An Exploratory Study Examining the Feasibility of Using Bayesian Networks to Predict Circuit Analysis Understanding

Gregory K. W. K. Chung and Gary B. Dionne

CRESST/University of California, Los Angeles

William J. Kaiser

Department of Electrical Engineering/University of California, Los Angeles

Paper presented at the annual meeting of the
National Council on Measurement in Education (NCME)
San Francisco, CA, April 2006

University of California, Los Angeles (UCLA)
Graduate School of Education & Information Studies
National Center for Research on Evaluation, Standards, and Student Testing (CRESST)
10945 Le Conte 1400C
Peter V. Ueberroth Building / Mailbox 957150
Los Angeles, CA 90095-7150
310.794.4392; greg@ucla.edu

Abstract

Our research question was whether we could develop a feasible technique, using Bayesian networks, to diagnose gaps in student knowledge. Thirty-four college-age participants completed tasks designed to measure conceptual knowledge, procedural knowledge, and problem-solving skills related to circuit analysis. A Bayesian network was used to model the knowledge dependencies among the circuit analysis concepts. Preliminary results suggested that the Bayesian network was generally working as intended. When high- and low-performing groups were formed on the basis of posterior probabilities, significant group differences were found favoring the high-performing group with respect to circuit definitions and circuit analysis problems, for both actual and self-assessments, and higher major GPA. The Bayesian network was able to predict participants' performance on a problem-solving item on average 75% of the time. The findings of this study are promising for our long-term goal of developing scalable and feasible online automated reasoning techniques to diagnose student knowledge gaps.

An Exploratory Study Examining the Feasibility of Using Bayesian Networks to Predict Circuit Analysis Understanding

Renewed interest in individualizing instruction, particularly with the use of technology, has resulted in a search for methods that can accurately diagnose student knowledge gaps and prescribe appropriate remediation. While the idea of individualized instruction has its roots in the programmed instruction movement of 40 years ago, the major difference now is the availability of far more sophisticated, affordable, and accessible delivery technologies (e.g., Internet/Web, low-cost personal computers) and technologies to support knowledge representation and automated reasoning (e.g., Bayesian networks). Together, these technologies provide the mechanism to make feasible and practical individualized assessment and instruction.

In this study, we explored the feasibility of using Bayesian networks to estimate students' understanding of introductory circuit analysis topics (e.g., Kirchoff's current law). Our work focused on two major activities: (a) modeling the knowledge dependencies among the various concepts, and (b) gathering validity evidence on the quality of the model. This research directly supports our long-term goal of developing online assessment and instruction for *individual* students in distributed and distance learning settings. In such contexts, we believe automated reasoning techniques to be the only feasible method for diagnosing knowledge gaps for large numbers of students. Note that our interest is not in whether automated reasoning techniques are superior to traditional techniques (e.g., an instructor diagnosis) but rather in the extent to which automated reasoning techniques can be used to diagnose knowledge gaps.

Bayesian networks are one of the most influential tools for modeling and reasoning under uncertainty (Jensen, 2001). A Bayesian network is a graphical probabilistic model, which comprises two parts: (a) a directed acyclic graph, in which nodes represent variables of interest and edges represent direct causal dependencies, and (b) a set of conditional probability tables, which quantify the dependencies between variables. Considerable past research has focused on the use of Bayesian networks for user modeling or assessment purposes (e.g., Anderson, Corbett, Koedinger, & Pelletier, 1995; Chung, Delacruz, Dionne, & Bewley, 2003; Martin & VanLehn, 1995; Mislevy & Gitomer, 1996; Mislevy, Steinberg, Breyer, Almond, & Johnson, 2002; O'Neil, Chuang, & Chung, 2003). In this study we use Bayesian networks to represent circuit analysis knowledge and gather validity evidence.

Our general approach to validating the use of Bayesian networks for diagnostic purposes consisted of the two components, a general modeling approach, and a validation approach.

Domain Modeling

We begin by modeling the domain in terms of knowledge dependencies. In our case, the domain was understanding the set of concepts related to node-voltage and mesh-current analyses in circuit analyses at the introductory level. Given that a student knows concept *X*, what are (a) the most directly related concepts, and (b) the extent to which students are likely to know those concepts? Conversely, given that a student does not know concept *X*, how likely is it that the student will know the related concepts? A key feature of Bayesian networks is the inclusion of uncertainty via probability estimates. This is critical as learning is inherently uncertain, occurring in many different ways and under many different conditions.

Validation Approach

Our general approach to validating the method was to gather evidence on the extent to which the probabilities yielded by the Bayesian network tracked with other measures of student learning. We interpreted the probabilities associated with each concept as an indication of how likely it was that a student understood that concept. Empirical evidence was gathered by examining the correlations between probabilities and other outcome measures, and background achievement measures. We also examined differences on various measures between high and low performers, as defined by Bayesian network probabilities.

Research Question

The main research question in this study was to what extent can Bayesian networks be used to diagnosis what students know and don't know?

Method

Participants and Design

Participants. The sample comprised 34 participants (27 males, 7 females) who were ethnically diverse (21 Asian-Americans, 4 White, 1 Latino/a, 1 Biracial, 1 African

American, and 6 unspecified) and with a mean age of 20.3 years old (*SD* = 2.1 years). Table 1 shows descriptive statistics and intercorrelations for GPA and self-reported SAT score measures; Table 2 shows the distribution of the sample by academic major and standing. In general, the sample was mostly male electrical engineering juniors with B-average GPAs. With respect to grades in the particular course (Electrical Engineering 10 [EE10]) that covered the concepts used in this study, 20 participants were currently enrolled in the course, 8 received an A in EE10, 3 received a B, 1 received a D or F, and 2 participants did not take EE10. In addition, for participants currently enrolled in EE10, we tracked when the participants participated in this study with when they received instruction in EE10 and could thus code whether students received relevant instruction or not. Fifteen students were classified as having not received instruction on the relevant topics and 16 were classified as having received instruction. Thus, the sample appeared to be typical with respect to background variables, and the sample also appeared to have differential exposure to the main concepts used in this study.

Table 1
Descriptive Statistics and Intercorrelations (Spearman) for Achievement and SAT Measures ($N = 34$)

Measure	*n*	*M*	*SD*	Min.	Max.	EE GPA	SAT I Verbal Score	SAT I Math Score
Overall GPA[a]	29	3.32	0.45	1.90	3.84	.77**	.07	.52*
EE GPA[b]	30	8.64	2.04	3.75	12	—	-.12	.45*
SAT I Verbal Score[c]	25	654.4	78.2	440	770		—	.09
SAT I Math Score[c]	26	752.3	48.8	610	800			—

[a]Max. = 4.0. [b]12 point scale. [c]Max. = 800.
*$p < .05$ (two-tailed). **$p < .01$ (two-tailed).

Table 2
Distribution of Participants by Academic Standing and Major (N = 34)

Academic standing	Electrical Engineering	Mechanical Engineering	Computer Science and Engineering	Computer Science	Other
Sophomore	3	0	2	1	0
Junior	13	1	7	0	0
Senior	2	0	2	0	2
Graduate	1	0	0	0	0

Design. A single group correlational research design was used to support examination of how the various measures related to each other.

Tasks and Measures

Participants were administered a variety of tasks intended to provide the basis for measuring their understanding of circuit analysis concepts.

Background Information. Participants were asked to complete a survey asking for their age, gender, and ethnicity. In addition, self-reported SAT I Verbal and SAT I Math scores, overall GPA, class standing, and major were gathered. Participants were asked for a list of courses they were currently taking, as well as their grades in all electrical engineering courses they had taken. The EE grades were computed to a 12-point scale (given that "+" and "-" were part of the actual student grade).

Knowledge Measures. Five tasks were administered to gather information on participants' knowledge of circuit analyses: conceptual definitions of circuit concepts, procedural understanding of node-mesh and node-voltage analyses, conceptual understanding of circuit analysis concepts, circuit problem solving, and self-ratings of understanding.

<u>Conceptual definitions of circuit analysis concepts</u>. Participants were asked to define, in a few sentences, 19 concepts related to circuit analysis. Participants were asked to state the concept in mathematical or descriptive form. The topics participants were asked to define were: *combination of sources, constraint equation, current division, dependent source, equivalent resistance, essential node, KCL, KVL, mesh, mesh current, node, node voltage, Ohm's law, parallel resistance, series resistance, sign convention, super mesh, super node,* and *voltage division*. This task was intended to provide information on participants' general familiarity of a concept.

Participants' written responses to each concept measure were scored dichotomously and a summary measure was computed as the sum of correct responses. A second measure was obtained, based on participants' self-assessment. Participants were asked to score themselves, on a scale of 0 to 10, on how well they understood each topic. The specific score ranges were:

0	=	no understanding of the topic at all, could use a lot of help with the content
1 – 5	=	some understanding, could use some help with the content
6 – 9	=	more understanding than not, may need some help with the content but probably not
10	=	complete understanding of the topic, don't need any help with the content

Circuit explanation essay. Participants were asked to write an essay to explain the importance of KCL, KVL, and Ohm's Law with respect to circuit analysis. The instructions emphasized that the goal of the essay was to convey their conceptual understanding, and one way to help them write the essay was to imagine that they were asked to give a guest lecture to new EE students about why these topics (KCL, KVL, Ohm's Law) are fundamental and important to circuit analysis. Some guiding questions were provided with the intention of helping participants start the essay and frame their response (*Why is it important to know these laws? What do these laws let you do? What role do KCL, KVL, and Ohm's Law play in electrical engineering in general – not only in solving simple circuits, but in the larger circuit analyses picture? Are KCL/KVL/Ohm's Law simply a set of mathematical relationships used to figure out voltage current and resistance, or is there something more fundamental to these laws?*). The specific prompt for this task was:

> Please explain the importance of KCL, KVL, and Ohm's Law. Why are these topics so fundamental and important to circuit analysis? Remember, the essay should convey your conceptual understanding.

The task was intended to provide information on participants' deep understanding of the interrelationships among the fundamental concepts of Ohm's Law, KVL, and KCL. Participants' responses were scored on a 5-point scale using the rubric shown in Table 3.

Table 3
Explanation Essay Rubric

Score	Scoring guidelines
1	No indication that the response shows understanding of any of the three concepts (Ohm's Law, KCL, KVL).
2	Basic understanding of one or more of the concepts. Includes only a description or a statement of the equations behind these concepts. No elaboration or insight into the concepts (i.e., why the concepts are important).
3	All concepts are mentioned, and at least one is explained on a more conceptual level (i.e., goes beyond stating the definition of the concept). Descriptions and definitions are partially correct. Minimal elaboration.
4	Two concepts discussed thoroughly or one in detail. Shows some principled understanding. Processes are elaborated. Response contains only a few minor misconceptions.
5	Complete response. All three concepts are discussed and elaborated. High level of detail. High level of discussion of concepts. Descriptions and definitions are accurate.

<u>Knowledge mapping of circuit problem-solving procedures</u>. A knowledge mapping task was administered which required participants to diagram the procedure to solve circuit problems using the node-voltage method and mesh-current analysis techniques. Seventeen steps (*choose solution approach (1), conduct mesh current analysis (2), conduct node voltage analysis (3), determine what is being asked (4), draw and label mesh currents (5), identify dependent sources (current) (6), identify dependent sources (voltage) (7), identify essential branch(es) (8), identify essential node(s) (9), label node voltages (10), list given information (11), report results (12), simplify circuit (13), solve equation (14), use KCL to write equations (15), use KVL and Ohm's law to write source equation (16), write constraint equations (17)*) and two links (*if mesh current analysis, if node voltage analysis*) were provided.

Scoring of participants' knowledge maps was done by counting the number of propositions (node-link-node) in the participant's map that were also in the criterion map (Herl, Baker, & Niemi, 1996). This measure was intended to provide information on participants' knowledge of the problem-solving procedure for node-voltage and mesh-current analyses. The criterion map was based on a written problem-solving procedure developed by the class instructor (Appendix A) and is shown in

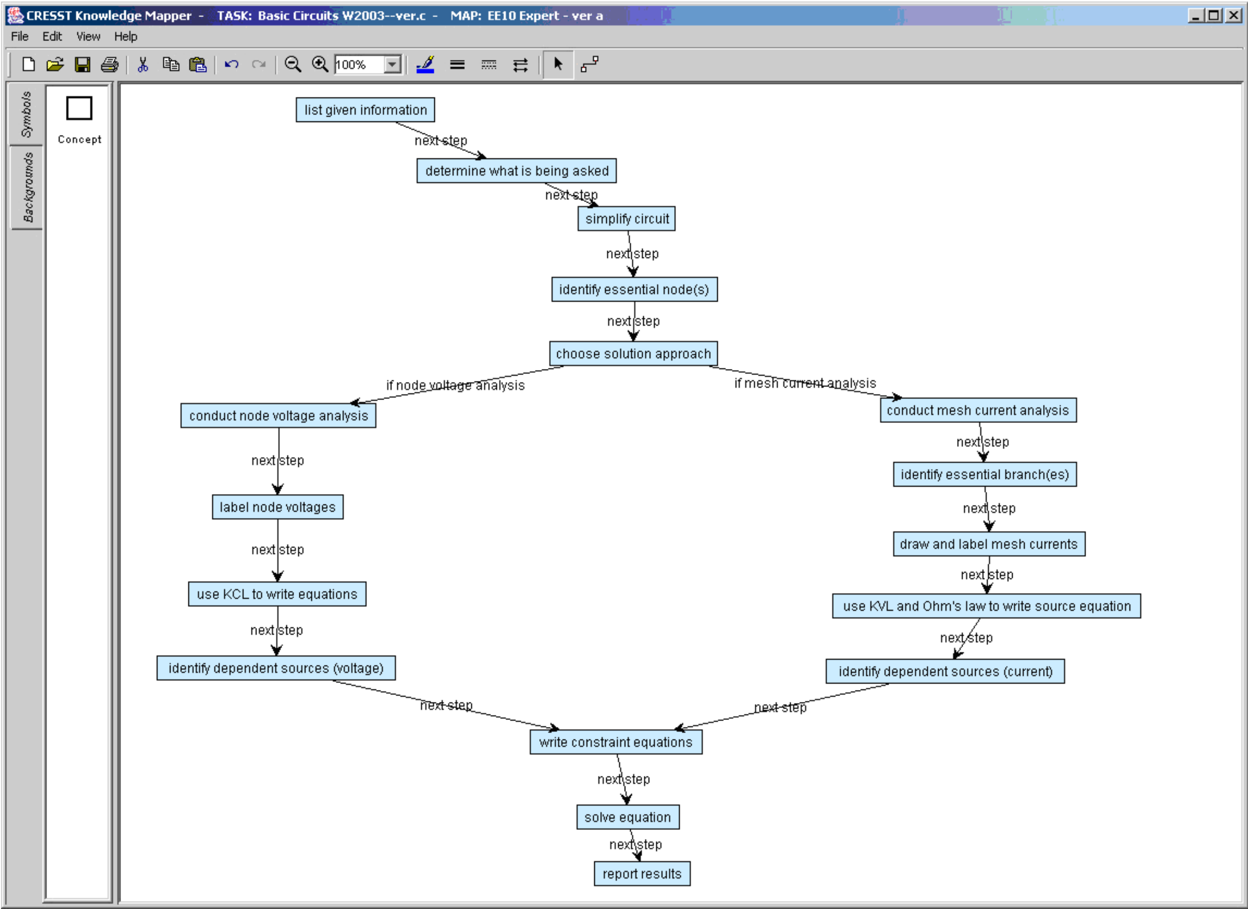

CRESST Knowledge Mapper - TASK: Basic Circuits W2003--ver.c - MAP: EE10 Expert - ver a
File Edit View Help
100%
Symbols
Backgrounds
Concept
list given information
next step
determine what is being asked
next step
simplify circuit
next step
identify essential node(s)
next step
choose solution approach
if node voltage analysis
if mesh current analysis
conduct node voltage analysis
conduct mesh current analysis
next step
label node voltages
next step
use KCL to write equations
next step
identify dependent sources (voltage)
identify essential branch(es)
draw and label mesh currents
use KVL and Ohm's law to write source equation
identify dependent sources (current)
write constraint equations
solve equation
report results

Figure 1.

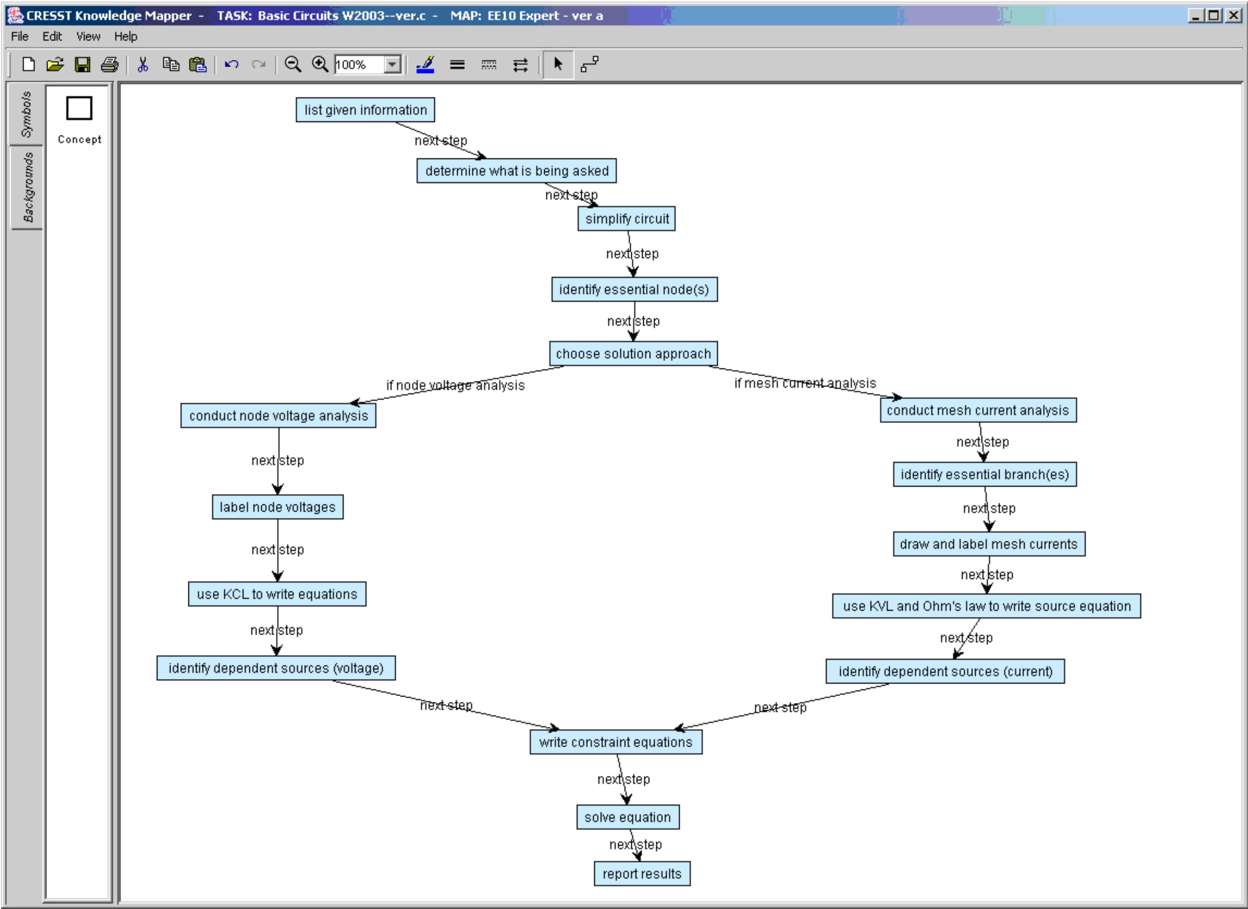
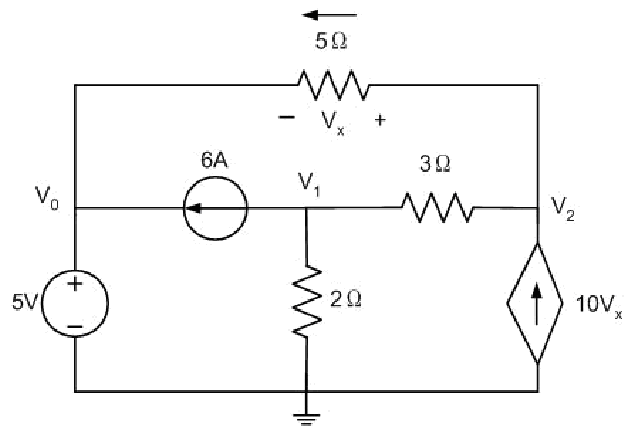


Figure 1. Criterion knowledge map for node-voltage and mesh-current analyses.

<u>Circuit problem solving.</u> Eight circuit problems were administered to participants. The problems were designed to require a range of concepts. Each problem required various circuit analysis concepts and was decomposed into subparts such that each subpart mapped to one concept. Thus, while each problem appeared as a task, the subparts actually mapped to different concepts. In this way we were able to sample different concepts across individual problems.

Participants were instructed to solve each subpart, explain each subpart in writing, and score themselves using the same 0-10 scale as in the definitions task on their understanding of that subpart. In addition, participants were instructed to score themselves holistically on their overall understanding of the entire problem. Figure 2 shows a sample problem-solving task. Each numbered question is a subpart.



Question	Answer	Score (0 – 10)
1. How many nodes are there?		
2. How many **essential** nodes are there?		
3. How many meshes are there?		

Problem	Explanation	Score (0–10)
4. Using KCL, solve for V_1 [in terms of V_o, V_2, or Vx]		
5. Using KCL, solve for V_2 [in terms of V_o, V_1, or Vx]		
6. As shown, $V_o = 5V$. What other constraint equation(s) are needed to fully solve the previous two equations?		

Figure 2. Sample circuit problem-solving task. Layout of figure and prompts was compressed for display purposes.

Table 4
Rubric for Assigning Probability Values to the Conditional Probability Table

Parent node state: *Understands*			Parent node state: *Does not understand*		
State			State		
R	W	Explanation	R	W	Explanation
1.00	.00	Absolute correlation,; if you understand the concept you **must** get the question right. People will occasionally make mistakes even with a perfect understanding of the topic. An exception is a "true/false" question that is directly questioning the targeted concept with no computation.	.00	1.00	It is impossible to get this question right if you do not understand the targeted concept. There is no chance of guessing the question correctly, so there must be significant, necessary computation involved. An undesirable question, since the extra computation could be responsible for the incorrect answer. Using the targeted concept must be the **only** way to arrive at the correct answer.
.95	.05	Essentially a perfect question. If you understand the concept you will get the question right; this probability allows for "stupid" mistakes.	.05	.95	It is unlikely that the correct answer could be guessed, so there must be more than a few possible answers. Using the targeted concept must be the **only** way to arrive at the correct answer. Incorrect ways of attempting the problem must not coincidentally result in the correct answer.
.85	.15	This is a good question. It is a question designed to test only the targeted concept so there is very little chance of a computational error. No necessary computation. Very few false negatives.	.15	.85	This is a question that may have only a handful of possible answers, such as a "How many?" question. Or, it is a question with no apparent ways to solve the problem that do not directly test the targeted concept.
.70	.30	This question tests the desired concept but there may be a less significant influence from another concept that may cause an incorrect answer. Also, there may be minor, necessary computations.	.30	.70	This question has just a few plausible options for answers. Guessing the right answer is possible, but not likely. There shouldn't be any easy or obvious way to get the answer without using the targeted concept.
.60	.40	This is a fairly poor question. The targeted concept is still being tested, but there are many other concepts that may influence the answer of the question or there are significant necessary computational steps that may result in an incorrect answer.	.40	.60	A question whose answer is easily guessed. An example is a "true/false" question. Or a question that can be easily solved using knowledge of a non-targeted concept.
			.45	.55	A poorly designed question. Attempting to solve the problem with a commonly held, incorrect assumption coincidentally yields the correct answer.
.50	.50	A meaningless question. There is no correlation between knowing the concept and answering the question correctly.	.50	.50	A meaningless question. There is no correlation between knowing the concept and answering the question correctly.

R = right or correct. W = wrong or incorrect.

Bayesian Network-Based Measures. The basic measure from the Bayesian network (BN) used was the probability values associated with each hypothesis (i.e., unobserved) variable. The graphical model is shown in the Appendix. The construction of the network was based on the following:

1. The hypothesis nodes were defined as "understanding *C*," where *C* was a circuit analysis concept.
2. The conditional probability table was based on a node being in two states: *understands* or *does not understand*. The guidelines used for assigning specific a priori probability values are given in Table 4.
3. The BN was constructed using a causal framework. Understanding a concept *C* directly influences understanding C_{child}, where C_{child} is a descendent of *C*. For an observable node *O* connected to *C*, *C* directly influences whether the participant answers *O* correctly (or incorrectly).
4. An observable node was an item, scored correct or incorrect, from a subpart in the circuit problem-solving questions. Each subpart targeted a specific concept.

Concept understanding. The probability for each hypothesis node was interpreted as the probability that the participant understood circuit analysis concept *C*. The probability was treated as a score and was used as a measure of understanding of concept *C*. The higher the probability for *C*, the higher the understanding of *C*.

Decision. Because of how we interpreted the probability values in each node, and how we intended to use the BN in a practical setting (to detect and administer feedback), we derived from the probabilities an understand/does not understand dichotomous measure. In a practical setting this measure would be used to detect understanding (or not understanding) and serve as the basis for remediation; thus, it was important to examine the characteristics of this measure.

Manipulation check measures. Self-reports were gathered about participants' perception of the knowledge mapping, definitions, explanation essay, and circuit problem-solving tasks. For each task, participants were asked three questions: (a) How difficult they found the task (1 = *not difficult*, 2 = *somewhat difficult*, 3 = *moderately difficult*, 4 = *very difficult*); (b) In general, how well they thought their conceptual understanding of circuit analyses was reflected by the tasks (1 = *not well at all*, 2 = *somewhat well*, 3 = *moderately well*, 4 = *very well*); and (c) In general, how much effort (i.e., trying to really do their best to answer the question) did they put into the task (1 = *not much at all*, 2 = *some amount*, 3 = *moderate amount*, 4 = *a lot*).

Reliability of measures. Cronbach's alpha was computed for each measure. Reliability of the measures was generally in the .80s and .90s. Three scales in the problem-solving scales were low or contained two items (combination of sources, parallel resistance, current division, Ohm's law, constraint equations, and KVL). These scales were dropped from subsequent analyses at the scale level; however, the items were retained in total score analyses. Interestingly, the self-ratings of understandings were uniformly high, with α ranging from .82 to .97. Table 5 and Table 6 show the reliabilities for the measures used.

Table 5
Reliability of Measures

Measure	*n*	No. of items	Score	*n*	No. of items	Self-rating
Circuit definitions	24	19	.85[a]	19	24	.92[b]
Overall problem solving						
Total score	24	9	.81[c]	19	9	.92[d]
Total holistic self-rating			--	19	9	.93[d]

[a]no. of items = 19 and n = 24. [b]no. of items = 24 and n = 19. [c]no. of items = 9 and n = 24. [d]no. of items = 9 and n = 19.

Table 6
Alpha Coefficient for Circuit Problem Solving Topic Scales, Scores and Self-rating

Topics	*n*	No. of items	Score	Self-rating
Combination of sources	24	5	.41	.82
Parallel resistance	34	2	.13	.87
Voltage division	33	3	.75	.88
Current division	33	2	.76	.85
Node	27	--	.78[a]	.92[b]
Essential nodes	34	3	.79	.95
KCL	33	4	.62	.92
Ohm's law	34	2	.42	.97
Constraint equations	33	2	.61	.90
Mesh	34	4	.91	.97
KVL	33	2	.57	.85

[a]no. of items = 7. [b]no. of items = 5.

Procedure

Participants were recruited via class announcements in introductory circuit classes. Participants who participated in prior research were also recruited via email. Participants were administered the tasks individually or in small groups. The set of tasks, order, and allotted time is shown in Table 7. Pilot testing of the measures provided the basis for the time; however, participants were allowed to complete the task at their own pace. In general, participants finished all the tasks within the allotted time.

Table 7
Administration Schedule

Task	Time allotted
Introduction to study	5
Knowledge mapping	25
Conceptual definitions	30
Circuit analysis essay	20
Background survey	10
Circuit problem solving	60

Results

Manipulation Check

Prior to conducting the analyses, participants' responses to manipulation check questions with respect to the knowledge map, the definitions, essay, and circuit problems were examined. Table 8 shows self-reported perception of task difficulty, utility, and effort. In general, participants perceived all tasks as having some difficulty, with the circuit problems rated as being the most difficult. With respect to how well participants perceived the measures as reflecting their conceptual understanding of circuit analysis, participants reported that they perceived the measures as moderately capturing their conceptual understanding of circuit analyses with the circuit problem-solving questions being rated the highest. Finally, most participants reported a moderate amount of effort at completing the tasks.

Table 8
Descriptive Statistics for Participant Self-Reports of Task Difficulty, Utility, and Effort (n = 33)

Measure	*M*	*SD*
Difficulty of task[a]		
Knowledge mapping	1.36	0.49
Definitions	1.79	0.65
Essay	2.33	0.99
Circuit problems	2.25	0.95
How well task captured conceptual understanding[b]		
Knowledge mapping	2.64	0.82
Definitions	2.61	0.70
Essay	2.36	1.17
Circuit problems	3.03	0.88
How much effort put into task[c]		
Knowledge mapping	2.94	0.83
Definitions	3.15	0.67
Essay	2.70	0.77
Circuit problems	3.30	0.88

[a]1 = *not difficult*, 2 = *somewhat difficult*, 3 = *moderately difficult*, 4 = *very difficult*. [b]1 = *not well at all*, 2 = *somewhat well*, 3 = *moderately well*, 4 = *very well*. [c]1 = *not much at all*, 2 = *some amount*, 3 = *moderate amount*, 4 = *a lot*.

For the knowledge mapping task, a significant correlation was found between participants' ratings of how well the task represented their conceptual understanding and their knowledge map score ($r = .46$, $p = .001$). The higher participants reported that their conceptual understanding was reflected by their knowledge map, the higher their knowledge map score. For the essay task, significant correlation was found between participants rating of how much effort they point into the essay and their essay score ($r = .35$, $p = .05$). The more effort participants reported, the higher their essay scores.

For the circuit problems, perceived difficulty of the circuit problems was negatively related to their problem-solving score across all problems ($r = -.57$, $p = .001$), and negatively related to their overall self-rating of each circuit problem ($r = -.38$, $p < .06$). In general, the more difficult the task was perceived, the lower the scores and self-ratings. In addition to perceived difficulty, effort was positively associated with problem-solving scores ($r = .35$, $p < .06$). As participants put more effort into solving the problems, the higher their scores.

These data are consistent with the idea that participants were taking the tasks seriously and expended reasonable effort. The pattern of correlations, particularly between performance and perceived difficulty and perceived utility, is consistent with prior research (e.g., Chung & Baker, 2003; Chung et al., 2003). Thus, we concluded that the data were suitable for subsequent analyses.

Validity Analyses

Validity evidence was gathered by examining three questions:

- What was the relation among the measures of knowledge? We expected to observe positive correlations among measures of circuit knowledge and participants' self-reports.
- To what extent is the model capturing systematic differences in student responses? We addressed this question by examining (a) the Bayesian network probabilities when random student performance data were entered, and (b) examining the relation between the Bayesian network probabilities and other measures (i.e., performance on the other knowledge measures, self-ratings, and EE GPA). We expected to observe null correlations when random student data were entered, and we expected to observe positive correlations between the Bayesian network probabilities and the other knowledge measures.
- How accurate is the Bayesian network? One indicator of the quality of the model is the degree to which it can predict student performance. We used a variation of a "leave-one-out" analysis to examine this issue.

What was the relation among the measures of knowledge? Overall, the results shown in Table 9 are consistent with the idea that the content and instruction are highly problem-focused. That is, one of the main desired instructional outcomes of a course at this level (EE10) is being able to solve a variety of different circuit analysis problems. Informal discussions with the instructor, the personal experience of all authors (all have electrical engineering backgrounds), a review of the instructor's lecture notes, the textbook, and observation of discussion sections and lectures support this idea.

With this context in mind, it is unsurprising that the EE GPA related most strongly to the circuit problem-solving measures. Much of the coursework in electrical engineering at the undergraduate level is developing competency in solving problems, and is reflected in the correlation with performance on the circuit problem-solving task (r_{sp} = .65, p < .01). Similarly, the correlation between conceptual knowledge of circuit concepts and success in solving problems is unsurprising (r_{sp} = .45, $p < .05$) as basic knowledge of the concepts was required to solve the circuit problems.

Participants' self-ratings of their performance were significantly related to their actual performance, for both the circuit definition (r_{sp} = .45, $p < .05$) and circuit problem-solving (r_{sp} = .77, $p < .01$) tasks. This is an important result for two reasons. First, the positive correlations suggest that participants were capable of evaluating their understanding of their own responses, and especially so when they were solving problems. That is, in general, participants knew when they got something right and when they didn't. Our speculation as to why the self-reports were so accurate for the problem-solving task lay in how the task was structured: Participants were asked to judge their understanding immediately after attempting to solve a subpart of the problem (see Figure 2), resulting in a response that was highly contextualized and specific.

The second reason the results are important is related to our long-term goal of developing automated methods for diagnosis of knowledge gaps (and remediation) in distance learning contexts. The use of self-ratings may provide (a) a short-term solution to the automated scoring of student responses by serving as proxy scores, as long as the actual response is still required; and (b) a quick measure of participants' metacognitive skills.

<u>Unexpected results.</u> The emphasis on solving problems may also explain the results related to the essay measure. In this case, it may have been too much to ask of participants who may have been too inexperienced in general to have developed much insight into the significance of the concepts the task was targeting. We also expected procedural knowledge to relate to circuit problem-solving performance, particularly because the knowledge mapping task asked participants to explicitly lay out their circuit problem-solving steps.

Table 9
Descriptive Statistics and Non-parametric (Spearman) Intercorrelations Among Circuit Knowledge Measures

						Conceptual knowledge				Circuit problem solving	
							Circuit definitions				
Measures	*n*	*M*	*SD*	Min.	Max.	Essay	Total score	Total self-rating	Procedural knowledge	Total score	Total self-rating
Mean EE GPA	30	8.64	2.04	3.75	12	.06	.45*	.29	-.01	.65**	.63**
Conceptual knowledge											
Essay	33	3.09	0.98	1	5	--	.20	.55*	.12	.15	.27
Definitions—total score	25	14.68	3.93	0	19		--	.53*	.21	.62*	.72**
Definitions—total self-rating	30	186.80	34.85	96	239			--	.31	.52*	.71**
Procedural knowledge	34	4.79	2.11	1	8				--	.28	.40*
Circuit problem solving											
Total score	33	13.27	5.66	1	21					--	.77**
Total self-rating	26	166.73	52.84	5	218						--

*p < .05 (two-tailed). **p < .01 (two-tailed).

To what extent is the model capturing systematic differences in student responses? We conducted three analyses to examine how well our Bayesian network was capturing systematic differences among participants' knowledge. Our first analysis was essentially a verification that the Bayesian network did not contain any unusual dependencies. We replaced each participant's actual responses (i.e., the correct/incorrect value that served as inputs to the Bayesian network) with randomly generated correct/incorrect responses. Our assumption was that with a random set of responses the Bayesian network should yield probabilities that show no relationship with the other outcome variables. Results confirmed this assumption. When a set of analyses (parallel to all analyses in this report) were conducted using the randomized data, there were no statistically significant relationships or differences on any group comparisons. This is an important piece of validity evidence as it verifies that the "machinery" was working as intended.

The second analysis was also a check to verify that our Bayesian network was computing probabilities consistent with our interpretation of how we viewed the dependencies among the concepts. This analysis examined the relation between probabilities yielded by the Bayesian network and scores on our circuit problem-solving measures. As expected, there was a positive and statistically significant relation between the Bayesian network concepts and corresponding circuit problem solving measures as shown in Table 10. A more interesting result was the significant relationships between the Bayesian network scales with the corresponding self-ratings of understanding. In this case, most of the correlations were significant but of moderate magnitude, r_{sp} = .40 – .83, compared to the aggregate self-rating scores (see Table 9).

Table 10
Non-parametric (Spearman) Correlations Between Bayesian Node Posterior Probabilities and Corresponding Circuit Problem-Solving Measures

		Circuit problem solving	
Concept	n	Scale score	Scale self-rating
Combination of sources	24	--[a]	.52**
Voltage division	33	.95**	.83**
Node	27	.99**	--
Essential nodes	34	.87**	.51**
KCL	33	.97**	.65**
Mesh	34	.68**	.40*

[a]Dropped due to low reliability ($\alpha < .50$).

The third analysis examined the condition closer to our expected operational of Bayesian network to infer degree of learner understanding. We examined whether differences existed on various background measures when participants were classified into high- and low-performing groups, based on the posterior probabilities for each concept in the Bayesian network. The classification was based on the top and bottom thirds of the sample, when sorted by the total number of concepts in the Bayesian network that had posterior probabilities of understanding greater than .50. We reasoned that if the Bayesian network was detecting systematic differences in participants' level of understanding, then performance on measures of knowledge should favor participants in the high-performing group. A similar result should be observed on background measures bearing on circuit analysis knowledge.

Results confirmed significant differences between high and low groups in favor of the high-performing group. As shown in Table 11, high performers scored significantly higher on the definitions and circuit analysis problems, for both actual and self-assessments, and had higher grades in EE courses. We interpret these results as evidence that our Bayesian network was sensitive in detecting overall differences in knowledge. (Note that the difference in the circuit problem solving total score measure is expected, as these scores served as inputs to the Bayesian network.)

Table 11
Non-parametric (Mann-Whitney) Test of Group Differences Between Bayesian Network Inferred High and Low Groups

Measure	Bayesian Network Inferred Groups: Low *n*	*M*	*SD*	High *n*	*M*	*SD*	Mann-Whitney test of group difference: *U*	*p* value	r_{pb}
Mean grade in EE courses	9	6.84	1.88	11	9.76	1.23	7.50	<.00	.70**
Conceptual knowledge									
Essay	11	2.91	1.14	13	3.15	.99	62.50	.59	.12
Definitions—total score	7	12.57	2.15	9	15.33	5.85	11.00	.03	.30
Definitions—total self-rating	9	161.33	47.50	12	200.92	24.92	27.00	.05	.49*
Procedural knowledge	11	4.27	2.00	13	5.31	2.25	52.00	.25	.24
Circuit problem solving									
Total score	11	6.55	3.01	13	18.54	1.81	.00	<.00	.93**
Total self-rating	7	111.14	65.64	12	189.17	31.76	11.50	.01	.65**

*p < .05 (two-tailed). **p < .01 (two-tailed).

How accurate is the Bayesian network in predicting performance? The last analysis examined the accuracy of the Bayesian network with respect to predicting performance using a "leave-one-out" analysis. Given all data less one response (the item left out), how accurate is the Bayesian network's prediction of the left-out response? This question was asked for all 39 items across all circuit problem-solving items. Accuracy was computed as the percent of correct predictions. For example, 100% accuracy would indicate the Bayesian network correctly predicted a participant's performance (correct or incorrect) on all 39 items. The mean percent correct, across all participants, was 79% (*SD* = 12%). The lowest accuracy rate was 43%, and the highest accuracy was 90%.

When accuracy of prediction was examined by item (i.e., on a given item, for how many participants did the Bayesian network accurately predict performance?), results were similar. The mean was 72% (*SD* = 14%), the lowest accuracy was 36%, and the highest 94%. We interpreted these results as additional evidence that our Bayesian network was measuring systematic understanding. In addition, these results provide evidence that the network was reasonably robust against missing information.

Table 12
Percent of Responses Accurately Predicted by the Bayesian Network (correct/incorrect)

	n	*M*	*SD*	Min.	Max.
Across participants	33	72	12	43	90
Across circuit problem-solving items	39	72	14	36	94

Summary and Discussion

The focus of this work has been on gathering validity evidence to help us understand how well a particular automated reasoning technique, in our case Bayesian networks, could model the knowledge dependencies associated with node-voltage and mesh-current analyses. Said another way, to what extent can Bayesian networks be used to infer students' understanding of various circuit concepts given their performance on various circuit problem-solving tasks?

We found preliminary evidence in support of our general approach. Results suggested that the Bayesian network was working as intended. Interestingly, the posterior probabilities correlated significantly with participants' self-ratings of understanding (r = .40 – .83). Further, when the posterior probabilities were used as the basis for forming high- and low-performing groups, significant differences were found favoring the high-performing group with respect to circuit definitions and circuit analysis problems, for both actual and self-assessments, and the high-performing group also had higher grades in EE courses. Additional evidence of the quality of the Bayesian network was found in the accuracy of the prediction of participants' performance. The Bayesian network was successful on average 75% of the time in predicting whether a participant was going to get the item correct. However, we did not find evidence of an association between either conceptual knowledge (essay) or procedural knowledge and circuit problem-solving performance. This was surprising as we expected higher performance on the circuit problem solving to be linked to deeper understanding of the content or with more knowledge of how to solve the problems.

The results of this study, while exploratory, are provocative for they suggest that probabilistic reasoning approaches can be fruitful in diagnosing students' understanding of different circuit analysis concepts at the group level. We tested a method that is feasible: (a) simple scoring (dichotomous scoring of problem solving steps) and thus a candidate for online scoring, and (b) modeling of the knowledge dependencies via Bayesian networks.

Future Directions

A clear next step is to gather additional validity evidence that would shed light on the accuracy of the network at the individual concept level in the Bayesian network. In the current study, we aggregated all concepts in the Bayesian network. Whether the probabilities at the individual concept level are in fact accurate in their diagnosis remains unknown. Thus, while probabilities of whether someone understands a concept are easy to compute given performance data, one is left wondering about the accuracy of the probabilities computed for concepts where there is no direct evidence (i.e., actual observable performance). This is clearly an important issue for individualized diagnosis, remediation, and instruction, where inferences are likely to be made not only about concepts directly tested (with the assessment items), but also about antecedent concepts.

A second area that deserves attention is on examining the sensitivity of the a priori probabilities specified in the conditional probability tables. In this study we used a general rubric to guide specification of the probabilities. Further work is needed to investigate the robustness of the network against swings in the probability specifications. This issue becomes important for practical reasons. If such methods are to scale well, there needs to be a simple and feasible method to elicit the probabilities from subject-matter experts (e.g., deriving numerical values from qualitative statements [Druzdzel & van der Gaag, 2000; Renooij & Witteman, 1999]).

Implications for Assessment and Instruction

Over two generations have passed since the ideas of programmed, adaptive, and individualized instruction were introduced by Lumsdaine and Glaser (1960). Current technologies, particularly advances in delivery system (e.g., distributed learning technologies) and automated reasoning capabilities (e.g., Bayesian networks), make feasible and seamless many of the techniques and ideas that by today's standards seem cumbersome and impractical. Having the capability to model and quantify the knowledge dependencies of a domain is a necessary step in establishing a credible link between assessment and instruction. Assessments provide data on a sampling of to-be-learned content, the domain model articulates the dependencies among and between the antecedent concepts and the to-be-learned content, assessment data fusion occurs via probabilistic statements about performance on the assessments and whether concepts are learned, and inferences are drawn about what a student knows about various concepts. Echoing the optimisms of Lumsdaine and Glaser (1960), we are well on the way to developing the means for detecting gaps in learning and understanding, and the methods for implementing individualized and dynamic instruction and assessment.

While this work has been limited to group-level comparisons, our long-term goal is to develop interactive, individualized assessment and instruction to support student learning in distance and distributed learning settings. In such settings individualized diagnosis and prescription become a clinical judgment about knowledge of particular concepts, and much more work is needed to develop methods that are accurate on an individual basis. The findings of this exploratory study suggest that automated reasoning offers real opportunities to meet this challenge.

References

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4*, 167-207.

Chung, G. K. W. K., & Baker, E. L. (2003). An exploratory study to examine the feasibility of measuring problem-solving processes using a click-through interface. *Journal of Technology, Learning, and Assessment, 2*(2). Available from http://jtla.org

Chung, G. K. W. K., Delacruz, G. C., Dionne, G. B., & Bewley, W. L. (2003). Linking assessment and instruction using ontologies. *Proceedings of the I/ITSEC*, *25*, 1811–1822.

Druzdzel, M. J., & van der Gaag, L. C. (2000). Building probabilistic networks: `Where do the numbers come from?' Guest editors' introduction. *IEEE Transactions on Knowledge and Data Engineering*, *12*, 481–486.

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.

Kaiser, W. J. (2003). Node-Voltage Analysis Problem-solving procedure. EE10 lecture notes, Winter 2003, Lecture 5. Los Angeles, CA: UCLA.

Lumsdaine, A. A., & Glaser, R. (Eds.) (1960). *Teaching machines and programmed learning: A source book.* Washington, D.C.: National Education Association.

Martin, J., & VanLehn, K. (1995). Student assessment using Bayesian nets. *International Journal of Human-Computer Studies, 42*, 575–591.

Mislevy, R. J., & Gitomer, D. H. (1996). The role of probability-based inference in an intelligent tutoring system. *User Modeling and User-Adapted Interaction, 5*, 253–282.

Mislevy, R. J., Steinberg, L. S., Breyer, F. J., Almond, R. G., & Johnson, L. (2002). Making sense of data from complex assessment. *Applied Measurement in Education, 15*, 363–389.

O'Neil, H. F., Chuang, S., & Chung, G. K. W. K. (2003). Issues in the computer-based assessment of collaborative problem solving. *Assessment in Education, 10*, 361-373.

Renooij, S., & Witteman, C. (1999). Talking probabilities: Communicating probabilistic information with words and numbers. *International Journal of Approximate Reasoning, 22*, 169–194.

Appendix A

Node-Voltage Analysis Problem-Solving Procedure (Kaiser, 2003)

Procedure
1. List the information that is requested by the problem.
2. Examine the circuit to determine the approach for a solution.
3. Determine if a circuit simplification may be accomplished using an equivalent circuit, for example, a parallel, series, delta, or Y, circuit structure.
4. List known values of circuit variables.
5. Identify and label the N_E Essential Nodes.
6. Choose one Essential Node and label it with a Reference Potential Symbol. The choice of this node will determine the level of simplicity of the calculation. However, *any* choice of an Essential Node will yield the same problem results. You should select in the circuit, that Essential Node that is connected to the most branches.
7. Identify and label the non-reference node voltages.
8. Each non-reference node voltage is labeled as positive.
9. Use KCL to write down an equation for each non-reference node, writing the equations in terms of the resistances, and node voltages.
10. Write down constraint equations associated with dependent sources
11. If a voltage source exists between two essential nodes (and is not in series with any other elements), then introduce a current variable associated with this source that will appear in node voltage equations at each node at each terminal of this source.
12. Write down N_E – 1 equations.
13. Solve the set of equations.

Appendix B

Bayesian Network