.1 The Contractor shall develop Use Cases for Sero! advancements. 2.2.2 The Contractor shall develop requirements for the alignment of Sero! with ADL's TLA.

ACHIEVED.

2.2.3 The Contractor shall conduct two design checkouts / usability reviews of Sero!.

ACHIEVED.

2.2.3.1 The Contractor shall establish and gain Institutional Review Board (IRB) approval for Human Research Protection Plans for Task 2.2.3.

ACHIEVED. IRB APPROVAL WAS GRANTED UNDER IDA IRB.

2.2.4 The Contractor shall use the results from Task 2.2.3 to revise the design of relevant Sero! features and functions.

ACHIEVED AND ONGOING.

2.3 The Contractor shall extend Sero! to enable assessors and learners to conduct advanced Concept Map-based assessment.

OVERALL TASK ACHIEVED.

2.3.1 The Contractor shall equip Sero! to enable assessors to access embedded assessment design instruction.

NOT ACHIEVED. FOCUS OF OPTIONAL POP.

2.3.2 The Contractor shall extend Sero! to enable assessors to directly import triples and to build and extend Concept Maps.

ACHIEVED AND IMPLEMENTED.

2.3.3 The Contractor shall extend Sero! to enable assessors to select and use multiple types of assessment items.

ACHIEVED AND IMPLEMENTED.

2.3.4 The Contractor shall extend Sero! to enable more complex, semi-automated analysis.

EXPLORATION ACHIEVED. NO SIGNIFICANT IMPLEMENTATION.

2.3.5 The Contractor shall equip Sero! to enable assessors to provide assessor feedback.

EXPLORATION ACHIEVED. NO SIGNIFICANT IMPLEMENTATION.

2.3.6 The Contractor shall equip Sero! to enable assessors to access reporting of learners' performances.

NOT ACHIEVED. FOCUS OF OPTIONAL POP.

2.3.7 The Contractor shall extend Sero! to enable learners to access embedded assessment task instruction.

ACHIEVED. FOCUS OF OPTIONAL POP.

3.3.8 The Contractor shall extend Sero! to enable learners to complete multiple types of assessment.

ACHIEVED.

2.3.9 The Contractor shall extend Sero! to enable learners to access assessment feedback reporting.

IMPLEMENTED BASIC SCORING. FOCUS OF OPTION YEAR.

2.3.10 The Contractor shall equip Sero! to integrate with the TLA.

ACHIEVED.

2.4 The Contractor shall develop learning content for use in Sero! for the purposes of supporting Tasks 2.2, 2.3 and 2.5.
2.4.1 The Contractor shall develop a set of Concept Maps describing expertise on the topic of social engineering in cyber-security.

ACHIEVED.

2.5 The Contractor shall engage in the planning, execution and analysis of ADL's Spring 2017 integration, test, and demonstration (ITD event).
2.5.1 The Contractor shall conduct planning for the ITD event.
2.5.2 The Contractor shall engage in the execution of ITD event.
2.5.3 The Contractor shall conduct an analysis of the ITD results for the purpose of informing continuation of 2.2 and 2.3.

ACHIEVED.

Appendix B. Report of findings from the ITD event

Introduction

Perigean conducted an analysis of the available user performance data from the ITD event. The analysis assumes some familiarity with the ITD event and its structure.

Sero! was included as an activity; specifically, an assessment activity comprising 53 Assessment Maps concerning cybersecurity content. Participants were introduced to Sero! on the first day when the ITD lead described badging activities. Participants were provided a one-page overview of Sero!'s functions, were informed about instructional videos available through the GUI, and were permitted to ask questions about its functionality if needed. No additional guidance was provided. The badging activity was stressed as the main goal of the week, with other TLA activities included to assist in learning the necessary information to complete the badging activities. Two Sero! Maps served to focus the badging activity: Secure Researcher and Social Engineer. The majority of Assessment Mapping activity occurred on days two and three.

Participants could select the Assessment Maps through several TLA-enabled pathways. They could launch Sero! through a user dashboard, i.e., an HTML page with a list of each activity arranged by ELO; and the PERLS system that learners had on their iPods. On both systems, Sero! was positioned after learning material of the same ELO. Once in Sero!, which was only available through a desktop browser, learners could see all their available assigned Maps. If a Map was failed the learner could launch it again through either service and resubmit as many times as necessary. The score report organized the subsequent attempts by timestamp to look at differences in score and duration.

Summary performance data across all Sero! assessments is shown in Table 1, including the Number of Items in each Assessment Map, the number of attempts at each Map, and the average duration across all attempts for each Map.

Table 1: Summary performance data

		Number Items	Attempts	Average Duration (mins)
TLO1	ELO1N	11	10	5.4
	ELO1I	14	16	13.0
	ELO1A	12	12	4.4
	ELO2N	13	8	5.9
	ELO2I	8	6	4.6
	ELO2A	15	4	2.5
	ELO3N	5	15	10.0
	ELO3I	N/A	N/A	N/A
	ELO3A	6	19	3.9
TLO2	ELO1N	11	2	8.1
	ELO1I	7	7	2.7
	ELO1A	6	4	1.5
	ELO2N	8	3	3.4
	ELO2I	6	2	2.9
	ELO2A	11	10	5.4
	ELO3N	10	2	5.4
	ELO3I	9	6	3.8
	ELO3A	11	2	8.1
TLO3	ELO1N	11	2	5.3
	ELO1I	7	1	5.0
	ELO1A	N-C	N-C	N-C
	ELO2N	7	4	1.3
	ELO2I	7	3	1.8
	ELO2A	N-C	N-C	N-C
	ELO3N	11	1	0.3
	ELO3I	10	4	2.7
	ELO3A	N-C	N-C	N-C
TLO4	ELO1N	5	6	1.6
	ELO1I	6	1	13.0
	ELO1A	N-C	N-C	N-C
	ELO2N	9	5	13.2
	ELO2I	11	3	5.1
	ELO2A	N-C	N-C	N-C
	ELO3N	11	7	9.2
	ELO3I	7	6	8.1
	ELO3A	N-C	N-C	N-C
TLO5	ELO1N	11	3	2.5
	ELO1I	8	3	17.1
	ELO1A	6	1	5.3
	ELO2N	6	18	3.3
	ELO2I	N-C	N-C	N-C
	ELO2A	N-C	N-C	N-C
	ELO3N	7	15	4.0
	ELO3I	7	9	4.9
	ELO3A	N-C	N-C	N-C
TLO6	ELO1N	7	4	12.5
	ELO1I	7	1	1.1
	ELO1A	N-C	N-C	N-C
	ELO2N	N-C	N-C	N-C
	ELO2I	10	1	2.0
	ELO2A	N-C	N-C	N-C
	ELO3N	4	24	1.7
	ELO3I	6	4	5.1
	ELO3A	N-C	N-C	N-C
AVERAGES		8.6	6.35	5.4

For scoring during the ITD event, the Error item was not scored. Fill-In items were updated on the second day to only check if the user answered the question, not how accurate the answer was. Even with these changes there are many failing grades.

Dataset

The starting point was all data collected from the 73 participants. Of these, 67 completed both the pre-test and post-test administered as part of the study. However, many of these participants did not take a Sero! assessment, and were therefore eliminated from the dataset, resulting in 45 participants who had completed the pre- and post-tests and at least one Sero! assessment.

By the close of the ITD event, 643 Assessment Maps were registered as having been accessed in couchDB, with 270 of these resulting in completed – i.e., scored – Maps. Some of the registered but not scored Maps were used to manage a badging activity. For the purposes of analysis, some of these attempts were discarded due to (1) the unreasonably short duration – which probably reflected a participant opening the assessment, then closing it – or (2) a 100% incompletion rate. Attempts that lasted less than 60 seconds were evaluated based on the completion rate; if very few or none of the questions had been completed, the attempt was considered invalid and discarded.

Thus, the following analysis spotlights participant performance in Sero! assessments; specifically, completion rates, scores across TLOs and time, and scores compared to the pre- and post-tests.

Completion Rates

A key goal for Base POP development was to enable learners to take assessments comprised of multiple assessment item types. Table 2 shows the completion rates by question type across all participants.

Table 2: Completion rates by item type

TLO	Multi-choice	Drag and drop-on	Connect-to	Fill-in	Click and correct
1	87%	74%	84%	91%	0%
2	80%	92%	86%	83%	0%
3	94%	88%	100%	40%	0%
4	79%	91%	0%	90%	0%
5	98%	83%	73%	43%	0%
6	100%	71%	75%	N/A	20%
AVERAGE	90%	83%	70%	69%	3%
STDEV	9%	9%	36%	26%	8%
MEDIAN	90%	86%	80%	83%	0%
MODE	N/A	N/A	N/A	N/A	0

Analysis. Consistent with findings from the usability tests, participants completed four of the five items types at reasonable rates. While the goal is near-100% completion, it is reasonable to expect that first-time users with little introduction may not complete every item. The Error

items have resulted in poor completion rates throughout all testing events. The likely reason is that they do not provide any particular cue to the learner, whereas the other items are cued by use of color, "???", an icon, and distinction from the graph.

Design seeds that follow from the completion rates, and will be explored in the Optional POP, include:

- indicators of remaining items,
- separation of error items,
- redirection to incomplete items after Submit.

Current and future item types may also introduce a trade-off for completion rates. For example, if a learner is cued to the presence of Error items, non-completion of an error item should result in the item marked as incorrect. Thus, scoring algorithms will need to consider completion rates in the analysis of learner performance.

Scores across TLOs and time

Analysis of the data began at the ELO/NIA level for each TLO, where statistics such as the amount correct, percentage complete, percentage correct out of completed questions, and average time elapsed during the test were calculated. The answers to each question were reported along with the correct answer. A binary system was assigned to label correctness to calculate scores later (1 for correct, 0 for incorrect). Another binary set was used to calculate completeness (1 for incomplete, 0 for complete), and combined with the correctness scores to calculate each participant's final score.

The composite scores were then broken down to the ELO/NIA level for every participant who had completed a Sero! assessment. First, the raw score was calculated by dividing the number of correct questions by the total number of questions. However, taking the number of questions that were actually attempted into account provides a much more accurate measure of the participant's score. Thus, the completion rate of each attempt was calculated (number of questions on a particular Map minus the number of incomplete questions, all divided by the number of questions on that Map). Then, the raw score was converted to the score out of the attempted questions (raw score divided by the difference between the total number of questions and the number of incomplete questions). Then the ELO-NIA scores were averaged for each participant to obtain the participant's ELO score. The TLO score was a subsequent averaging of the ELO scores, and the total score was the subsequent averaging of TLO scores. Figure 1 shows all Sero! scores, by TLO and day of the ITD event.

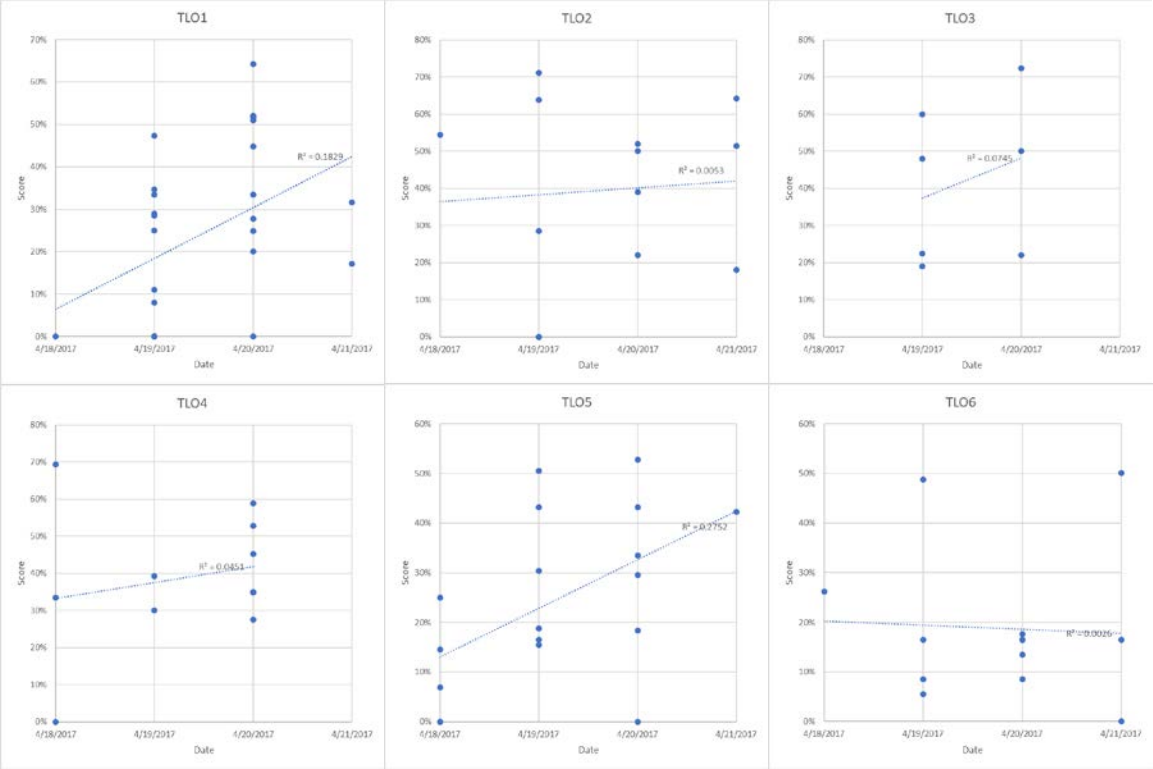

Figure 1: Sero! scores, by TLO and day of the ITD event

The TLO scores were then extracted and compared across all three tests. Perigean found that Sero! fell roughly between the scores of the pre- and post-tests. The R^2 value for the regression line of the averages ($R^2 = 0.8648$) shows that Sero! fits the validity curve well. The R^2 value in this study is merely meant to indicate how close Sero! lies to the validity curve. It is not intended to show that Sero! affects the average pre- and post-tests at all. Figure 2 shows the scores of all 45 participants (blue set, Series TOTAL) and the averages from each test (orange set, Series AVERAGE)

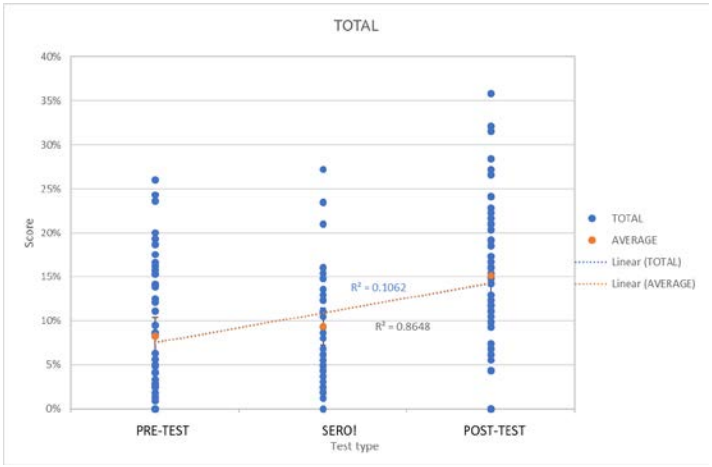

Figure 2: Pre-test, Sero!, and Post-test scores for participants who took all three

Figure 3 shows the comparison of average scores across all participants.

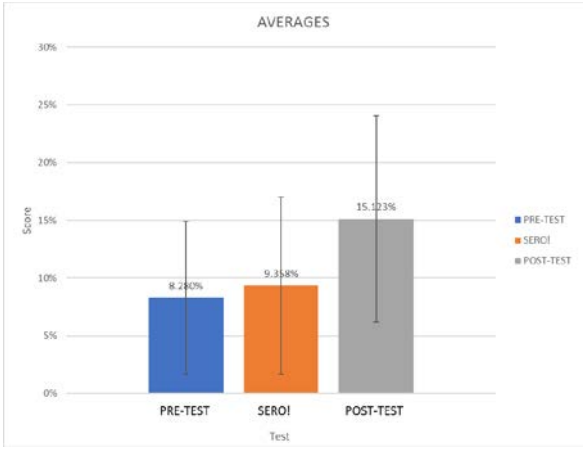

Figure 3: Comparison of average scores across all participants

Analysis. While the data from the ITD event presented many opportunities for confounds, taken as a whole the results offer foundational support for the validity of Sero! assessments. Learner performance across days shows expected improvement, suggesting that Sero! can provide insight into and tracking of ongoing learner achievement at a micro level. Overall performance on Sero! assessments showed as expected to fall between the pre- and post-test performance, suggesting that Sero! assessments can also show summative achievement at a macro-level.

Focus group comments

During the focus group review session on the last day of the ITD, participants expressed several considerations about Sero!. A number of them stated that Sero! was too complicated; one participant called out the GUI as being "visually noisy." Not seeing which items were wrong on each attempt was a cause of frustration. Some participants noted that felt it was harder to read the sentences in the graph then it would if they were paragraphs. Some mentioned they wanted the ability to collapse different sections of the Map so they could focus on others. One user felt the experience would make more sense if the Map began collapsed and as the user answers the available questions the Map should expand.

Compared to the other activities (CyberScorpion and Ares), Sero! was not developed solely to work with cybersecurity information. The other activities simulated computer environments and had the participants execute actual commands relevant to the course learning material. Participants preferred these activities and felt Sero! forced them to review the learning material to find the right answers. One participant said if the material was on organic chemistry the graph structure would be more effective. Participants suggested that the Map should direct users to relevant learning material on questions they got incorrect.

Participants stated that the learning material they were directed to before taking Sero! assessments did not contain the info needed to complete the assessment. This was a source of frustration for the users who arrive to assessments about which they had not reviewed learning material. This order was dictated by PERLS or by the user clicking on links in the dashboard.

Analysis. Participant experience during the ITS tracked well with usability testing. The feedback will be incorporated into future design and development, particularly comments regarding the instructional material. Perigean views the comments regarding complexity of the assessment as promising for the capacity of Sero! to assess higher order thinking skills.

Appendix C. Technical Documentation

The following presents a description of Sero!'s architecture, and should be considered as a draft of the technical documentation. A complete technical documentation will be provided by the end of the Optional POP.

Architecture description

Sero! GUI. The Sero! GUI is engineered using the following framework and library:

- Angular.js – A javascript framework for displaying and managing the HTML interface as well as the data calls processed in the application.
- D3.js – A javascript library for rendering data in visualizations. Sero uses this to create the Concept Maps and Assessment Maps in SVG, a display format that works well in HTML pages.

Cloudant. A cloud-hosted JSON data store (CouchDB), Sero! uses Cloudant to store user account information for assessors and learners, Assessment Maps, and assignments. Cloudant is managed by IBM and has many scalable features for inserting and retrieving documents.

AWS. Sero! is hosted in an EC2 instance. This allows for better support on the application in case of high load. The future implementation of Sero! will take advantage AWS services, including:

- Cognito, a service for managing user accounts and password authentication. This offers scalable secure account access to the application.
- Elastic Beanstalk is a load balancing service to allow for better tracking of resources, logging, versioning and general system administration that we do not have.

Login. In the current form, a user navigates to the Sero! URL to see the application home page and login with their username. After logging in as a learner, the user sees the learner page; as an assessor, the assessor page. During the ITD event, a user could navigate to the serourl + "/adl/<activityid>" and generate an ad hoc assignment for that activity and redirect the user to their learner page with the keycloak information.