A CASE FOR THE SUPERIORITY OF CONCEPT MAPPING-BASED ASSESSMENTS FOR ASSESSING MENTAL MODELS

Brian Moon, Charles Johnston & Skyler Moon
Perigean Technologies LLC, USA
Email: {brian, chip, skyler}@perigeantechnologies.com

Abstract. Mental models have long been considered an important enabler of cognitive performance and a key to understanding educational progress. Yet assessing mental models using straightforward, valid, reliable, and efficient methods remains an elusive challenge. Research suggests that concept mapping holds the promise of the direct analysis of mental models. We stress mapping to emphasize a goal of assessing a dynamic thing (mental models) through the use of a process (concept mapping). Our supposition is that concept mapping-based methods can be used to assess mental models. However, the practicality and feasibility of conducting concept mapping-based assessment (CMA) has to date stifled widespread application of the approaches. To meet this challenge, we have developed a system, Sero!, that implements a robust CMA process in a cloud-based platform. This paper presents an overview and advantages of mental model assessment, a rationale in support of the use of CMA, our development of Sero!, and makes a case for the use of Sero!'s specific instantiation of a CMA as a scalable approach that can support many applications in the assessment of mental models. A demonstration of the implementation in a large training event is presented to bolster the case.

Keywords*:* concept mapping, mental models, assessment

1 Introduction

Mental models have long been considered an important enabler of cognitive performance and a key to understanding educational progress (Johnson et al., 2006; Seel, 1997). Yet assessing mental models using straightforward, valid, reliable, and efficient methods remains an elusive challenge (Schute & Zapata-Rivera, 2008; Smith, 2009).

Research suggests that concept *mapping* holds the promise of the direct analysis of mental models. We stress *mapping* to emphasize a goal of assessing a dynamic thing (mental models) through the use of a process (concept mapping). Our supposition is that concept mapping-based methods can be used to assess mental models. However, the practicality and feasibility of conducting concept mapping-based assessment (CMA) has to date stifled widespread application of the approaches.

To meet this challenge, we have developed a system, Sero!, that implements a robust CMA process in a cloud-based platform. This paper presents an overview and advantages of mental model assessment, a rationale in support of the use of CMA, our development of Sero!, and makes a case for the use of Sero!'s specific instantiation of a CMA as a scalable approach that can support many applications in the assessment of mental models. A demonstration of the implementation in a large training event is presented to bolster the case.

2 Mental Model Assessment: Overview and Advantages

The notion of mental models was described in earliest form by Charles Sanders Peirce (1896) as a set of premises formed into a diagram serving to enable "mental experiments" that conclude in their necessary or probable truth. Scottish psychologist Kenneth Craik (1943) is widely cited as reifying the notion with his concept of "small-scale models" of reality that humans use to anticipate, reason and explain. Slight variations on the theme have yielded notions of "structural knowledge" (Jonassen et al., 1993), "frames" (Klein, Moon & Hoffman, 2006), and "mental pictures" (Alexander, 1963). Generally speaking, the functions of most of these concepts have been to serve similar purposes – to describe, explain and predict the way the world works (Rouse and Morris, 1986).

Since their introduction as a construct, mental models have been considered an important enabler of cognitive performance. "[M]ental model assessment is diagnostic of knowledge acquisition for a complex task and mental model accuracy is related to accuracy in metacognitive processes" (Scielzo et al., 2003). Numerous researchers have also noted the potential value of studying the degree to which teams have and use shared mental models (e.g., Mohammed & Hamilton, 2010). Importantly, Schute (2008) notes that "figuring out how to help people develop and hone good mental models are important goals with potentially large educational and economic benefits" (p. 2).

The intuitive reality of mental models and various attempts to operationalize them (e.g., Jonassen, 1995), however, has belied the challenge of assessing them and more importantly when and how they change (Johnson-Laird, 2013). Straightforward, valid, reliable, and efficient methods for assessing mental models remain elusive.

3 The Potential Value of CMA for Mental Model Assessment

Numerous techniques for assessing mental models have been studied. These include think-aloud protocols, narrative text and causal diagrams (Smith, 2009, pp. 30-38). Schreiber et al (2006), Rowe et al (2007), and Tossell et al (2009) have all explored mental model assessment through the use of networks of knowledge structures using pairwise comparison of concepts that can be represented via Pathfinder network analysis. Widely used assessment techniques have also been implicated for assessing mental models. Seel (1995) attempted to reveal the extent of mental model change through the use of "pretest–posttest comparison on the basis of multiple-choice tasks". Lindel and Olson (2002) developed a "Lunar Phases Concept Inventory (LPCI) … to aid instructors in assessing students' *mental models* of lunar phases," comprising of a set of multiple-choice items.

While these techniques offer many advantages for assessment, they often fall short of the necessary requirements for the assessment of mental models and/or are not feasible for practical implementation. Bennett (2017) reports that analyzing "shared mental models using…Pathfinder…take(s a) long time for the data collection with a large set of concepts." Moreover, while the validity of the approach has been demonstrated, the pairwise comparison method fails to incorporate a critical element in shared mental models – namely, the *specific* nature of the relations *between* concepts – and evaluates individual judgments, rather than *holistic* and *interdependent* relationships *among* concepts. The same criticism can be put to multiple choice items that are presented as a linear series of discrete questions that typically fail to assess the nature of the relationships *between* the concepts introduced in separate questions. Situational judgment tests offer indirect assessment of mental model.

Concept mapping holds the promise of advancing assessment of mental models, with their capacity to blend recall, recognition and reasoning techniques in the context of a nonlinear assessment. Concept maps are organized sets of propositions, or statements about the universe, comprising two concepts connected by a link that posits the nature of their relationship. An example concept map describing Applied Concept Mapping (Moon et al., 2011) is shown in Figure 1. The example shows the semi-hierarchical shape common to concept maps, which reinforces and reflects the mostly hierarchical shape of knowledge.

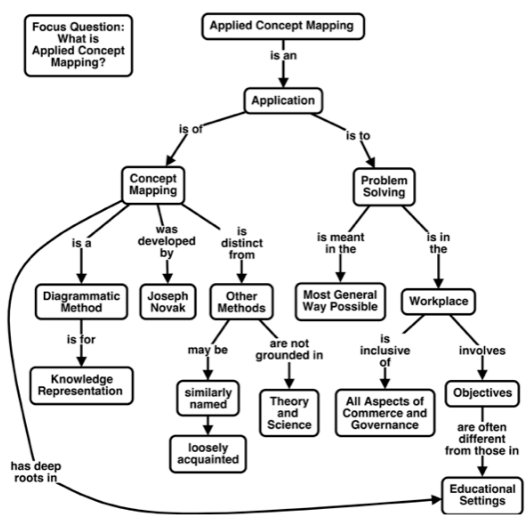

Figure 1. Example Concept Map.

The validity and reliability of concept mapping tasks for learning assessment has been thoroughly demonstrated (Ruiz-Primo and Shavelson 2001; Schaal, 2008), as most clearly evidenced by its inclusion in the U.S. National Assessment of Educational Progress' Science Framework (NAGB, 2010).

Ideally, during an assessment, a learner practices cognitive higher-order thinking skills to analyze, compare, infer, interpret, and evaluate the challenge set in front of them. Deeper critical thinking dives past cognition and into metacognition components, which demands the learner to plan, monitor, review, and revise (Quellmalz, 1991). These cognitive and metacognitive skills are used by successful learners across disciplines and tasks when working towards conceptual understanding of a topic or when solving a complex problem. Concept mapping-based assessment provides a capability to blend recall, recognition and reasoning techniques in the context of a nonlinear assessment. A concept mapping-based assessment offers a unique opportunity to blend recall and recognition in a single assessment, thereby encouraging learners to exercise metacognitive level thinking through review and revision of the maps' concepts. Rebich and Gautier (2005) have highlighted other key advantages for concept map-based learning assessment that differentiate it from other assessment approaches:

> The usefulness of concept mapping for assessment is partially due to its level of complexity, which distinguishes it from more conventional evaluation techniques such as multiple-choice tests… [T]hese traditional unidimensional assessment measures represent a failure to recognize that much disciplinary knowledge is based on an understanding of relationships among concepts…[C]oncept mapping allows for more efficient data collection than interviews do, and presents an advantage over writing-based assessments in that it is *inherently non-linear* and facilitates self-monitoring…Concept maps may be useful in revealing thought processes that generally remain private to the learner, and it has been suggested that they may be more sensitive to developmental changes than traditional testing in which questions often focus on isolated ideas (p. 358).

4 The Case for a Specific Instantiation of a CMA

Given the potential value for CMA in assessing mental models, many approaches for using concept mapping as an assessment have been proffered and studied, including as a method for assessing mental models (Chang, 2007). A number of taxonomies of the approaches have also been developed to guide the assessment process (e.g., Strautmane, 2012). In order to make a case for a specific instantiation of a CMA, it is useful to review the CMA approaches and other assessment approaches in the abstract.

4.1 Consideration of CMA Approaches

As an oversimplified but reasonable way to characterize the assessment task approaches, it is useful to consider two ends of a spectrum. On one end is a blank canvass onto which a Taker (of the assessment) may, using concept mapping techniques and conventions, create a concept map that is representative of their mental model. At the other end of the spectrum would be a completed concept map, about which a Taker could offer a statement(s) of agreement with their own mental. Both ends offer valuable advantages and severe drawbacks. The blank canvass approach is a highly complex task that leaves open questions about comprehensiveness and completion (Schau, et al., 2001) and results in individual representations that multiply quickly to the point of impracticality for an assessor. Though their potential as an "excellent diagnostic device" has been known for some time (Lay-Dopyera & Beyerbach, 1983), the time to manually analyze the maps, or transcribe their content for analysis, can be considerable (Cañas *et al*., 2010). The completed concept map approach can over-constrain the representation process to the point of reification and limits the insight possibilities into individual mental models. But capturing assessment results is straightforward and simple and can be achieved across many Takers. The key tradeoff between the ends of spectrum is the depth of insight and efficiency of assessment process.

For these reasons, the majority of work in CMA has targeted approaches that fall somewhere in the middle of the spectrum. Some such approaches advocate for scaffolding the concept mapper with a few concepts and (sometimes) linking phrases, instructing the Taker to build out the map. Novak (1998) recommend such an approach for complex topics:

> By providing a small concept map, perhaps with 6-12 concepts and appropriate linking words, we can activate recall of pertinent concepts known by the learner and/or model appropriate structuring of these

> concepts. This skeleton map can also function as an advance organizer of proceeding to build a more detailed knowledge model by supplying ideas, at least some of which would be familiar to the learner. Since it is also common that learners have some misconceptions or faulty knowledge structures, the expert skeleton map can encourage rethinking already-held propositions. An expert skeleton concept map may also contain several suggested concepts in what we refer to as a "Parking Lot." These are concepts that the learners might incorporate into the concept map, thus providing them some further scaffolding of learning (pp. 264).

This approach seeks to overcome the most challenging aspect of the blank sheet approach – i.e., where to start – and even suggests that providing propositions may have learning value in correcting misconceptions. Correia et al (2016) and Correia and Moon (2018) have advanced on this idea in an assessment context by explicitly introducing misconceptions as a means for assessing whether such misconceptions exist. Yet this skeleton/scaffolding approach does not overcome the scalability challenge of reviewing many maps.

Other approaches suggest a "Swiss cheese" approach of providing concept map with a predetermined structure but missing content (concepts and/or linking phrases) that may be filled in, typically from a bank of items (Torre et al., 2013) or by filling in content (Schau, 1997) or by providing options for selecting the most appropriate content (Moon et al., 2010; Sas, 2010). These approaches enable the Taker to impart some of their own representation as they consider what is available to them in the map.

A key advantage of these middle-way approaches is that they enable the exercise of the higher order thinking skills outlined by Bartlett (1955):

- interpolation : filling in information that is missing from a logical sequence,
- extrapolation : extending an incomplete argument or statement,
- reinterpretation : rearrangement of information to effect a new interpretation.

Thus, it can be inferred that Takers of such assessments that perform well can be thought to have achieved such higher order thinking skills.

Once Takers complete the assessment task, their artefacts must be reviewed for any assessment, let alone feedback and thus learning, to occur. Work in this area can be characterized (again through oversimplification) by two approaches that line up roughly with the ends of the spectrum in the assessment tasks. Toward the end of the 'blank sheet' end are review approaches that seek to assess the structure, size and other attributes of the concept map. Toward the other end of the spectrum, where attributes of the map are already given, are approaches that seek to assess the "correctness" of the propositions. In many implementations, both approaches are implicated. With all of the approaches, a number of challenges with scoring reliability have been studied (Ruiz-Primo, Schultz, Li, & Shavelson, 2001).

In considering the variety to middle-way approaches to CMA, several common principles emerge that suggest best practice:

- Start with *some* content,
- Task Takers with conducting concept mapping *processes*, to include:
  - Generating content, considering and using content that is available, making explicit connections between concepts,
- Control the *difficulty* of the task by varying aspects, to include:
  - The given content of the domain, the level of exposure if any to an 'initial' concept map, the number and types of concept mapping process tasks,
- *Compare* the created artefact to something.

The last principle is worthy of additional comment. Ultimately, in order to glean value from the concept mapping process, the concept map must be compared to something. This matters regardless how the concept map was created – i.e., from a blank sheet or highly constrained task. Comparison may occur throughout the concept mapping process or at a final stage of creation. But it is the comparison of the concept map as an artefact – i.e., a representation of a mental model – that yields value for assessment and learning. Comparison may be with a teacher/instructor/expert mental model – whether explicitly represented or not – or with a standardized model that represents a consensus of the state of knowledge. Even assessment approaches that use attributes of the map must ultimately compare maps to some sense of what is 'appropriate' and 'sensible.' For any concept could be 'unpacked' into a set of propositions (yielding higher numbers of elements in the map), and any concept could be crosslinked to any other concept in a given map (yielding higher numbers of crosslinks). Thus, ultimately a comparison to some representation of a given domain that represents 'truth about the aspect of life under consideration' must be made in order to gain value in assessment and, in turn, learning.

4.2 Comparison to other Assessment Approaches

Because they make use of similar item types, CMAs can provide the same diagnostic benefits as traditional assessment item types. In addition, we assert that CMAs offer a fundamentally different assessment than traditional assessment items that are, by definition, delivered in linear fashion. Table 1 outlines the shared and unique advantages.

	Shared with Traditional Assessment Items	**Unique to Concept Mapping-based Assessment**
Authoring	Challenging distractors, (ideally but rarely) logic, selective and/or meta coverage of content	Meaningful propositions about cohesive & coherent abstraction of content
Taking	Reasoning about discrete, sequenced questions and answers	Reasoning about ***and across*** propositions using dependencies, context, sequence
	Recognition and recall	Higher Order Thinking Skills interpolation : filling in information that is missing from a logical sequence extrapolation : extending an incomplete argument or statement reinterpretation : rearrangement of information to effect a new interpretation
	1-to-1, 1-to-many matching / placement	1-to-Map
Analysis	Pass/Fail, Correct/Incorrect, Duration between answers, Answer revision, Sequence of answers (including first answer), Correct per item types	Assertions about cognitive performance made across items, by Proposition, and by Map
Reporting	Rapid scoring and reporting at individual and group levels	Visualization of targeted conceptual understandings

Table 1: Comparison of traditional assessment items to concept mapping-based assessment

5 Development of Sero!, an implementation of CMA

Sero! is a software platform for conducting CMA. It was developed by Perigean Technologies as a prototype assessment tool to advance the current capability and research of concept mapping as an assessment strategy. It is architected as a cloud-based software service and usable on any desktop web browser.

There are two intended user types for Sero!: Assessors and Takers. Assessors can create original concept maps and/or import concept maps, which become the MasterMaps (i.e., answer key) for the assessment. Assessors can assign their choice of AssessmentMaps to Takers, who complete the AssessmentMaps by clicking on, dragging, and editing assessment items to revise and complete the map. Once the Takers have completed the assessment, Sero! provides semi-automated support for comparing the Takers' maps to the MasterMaps, and visually, numerically, and textually report the results to the Assessor, who can review the Takers' maps and provide additional feedback and guidance, where appropriate. Figures 2 and 3 show Sero!'s Graphical User Interfaces for the Assessor and Taker Roles.

Sero! provides semi-automated support in authoring AssessmentMaps, based on the MasterMaps. Concepts, linking phrases, and propositions can be converted to Multiple Choice, Generate and Fill-In (GAFI), Select-And-Fill-In (SAFI), Error Detection/Correction and ConnectTo assessment item types – all of which have been demonstrated individually in the research literature as being valid and reliable methods (e.g., Moon et al., 2010; Correia, 2016; Schau, 2001). Similar to traditional assessment formats, fill in the blank and multiple-choice questions located throughout a concept map forces a Takers to retrieve knowledge to answer the question. While Sero! assesses this type of necessary declarative knowledge, it simultaneously allows a Takers to see, in real time, how an answer to one item may affect how they answer subsequent responses and the rest of the map. In this way, Sero's use of concept maps requires Takers to exercise reinterpretation skills as the rearrangement of information effects a new interpretation, impacting connected concepts, some of which were determined by the Taker, and how that may alter other relationships throughout the rest of the map.

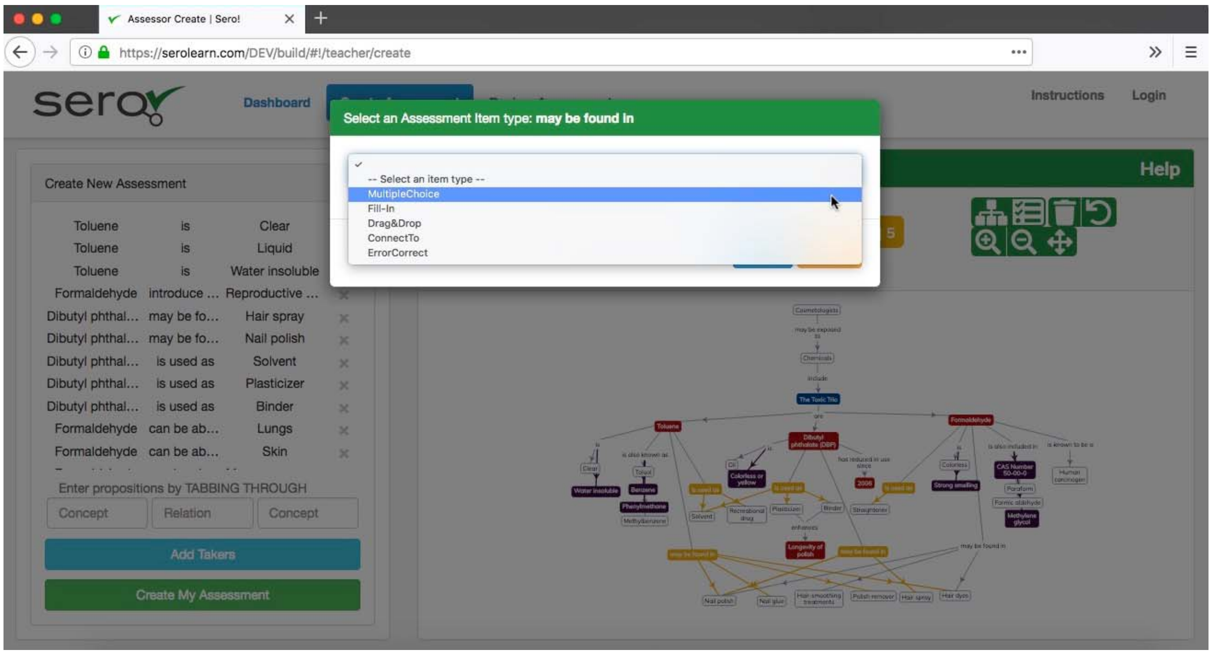

Figure 2. Sero! Assessor Graphical User Interface – as of July 2018.

6 Validation Demonstration

In 2017, Sero! was used in a five-day, live prototype and demonstration event ("event') aimed at demonstrating an adaptive learning paradigm. In addition, data were collected regarding the validity of CMA's implemented in Sero! for the assessment of evolving learner mental models.

6.1 Event Purpose, Participants, Materials and Task

The event focused on training in the cybersecurity domain. Six terminal learning objectives and 54 enabling learning objectives scoped the training event, which presented a large corpus of digital training materials and activities that included simulations/games and simulated cybersecurity activities. To simulate an adaptive learning environment, participants were enabled to select activities in the sequence they chose and/or were recommended to, based on their performance.

Participants in the event were 73 members of the U.S. military. Of these, the performance of 45 participants were selected for data analysis of Sero!'s capacity. The remaining participants did not complete enough of the presented tasks to be considered for the analysis.

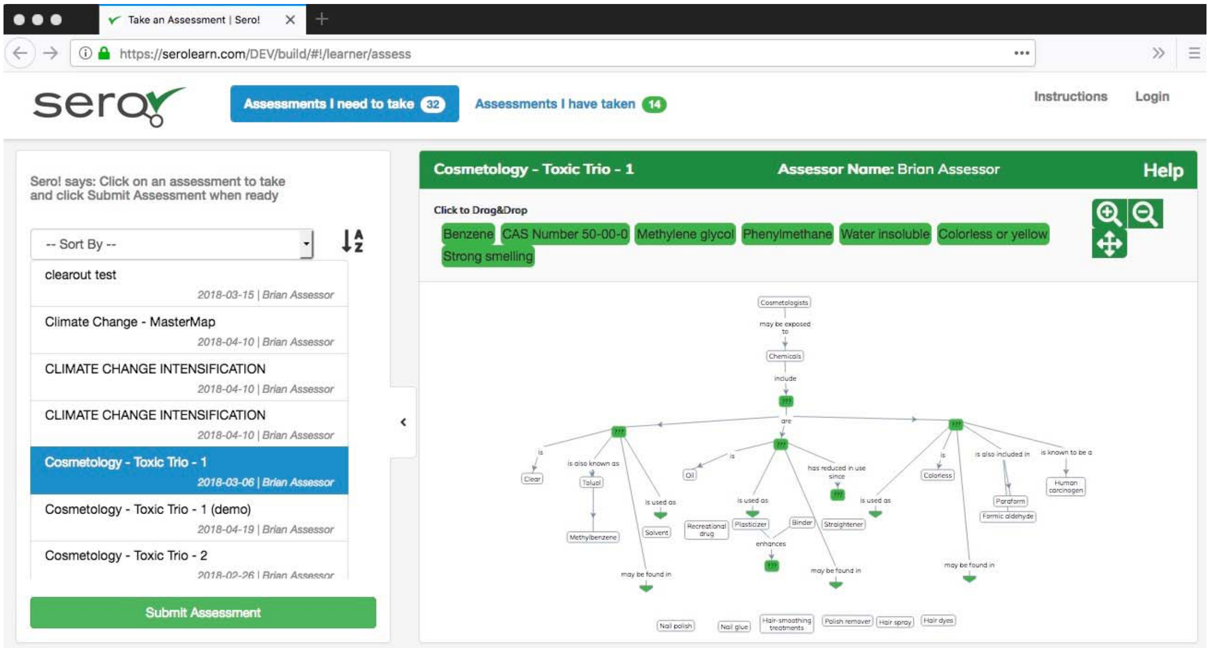

Figure 3. Sero! Taker Graphical User Interface – as of July 2018.

Fifty-three concept mapping-based assessments were developed for the event, matching to all but one of the learning objectives. An example assessment map is shown in Figure 4. The assessment maps were developed from a set of multiple-choice questions that also served as the pre- and post-test assessments for the event. As participants completed training activities, they were also recommended to complete Sero! assessments. The 45 participants completed the pre- and post-tests, some of the 53 Sero! assessments, and some of the training materials and activities.

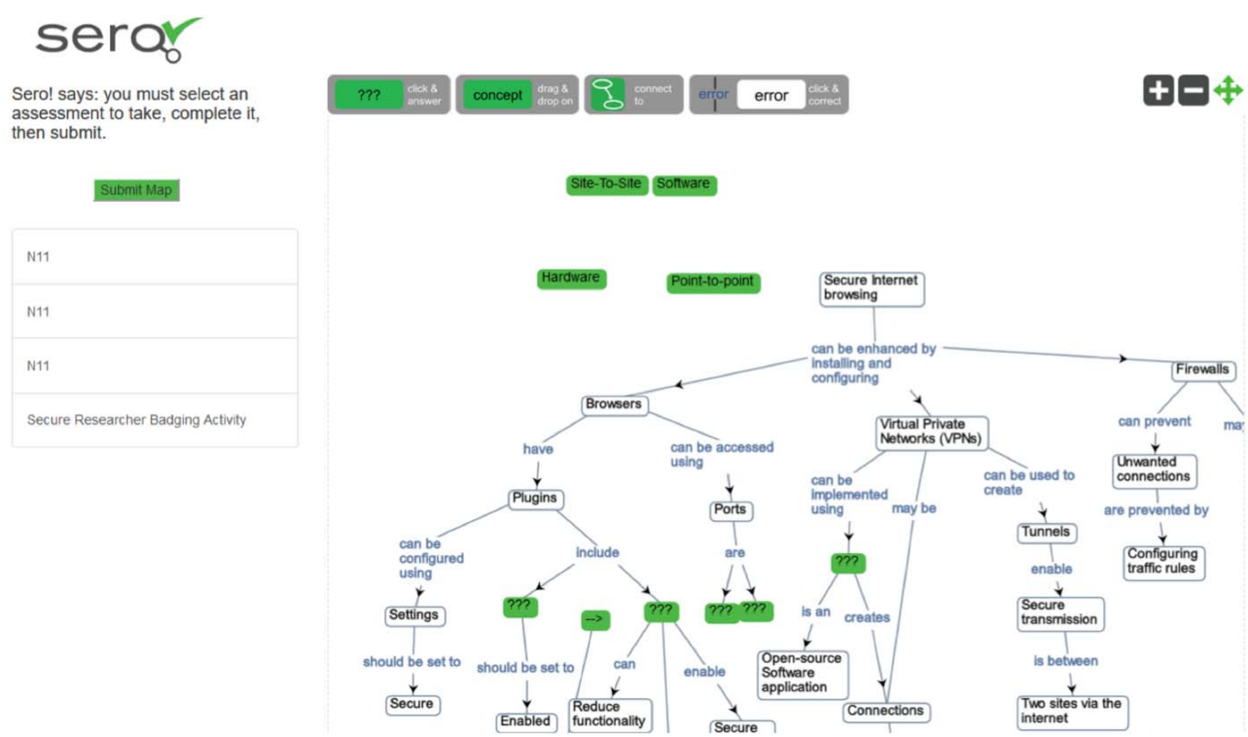

Figure 4. Sero! Taker Graphical User Interface for Demonstration Event – July 2017.

6.2 Results

The validity of the Sero! assessments was analyzed by plotting scores from the pre- and post-tests and Sero! assessments, and establishing a "validity curve," i.e., the line between the average pre- and post-test scores on which the scores from the assessment should fall if they were taken chronologically in between the two other tests. Figure 5 shows the results. The black oval indicates the area within which a valid test's scores should fall. Sero! did indeed fall between the scores of the pre- and post-tests. The R2 value for the regression line of the averages (R2 = 0.8648) shows that Sero! fits the validity curve well. Figure 6 shows the average scores.

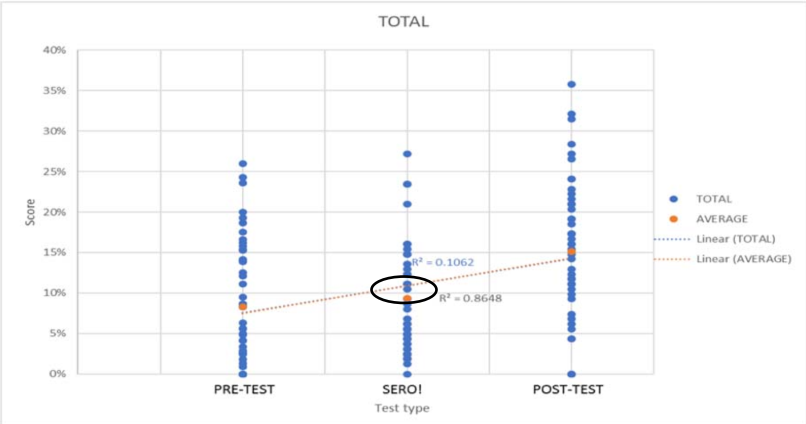

Figure 5. All Scores and Validity Curve.

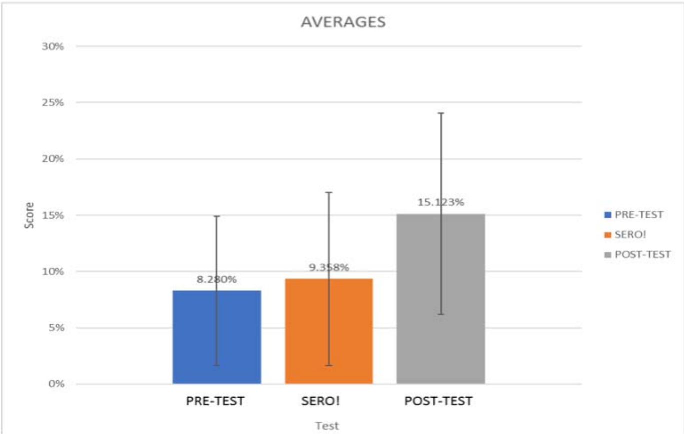

Figure 6. Average Scores.

In addition, scores were also analyzed across days of the training event. Because participants engaged the training material over the course of several days, a valid assessment should expect to show scores increasing across time. The sample in Figure 7 from one of terminal learning objectives (TLOs) shows a marked improvement from the first day of Sero! assessments to the last, demonstrating that Sero! effectively gauged learning progress. All but one of the TLOs showed performance improvement across time – the degraded performance in the most difficult TLO was shown to be only slight.

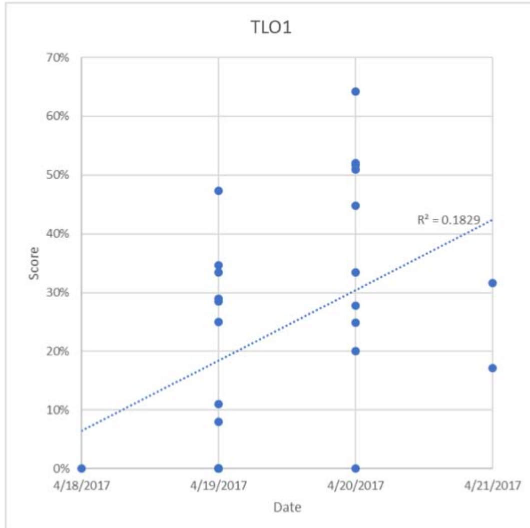

Figure 7. Sero! Assessment Performance Scores across Time.

In addition to scores, data regarding the completion of individual assessment items with Sero! assessments were collected and analyzed to mark invalid attempts. An assessment completion attempt was considered invalid if very little time (<60 seconds) elapsed for the entire assessment and/or the completion rate for all items was close to 0%. However, in later, broader analysis of the completion rates, it was found that some question types had an unusually low completion rate, possibly leading to a higher than appropriate rejection rate. In particular, the Error item type completion rate was very low, reflecting performance seen during usability testing. Unlike the four other types, this item type is not visually prompted.

TLO	Multiple-choice	Drag and Drop	ConnectTo	Fill-in	Error
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%**

Table 2: Assessment Item Completion Rates

6.3 Discussion

The results demonstrated the validity of CMA's implemented in Sero! for the assessment of evolving learner mental models, and raised questions about the feasibility of presenting mixed, multiple assessment items in a CMA. The demonstration was limited in the number of participants and less-than-desirable controls over the demonstration.

7 Summary

Mental model assessment is a key component for learning. Concept mapping-based assessments have repeatedly shown potential for more accurately assessing mental models than other assessment techniques. Our CMA implementation, Sero!, increases the potential for the feasibility of CMA. Future studies should capitalize on the superiority of technology-facilitated CMA for learning and other applications, such as personnel selection.

8 Acknowledgements

Development of Sero! and participation in the event was supported by Contract # W911QY-16-C-0081, sponsored by the U.S. Department of Defense.

References

Alexander, C. (1964). *Notes on the Synthesis of Form*. Cambridge, MA: Harvard University Press.

Bennett, W. (2017). Email communication regarding BAA-RQKH-2015-0001. June 23, 2017.

Cañas, A. J., Bunch, L. & Reiska, P. (2010) CmapAnalysis: An Extensible Concept Map Analysis Tool. In J. Sánchez, A. J. Cañas & J. D. Novak (Eds.), *Concept Maps: Making Learning Meaningful. Proceedings of the Fourth International Conference on Concept Mapping* (Vol. 1). Viña del Mar, Chile: Universidad de Chile.

Chang, S. N. (2007). Externalising Students' Mental Models through Concept Maps. *Journal of Biological Education*, *41*(3), 107-112.

Correia, P., Cabral, G., & Aguiar, J. (2016). Cmaps with Errors: Why not? Comparing Two Cmap-Based Assessment Tasks to Evaluate Conceptual Understanding. In *International Conference on Concept Mapping* (pp. 1-15). Springer International Publishing.

Correia, P. and Moon, B. (2018). Using Concept Maps with Errors to Identify Misconceptions: The Role of Instructional Design to Create Large-Scale On-Line Solutions. IGI Global

Craik, K. (1967). *The Nature of Explanation*. 1943. Cambridge University, Cambridge UK.

Johnson-Laird, P. N. (2013). Mental Models and Cognitive Change. *J. of Cognitive Psychology*, *25*(2), 131-138.

Jonassen, D. H., Beissner, K., & Yacci, M. (1993). *Structural knowledge: Techniques for Representing, Conveying, and Acquiring Structural Knowledge*. Psychology Press.

Jonassen, D. H. (1995, October). Operationalizing Mental Models: Strategies for Assessing Mental Models to Support Meaningful Learning and Design-Supportive Learning Environments. In *The first international Conference on Computer Support for Collaborative Learning* (pp. 182-186). L. Erlbaum Associates Inc.

Klein, G., Moon, B., & Hoffman, R. R. (2006). Making Sense of Sensemaking 2: A Macrocognitive Model. *IEEE Intelligent Systems*, *21*(5), 88-92.

Lay-Dopyera, M., & Beyerbach, B. (1983). Concept Mapping for Individual Assessment.

Lindell, R. S., & Olsen, J. P. (2002, August). Developing the Lunar Phases Concept Inventory. In *Proceedings of the 2002 Physics Education Research Conference*. New York: PERC Publishing.

Moon, B., Ross, K., & Phillips, J. (2010). CMLA for Adult Learners. In J. Sánchez, A. J. Cañas & J. D. Novak (Eds.), *Concept Maps: Making Learning Meaningful. Proceedings of the Fourth International Conference on Concept Mapping* (Vol. 1). Viña del Mar, Chile: Universidad de Chile.

Moon, B., Hoffman, R., Novak, J. & Cañas, A. J. (2011). *Applied Concept Mapping: Capturing, Analyzing, and Organizing Knowledge*. New York: CRC Press.

Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor No More: A 15-year Review of the Team Mental Model Construct. *Journal of Management*, *36*(4), 876-910.

National Assessment Governing Board. (2010). Science Framework for the 2011 National Assessment of Educational Progress.

Novak, J. D. (1998). *Learning, Creating, and Using Knowledge*. Mahwah, NJ: Erlbaum.

Peirce, C. S. (1986). Die Kunst des Räsonierens (1893). na.

Quellmalz, E. S. (1991). Needed: Better Methods of Testing Higher-Order Thinking Skills. *Developing Minds.: A Resource Book for Teaching Thinking*, 338-343.

Rebich, S., & Gautier, C. (2005). Concept Mapping to Reveal Prior Knowledge and Conceptual Change in a Mock Summit Course on Global Climate Change. *Journal of Geoscience Education,* 53(4), 355-365.

Rouse, W. B., & Morris, N. M. (1986). On Looking into the Black Box: Prospects and Limits in the Search for Mental Models. *Psychological Bulletin, 100*(3), 349.

Rowe, L. J., Schvaneveldt, R. W., & Bennett Jr, W. (2007). Measuring Pilot Knowledge in Training: The Pathfinder Network Scaling Technique. L-3 Communications, Mesa, AZ.

Ruiz-Primo, M.A., Schultz, S.E., Li, M. & Shavelson, R.J. (2001). Comparison of the Reliability and Validity of Scores from Two Concept Mapping Techniques. *Journal of Research in Science Teaching*, 38(2), 260-278.

Sas, I. C. (2010). The Multiple-choice Concept Map (MCCM): An Interactive Computer-Based Assessment Method.

Schaal, S., (2008). Concept Mapping in Science Education Assessment: An Approach to Computer-Supported Achievement Tests in an Interdisciplinary Hypermedia Learning Environment. In A. J. Cañas, P. Reiska, M. Åhlberg & J. D. Novak (Eds.), *Concept Maps: Connecting Educators*. Proc. of the Third Int. Conference on Concept Mapping. Tallinn, Estonia: Tallinn University.

Schau, C. (1997). Use of Fill-in Concept Maps to Assess Middle School Students' Connected Understanding of Science.

Schau, C., Mattern, N., Zeilik, M., Teague, K. W., & Weber, R. J. (2001). Select-and-Fill-In Concept Map Scores as a Measure of Students' Connected Understanding of Science. *Educational and Psychological Measurement*, *61*(1), 136-158.

Schreiber, B. T., DiSalvo, P., Stock, W. A., & Bennett Jr, W. (2006). Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study. Volume 5. *Using the Pathfinder Methodology to Assess Pilot Knowledge Structure Changes*. Lumir Research Inst Grayslake Il.

Scielzo, S., Fiore, S. M., Cuevas, H. M., & Salas, E. (2004). Diagnosticity of Mental Models in Cognitive and Metacognitive Processes: Implications for Synthetic Task Environment Training. *Scaled worlds: Development, Validation, and Applications*, 181-199.

Seel, N. M. (1995). Mental Models, Knowledge Transfer, and Teaching Strategies. *Journal of Structural Learning, 12*(3), 197-213.

Shute, V. J., & Zapata-Rivera, D. (2008). Using an Evidence-Based Approach to Assess Mental Models. In *Understanding Models for Learning and Instruction* (pp. 23-41). Springer, Boston, MA.

Smith, L. J. (2009). Graph and Property Set Analysis: A Methodology for Comparing Mental Model Representations. The Florida State University.

Strautmane, M. (2012). Concept Map-Based Knowledge Assessment Tasks and their Scoring Criteria: An Overview. In A. J. Cañas, J. D. Novak & J. Vanhear (Eds.), *Concept Maps: Theory, Methodology, Technology. Proceedings of the Fifth International Conference on Concept Mapping* (Vol. 2, pp. 80-88). Valletta, Malta: University of Malta.

Torre, D. M., Durning, S. J., & Daley, B. J. (2013). Twelve Tips for Teaching with Concept Maps in Medical Education. *Medical teacher*, 35(3), 201-208.

Tossell, C. C., Smith, B. A., & Schvaneveldt, R. W. (2009, October). The Influence of Rating Method on Knowledge Structures. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, No. 26, pp. 1893-1897). Sage CA: Los Angeles, CA: SAGE Publications.