Linking Assessment and Instruction Using Ontologies

Gregory K. W. K. Chung, Girlie C. Delacruz, Gary B. Dionne, William L. Bewley

UCLA/CRESST

Los Angeles, CA

greg@ucla.edu, gdelacruz@cse.ucla.edu, gdionne@ucla.edu, bewley@cse.ucla.edu

ABSTRACT

In this study we report on a test of a method that uses ontologies to individualize instruction by directly linking assessment results to the delivery of relevant content. Our sample was 2nd Lieutenants undergoing entry-level training on rifle marksmanship.

Ontologies are explicit expressions of the concepts in a domain, the links among the concepts, and the governing constraints of these links. We have developed an ontology for the domain of rifle marksmanship. The ontology contains over 160 concepts and over 160 relationships that capture the different types of relations among the concepts (e.g., causal, part-whole, classifying, functional). The content was drawn from Marine field manuals, and interviews with snipers and coaches. Concepts were tagged with instructional content (e.g., definitions, explanations, elaborations, multimedia examples). Relations were tagged with an explanation of why the particular relation holds under particular conditions.

Assessment is tied to instruction via influence (Bayesian) networks. Performance on assessment items determines what content is pulled from the ontology for delivery. For example, if a Marine scores poorly on all assessment items related to breathing control, then instructional content tied to the ontology concept "breathing control" (and any linked concepts) could be delivered. Conversely, if a Marine scores low on items that suggest poor knowledge of the shot group associated with poor breathing control, then only a shot group related to breathing might be delivered.

Our test of this approach appears feasible and promising. The Bayesian network appeared to be successful in identifying knowledge gaps, and relevant and targeted content was served to Marines. Learning appeared to be occurring at a faster rate over time for Marines who received targeted instruction compared to Marines in a control group. Implications are discussed.

ABOUT THE AUTHORS

Dr. Gregory K. W. K. Chung is a Senior Research Associate at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST). His current work at CRESST involves developing problem-solving assessments for computer-based assessments and developing Internet-based assessment tools for diagnostic and embedded assessment purposes. Dr. Chung earned a Ph.D. in Educational Psychology from the University of California at Los Angeles, an M.S. degree in Educational Technology from Pepperdine University at Los Angeles, and a B.S. degree in Electrical Engineering from the University of Hawaii at Manoa.

Girlie C. Delacruz is a Research Associate at CRESST. Her current work at CRESST involves developing Internet-based assessment tools for diagnostic and embedded assessment purposes, evaluating technology-based learning environments, and developing problem-solving assessments for computer-based assessments. Ms. Delacruz earned her B.A. degree in Psychology with a minor in Applied Developmental Psychology from the University of

California at Los Angeles, and is pursuing her Ph.D. in Psychological Studies in Education at the University of California at Los Angeles.

Gary B. Dionne is a software developer and systems analyst at CRESST. His work encompasses a wide range of disciplines, including algorithm and systems development, Internet-based programming, and natural language processing, and their uses in novel assessment and evaluation vehicles. Mr. Dionne is currently pursuing his B.S. degree in Electrical Engineering and Computer Science from the University of California at Los Angeles.

Dr. William L. Bewley is an Assistant Director of CRESST. His experience combines research on human learning and performance, instructional systems design, education and training product development, software development, and management. His current work is concerned with research on applications of advanced technology to human performance assessment in distance learning and team problem-solving environments and evaluation models for e-learning. He has advised the National Science Foundation, commercial companies, and non-profit organizations on educational technology and integration of technology into curricula. Dr. Bewley earned a Ph.D. and M.S in Psychology and Computer Science from the University of Wisconsin and a B.A. in Psychology from Stanford University.

Linking Assessment and Instruction Using Ontologies

Gregory K. W. K. Chung, Girlie C. Delacruz, Gary B. Dionne, William L. Bewley

UCLA/CRESST
Los Angeles, CA
greg@ucla.edu, gdelacruz@cse.ucla.edu, gdionne@ucla.edu, bewley@cse.ucla.edu

CONTEXT OF STUDY

The focus of this research was on evaluating an automated approach to link assessment information (culled from tests of knowledge) to individualized instructional recommendations. That is, given that assessments results suggest a gap in someone's knowledge, can an automated method be developed to provide remediation that targets the individual's specific knowledge gaps?

This work is embedded in a larger research program to develop assessment models and tools for Naval distributed learning. CRESST is under contract to the Office of Naval Research (ONR) and the first application of our work is for U.S. Marine Corps (USMC) marksmanship training. Our USMC work is focused on developing online assessments of Marines' knowledge of rifle marksmanship.

Our approach was to use Bayesian networks and assessments of knowledge to first infer an individual's knowledge gap, and then deliver remediation content (pulled from an ontology) that was targeted to address only that knowledge gap.

Definition of an Ontology

An ontology provides a shared and common understanding of a domain that can be communicated among people and computational systems (Fensel, Hendler, Lieberman, & Wahlster, 2003). The ontology captures one or more experts' conceptual representation of a domain expressed in terms of concepts and the relationships among the concepts. An ontology is a commitment to a point of view of how a domain is structured, but there can be multiple representations (Chandrasekaran, Josephson, & Benjamins, 1999; de Clercq, Hasmon, Blom, & Korsten, 2001; McGuinness, 2003). Ontologies are important because they provide a common, explicit framework for sharing and using knowledge. More concretely, an ontology standardizes the terms and structure of the domain. The standardization makes possible sharing of the ontology; thus, the knowledge contained therein is used across multiple computer platforms for different applications (Gruber, 1995). Ontologies were first developed as part of the AI research effort to facilitate knowledge sharing and reuse. The use of ontologies has extended recently to fields such as information retrieval, knowledge management, medical guidelines, military, and e-commerce. CRESST is now applying ontologies to assessment.

ONTOLOGIES TO SUPPORT ASSESSMENT AND INSTRUCTION

For assessment and instructional purposes, the capability to express the concepts in a domain, the links among the concepts, and the governing constraints offers clear advantages over relational or highly structured data models. Usually, the representation of a domain is best represented as a network (vs. a strictly hierarchical representation, for example), especially in knowledge-rich applications.

The existence of computational tools to create, edit, maintain, and exchange ontologies makes feasible the use of ontologies in assessment and instruction. Protégé is one such computational tool, originally developed in 1987 at Stanford University and now in its third generation (Gennari et al., 2002). Protégé has an easy-to-use graphical user interface, Java implementation, and an active developer community. Similar products are available from both academic and commercial vendors.

In the following sections, we describe an ontology we developed on rifle marksmanship for the USMC.

Ontology of U.S. Marine Corps Rifle Marksmanship Knowledge

The overall purpose for developing an ontology was to capture the knowledge and structure of the domain in a way that would allow exploration of the use of ontologies for assessment and instructional purposes. We judged the domain of rifle marksmanship to be an ideal candidate to represent in an ontology because the domain is bounded, and domain experts agreed on the set of important topics.
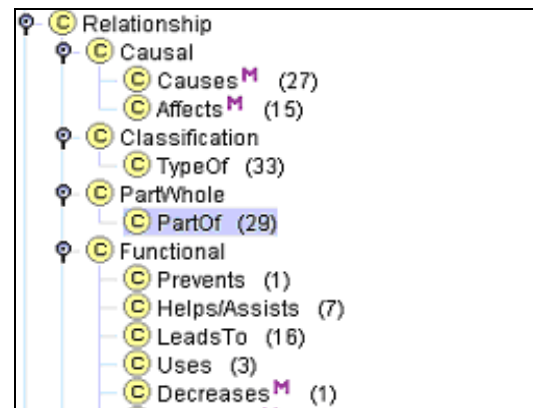
Domain Structure

Our knowledge engineering strategy was to capture knowledge in two representations: (a) as outlined by doctrine (e.g., USMC field manuals), information which could be organized as a hierarchically structured body of knowledge; and (b) as perceived by experts (e.g., coaches, snipers, rifle team members), information which could be organized conceptually (i.e., as a network) to reflect how domain experts perceived the knowledge to be interrelated.

Currently, our rifle marksmanship ontology contains 168 different concepts that cover seven fundamentals of rifle marksmanship and 160 relationships among the concepts using 16 relationship types. Figure 1 shows a portion of the hierarchy of the ontology. The structure of the content is captured by the *Knowledge* class. The hierarchical structure shows the taxonomy of class and subclass relationships among the topics.
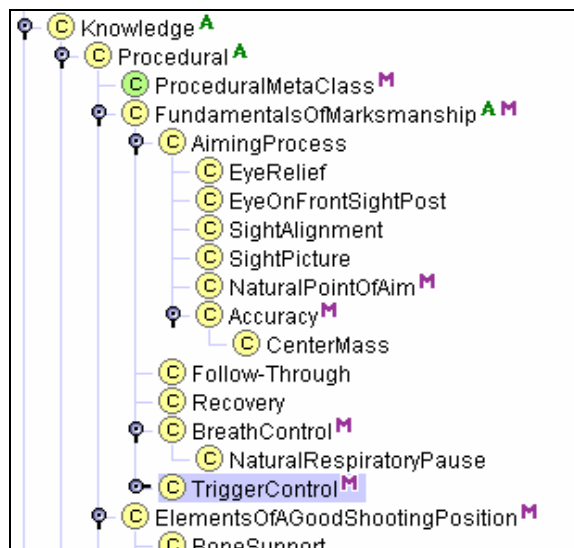
Figure 1. Example of the rifle marksmanship content organized hierarchically

Figure 2 shows how the content is organized as perceived by our domain experts. In this case, the organization is a network and represented by the *Relationship* class. The *Relationship* class is made up of subclasses that represent high-level relation types (e.g., causal, part/whole). Subclasses of each relation type represent increasingly specific relations (e.g., *PartOf* is a particular kind of relation within the *PartWhole* class). Figure 2 shows specific instances of the *PartOf* relation that directly connect different topics shown in Figure 1. Our assumption is that the hierarchical representation reflects the organizational structure of the content similar to a table of contents, and the relational structure captures the detailed relations that presumably underlie deep understanding of the content.


Figure 2. Example of relationship classes. The relationship class specifies how the content is related conceptually

Binding Content to the Ontology Structure

Many ontologies typically capture only the structure of the domain (e.g., Figure 1). However, to be useful instructionally, content would ideally be bound to the structure. For example, Figure 3 shows an example of how content is related directly to objects in the ontology. For each topic, we have defined different knowledge types—conceptual (or declarative) knowledge and procedural knowledge. Further, we have partitioned the information into subtypes: definition, explanation (i.e., why the topic is important), and elaboration (i.e., supplemental information). Although not shown in Figure 3, we have also allowed for the inclusion of different media types (e.g., video, picture, URL). For example, for the topic *BreathControl* we have a video demonstrating the effects of breathing on the position of the rifle muzzle and bullet strike (breathing causes the rifle to move vertically; firing while breathing results in a vertical dispersion of shots).

Source material was drawn from the U.S. Marine Corps rifle marksmanship manual (USMC, 2001). Marksmanship training is derived from this manual. For concepts, the instructional content is delineated in terms of definition, explanation, elaboration, and multimedia examples (e.g., a picture of the trigger) where appropriate. For relations, the instructional content was an explanation of why the particular relation holds under the particular conditions.
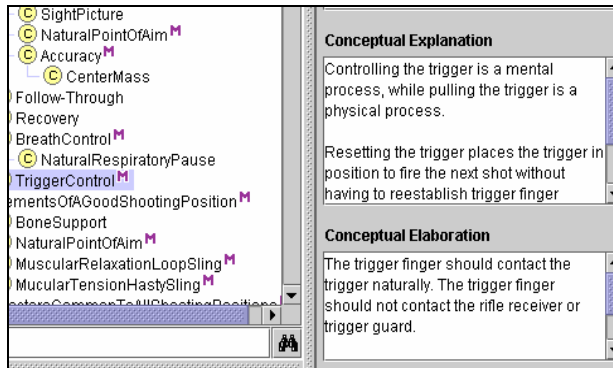
Figure 3. Example of the rifle marksmanship content bound to the topic *TriggerControl*

RECOMMENDING INDIVIDUALIZED INSTRUCTIONAL CONTENT

Because of how we have structured the ontology (i.e., hierarchical and network/conceptual representations) *and* because we have bound content at different grain sizes to specific topics in the ontology, we now have the means to deliver content at different grain sizes depending on the application. In this section we describe our technique for identifying knowledge gaps and delivering individualized content.

Identifying Knowledge Gaps Using Bayesian Networks

The first step in recommending individualized content is to identify an individual's knowledge gaps. Once the gaps are identified, relevant content needs to be retrieved and delivered to the individual.

Identifying what students know and do not know is accomplished by diagnostic assessments. For example, our strategy for assessing Marines' understanding of rifle marksmanship is to use a range of measures that reflect different cognitive demands. For example, we broadly sample their knowledge of marksmanship using selected-response multiple-choice tests.

This assessment information is then fused together using a Bayesian network to yield probabilities on the degree to which a Marine understands different topics of rifle marksmanship. A Bayesian inference network, also known as an influence or probabilistic causal network, depicts the causal structure of a phenomenon in terms of nodes and relations (Jensen, 2001). Nodes represent states, and links represent the influence relations among the nodes. Node states can be observable or unobservable.

The utility of a Bayesian inference network is that it yields the probability that an unobservable variable is in a particular state (e.g., understands trigger control) given observable evidence (e.g., whether the participant knows the definition of trigger control). The probability of the unobservable variable being in a particular state is the inference made about student understanding.

Recommending Instructional Content

Linking the Bayesian network and the ontology is conceptually equivalent to the link between assessment and instruction. That is, the (unobservable) nodes in the Bayesian network was conceptualized to represent a concept in the domain of rifle marksmanship. The probability values for the nodes (or concepts) was taken to reflect the probability that the Marine understood that concept. For each concept for which we had content, if the probability fell below the threshold (set to .65 after inspecting the probability distribution), then the software pulled content from the ontology and made it available to the Marine. There was a one-to-one mapping between the concepts in the Bayesian network and concepts in the ontology.

RESEARCH QUESTIONS

Our research questions focused on examining the feasibility of individualizing content delivery based on a model of knowledge dependencies:
• To what extent does our Bayesian network detect knowledge gaps in individual participants with respect to the domain of rifle marksmanship?
• How effective is individualized content delivery on learning when a Bayesian network is used to detect knowledge gaps and an ontology is used to provide relevant and detailed content?

METHOD

Participants

Fifty-three 2nd Lt. Marines undergoing entry-level rifle marksmanship training were recruited for this study. Of the 53 Marines, 16 participants were randomly assigned to the experimental condition (individualized-content delivery study), and the remaining 37 assigned to a control condition.

Design

We used a two-group pretest, treatment, posttest design. The treatment condition received feedback of our estimates of their knowledge on different topics of rifle marksmanship, based on the Bayesian network probabilities. Participants were then given online access to relevant content on those topics. The control group did not receive the feedback or access to the content. Pretest and posttest measures are described in the measures section.

Tasks

The primary task for participants in the treatment condition was to first complete the assessment measures (described next) and then receive a "report card" on rifle marksmanship topics the system "scored."

Given the score, Marines were instructed to learn as much as they could about the topics on which they received a low score. In this way, we approximated the assessment-instruction cycle. The entire system was administered in an online format. Marines were given access to information about topics on which they scored low. The content for these topics was drawn directly from the marksmanship ontology, and included text explanations, digital photographs, or digital videos.

An example screenshot is shown in Figure 4. For each Marine, information was made available on topics for which a Marine scored 6 or lower. Also, different kinds of information was made available depending on the Marine's performance on the various assessment items. For example, if a Marine got a definition of a topic correct but performed poorly on more complex assessment items covering the same topic, the definition of the topic was not delivered. The intent was to deliver only the information needed, no more and no less.

Figure 4. Fragment of the screen shown for a particular Marine who scored low on the topic of Stock Weld Placement.

Measures

Qualification Score
The qualification score was the Marines' score of record. The qualification score is the primary performance measure.

Background Information
The following information was collected from participants: age, ethnicity, sex, rank, ASVAB general technical score, occupational specialty, and type of unit.

Knowledge Mapping
Knowledge maps were used to measure participants' conceptual knowledge (Herl, Baker, & Niemi, 1996).

The task required participants to graphically depict their understanding of rifle marksmanship in terms of a network. The nodes in the network represented concepts, and labeled links represented the relationships among concepts. Twenty-five concepts and 10 links were provided to participants and the knowledge map task was administered online.

Prior Knowledge

The prior knowledge measure was designed to survey participants' knowledge of rifle marksmanship. Participants were given a 41-item multiple-choice test that sampled the following topics: sight picture, sight adjustment, sight alignment, weapons safety, breathing, trigger control, stock weld, eye relief, bone support, firing hand placement, follow-through, forward hand placement, grip of firing hand, and muscular relaxation.

Shot Group Depiction

The shot group depiction task was designed to measure participants' knowledge of the shot groups associated with common shooter problems. Participants were instructed to draw a 5-shot group for problems with breathing, sight adjustment, flinching, bucking, and focusing on the target.

Evaluation of Shooter Positions

This task was intended to measure participants' skill at identifying proper and improper firing positions of a shooter posing in proper and improper positions. The shooter was shown in QuickTime VR, and participants could rotate the image to view the shooter from different angles. Participants were asked to judge how proper or improper the shooter's position was on the following elements: placement of firing hand, placement of forward hand, forward elbow placement, stock weld placement, rifle butt placement, leg placement, feet placement, and body placement.

Scientific Reasoning

Lawson's Classroom Test of Scientific Reasoning (CTSR) (revised 24-item multiple choice edition) was used to measure scientific reasoning (Lawson, 1987). All items were multiple choice. The purpose for including the CTSR was to gather information on participants' reasoning; this measure was used as a proxy for aptitude.

Level-of-Knowledge Survey (Experimental Condition Only)

Participants in the experimental condition were instructed to rate their knowledge on a scale of 0-10 on various rifle marksmanship concepts. The list of concepts comprised the Bayesian network and included top-level concepts (e.g., aiming) and low-level (e.g., grip of firing hand) concepts.

Procedure

Data collection occurred over a 2-week period. Prior to any training on marksmanship or the treatment, all participants were administered a pretest knowledge map and the CTSR measure. Participants in the control condition were also administered the prior knowledge measure. Participants then attended classroom lectures for 1.5 days on rifle marksmanship. Following the classroom lectures, participants in all conditions were administered a second mapping task where they were instructed to improve their maps. In addition, the experimental condition received the prior knowledge measure—the purpose of administering this measure after instruction was to have a range of performance with which to update the Bayesian network.

A third mapping task was administered a day later after participants received firing practice and coaching; however, participants in the experimental condition first received the intervention (i.e., feedback on their level of knowledge and individualized content delivery). Participants receiving the feedback were instructed to learn as much as they could on the topics they scored low on. Following the intervention, the experimental condition then received a posttest prior knowledge task and a posttest knowledge mapping task.

Two additional knowledge mapping tasks were administered throughout practice firing, and a final knowledge mapping task was administered at the end of the training sequence (i.e., after the participants fire for "record score"). The final mapping task required participants to start with a blank map and participants in the control condition were administered posttest prior knowledge surveys.

RESULTS

Two sets of analyses are presented, organized by research questions. The first set of analyses examines the fidelity of the Bayesian network model with respect to detecting knowledge gaps in individuals. The second set of analyses examines the instructional effect of individualized content delivery.

To what extent does the Bayesian network model of the dependencies, among rifle marksmanship knowledge, detect knowledge gaps in individual participants?
Individual items from our prior knowledge, shot-group, and QuickTime VR assessments were used as input

(i.e., evidence) to the Bayesian network. Given the evidence, the network was updated and probabilities were obtained for each "hypothesis" node. The hypothesis node represents the inference that a participant knows a concept given his/her performance on the assessments.

The probabilities from the hypothesis variables were used as scores and were rescaled from 0 to 1.0 to 0 to 10, to correspond to the scale of the level-of-knowledge survey administered to participants.

Because of the small number of participants who received the level-of-knowledge survey (i.e., $n = 16$), we dichotomized level-of-knowledge scores into two categories: low and high knowledge. Thus, scores from 0 to 5 were considered low, and scores from 6 to 10 were considered high. This transformation was done on participants' self-reports of their level of knowledge and on the scores derived from the Bayesian network.

The first set of analyses examined the correspondence between the level-of-knowledge scores derived from participants' self-reports and the scores derived from the Bayesian network. As shown in Table 1, in general most participants rated their knowledge of the different concepts as high on nearly all of the concepts. The Bayesian network scores consistently agreed with participants' perception. The overall agreement percentage across all concepts is 79%. While these results appear favorable with respect to the Bayesian network model of knowledge dependencies, caution should be used when interpreting these results: there was a skewed distribution across low and high categories (i.e., it is unclear what the agreement would be if there were more participants who rated their knowledge as low). A second caution is that the validity of these results depends on the accuracy of participants' perceptions of their level of knowledge.

Table 1. Agreement between participants' and Bayesian network level-of-knowledge scores (high or low) ($n = 16$)

Concept in Bayesian network	No. of matches		No. of mis-matches
	Low	High	
Aiming process[a]	1	13	2
Breath control	0	9	7
Trigger control	0	13	3
Bone support	1	10	5
Elbow placement	2	10	4
Eye on front sight post[a]	1	12	3
Eye relief	1	12	3
Feet placement[a]	1	13	2
Firing hand placement[a]	1	13	2
Finger placement[a]	1	14	1
Follow-through	2	12	2
Forward hand placement	1	11	4
Grip of firing hand	1	11	4
Leg placement[a]	1	13	2
Muscular relaxation[a]	1	13	2
Natural point of aim[a]	1	12	3
Rifle butt placement	0	12	4
Natural respiratory pause	0	15	1
Sight adjustment[a]	6	0	10
Sight alignment	1	12	3
Sight picture	0	13	3
Stockweld placement	2	9	5
Trigger control procedure[a]	0	13	3
Trigger squeeze	0	13	3

[a]These concepts were part of the Bayesian network but content was not available for these concepts and thus they were not part of the content delivery.

Table 2. Non-parametric correlations (Spearman) between probabilities for high-level concepts in the Bayesian net and knowledge and performance measures ($N = 53$)

Concept in Bayesian network	CTSR	Knowledge map	Prior knowledge of rifle marksmanship	Shot group	Evaluation of shooter positions	Qualification score
Fundamentals of rifle marksmanship	.28*	.08	.73**	.27*	.32*	.22§
Aiming	.35**	.06	.68**	.24§	.38**	.20
Breath control	.24	.08	.66**	.48**	.17	.16
Trigger control	.36**	.20	.50**	.30*	.30*	.40**
Position	.17	.14	.59**	.17	.36**	.32*

§$p < .10$ (two-tailed).
*$p < .05$ (two-tailed).
**$p < .01$ (two-tailed).

The next set of analyses examined the associations between major concepts in the Bayesian network and external measures (Table 2). Presumably, if the dependencies have been modeled accurately, then the scores should be correlated. For this analysis, the full sample of participants was available, and probabilities were used as scores. The non-parametric procedure (Spearman) was used due to the skewed distribution of the probabilities.

The results shown in Table 2 are interesting. The correlations between the prior knowledge measures and the probabilities in the Bayesian network are to be expected—the network is updated with information from the prior knowledge, shot group, and shooter evaluation measures. The relationship with the CTSR (an ability proxy) is also promising. We interpret this as the Bayesian network being moderately sensitive to the cognitive demands of learning the domain. However, the null correlations between the Bayesian network and the knowledge map score are unclear. That is, knowledge maps have been used as measures of conceptual understanding. The Bayesian network is intended to reflect the knowledge dependencies among the different concepts—presumably a conceptual structure; thus, it is unclear why there is essentially no relationship between the measures.

The significant relationships between concepts in the Bayesian network and qualification score is interesting because it suggests a link between knowledge, as measured by our assessments and modeled in the Bayesian network, and the outcome performance of interest.

How effective is individualized content delivery on learning when a Bayesian network is used to detect knowledge gaps and an ontology is used to provide relevant and detailed content?

In this section we attempt to answer this question by first examining the sensitivity of our Bayesian network to instructional effects. Our assumption is that if participants learn something from the content, they will perform well on parts of the assessments that call for the knowledge learned. Conversely, if participants did not learn a particular content, we would not expect to see any changes in performance on the assessment. Because the Bayesian network is updated directly with assessment information, we expect to observe the same properties.

Analysis of Individual-Level Effects: Comparing Bayesian Network Probabilities to Detect the Local Effects of Individualized Content Delivery

To determine how effective the targeted content delivery was, an analysis of the change in the Bayesian network probabilities was done, with respect to the pre-instruction and post-instruction administration of particular content nodes. The change in probabilities between the pretest and posttest was computed for each content node across all 16 participants. This procedure yielded a matrix of 224 cells, where rows represented participants ($n = 16$ participants) and columns represented concepts (14 concepts). Fifty-two cells were dropped because of a technical problem in the software used to compute the probabilities.

Based on the Bayesian network probabilities computed from the pretest assessments, we identified all the participant × concept combinations for which content was served (33 cells in the matrix). We also identified

concepts for which content was not served (139 cells in the matrix).

We reasoned that if our Bayesian network accurately identified knowledge gaps, and if we were successful in binding relevant content from the ontology to the Bayesian network concepts, then the content served to participants would be relevant and targeted. To the extent that the participant engaged the content, we assumed they would learn the content. Participants' learning would be reflected in their posttask performance on our assessments. Because the assessment performance information is used to update the Bayesian network, we could update the Bayesian network with the posttask assessment information and obtain a second set of probabilities that reflected participants' increases in learning. For concepts that were not served up, we did not expect any learning to occur.

To test this assumption, we conducted a paired t test between the posttask probabilities and pretask probabilities. There was a significant difference between the posttask and pretask probabilities when content was served, $t(32) = 7.36$, mean gain = .34, SE= .05. In contrast, there were no significant differences when content was not served, mean gain = .003, SE = .009, $n = 138$.

Further, it appears that participants were engaged in the task. The more concepts that were served to participants, (a) the more effort they reported putting into learning the information ($r_{sp} = .73$, $p < .01$, $n = 15$); (b) the more often participants reported attempting to learn the information ($r_{sp} = .89$, $p < .001$, $n = 15$); and (c) the more participants reported video as being useful ($r_{sp} = .53$, $p < .05$, $n = 15$). Interestingly, this relationship was not found for pictures or text.

Analysis of Group-Level Effects: Comparing Knowledge Map Scores Over Time to Evaluate the Conceptual Effects of Individualized Content Delivery

Detecting significant differences in the changes in probabilities from pre- to posttask supports the idea that our Bayesian network representation is capturing aspects of knowledge dependencies. Targeted delivery of content, based on estimates of an individual's knowledge gaps, appears to result in increases in knowledge related to the delivered content. However, there remains the question of degree of knowledge: To what extent does individualized content delivery affect increases in conceptual knowledge?

To answer this question, we examined participants' knowledge map scores over six occasions, across the experimental and control conditions. The first five mapping occasions were cumulative: Participants started with a blank map on the first occasion and modified their maps on subsequent occasions. The sixth and final mapping occasion was done with a blank map. For the purposes of this analysis, the first 5 mapping occasions are treated as repeated measures, and the final mapping occasion is treated as an independent measure.

Knowledge mapping performance was analyzed with a 2(condition) × 5(mapping occasion) ANOVA, mapping occasion (occasion 1 to 5) as the within-subjects factor and condition (individualized content delivery, control) as the between subjects factor. A significant main effect was found for mapping occasion, $F(2.1, 580.7) = 18.1$, $p < .001$. Because the interaction term did not meet the sphericity assumption, the Huynh-Feldt correction was applied. This result shows differences in map scores across occasions. Participants' map scores increased across occasions. Pairwise comparisons show a significant increase in map scores between the first and all subsequent occasions (see Table 3). In addition, a significant difference was found between map scores of the second and fourth occasions.

Table 3. Knowledge Map Scores by Occasion

Condition					

	Knowledge Mapping Occasion				
Condition	**1**	**2**	**3**	**4**	**5**
Experimental					
M	18.42	24.50	26.67	27.33	26.42
SD	10.03	11.77	11.93	11.77	12.46
Control					
M	12.26	16.43	18.26	18.65	17.91
SD	10.14	13.09	14.26	13.81	13.49

Note. Experimental, $n = 12$; Control, $n = 23$.

A main effect for condition was also found, favoring the individualized content delivery condition, $F(1, 33) = 3.46$, $p = .07$. Because of the exploratory nature of this study, we included the condition term in subsequent simple effects analyses. No interaction effects were found. Follow-up pairwise comparisons showed a significant difference between the experimental condition and the control condition at fourth and fifth occasions. The experimental condition at mapping occasion 4 had significantly higher scores than the control condition, $t(45) = 2.4$, $p < .05$. Similarly, there was a trend favoring the experimental condition at occasion 5, $t(44) = 1.82$, $p < .08$.

An independent *t* test was performed on the posttest knowledge map. This mapping activity was separate and distinct from the repeated mapping activity. Participants created a knowledge map from scratch. There was a difference that approached significance, $t(49) = 1.95$, $p < .06$. The experimental condition ($M = 26.0$, $SD = 14.0$) outperformed the control condition ($M = 18.6$, $SD = 11.8$). We interpret this result as a possible effect due to the targeted remediation.

Finally, when the posttest prior knowledge measures were compared, no significant differences were found.

DISCUSSION

In this study we tested an approach to explicitly link assessment and instruction via the use of (a) an ontology to provide the structure and content for the domain of rifle marksmanship, and (b) a Bayesian network model of the knowledge dependencies underlying the understanding of the domain. Assessments of knowledge of rifle marksmanship were administered, and participants' performance on the assessments were used to update the Bayesian network. The Bayesian network was used to estimate participants' understanding of the domain given the assessment results. Individualized content delivery was implemented by first identifying knowledge gaps (as measured by [low] probabilities in the Bayesian network), and then related content from the ontology was pulled and delivered to the participant. Each participant was provided access to an individualized set of content.

Our results are to be taken as exploratory and limited by the small sample size; however, our findings are extremely provocative given these limitations. First, our Bayesian network model appears to agree at an aggregate level with participants' perception of their level of knowledge. The overall agreement is about 80%. This finding suggests that our Bayesian model—the set of concepts and how the concepts influence each other—is doing a reasonable job of capturing the knowledge dependencies. While the model is imperfect and the results very tentative, the general approach appears promising.

Achieving agreement with participants' perception of level of knowledge is a first step in establishing the validity of the approach. However, this evidence alone is insufficient for a variety of reasons (e.g., participants may not be a good judge of what they don't know). Additional evidence that would support the general approach is seen in the impact of the individualized delivery of content on participants' learning. When individualized content is provided to participants, they appear to engage the material and learn from it, as evidenced by (a) increases in the probability estimates of their knowledge only on the very specific and relevant concepts in the Bayesian network and no increases in the probabilities for non-related nodes; and (b) higher performance than participants in a control condition on an independent measure that purports to measure knowledge at a conceptual level (i.e., a coherent network of ideas).

It is this latter finding that is the most interesting and compelling. First, there existed no differences on the knowledge map scores prior to the treatment. However, after the provision of individualized content, participants in the experimental appeared to accelerate. Further, the finding of no difference on the posttest prior knowledge test is remarkable for the following reason: the evidence used to update the Bayesian network is in large part taken from performance on the prior knowledge measure (a selected-response measure that samples surface knowledge of rifle marksmanship), yet the learning impact is reflected in participants' conceptual understanding and not at the surface level.

While it appears we have been moderately successful in identifying knowledge gaps, more direct evidence is needed (e.g., as provided by think-aloud protocols or other in-depth measurement). Such efforts will guide us on the refinement of the approach. Future work should also examine in more depth the relationship between learning due to the targeted instructional remediation and differences in the outcome (i.e., shooting) performance.

Linking assessment and instruction is the *sin qua non* of education and training. To date, attaining this linkage has been difficult, elusive, and unscalable. The approach we have explored in this paper is grounded in cognition and instruction, and demonstrates an integration of online assessments of complex learning, domain modeling that begins with cognitive demands, and data fusion methods that enable principled ways to synthesize and use assessment information.

ACKNOWLEDGEMENTS

REFERENCES

Chandrasekaran, R., Josephson, J. R., & Benjamins, V. R. (1999). What are ontologies, and why do we need them? *IEEE Intelligent Systems, 14*, 20–26.

de Clercq, P. A., Hasmon, A., Blom, J. A., & Korsten, H. H. M. (2001). The application of ontologies and problem-solving methods for the development of shareable guidelines. *Artificial Intelligence in Medicine, 22*, 1–22.

Fensel, D., Hendler, J., Lieberman, H., & Wahlster, W. (Eds.). (2003). *The semantic web: Why, what, and how*. Cambridge, MA: MIT Press.

Gennari, J., Musen, M. A., Fergerson, R. W., Grosso, W. E., Crubézy, M., Eriksson, H., Noy, N. F., & Tu., S. W. (2002). *The evolution of Protégé: An environment for knowledge-based systems development* (Stanford Medical Institute Tech. Rep. No. 2002-0943). Stanford University: Palo Alto, CA.

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies, 43*, 907–928.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.

Jensen, F. V. (2001). *Bayesian networks and decision graphs*. New York: Springer-Verlag.

Lawson, A. E. (1987). *Classroom test of scientific reasoning: Revised paper-pencil edition*. Tempe, AZ: Arizona State University.

McGuinness, D. L. (2003). Ontologies come of age. In D. Fensel, J. Hendler, H. Lieberman, & W. Wahlster (Eds.), *The semantic web: Why, what, and how* (pp. 171–191). Cambridge, MA: MIT Press.

USMC. (2001). *Rifle Marksmanship*. USMC Reference Publication (MCRP) 3–01A. Washington, DC: U.S. Marine Corps.