Automated Assessment of Domain Knowledge With Online Knowledge Mapping

CSE Technical Report 692

Eva L. Baker, Principal Investigator
UCLA/CRESST

July 2006

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
GSE&IS Building, Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported under the Office of Naval Research Award Number N00014-02-1-0179, as administered by the Office of Naval Research. The findings and opinions expressed in this report do not reflect the positions or policies of the Office of Naval Research.

Automated Assessment of Domain Knowledge With Online Knowledge Mapping

Gregory K. W. K. Chung, Eva L. Baker, David G. Brill, Ravi Sinha, Farzad Saadat, and William L. Bewley

UCLA/CRESST

ABSTRACT

A critical first step in developing training systems is gathering quality information about a trainee's competency in a skill or knowledge domain. Such information includes an estimate of what the trainee knows prior to training, how much has been learned from training, how well the trainee may perform in future task situations, and whether to recommend remediation to bolster the trainee's knowledge.

This paper describes the design, development, testing, and application of a Web-based tool designed to assess a trainee's understanding of a content domain in a distributed learning environment. The tool, called the CRESST Human Performance Knowledge Mapping Tool (HPKMT), enables trainees to express their understanding of a content area by creating graphical, network representations of concepts and links that define the relationships of concepts.

Knowledge mappers have been used for several years, almost always as an aid for organizing information in support of problem solving or in instructional applications. To use knowledge maps as assessments there must be a reliable scoring method and there must be evidence for the validity of scores produced by the method. Further, to be practical in a distributed learning environment, the scoring should be automated. The HPKMT provides automated, reliable, and valid scoring, and its functionality and scoring method have been built from a base of empirical research.

We review and evaluate alternative knowledge mapping scoring methods and online mapping systems. We then describe the overall design approach, functionality, scoring method, usability testing, and authoring capabilities of the CRESST HPKMT. The paper ends with descriptions of applications of the HPKMT to military training, limitations of the system, and next steps.

A critical first step in developing learner-centric systems is gathering quality information about an individual's competency in a skill or knowledge domain. Such information includes, for example, an estimate of what trainees know prior to training, how much they have learned from training, how well they may perform in a future target situation, or whether to recommend remediation content to bolster the trainees' knowledge.

The focus of this paper is on the development of a tool designed to assess a trainee's understanding of a content domain via graphical representation. The tool, referred to as the CRESST Human Performance Knowledge Mapping Tool (HPKMT), requires trainees to express their understanding of a content area by creating knowledge maps. Knowledge maps are network representations, where nodes represent concepts and links represent the relationships between two concepts. The questions that guided the development of the HPKMT were:

- What are the existing scoring methods for knowledge maps? While we already had an approach for scoring knowledge maps, we were interested in incorporating any new developments in scoring approaches.

- What are the existing software packages that can be used for assessment purposes? At a minimum, we required the package to support the knowledge mapping format. In addition, we were interested in a tool that could support other, unforeseen formats. In a distributed learning context, it would be extremely useful to have a common user interface with a single metaphor to deliver different assessment formats. Thus, we reviewed existing software to determine whether there existed commercial, off-the-shelf (COTS) products that could be used.

- What are the key functional requirements for such an assessment tool? In a preview of our findings, we concluded from our evaluation of COTS products that there existed no product that would satisfy our requirements for an assessment tool.

There are three main sections of this paper. In the first section, we situate the HPKMT within the general area of knowledge mapping for assessment purposes. We first review the different scoring approaches that have been used to evaluate the quality of knowledge maps. Our focus is on scoring because this aspect critically impacts the validity of inferences drawn about trainee performance. We then discuss briefly current online mapping systems and our judgment of their ease of use and scoring capability. The second major section discusses the HPKMT system, the design of which was based on the review of scoring, the review of online products, and our prior work with online knowledge mapping. We report on our experience with using the HPKMT in both a usability test and field test. The last major section is a description of the authoring prototype we have developed. We then discuss limitations of this work and identify next steps.

Overview of Scoring Methods

A presumed critical capability of an assessment in a distributed learning setting is automated scoring. A critical validity issue of an assessment is the scoring, regardless of automated capability. In this section we briefly describe the different types of scoring and provide examples of their use. For in-depth reviews of assessment issues related to knowledge maps, see Ruiz-Primo and Shavelson (1996).

In general, scoring knowledge maps can be referent-based or referent-free. Referent-based methods compare a student's map against a referent map (e.g., an expert's map or other gold standard). Referent-free methods evaluate the student's map against a rubric or with other criteria (e.g., judging the quality of the propositions [node-link-node tuple], or counting the number of concepts in the map). In either case, different scoring approaches use to different degrees the configural and semantic properties of the network.

Table 1 summarizes scoring methods.

Table 1

Simplified Summary of Knowledge Mapping Scoring Methods

	Configural	Semantic
Referent-free	Explicitly scores a map or elements of a map on its structural aspect (e.g., considering degree of hierarchical organization). Example application: Novak and Gowin (1984).	Explicitly scores a map or elements of a map on its semantic aspect (e.g., scoring quality of propositions). Example applications: Osmundson, Chung, Herl, and Klein (1999). Shavelson (Ruiz-Primo, Schultz, Li, & Shavelson, 2001).
Referent-based	Compares the network structure of a student's map and the referent map. Does not take into account the meaning of the relationships. Example application: Herl, Baker, and Niemi (1996).	Compares the semantic structure of a student's map and the referent map (e.g., proposition-by-proposition comparison between a student's map and an expert's map). Ignores the configural aspects of the network. Example applications: Herl et al. (1996). Osmundson et al. (1999).

Referent-Free Methods

The scoring procedure specified by Novak and Gowin (1984) is one of the earliest and most commonly used methods of scoring knowledge maps. Their method considers hierarchy as an important component of the scoring, as well as propositions, cross-links, and examples. In terms of hierarchy, credit is given for each hierarchical level showing subordinate concepts at a lower level as more specific than their parent concepts. Each valid and meaningful proposition is also credited, as are examples and cross-links. Cross-links are links between different hierarchical levels. Novak and Gowin's scoring scheme is weighted heavily towards the hierarchical structure of the map. The theoretical rationale for this scoring scheme is Ausabel's theory of learning, particularly the idea of subsumption (new ideas can be subsumed under more general concepts) and progressive differentiation (as learning occurs, there is more differentiation among the concepts, which is shown by the inclusion of more propositions and cross-links).

Evidence from several studies suggests that Novak and Gowin's (1984) scoring scheme can differentiate between high- and low-knowledge students in biology (Markham, Mintzes, & Jones, 1994) and between first-year and advanced pediatric residents studying seizures (West, Pomeroy, Park, Gerstenberger, & Sandoval, 2000). This scoring scheme also appears to be sensitive to learning, as student map scores increased over the course of instruction (Pearsall, Skipper, & Mintzes, 1997; West et al., 2000).

A second scoring scheme that is commonly used considers only the propositions contained in the map and not the configural aspects. This method is to rate the quality of the propositions in the map. Each proposition is evaluated in terms of its accuracy. For example, Ruiz-Primo and colleagues used a proposition accuracy score as one measure of the quality of students' knowledge maps (Ruiz-Primo, Schultz, Li, & Shavelson, 1997a, 1997b; Ruiz-Primo et al., 2001). Each proposition in a student's map was scored on a 5-point scale, ranging from 0 (invalid/inaccurate) to 4 (complete and correct and showing a deep understanding of the relation between two concepts). Ruiz-Primo and colleagues found that students' proposition accuracy scores differentiated high-knowledge students from low-knowledge students (e.g., Ruiz-Primo et al., 1997a, 1997b) and students' map scores were moderately correlated (r between 0.40 to 0.50) with other measures of content knowledge in other formats (e.g., essays, multiple choice tests). Similar relationships have been found between knowledge map proposition accuracy scores and classroom end-of-unit tests and standardized tests of reading, math, and science (Rice, Ryan, & Samson, 1998), and between knowledge maps and physics problem solving (Austin & Shore, 1995).

Referent-Based Methods

Referent-based methods compare a student's map against a criterion map. Example referents include a domain expert's map, a composite map of experts, or the instructor's map. The essential measure is the number of propositions in the student map that are also in the referent map. Several studies have investigated the technical properties of this approach. For example, Ruiz-Primo et al. (2001), in addition to using proposition accuracy scores, also scored students' maps against an expert's map. The correlation between the proposition accuracy score and expert-based score was sufficiently high for Ruiz-Primo et al. to conclude that an expert-based method was the most efficient scoring method (i.e., in terms of scoring time and reliability of scores). Similar results were found by Osmundson et al. (1999) and Chung, Harmon, and Baker (2001).

The findings of Ruiz-Primo et al. (2001) are consistent with earlier work by Herl (1995), Herl et al. (1996), and Osmundson et al. (1999). In general, scoring student knowledge maps using expert-based referents has been found to discriminate between experts and novices (Herl, 1995; Herl et al., 1996), discriminate between different levels of student performance (Herl, 1995; Herl et al., 1996), relate moderately to external measures (Aguirre-Muñoz, 2000; Herl, 1995; Herl et al., 1996; Klein, Chung, Osmundson, Herl, & O'Neil, 2002; Lee, 2000; Osmundson et al., 1999), detect changes in learning (Chung et al., 2001; Osmundson et al., 1999; Schacter, Herl, Chung, Dennis, & O'Neil, 1999), and be sensitive to language proficiency (Aguirre-Muñoz, 2000; Lee, 2000).

The final type of scoring is to simply compare the network topology of a student's map and the referent map. Herl et al. (1996) investigated the utility of this approach and found high correlations between scores based on a comparison of the network topology and scores based on the overlap of propositions between the student and expert map.

Summary

A variety of methods have been developed and researched for scoring knowledge maps. Using expert-based referent maps as scoring templates appears to be an efficient means of scoring and produces results comparable to proposition quality ratings. This finding has important implications for automated scoring purposes, as will be discussed in the next section.

Review of Existing Mapping Software

The purpose of this review was to identify COTS software that could be used for knowledge mapping for assessment purposes. Our guideline for inclusion was practical: If we were to adopt existing mapping software for our own assessment-focused applications, which software would we at least review for consideration? We reviewed 15 packages that were designed specifically for knowledge mapping or could be used to construct knowledge maps. Two raters judged the software on the following criteria: (a) intuitiveness of the interface, and (b) automated scoring capability. One rater conducted the usability study described in the next section, and the second rater had no experience with knowledge mapping but had experience using graphical packages. Each criterion is briefly described.

Intuitiveness of the Interface

One of the critical lessons learned from our field experience using computer-based assessments is that the user interface needs to be as easy to use as possible. Typical field conditions include limited time in a group instructional setting in a non-computer lab room (e.g., classroom, library, or side room). The configuration of the setting has important training implications. Usually group instruction is required on how to use the software and users are not seated in front of a computer using the software during training. If the user interface is complex, then questions arise during the task. Under conditions where there is ample time, questions are usually not a problem; however, when there are time constraints (e.g., a class period) and the purpose is assessing student knowledge, then user-interface complexities may introduce undesirable effects that are unrelated to the task. In a distributed learning setting, we expect that the interface will play an even more important role as there will be no face-to-face support. Thus, ease of use is an important criterion for practical and validity reasons.

Scoring Capability

A second criterion is that if the intention is to use knowledge maps for assessment purposes in a distributed learning context, then as a practical matter there needs to be automated scoring capability. Further, it is not enough to be able to "score" a map. There should be validity evidence associated with the scoring algorithm, either by implementing one of the approaches discussed in the previous section, or by documenting the relationship between the scoring algorithm and other external measures, theory, or the extant literature. A summary of our review is shown in Table 2.

Table 2
Summary of Product Evaluations

Product	**Intuitiveness of the user interface**		**Automated scoring capability**	**Comments**
	R1	**R2**		
Axon Idea Processor	M	M	No	Difficult to find how to add relation labels.
Concept Draw MindMap	M	M	No	Interface is intimidating. Difficult to link subcategories.
Decision Explorer	L	L	No	Too hard to use.
Fuzzy Thought Amplifier	Evaluation copy unavailable			
IHMC Concept Map	L	L	No	Too hard to use without over-relying on Help function.
Inspiration	H	H	No	Easy to use, impressive range of choices.
Knowledge Manager	Evaluation copy unavailable			
LifeMap	L	M	No	Hard to use. Icons are not clear.
Mind Manager	M	M	No	Not easy to use.
Mind Mapper	M	H	No	Difficult to find how to add relation labels. Icons are confusing.
Mind Matrix	H	H	Yes	Easy to use.
Semantica	Evaluation copy unavailable			
Smart Ideas	H	H	No	Easy to use. Linking in particular is very clear.
SmartDraw	H	H	No	Easy to use but it has too many features that are not related to making concept maps.
Thinking Maps	M	H	No	Cannot add relation labels.
VisiMap	L	L	No	Too hard to use. Difficult to create new concepts.
Visio	H	H	No	Easy to use and visually appealing.
Visual Mind	M	H	No	Cannot add relation labels.

Note. H = high, M = medium, L = low.

Summary

The majority of software that provide knowledge mapping capability appear to be focused primarily on instructional applications, including representing and organizing information, and planning. Aside from our original mapping software (Version 1.0, Chung & Baker, 1997), Mind Matrix is the only other product that has focused on assessment uses and the only other product that has automated scoring capability. However, the product appears to only support knowledge mapping representations. Our requirement was for a tool to support knowledge mapping at a minimum, with the capability to expand into other assessment formats as the need arose.

In general, products not specifically designed for knowledge mapping tend to be sophisticated drawing packages. For products that have been designed specifically for knowledge mapping, there is a range of support for different kinds of maps and media. With the exception of the CRESST system, there appears to be little empirical work directly associated with the use of these systems for assessment purposes.

From the review of scoring algorithms and the review of COTS mapping packages, it was clear that there was no existing product that was sufficiently compelling to adopt out of the box for assessment purposes. Thus, we decided to design and develop our own system. We drew on our prior experience and the software review to develop functional requirements.

HPKMT Functional Overview

The primary purpose of the Human Performance Knowledge Mapping Tool is to support the assessment of an individual's content understanding of a particular domain in a distributed learning environment. There were four design goals: (a) develop a research-based, empirically supported measurement tool; (b) develop an easy-to-use application that accommodates users of varying computer experience; (c) develop an architecture that can support the development of a networking tool as well as accommodate future, unforeseen assessment formats; and (d) develop an architecture that can accommodate automated scoring of student responses for all supported assessment formats.

The requirements for the HPKMT were generated from a variety of sources. First, we reviewed the knowledge mapping literature to get an overview of how knowledge maps were used for assessment purposes. We also reviewed existing software from commercial and academic labs that purported to provide the capability to create knowledge maps. Finally, we incorporated lessons learned from the CRESST research experience with online knowledge mapping.

The knowledge mapping system reported herein is a second-generation version of our online mapping system. Our original version, reported in Chung and Baker (1997), was used in numerous contexts, from elementary school science classrooms to adult learners, under a variety of conditions ranging from instructional use to tests of content knowledge. The current version incorporates lessons learned from this period. Key lessons learned were:

- A simple interface was critical for end-user use and for training end-users. Because our application was testing, it was important to minimize the total amount of time the assessment-related activities took away from classroom instruction.

- There needs to be a general architecture that would support the rapid development of different assessment formats. Because we could not predict the range of assessment formats for future uses, we designed a system to accommodate the general attributes of expected tasks (i.e., declarative knowledge, procedural knowledge, physical-based knowledge).

- Authoring capability is essential to allow multiple users to easily stitch knowledge mapping tasks together.

- The method of launching the mapper could not be sensitive to typical network security constraints (e.g., applets are typically blocked by firewalls).

- Automated scoring capability is critical and the capability to incorporate multiple scoring approaches is essential to gather information on the technical quality of the scoring approach.

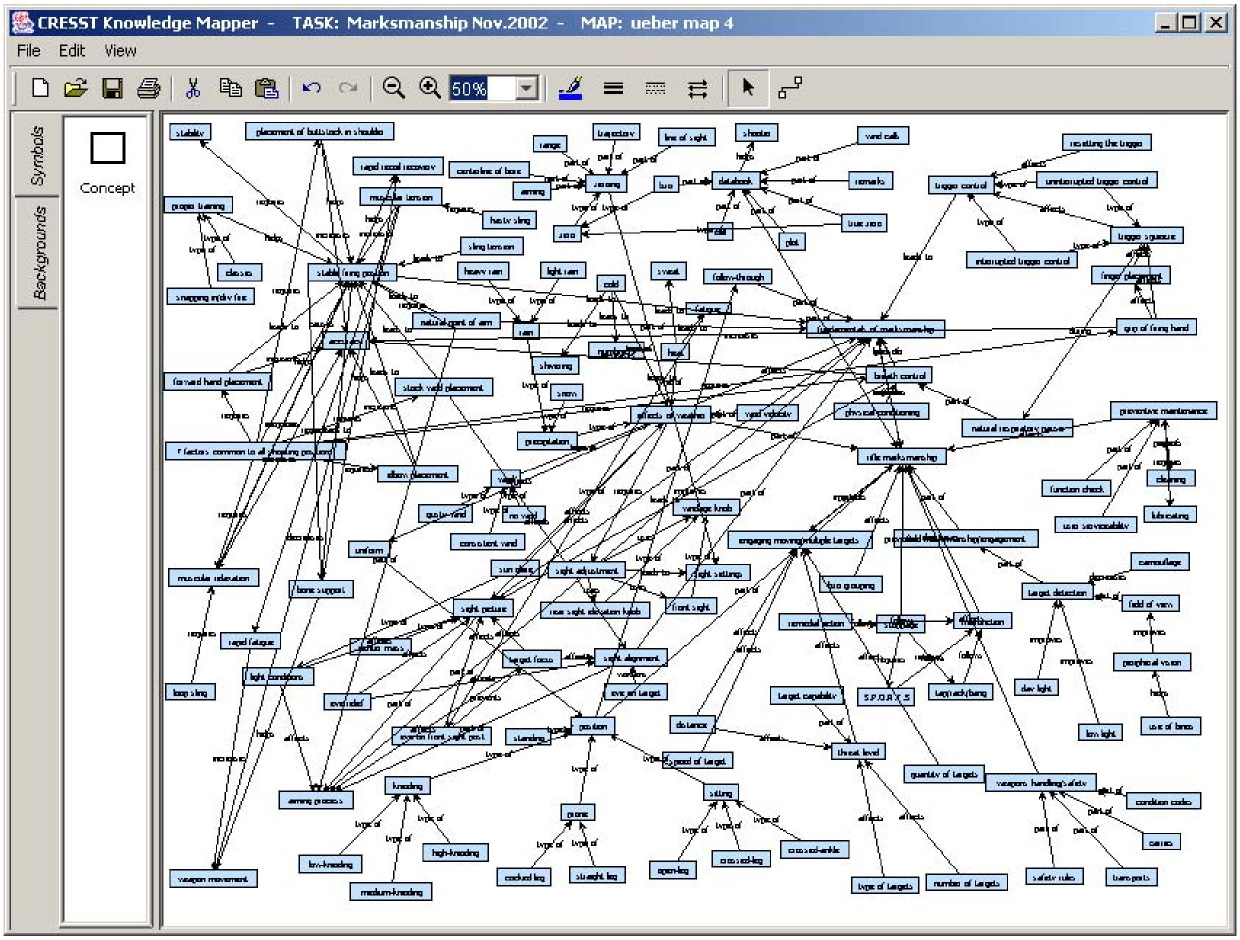

Figure 1. User interface of Version 2.0 of the CRESST knowledge mapper

User Interface

One of the clearest successes we have had with Version 1.0 is the usability of the software. Version 1.0 was simple, focused, and easily comprehendible by a wide range of users. There were clearly some user-interface problems, but these problems could be handled via training. Once the software was demonstrated, most users had little difficulty operating the software. We have repeatedly observed this result across different ages and contexts (e.g., Chung, Baker, & Cheak, 2002; Chung et al., 2001; Chung, O'Neil, & Herl, 1999; Herl, O'Neil, Chung, Bianchi et al., 1999; Herl, O'Neil, Chung, & Schacter, 1999; Klein et al., 2002; Osmundson et al., 1999; Schacter et al., 1999). Thus, our starting point for the design of Version 2.0 was to maintain the ease of use and simplicity of the user interface.

The main goal for the Version 2.0 user interface was for it to be simple and intuitive. Such an interface would translate into minimal training, explanation, and construct irrelevant variance—all important operational features when the tool is used for assessment purposes.

In Version 2.0, we adopted a modern interface for graphical drawing applications. As shown in Figure 1, the screen space was partitioned into two major areas: (a) symbol space, which housed icons; and (b) canvas space, which is where the drawing took place using the symbols from the symbol space. System functions common to applications were provided as well (e.g., buttons to open a file, save a file, zoom).

Basic Operation

The HPKMT was designed to be used with only the mouse. Concepts are added to the map by clicking and dragging the Concept icon onto the canvas. When a concept is added to the map, a pop-up menu of available concepts is provided for the user to choose from. Links between concepts are made by drawing a line from one concept to another. At the completion of the link operation, a pop-up menu of available links appears for the user to choose from. Concepts and links can be changed by either double-clicking or right-clicking on the concept or link to bring up the label menu.

Task Format

One goal for Version 2.0 was to develop a more general system in terms of the types of knowledge the system can assess. Increasingly, we found a single-representation format limiting (e.g., see Herl, O'Neil, Chung, Bianchi et al., 1999). Thus, for Version 2.0 we began with the assumption that assessment needs will evolve over time and that a tool to support a variety of unforeseen formats would be desirable.

The approach we adopted was to provide the capability for end-users to use different icon sets. An icon set is a collection of icons that represents the elements of the domain being assessed. For example, the HPKMT screen shown in Figure 1 has an icon set with one icon—a rectangle. A procedural knowledge icon set might have three icons—rectangle for process, diamond for decision making, and circle for begin/end.

The utility of this approach is twofold: (a) The same codebase can be used to store, retrieve, and display the representation. Changes occur at the icon-set level and not in the software. A library of icon sets can be developed to represent different assessments from the abstract (e.g., planning) to the concrete (as actual physical entities). (b) The assessment option is open-ended. If the domain can be represented using icons, the HPKMT shell can display it.

A second feature of Version 2.0 is to provide the capability to use different background sets. Like icon sets, the HPKMT can have images to display in the background. Different backgrounds can be used as insertable plates, depending on the purpose. For example, the way we typically use the HPKMT is with no background and the physical location of the nodes has no meaning. But another application may make use of the physical location. For example, this feature was used to assess Marines' knowledge of shot groups for a rifle marksmanship application. As shown in Figure 2, the same shell has been used to administer a "physical"-based representation. Physical placement of the icons relative to other icons and relative to the background image is used directly for scoring.

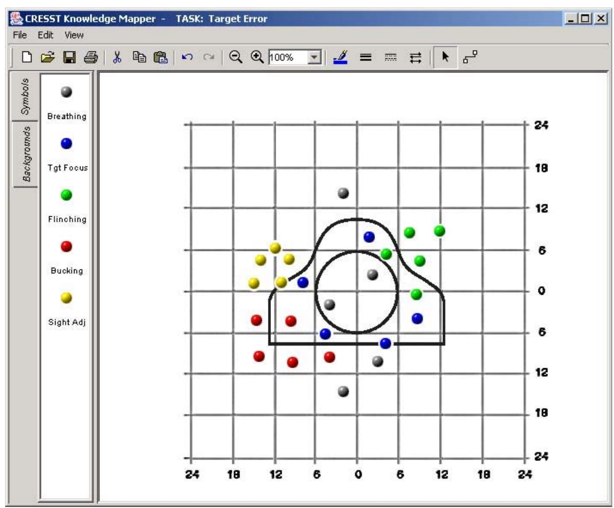

Figure 2. Use of background to depict physical-based layouts

Network Accessibility

One restriction imposed by firewalls is the blocking or filtering of Java applets and other objects (e.g., Active X). Our experience with Java applets in large school districts and commercial sites suggested that firewalls were becoming commonplace and highly restrictive. Given the high probability of the use of firewalls by our likely end-users, we designed the HPKMT to be client-based. We used a technology called Webstart, which allowed the HPKMT to be launched as an application on the user's machine via the user's browser. This is a key feature because the HPKMT is implemented as an application, not an applet (thereby avoiding firewall restrictions). Code maintenance is handled automatically by Webstart, which automatically checks for updates to the HPKMT and downloads components that have changed.

Automated Scoring Capability

Scoring of Knowledge Maps

We designed the scoring engine to be loosely coupled to the system, thereby allowing for maximal flexibility in the scoring routines. That is, knowledge mapping data are stored in tables and databases that are independently accessed by the scoring engine. This separation reduces the complexity of the mapping code, increases maintainability, and allows for scoring algorithms to be developed independently of the user interface.

Currently, we have implemented the following scoring routines for the knowledge mapping representation: (a) exact proposition matching and (b) synonym-based proposition matching. The exact-proposition-matching algorithm is based on the algorithm developed by Herl et al. (1996). Exact proposition matching involves counting the number of propositions in the trainee map that also exist in the referent map (e.g., an expert's map).

Synonym scoring is a more lenient method that considers a match if there exists an intersection between the set of synonyms for each term in the trainee map and each corresponding term in the referent map. Currently, the synonyms are stored in a database table and manually specified depending on the context of the particular task. We are currently integrating WordNet into the synonym scoring algorithm, as the WordNet tables contain synonyms for over 100,000 words.

Scoring of Shot Groups

The second automated scoring application we have developed within the HPKMT shell is to score shot-group patterns. As shown in Figure 2, trainees are asked to depict the prototypical shot groups of five types of shooter errors. The scoring algorithm compares the trainee's response against an expert's response and was based on work by Johnson (2001). The particular variables compared differ by shot group, but in general include dispersion area, dispersion radius, vertical displacement, horizontal displacement, and target quadrant.

Summary

Our design approach has already proved useful. The architecture of the system allowed us to deliver different assessment formats within the same delivery shell using the same codebase. Automated scoring has been implemented for each format. The use of the same shell to deliver assessments minimizes user-interface complexities.

We have developed a novel application that moves the complexity of software development to the scoring aspect. For each class of representation types, software needs to be developed to score that representation. But this is much easier to do than creating stand-alone software for each type of assessment. Thus, combinations of icons and backgrounds can be used to maximize the reuse and representational generalizability of the tool.

Usability Study

Because we anticipated a wide range of users of the mapper software, a usability study was conducted to (a) evaluate how intuitive the operation of the HPKMT was, (b) evaluate how learnable the HPKMT operations were, and (c) uncover error and atypical usage patterns.

In general, the usability study found that users of varying levels of computer experience learned how to use the HPKMT with minimal instruction. The most difficult function of the tool to learn was linking concepts, with the most users requiring suggestions or explicit instruction on how to link. Since participants were asked to use and explore the tool with no instruction outside of an introduction to knowledge mapping in general, the findings suggest that a short training session illustrating the basic functions of the tool would be sufficient for users with even limited computer experience. Once users learned how to use the tool, there were few recurring usability issues. The usability study yielded recommendations for changes to the HPKMT, including modifications to the linking function, and renaming menu items and toolbar button labels for clarity.

Fielded Use

We tested the HPKMT in an assessment context with 50 Marines at a training base in North Carolina. Most of the Marines were between 18 and 20 years old. Our informal observations, based on observing Marines using the software and Marines' questions and comments, suggested that the Marines had little trouble using the tool. The only notable question was how to delete objects, as some users attempted to delete objects with the backspace key instead of the delete key.

Prototype Authoring Functionality

To support the creation and maintenance of knowledge mapping tasks, a prototype authoring system was created. The design of the system was based on the work of Chung et al. (2002). The near-term goal of the authoring functionality was to support research activities. A long-term goal was to develop an interface that is suitable for use by a wide audience (e.g., trainers, trainees, researchers, course managers).

The authoring prototype was developed to facilitate the research use of the HPKMT. The prototype contains the minimal functionality required to allow users to create mapping tasks and assign tasks to students and other trainees.

The authoring process is illustrated in Figure 3. The metaphor adopted was that one of a course. A course consists of one or more knowledge mapping tasks. Each mapping task has properties associated with it, such as the set of concepts and links, the set of icons used, and the mode of operation (e.g., select-only, type-in, or both). Tasks can be created by using existing concepts and links, by creating new ones, or by editing an existing set. Trainees' names must be registered in the system before mapping can be accessed. Trainees are assigned tasks. In the following sections, each major functional area is described.

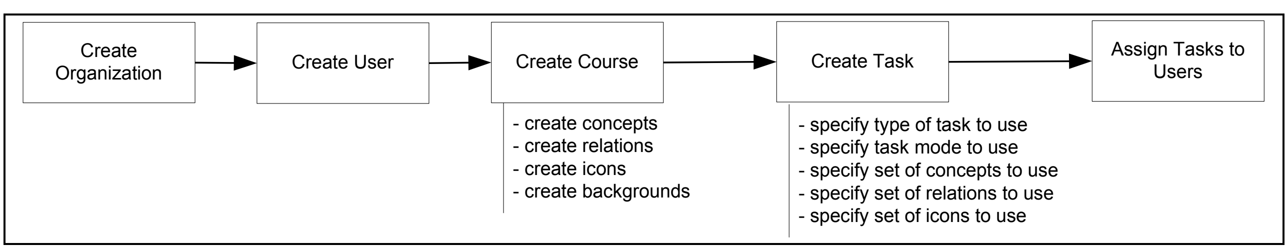

Figure 3. Top-level flow for creating a new knowledge mapping task

Top-Level Organization

Users can access all top-level functions from the main screen. The design of the Web favored breadth over depth, which allows users to reach their destination within 4 clicks (Chung et al., 2002).

Course Creation

A critical step in the authoring process is setting up a "course." A course bundles one or more tasks and can be assigned to trainees. The tasks within a course can be assigned to trainees.

Task Creation

Creating a task involves selecting the type of knowledge representation (e.g., declarative knowledge, procedural knowledge, physical) and specifying whether the student will be able to select terms and links, type terms and links, or both. Other important functions are specifying the concepts and relations to be used, as well as the set of icons to be used for mapping and the background.

Specifying Concept and Link Sets

Specifying the set of concepts and relations is done via a simple HTML form. The form allows the creation of sets of concepts and relations that can be reused in future tasks. In this way, link sets can be quickly created with a known set of links, and tailored to specific content areas.

Specifying Image Sets

Analogous to specifying concept and link sets, users can also specify image sets for the mapping symbol set and the background image sets. One purpose of this function is to allow rapid uploading of specific images that can be used in the HPKMT. For example, instead of the rectangle used in Figure 1, one could specify a custom-design icon created in a third-party graphics program (e.g., a digital image or symbol) and have the custom-designed icon function as the node. In fact, that is what was done in Figure 2—icons were created offline to represent bullet strikes, and the background target was a scanned GIF file of a target.

Discussion

We have developed a software application that can administer and score knowledge maps and other formats. Our usability testing and our field test suggest that we have developed an easy-to-use application. Our architecture was designed to allow different assessment formats by allowing end-users to upload icons and background images. In our review of existing mapping software, we found only one other application focused on assessment, and we found no other application that has multiple assessment capability. Our use of knowledge mapping as an assessment tool with automated scoring and our general capability to support unforeseen assessment formats are unique. The CRESST HPKMT appears to be the only empirically supported online mapping tool in existence.

Limitations of This Work and Next Steps

There are two major limitations of this work. First, the authoring component needs to be recast to make it more user-friendly and intuitive. Currently, specialized knowledge and training are needed before tasks can be created. There is no capability to specify tasks in real-time (e.g., using the HPKMT itself as the authoring tool). The second limitation is the lack of reporting. We currently have no unified method of reporting student performance. Scores can be reported on a single-user basis, but there is no capability for group-level reporting.

Our efforts at designing a usable knowledge mapping tool appear to be a successful first step. The software is reasonably stable and we have tested the architecture by implementing two different assessment formats with scoring capability. Clear next steps include retooling the authoring prototype interface. A second line of activity is to deploy the HPKMT in distributed learning contexts in different content areas to expand the suite of assessments.

REFERENCES

Aguirre-Muñoz, Z. (2000). *The impact of language proficiency on complex performance assessments: Examining linguistic accommodation strategies for English language learners.* Unpublished doctoral dissertation. University of California, Los Angeles.

Austin, L. B., & Shore, B. M. (1995). Using concept mapping for assessment in physics. *Physics Education, 30,* 41-45.

Chung, G. K. W. K., & Baker, E. L. (1997). *Year 1 technology studies: Implications for technology in assessment* (CSE Tech. Rep. No. 459). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Chung, G. K. W. K., Baker, E. L., & Cheak, A. M. (2002). *Knowledge Mapper authoring system prototype* (CSE Tech. Rep. No. 575). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Chung, G. K. W. K., Harmon, T. C., & Baker, E. L. (2001). The impact of a simulation-based learning design project on student learning. *IEEE Transactions on Education, 44,* 390–398.

Chung, G. K. W. K., O'Neil, H. F., Jr., & Herl, H. E. (1999). The use of computer-based collaborative knowledge mapping to measure team processes and team outcomes. *Computers in Human Behavior, 15,* 463–494.

Herl, H. E. (1995). *Construct validation of an approach to modeling cognitive structure of experts' and novices' U.S. history knowledge.* Unpublished doctoral dissertation, University of California, Los Angeles.

Herl, H. E., Baker, E. L., & Niemi, D. (1996). Construct validation of an approach to modeling cognitive structure of U.S. history knowledge. *Journal of Educational Research, 89,* 206-218.

Herl, H. E., O'Neil, H. F ., Jr., Chung, G. K. W. K., Bianchi, C., Wang, S-L., Mayer, R.E., Lee, C-Y., Choi, A., Suen, T., & Tu, A. (1999). *Final report for validation of problem-solving measures* (CSE Tech. Rep. No. 501). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Herl, H. E., O'Neil, H. F., Jr., Chung, G. K. W. K., & Schacter, J. (1999). Reliability and validity of a computer-based knowledge mapping system to measure content understanding. *Computers in Human Behavior, 15,* 315–334.

Johnson, R.F. (2001). Sta*tistical measures of marksmanship* (ARI Tech. Note TN-01/2). Alexandria, VA: U.S. Army Research Institute.

Klein, D. C. D., Chung, G. K. W. K., Osmundson, E., Herl, H. E., & O'Neil, H. F., Jr. (2002). *The validity of knowledge mapping as a measure of elementary students' scientific understanding* (CSE Tech. Rep. No. 557). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Lee, J. J. (2000). *The impact of Korean language accommodations on concept mapping tasks for Korean American English language learners.* Unpublished doctoral dissertation. University of California, Los Angeles.

Markham, K., Mintzes, J., & Jones, M. G. (1994). The concept map as a research and evaluation tool: Further evidence of validity. *Journal of Research in Science Teaching, 31,* 91-101.

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn.* New York: Cambridge University Press.

Osmundson, E., Chung, G. K. W. K., Herl, H. E., & Klein, D. C. D. (1999). *Concept mapping in the classroom: A tool for examining the development of students' conceptual understandings* (CSE Tech. Rep. No. 507). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Pearsall, N. R., Skipper, J., & Mintzes, J. (1997). Knowledge restructuring in the life sciences: A longitudinal study of conceptual change in biology. *Science Education, 81,* 193-215.

Rice, D., Ryan, J., & Samson, S. (1998). Using concept maps to assess student learning in the science classroom: Must different methods compete? *Journal of Research in Science Teaching, 35,* 1103-1127.

Ruiz-Primo, M. A., Schultz, S., Li, M., & Shavelson, R. J. (1997a). *Concept map-based assessment in science: Two exploratory studies* (CSE Tech. Rep. No. 436). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Ruiz-Primo, M. A., Schultz, S., Li, M., & Shavelson, R. J. (1997b). *On the validity of concept-map based interpretations: An experiment testing the assumption of hierarchical concept maps in science* (CSE Tech. Rep. No. 455). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Ruiz-Primo, M. A., Schultz, S. E., Li, M., & Shavelson, R. J. (2001). Comparison of the reliability and validity scores from two concept-mapping techniques. *Journal of Research in Science Teaching, 38,* 260-278.

Ruiz-Primo, M. A., & Shavelson, R. J. (1996). Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching, 33,* 569-600.

Schacter, J., Herl, H. E., Chung, G. K. W. K., Dennis, R. A., & O'Neil, H. F., Jr. (1999). Computer-based performance assessments: A solution to the narrow measurement and reporting of problem-solving. *Computers in Human Behavior, 15,* 403–418.

West, C. D., Pomeroy, J. R., Park, J. K., Gerstenberger, E. A., & Sandoval, J. (2000). Critical thinking in graduate medical education: A role for concept mapping assessment. *Journal of the American Medical Association, 284,* 1105-1110.