Paper presented at the Learning Progressions in Science (LeaPS) Conference, June 2009, Iowa City, IA

A BAYES NET APPROACH TO MODELING LEARNING PROGRESSIONS AND TASK PERFORMANCES

A major issue in the study of learning progressions is the linking of student performance on assessment tasks to the progressions. This paper describes the development and use of learning progressions (LPs) in the field of computer networking, centering on Bayes nets built around LPs. We discuss how Bayes nets can be built to 1) accommodate responses to tasks keyed to targeted levels of learning progressions and 2) define ordered student-model variables to characterize students' standing on LPs. The ideas are illustrated with exemplar Bayes nets built on Networking Academy LPs, and tasks designed to obtain evidence in their terms. Extensions to observations from complex tasks keyed to multiple learning progressions are described.

Patti West, Cisco
Daisy Wise Rutstein, University of Maryland
Robert J. Mislevy, University of Maryland
Junhui Liu, University of Maryland
Roy Levy, Arizona State University
Kristen E. DiCerbo, Cisco Learning Institute
Aaron Crawford, Arizona State University
Younyoung Choi, University of Maryland
John T. Behrens, Cisco

Introduction

Linking student performance on assessment tasks to learning progressions is essential in order to both validate the learning progressions and to provide information to students, teachers, and curriculum designers. In order to facilitate the desired inferences based on an assessment system, the statistical or psychometric model should be aligned with a substantive theory regarding cognition and the development of expertise in the learning progression (Borsboom, 2006; Mislevy, Steinberg, & Almond, 2003). Bayesian networks (Jensen, 1996; Pearl, 1988) represent a flexible approach to latent variable modeling of familiar and complex assessments (Almond, DiBello, Moulder, & Zapata-Rivera, 2007; R. Levy & Mislevy, 2004) and as such offer an appropriate approach to learning progressions of students within a given curriculum.

This paper details the construction, calibration, and application of Bayesian network models of assessments as they relate to learning progressions. We will provide a brief background on the development of the learning progressions in the curriculum and task design that enables this analysis. We will then discuss the results of modeling performance with an illustrative example on more traditional multiple choice assessments. We conclude with a brief comment regarding future directions of the work and nascent work on modeling performance on simulation-based assessments in the same curriculum.

The work takes place in the context of the Cisco Networking Academy, examining components of the four semester Cisco Certified Network Associate (CCNA) course sequence. The Cisco Networking Academy is a global program in which information technology is taught via a blended program with face-to-face classroom instruction, an online curriculum, and online assessments. Courses are delivered at high schools, 2-3 year community college and technical schools, and 4 year colleges and universities. Since 1997, the Networking Academy has grown to reach a diverse population of approximately 600,000 students each year in more than 160 countries. Murnane, Sharkey and F. Levy (2002) discuss the motivations and origins of the program, while F. Levy and Murnane (2004) describe issues related to technological application of the curriculum and assessment. Behrens, Collison, and DeMark (2005) discuss the assessment framework that drives the ongoing assessment activity in the program and provides the data for this work.

Identification of Learning Progressions

In 2007, the Cisco Networking Academy updated and redesigned the curriculum for their primary network course offerings. Previously, the Academy offered a single four course series that focused on specific individual networking technologies, each course focused on a specific technology: Physical networking & protocols, Routing, LAN switching, and Wide-area networking (WAN). Taken as a whole, the curriculum prepared students for entry level networking jobs and CCNA certification. As part of the redesign, two separate curriculum strategies were adopted. One strategy (used to create the Discovery course sequence) evolved from a whole task practice (van Merriënboer, 1997) design in which students were presented with the opportunity to build functional networks of increasingly larger and more complex designs as they progressed through each of the four courses. The other strategy (used to create the Exploration course sequence) updated the previous course offerings while maintaining the focus within each technology silo.

Informing the design of both curricula were the results of statistical analyses of literally millions of student exams taken over the life of the previous four-course curriculum. From this analysis (Classical and IRT at the chapter exam level and at the summative final exam level), patterns emerged that indicated that the placement of certain assessment tasks targeting specific Knowledge, Skills, and Abilities (KSAs), at different points within the curriculum affected the performance (difficulty) of the task. These patterns, in combination with subject matter expert input, helped create the initial learning progression framework. The Networking Academy participated in ongoing research into methods to identify the features that differentiated between novice and expert performance within the curriculum domain (DeMark & Behrens, 2004; Behrens, Frezzo, Mislevy, Kroopnick, & Wise, 2007). In addition to the Networking Academy input, external research highlighting the real-world skill and knowledge necessary for various job levels was used to validate the SME opinion and analysis results.

Curriculum maps for the new courses used this initial learning progression framework as a basis for developing chapter and course objectives. Within each chapter, the learning material, practice activity and formative assessment opportunities were designed to build on each other to present and reinforce KSAs within a section of the overall LP framework. Items in chapter tests focus on these ideas. Thus chapters may be seen in themselves as simple learning progressions. Looking across chapters, however, revealed dependencies in data that suggest dependencies in the ordered acquisition of the required knowledge and skills as the student moves through the course of instruction. The expert-defined learning progression analyses identify conceptual development of strands of increasing complexity/sophistication, which, when ideas across chapters are considered, suggest more elaborated learning progressions.

Each item on a chapter test is generally aimed at one level of one LP. Tasks are built to do this by means of task features that are keyed to the targeted level in the LP. Analyses of data from exams that accompany the new curricula are enabling us to refine both the learning progressions and the assessment design. We found that some unexpectedly difficult items incorporated ideas from an additional LP, which increased the difficulty (in the discussion we look ahead to modeling tasks that are intended to tap knowledge and skills from multiple LPs in concert). Subject matter expert review of similarly designed tasks that performed differently isolated the features of tasks that affected the difficulty. This helped us find the appropriate grain size to define the LPs and to create assessment task design patterns to target the KSAs at each LP level.

Task Design

Evidence centered design (ECD; Mislevy et al., 2003) is a framework for designing assessments to support desired inferences about students similar to other approaches that explicitly incorporate theories of cognition into the design process (e.g., Embretson, 1998). ECD guides the assessment design process via addressing a series of questions: "What claims or inferences do we want to make about students?", "What evidence is necessary to support such inferences?", and "What features of observable behavior facilitate the collection of that evidence?" ECD is applied to develop tasks and scoring rules for measuring student's proficiencies through the perspective of learning progressions.

This framework enables us to explicitly build LPs into task design. Tasks can reflect the LP knowledge and skills required for these tasks. Tasks can also differ in terms of previous knowledge required based on the LPs. Scoring rules can be defined based on the LPs as specified by a given design pattern, to capture evidence as to where a student is along this progression. Careful attention to task design shows us how two seemingly similar items actually assess different levels of a learning progression. Below is an example of two such items:

Variant A	Variant B
It is necessary to block all traffic from an entire subnet with a standard access control list. What IP address and wildcard mask should be used in the access control list to block only hosts from the subnet on	It is necessary to block all traffic from an entire subnet with a standard access control list. What IP address and wildcard mask should be used in the access control list to block only hosts from the subnet on

which the host 192.168.16.43/28 resides?	which the host 192.168.16.43/24 resides?
A.192.168.16.0 0.0.0.15	A.192.168.16.0 0.0.0.15
B.192.168.16.0 0.0.0.31	B.192.168.16.0 0.0.0.31
C.192.168.16.16 0.0.0.31	C.192.168.16.16 0.0.0.31
**D.192.168.16.32 0.0.0.15	D.192.168.16.32 0.0.0.15
E.192.168.16.32 0.0.0.16	E.192.168.16.32 0.0.0.16
F.192.168.16.0 0.0.0.255	**F.192.168.16.0 0.0.0.255

The change in the stem from /24 to /28 requires students to perform a more advanced IP addressing skill, subdividing one of the octets. This moves the question from one that distinguishes novices from those who know nothing to one that distinguishes individuals who are at a higher level (level 3 in terms of the learning progression described below) from those at lower levels. Even changes such as this that seem minor on the surface must be accounted for in task design when they affect demands related to the learning progression.

Of particular assistance in this task is an ECD approach called design patterns (Frezzo, Behrens, & Mislevy, in press), which were used to develop tasks and scoring rules for various types of assessments in the Cisco Networking Academy. Design patterns are useful tools when developing assessments as they provide a structured model of the knowledge and skills required as needed in a particular task. A design pattern outlines the knowledge, skills and abilities to be measured, the type of evidence needed to measure these skills and the methods for determining how this evidence reflects on the skills. While a design pattern may specify the requirements of a particular assessment it also leaves room so that similar tasks can be developed based on any given pattern. While these tasks may be similar, they can vary in difficulty and other aspects in order to reflect the purpose of the assessment.

Overall, design patterns and the tools in ECD can aid in developing an assessment that will provide insight on where a student is located on the scale of learning progressions. This information can be used to determine the skills a student has as well as where the student lies along a given learning progression. Bayesian networks provide the structure in which this evidence is synthesized and analyzed in a rigorous statistical framework.

Use of Bayesian Networks

The proposed learning progressions and the ECD-based assessments lead naturally to a desire to make inferences from the assessments about the learning progressions. We used Bayesian networks (BNs) as a tool to help make these inferences. BNs are models that combine probability theory and graphical theory to represent the probabilistic relationships among a large number of variables. They have many uses, and have been applied in a broad range of settings including diagnostic and expert systems (Spiegelhalter, Dawid, Lauritzen, & Cowell, 1993). In education, BNs have been used in complex assessment systems (Almond et al., 2007; R. Levy & Mislevy, 2004; Reye,

2004) and frequently have been used in the context of intelligent tutors to create models of an individual students' knowledge and provide information based on that model (Conati, Gertner, & VanLehn, 2002; Murray & VanLehn, 2000). Important for this application is that they allow for a representation of the theory of the relationships in a domain and use probability theory to examine the strength of those relationships.

BNs are based on conditional probabilities in which the probability of one event, for example success on a given assessment task, is conditional on the probability of a previous event, for example success on a previous task. However, instead of focusing only on the relationship between two variables, BNs and related graphical models structure relationships across multivariate systems. A BN (Jensen, 1996; Pearl, 1988) models the relationships among a set of variables by specifying recursive conditional distributions in order to structure the joint distribution. The networks are so named because they support the application of Bayes' theorem across complex networks by structuring the appropriate computations (Lauritzen & Spiegelhalter, 1988; Pearl, 1988).

A BN may also be represented as a graphical model (see Figure 1), consisting of the following elements (Jensen, 1996):

- A set of discrete variables represented by ellipses or boxes and referred to as nodes. Each variable has a set of exhaustive and mutually exclusive states.
- A set of directed edges (represented by arrows) between nodes indicating the probabilistic dependence between variables. Nodes at the source of a directed edge are referred to as parents of nodes at the destination of the directed edge, their children.
- For each exogenous variable (i.e., a variable without parents such as Connectivity and IP Addressing in Figure 1), there is an associated unconditional probability distribution where the probabilities over the states sum to one.
- For each endogenous variable (i.e., a variable with parents such as ConTask1 in Figure 1), there is an associated set of conditional probability distributions corresponding to each possible combination of the values of the parent variables, where the probabilities of the states in each conditional distribution sum to one (see Figures 2 and 3).

The variables and the directed edges together form an acyclic directed graph (Brooks, 1998; Jensen, 1996; Pearl, 1988). These graphs are directed in the sense that the edges follow a "flow" of dependence in a single direction; in contrast to other graphical modeling traditions (e.g., Bollen, 1989) the arrows are always unidirectional rather than bi-directional. The graphs are acyclic in that following the directional flow of directed edges from any node it is impossible to return to the node of origin. The structure of the graph conveys the patterns of dependence and (conditional) independence among the variables in the joint distribution and corresponds to the computations involved in constructing the joint distribution of the variables in the system and subsequently conducting Bayesian inference to yield posterior distributions for the unknown variables once data have been observed (Lauritzen & Spiegelhalter, 1988; Pearl, 1988). In connection with this last point, BNs properly and efficiently quantify and propagate the evidentiary import of observed data on unknown entities, thereby facilitating evidentiary reasoning under uncertainty as is warranted in psychometric and related applications

(Almond & Mislevy, 1999; Almond et al., 2007; R. Levy & Mislevy, 2004; Mislevy, 1994; Mislevy & R. Levy, 2007; Spiegelhalter et al., 1993).

The graphs give rise to conditional probability tables. Figure 2 shows conditional probabilities for a hypothetical Observed Variable (ConTask1) from Figure 1 at different states of the Student Model Variable labeled Connectivity. A Student Model Variable (SMV) describes the characteristics of examinees we want to make inferences about, in this case their knowledge about network connectivity. The states in the SMVs in the Bayes net correspond to levels in the LPs marked by conceptual "jumps." In assessment applications of Bayes nets, SMVs are latent variables that characterize the state of the learner, Observable Variables (OVs) characterize features of task performances (such as correctness, efficiency, or consistency), and OVs are modeled as depending stochastically on SMVs through the conditional probability tables.

BNs may be employed to model the hypothesized structure of multiple learning progressions, where discrete latent variables correspond to the skills and the categories of latent variables correspond to the different levels of the skills. The pattern of dependence of the observables on the latent variables reflects the hypothesized structure of the manner in which performance depends on the student's status with respect to the progression. Possible sources by which to model the relationships among the latent variables include exploratory path analyses of scores on the exams and subject matter experts' beliefs about the domain and students' learning progressions. Figure 3 displays conditional probabilities for an OV in a multidimensional BN. This item has two SMV parents (IP Addressing and Connectivity), which combine to form the probability distribution for the OV. Building out networks of combinations of OVs and SMVs allows us to map out complex and interrelated learning progressions.

A Worked Example

The following is a worked example of the creation and validation of a single dimensional LP for the skill of IP Addressing using BNs. This section will describe the expert-generated skill levels, the data used in the analyses, and the analyses resulting in a method for validation and revision of the progression. The results of this example provide information about the assessment items used to identify the levels of the learning progression as well as identification of the levels of specific students.

Expert Identification of a Learning Progression

To begin the process, a learning progression for the skill of IP Addressing was developed by an expert in the field. This progression, presented in Appendix A, contains five levels, from Novice to Expert. Within each level is a set of KSAs that individuals at that level would be expected to have.

Next, the same content expert examined the end of chapter exams for the first course in the Discovery course sequence in order to identify items that map to the levels in the IP Addressing progression. These end of chapter exams are more traditional multiple choice exams that average around 20 questions per exam. The first Discovery course contains 9 chapter exams. The analysis led to the identification of 4 items at the novice level, 9 items at the basic level, 12 items at the intermediate level, and 11 items at the advanced level. The items at each level are those that should differentiate best between that level and the one below it. An individual at the basic level should have a lower probability of mastering the intermediate items than an individual at the intermediate level. It is not surprising that no items were identified at the expert level given that the course being examined is the first in a series of four. The items came from five different chapter exams, and in most cases a given chapter exam yielded items at multiple levels.

Data Selection and Preparation

Our goal in this analysis was to use end of chapter exam data to validate the number of expert-identified skill levels and identify the exam items that best discriminate between these levels. As such, a cross-sectional sample of data was taken, as opposed to a longitudinal sample. In the future, in which a goal might be to model individual students' progressions through the LP, a longitudinal sample might be taken. However, in this case, data from all of the end of chapter exams taken in November 2007 were included in the analysis. This month was selected due to the high volume of exams taken. In cases where a student took exams on multiple days in the month, only the exam(s) taken on the first day were included in the data. This resulted in a sample of 3827 student records.

The number of data points for each chapter is shown in Table 1. For any given chapter data from at least 198 students were taken. In addition 86 students took all of the chapter exams on the same day. Since it is assumed that no learning occurred during that day, all items taken in one day should reflect the student's appropriate level of the learning progression and therefore all data for these students were used.

Some initial analysis was performed to provide insight into the nature of these items and their relationship with each other. Difficulty values (see Table 2) were calculated in order to identify items that might not be appropriate for further analysis. One level 1 item (from chapter 6) was found to have a difficulty value of 1 (everyone obtained a correct answer for that item) and was therefore not used in further analysis. Average difficulty values were obtained, and it was found that on average the items increased in difficulty as they increased in levels. While this is to be expected, caution should be taken in the interpretation of this finding, as the people who took each of these items may differ and therefore comparisons across items may reflect differences in the population as well as differences in the items.

Polychoric correlations were calculated and examined to determine if there were patterns in the data. It was expected that items that measured the same levels would be found to have higher correlations than items that measuring different levels. While a few items followed this expected pattern there was a high level of correlations across items from all levels. On average the correlations between any two groups of items were between .35 and .44 and it was not always the case that items of the same level had the highest average correlation with other items of the same level but there were relatively higher correlations between adjacent levels than remote levels. For example, level 3 items on average have higher correlation with level 2 and level 4 items than with level 1 items (see figure 2) (see Figure 4). The similarities across levels of correlations may in part be due to the fact that all of the items should be measuring the same underlying skill. Items were then examined to determine if items from the same chapter had higher correlations than items from differing chapters. Again no strong patterns of correlations were found (see Figure 5).

In addition, a factor analysis was run using the polychoric correlations. While there were eight eigenvalues over the value of 1 it was found that there was strong evidence for 1 dominant factor (see Figure 6). This was not entirely unexpected as all of the items should be measuring the same underlying skill.

In general, while the exploratory data analysis provided evidence that the items were related to each other and that they were all measuring the same general skill set, there still did not seem to be evidence one way or another regarding whether the items themselves were labeled at the appropriate level. For this, further analysis using BNs was conducted.

Creation of Bayesian Networks

In this example, the BN contains a single discrete latent variable modeled as the parent for the discrete OVs (i.e., scored item responses), the children, as graphically depicted in Figure 7. This BN model is equivalent to a latent class model (Dayton & Macready, 2007; Lazarsfeld & Henry, 1968). A latent class analysis was conducted using the poLCA package (Linzer & Lewis, 2007) in R (R Development Core Team, 2008) using multiple start values to determine the optimal solution. Given that the item pool contained items corresponding to four levels of the expert-based learning progression pool, it was anticipated that ideally a 5-class model would be supported, where, in addition to the four levels defined by the expert represented in the items, a fifth class would emerge representing a knowledge state below the first (novice) level. The implication for the data analysis is that, by defining the levels in terms of what students know or can do (Appendix A), there is an additional, implicit baseline level (level 0) representing essentially below novice knowledge. That is, the novice level items would discriminate between students at the novice level (or beyond) and students who had not yet learned the novice material (i.e. level 0). Similarly, items at the advanced level would discriminate between students at the intermediate level and students who were at the advanced level (or beyond). The lack of items at the expert level in this sample precluded us from expecting the 6-class model that might otherwise have been suggested theoretically.

To allow for the possibility that the data provided better support for a model with a different number of levels, we compared 2-Class, 3-Class, 4-Class, 5-Class, and 6-Class models. The 4-Class model demonstrated the best fit to the data, based on statistical fit in terms of the BIC (Schwarz, 1978) and the bootstrapped likelihood ratio test (McLachlan & Peel, 2000; Nylund, Asparouhov, & Muthén, 2007) conducted in M*plus* (Muthén & Muthén, 1998-2006). In addition, this model offered the best interpretability of the classes in terms of class membership proportions and consistently ordered patterns of class performance across items. That is, the four classes identified in the analysis correspond to increasing levels of performance on the items and are interpreted as increasing levels of knowledge, skills, and proficiency. A BN representation of the 4-Class model was then constructed in Netica (Norsys Software Corp, 2007).

The lack of support for a 5-Class model was apparently due to the small number of items at the novice level (level 1), as well as an absence of students below the novice level. This is unsurprising, given that the items used in this analysis are drawn from assessments administered after instruction has occurred. In other words, to discriminate well between students who are essentially ignorant of the material and those who have achieved novice understanding of the material, students would need to be measured earlier, perhaps with a pre-test that included more novice items. As discussed below, many of the items functioned in ways consistent with the expert-based expectations of the learning progression. Thus, the four classes are interpreted as the first four levels of the learning progression, where the first class is perhaps a mixture of students at and below the first level of the progression.

Inferences Regarding Assessment Items.

One goal of the analysis included the modeling of the items' locations along the learning progression. Specifically, an item was classified as being "at the level" of a certain class if it supported an interpretation that students reaching that level would be able to solve or complete the task whereas students at lower levels would find be unlikely to be successful. To classify items, the conditional probability tables were examined. For each item, the odds of answering the item completely correctly were calculated in each class and odds ratios were calculated to compare adjacent classes. These odds ratios capture the power of the items to discriminate between classes. To construct an odds ratio for the first class, the probability that a complete novice would get the item right was defined as the probability of getting the item right by guessing. Each item was assigned to a level based on considerations of (a) the size of these odds ratio between successive classes, (b) the criterion that the probability of responding correctly at the assigned level exceeded .50 for dichotomously scored items, and (c) the distribution of probability across the response categories for polytomously scored items.

The results indicate that many of the items discriminate strongly between classes. For example, Figure 8 contains the conditional probability table for an item where it is clearly seen that only students in the fourth (highest) class are likely to successfully solve it. Statistically, this item aids in distinguishing students in the fourth latent class (level) from the remaining classes. Substantively, the item captures one aspect of what it means to be at the fourth level of the learning progression. Students at the fourth level have learned the knowledge and skills necessary to correctly answer this item; students at the lower levels have not.

Still other the items were more ambiguous in terms of their levels. For example, Figure 9 contains the conditional probability table for an item where it is seen that students in the second class have a probability of .88 for earning partial or full credit, but only a .52 probability of earning full credit, whereas students in the third class have a .86 probability of earning full credit. A simple classification of this item in terms of one level is insufficient to fully capture its connection to the classes. A richer characterization of the item, recognizing that it discriminates well between multiple adjacent classes, states that once a student reaches class two, she is very likely to earn at least partial credit but needs to reach class three (or four) in order to be as likely to earn full credit.

The results were largely consistent with the expert-based expectations regarding the items. Ten items exhibited clear and distinct patterns in which they distinguished between classes exactly as predicted by experts. That is, these items were "located" at the level as expected. Figure 8 is an example of one such item; the expert prediction of this item as a level 4 item is strongly supported by the results. Five items distinguished roughly well at the level predicted by experts and at one other level; that is, they appeared to be located at the expected level and one other level. Eighteen of the items were located at a level adjacent to where they were predicted to be located (e.g., an item expected at level 4 was located at class 3). One item was located at one class adjacent to the predicted to the class and another class not adjacent. The results for this item are given in Figure 9. This item was expected to be a level four item. As discussed above, the polytomous scoring of this item makes it possible to view it as being located at class 2 or class 3. Only one item was clearly located at a class that was not equal to or adjacent to the predicted level.

Inferences Regarding Students.

The conditional probability tables also reveal how inferences regarding students are conducted in the BN. For example, observing a correct response for the item in Figure 8 is strong evidence that the student is in class 4; observing an incorrect response for the item in Figure 8 is relatively strong evidence that the student is not in class 4. The use of a BN approach supports inferences regarding students by collecting and synthesizing the evidence in the form of observed values of variables. That information is then propagated through the network via Bayes' theorem to yield posterior distributions for the remaining unknown variables (Pearl, 1988), including the latent class variable corresponding to the skill level. For example, Figure 7 contains the BN for a student who has completed four of the items. The student correctly answered the first two items and incorrectly answered the latter two items. On the basis of this evidence, the posterior distribution for their latent skill variable indicates that this student has a probability of being in classes 1-4 of .476, .332, .172, and .021, respectively. On this basis, we may infer that the student is most likely in one of the first two classes (i.e., is at one of the first two levels of the skill progression) but that there still remains considerable uncertainty. The collection and inclusion of more data would lead to a more refined inference, as illustrated in Figure 10, which contains the BN for another student who has completed all of the items. The posterior distribution for this student is quite clear in supporting an inference that the student is in class 3 (posterior probability equals .997); that is, the student is in the third level of the learning progression.

Comment on the Example

This example demonstrates how assessment data can be used to validate a learning progression using statistical modeling in the form of a BN. Assessment items that discriminated between various levels in the progression were identified. In addition, it was demonstrated how students could be classified into levels based on their assessment results.

The results of the modeling offer a data-based interpretation of the development of skills that constitute the learning progression. In some cases, the results for items serve to confirm the expert-based expectation. For other items, the results are more ambiguous or offer an alternative interpretation to that from the experts. Taking a comprehensive perspective on assessment of learning progressions, the results of the statistical analyses will be taken back to the subject matter experts for consultation and possible refinements in terms of the definition of the learning progression (Appendix A), the items that assess the aspects of the learning progression, and the utility of additional items for modeling students' progression.

The BN modeling approach facilitates probability-based reasoning about students in terms of their learning progression. Assessment data (e.g., scored item responses) enter the network in the form of OVs. Synthesizing the evidentiary import of the data, the posterior distribution of class membership, interpreted as the level of the learning progression, governs the inferences regarding the student.

Conclusions and Looking Forward

In summary, this paper describes the application of a variety of techniques centered around BN modeling to a real-world example of learning progressions. LPs defined by experts matched with ECD-based assessment tasks completed by thousands of students provide the basis for this analysis. It is argued that BNs are well positioned to support inferences at fine-grained levels aligned with rich substantive theories, and as such are powerful statistical tools for modeling and structuring substantive inferences and feedback to students, instructors, and curricular designers.

Growing the Progression

The worked example shown in this paper focused on OVs related to one SMV (IP Addressing). The results from this analysis, however, can be "plugged in" to a much larger model that displays the relationships between SMVs (see Figure 11). This allows for the modeling of the influence of mastery of one area on the mastery of another area. For example, in this figure, if students are having difficulty configuring a router, we can determine the areas further back in the learning progression that they most likely did not master. We can also show instructors which areas it is most important for students to master in order to master future topics, helping the instructor make decisions about where to focus time in the class.

Instructional Implications

The Learning Progression in the worked example and the larger LP connecting the SMVs apply to both of the Networking curricula developed for the Cisco Networking Academy. That is, the students must gain these knowledge, skills, and abilities to become expert networkers. However, the two curricula differ in the pathways that students take through these progressions. In the Exploration curriculum, all of the routing information is presented in the second course of the four course sequence. Students would be expected to move through the larger LP during that course. Alternately, the Discovery curriculum uses more of a spiral technique, introducing early elements of the LP in the first course (as we saw with the worked example here) and then building on those in the second through fourth courses.

Additional Assessment Types

This paper focused on analysis using traditional multiple choice exams. In the future, we will be using the same process in the analysis of assessments from a Networking Academy tool called Packet Tracer (PT). PT is a comprehensive simulation, visualization, collaboration, and micro-world authoring tool for teaching networking concepts (Frezzo, Behrens, Mislevy, West & DiCerbo, in press). PT assessments are being constructed using design patterns and task templates to create complex tasks at appropriate levels (Wise-Rutstein, 2005; Frezzo et al., in press). These design patterns and templates additionally provide structure for conditional probabilities in the Bayes nets and thus cast the interpretation of performance in terms of the LPs through the SMVs.

In the next phases, we will be seeing the larger-scale deployment of PT as an assessment tool. We will be applying these methods, and others as needed, to the assessment information resulting from students completing those tasks. We anticipate that the inclusion of this rich information will provide new insight into students' learning progressions.

In this project, teacher and subject matter experts have served as inputs into the modeling process. In the future, we need to continue to close the loop so that information resulting from the modeling then feeds back to inform future instruction and curriculum design. As such, methods of communication of both student-level and aggregate results need to continue to be refined. With these and other improvements, learning progressions will play an important role in understanding and improving student outcomes.

References

Almond, R. G., DiBello, L. V., Moulder, B., & Zapata-Rivera, J. D. (2007). Modeling diagnostic assessments with Bayesian networks. *Journal of Educational Measurement, 44,* 341-359.

Behrens, J. T., Collison, T. A., & Demark, S. F. (2005) The Seven Cs of comprehensive assessment: Lessons learned from 40 million classroom exams in the Cisco Networking Academy Program. In S. Howell and M. Hricko (Eds.), *Online Assessment and Measurement: Case Studies in Higher Education, K-12 and Corporate.* (pp 229-245). Hershey, PA: Information Science Publishing.

Behrens, J. T., Frezzo, D. C., Mislevy, R. J., Kroopnick, M., & Wise, D. (2007). Structural, Functional and Semiotic Symmetries in Simulation-Based Games and Assessments. In E. L. Baker, J. Dickieson, W. Wulfeck, & H. F. O'Neil (Eds.) *Assessment of Problem Solving Using Simulations* (pp. 59-80). New York: Erlbaum.

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika, 71,* 425-440.

DeMark, S. F. & Behrens, J. T. (2004). Using statistical natural language processing for understanding complex responses to free-response tasks. *International Journal of Testing, 4*, 371-390.

Dayton, C. M., & Macready, G. B. (2007). Latent class analysis in psychometrics. In C. R. Rao and S. Sinharay (Eds.), *Handbook of statistics, Volume 26* (pp. 421-446). North-Holland: Elsevier.

Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods, 3,* 380-396.

Frezzo, D. C., Behrens, J. T., Mislevy, R. J. (in press). Design Patterns for Learning and Assessment: Facilitating the Introduction of a Complex Simulation-Based Learning Environment into a Community of Instructors. *Journal of Science Education and Technology.*

Frezzo, D. C., Behrens, J. T., Mislevy, R. J., West, P., & DiCerbo, K. E. (in press) *Psychometric and Evidentiary Approaches to Simulation Assessment in Packet Tracer Software.* Proceedings of the Spanish Meeting of the IEEE.

Jensen, F. V. (1996). *An introduction to Bayesian networks*. New York: Springer-Verlag

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin, 1968.

Leighton, J.P. & Gierl, M. J. (Eds.) (2007). *Cognitive Diagnostic Assessment: Theories and Applications.* Cambridge: Cambridge University Press.

Levy, F., & Murnane, R.J. (2004). The new division of labor: How computers are creating the next job market. Princeton, NJ: Princeton University Press.

Levy, R., & Mislevy, R. J (2004). Specifying and refining a measurement model for a computer-based interactive assessment. *International Journal of Testing, 4,* 333-369.

Linzer, D. A., & Lewis, J. (2007). poLCA: Polytomous variable latent class analysis. R package version 1.1. http://userwww.service.emory.edu/~dlinzer/poLCA.

McLachlan, G., & Peel, D. (2000). *Finite mixture models.* New York: John Wiley.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*, 3-62.

Murnane, R. J., Sharkey, N. S. & Levy, F. (2004). A role for the internet in American education? Lessons from the Cisco Networking Academies. In P. A. Graham & N. G. Stacey (Eds.), *The Knowledge Economy and Postsecondary Education: Report of a Workshop.* (pp.127-158). Committee on the Impact of the Changing Economy on the Education System, National Research Council. Washington, D.C.: National Academies Press.

Muthén, L. K., & Muthén, B. O. (1998-2006). Mplus User's Guide. Fourth Editions. Los Angeles, CA: Muthén & Muthén.

Norsys Software Corp (2007). *Netica manual*. http://www.norsys.com

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*: *A Multidisciplinary Journal, 14,* 535-569.

Pearl, J. (1988). Probabilistic reasoning in intelligent systems: Networks of plausible inference. San Mateo, CA: Kaufmann.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Reye, J. (2004). Student modelling based on belief networks. *International Journal of Artificial Intelligence in Education, 14,* 63-96.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461-464.

Spiegelhalter, D. J., Dawid, A. P., Lauritzen, S. L., & Cowell, R. G. (1993). Bayesian analysis in expert systems. *Statistical Science, 8,* 219-247.Van Merriënboer, J. (1997). Training Complex Cognitive Skills: A Four-Component Instructional Design Model for Technical Training. Englewood Cliffs, NJ: Educational Technology Publications.

Wise-Rutstein, D. (2005, April). Design patterns for assessing troubleshooting in computer networks. Presented at the annual meeting of the American Education Research Association, San Francisco, CA.

Table 1. Number of examinees for each chapter and each chapter grouping.

Chapter	Chapter				
	3	4	5	6	9
3	1992				
4	374	1621			
5	217	331	745		
6	140	154	247	336	
9	86	89	99	113	198

Note that the number in any cell corresponds to the number of people who took the column chapter through the row chapter (so 140 people took chapters 3 through 6).

Table 2. Item difficulty value

chapter	level	item	difficulty
4	1	1	0.924
6	1	2	1.000
6	1	3	0.833
6	1	4	0.818
6	2	5	0.732
5	2	6	0.651
3	2	7	0.619
3	2	8	0.630
9	2	9	0.692
5	2	10	0.894
5	2	11	0.615
5	2	12	0.855
3	2	13	0.699
3	3	14	0.738
3	3	15	0.354
3	3	16	0.611
3	3	17	0.467
3	3	18	0.841
5	3	19	0.741
5	3	20	0.734
5	3	21	0.850
5	3	22	0.643
5	3	23	0.710
5	3	24	0.711
9	3	25	0.833
9	4	27	0.778
3	4	28	0.654
5	4	29	0.631
5	4	30	0.773
5	4	31	0.647
5	4	32	0.387
5	4	33	0.532
5	4	34	0.790
5	4	35	0.556
5	4	36	0.514

Figure Captions

Figure 1. A sample Bayes net with two student model variables (SMVs: Connectivity and IP Addressing), each embodying a four-level learning progression, and eight observable variables (OVs). By construction around salient task features and requirements, the OVs depend on one or both SMVs and are targeted to discriminate at specified values. (Figure obtained using the *Netica* Bayes net program)

Figure 2. Sample Netica Output - Conditional probabilities of the observable variable ConTask1. Its possible values are 0, 1, and 2 from a partial credit scoring scheme. Connectivity is the SMV parent of this Observed Variable. Each row is the conditional probability distribution for the Observed Variable given a value of the SMV Connectivity. This is a task meant to discriminate best at Level 2, the level at which there is an 70% probability of scoring at least a 1.

Figure 3. Conditional probabilities of the observable variable ConAddTask1. Its possible values are 0 and 1 (unsuccessful and successful solution). Both Connectivity and IP Addressing are the SMV parents of this OV. Each row is the conditional probability distribution for the OV given a combination of value of the two parents. By construction, this task has features for which understanding and carrying out a solution uses concepts at Level 1 of the Connectivity learning progression and Level 2 of IP Addressing. The conditional probability distributions thus show only 10% probability of a successful solution for all SMV combinations in which these levels are not reached, and 80% probability at all combinations with at least these levels.

Figure 4. Average correlations of items across levels. Each point is the average correlation of items from the level specified on the x axis with the items from the level of the line the point is on.

Figure 5. Average correlations of items across chapters. Each point is the average correlation of items from the chapter specified on the x axis with the items from the chapter of the line the point is on. Chapter 4 does not appear as there was only 1 item from that chapter. (Other chapters did not have items for this skill).

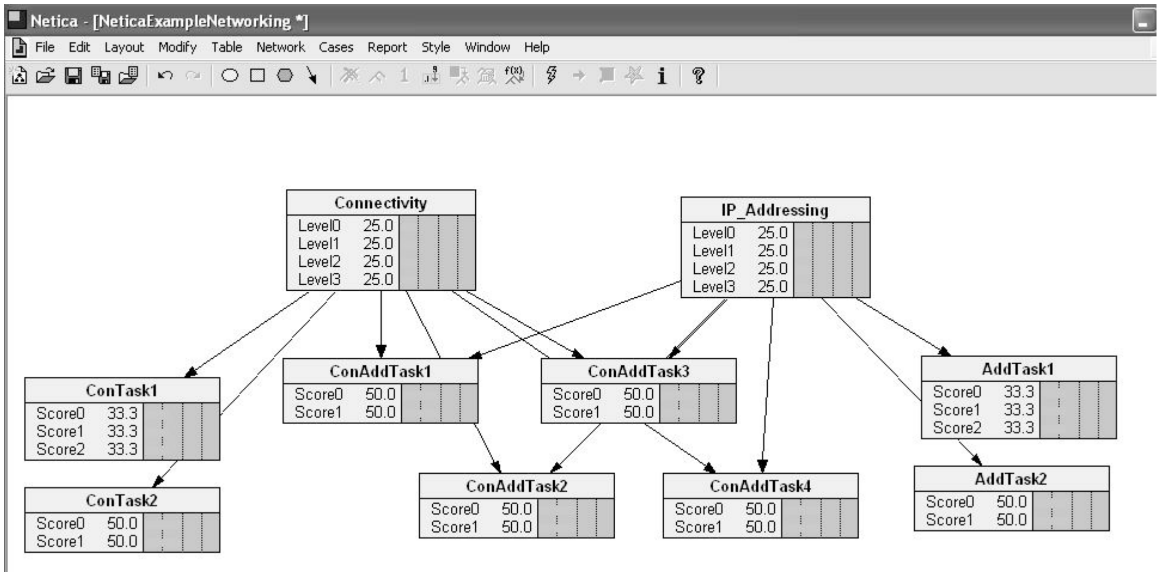
Figure 6. Scree plot for factor analysis – There is a line at 1 on the y axis

Figure 7. Little Johnny Bayes net. Bayes net for Student 320 with data on the first four items only. One SMV: IP Addressing, which embodies a four-level learning progression, and 35 observable variables (OVs). The posterior distribution indicates that this student is more likely to be a member of Class1 than any one of the other classes. (Figure obtained using the *Netica* Bayes net program)

Figure 8. Conditional probabilities of a clearly discriminating item (Item31). Its possible values are 0 and 1from a dichotomous scoring scheme. IP Addressing Skill is the SMV parent of this Observed Variable. Each row is the conditional probability distribution for the Observed Variable given a value of the SMV IP Addressing Skill. This is a task that discriminates best at Level 4, the level at which there is a 87.5% probability of scoring 1.

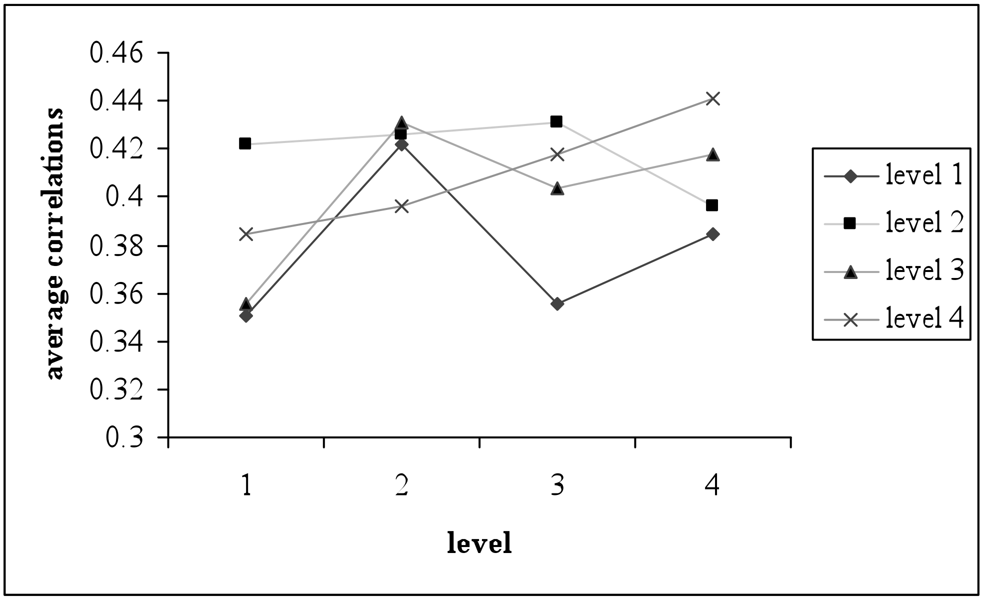
Figure 9. Conditional probabilities of a more ambiguous item (Item 33). Its possible values are 0, 1, and 2 from a partial credit scoring scheme. IP Addressing Skill is the SMV parent of this Observed Variable. Each row is the conditional probability distribution for the Observed Variable given a value of the SMV IP Addressing Skill.

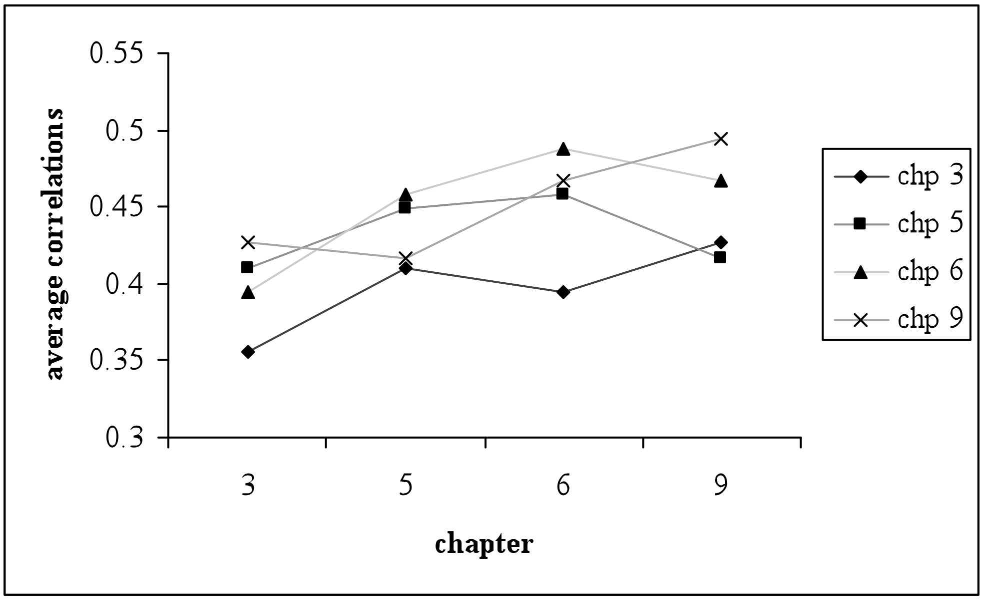
Figure 10. Little Sally Bayes net. Bayes net for Student 67 with data on all 35 of the items. One SMV: IP Addressing, which embodies a four-level learning progression, and 35 observable variables (OVs). The posterior distribution indicates that this student is almost certainly a member of Class3. (Figure obtained using the *Netica* Bayes net program

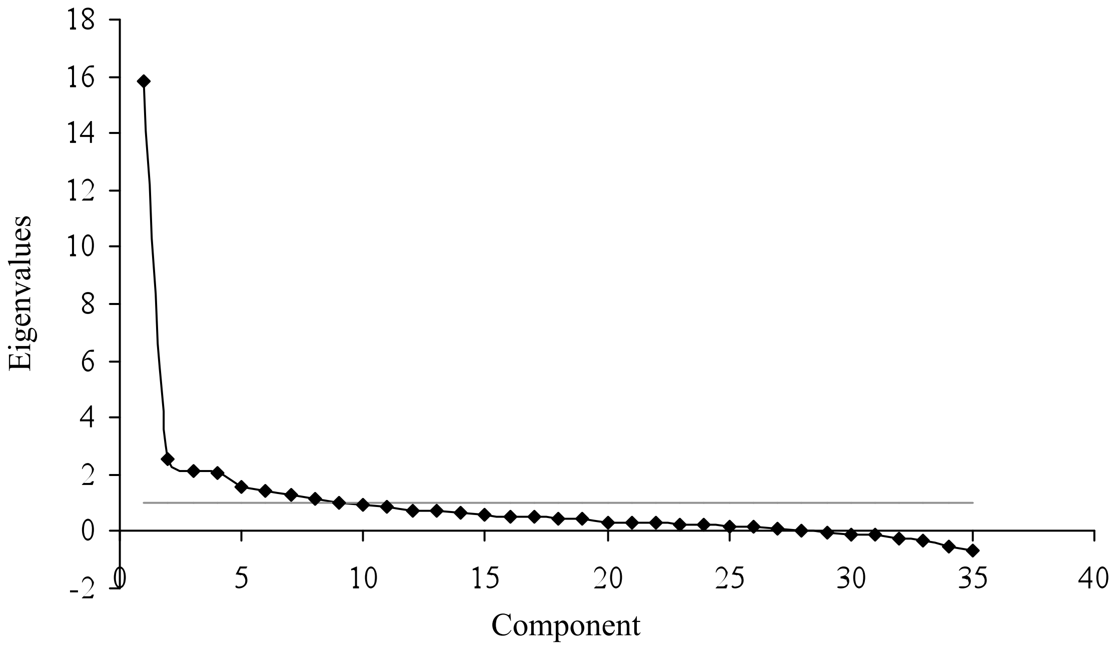

Netica - [NeticaExampleNetworking *]
File Edit Layout Modify Table Network Cases Report Style Window Help
Connectivity
Level0 25.0
Level1 25.0
Level2 25.0
Level3 25.0
IP_Addressing
Level0 25.0
Level1 25.0
Level2 25.0
Level3 25.0
ConTask1
Score0 33.3
Score1 33.3
Score2 33.3
ConAddTask1
Score0 50.0
Score1 50.0
ConAddTask3
Score0 50.0
Score1 50.0
AddTask1
Score0 33.3
Score1 33.3
Score2 33.3
ConTask2
Score0 50.0
Score1 50.0
ConAddTask2
Score0 50.0
Score1 50.0
ConAddTask4
Score0 50.0
Score1 50.0
AddTask2
Score0 50.0
Score1 50.0

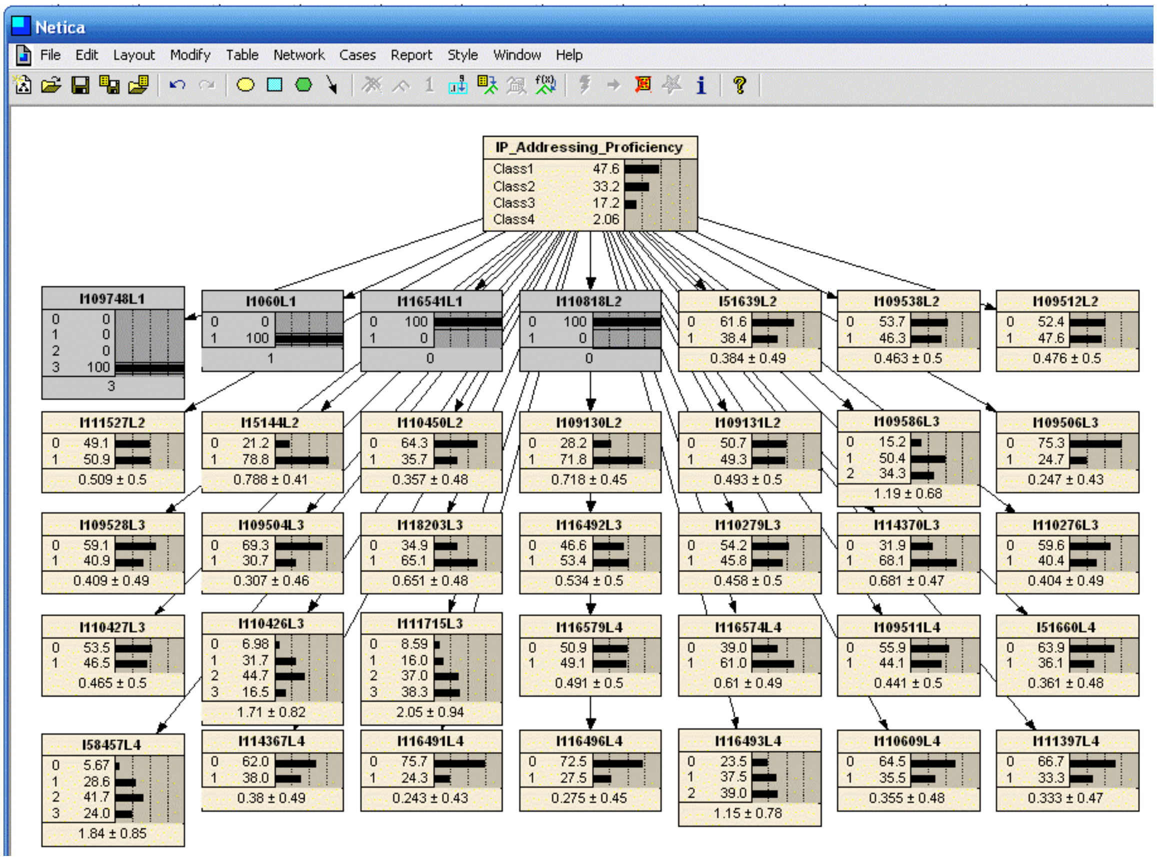
Connectivity	Score0	Score1	Score2
Level0	85.000	10.000	5.000
Level1	60.000	30.000	10.000
Level2	30.000	50.000	20.000
Level3	10.000	60.000	30.000

Connectivity	IP_Addressing	Score0	Score1
Level0	Level0	90.000	10.000
Level0	Level1	90.000	10.000
Level0	Level2	90.000	10.000
Level0	Level3	90.000	10.000
Level1	Level0	90.000	10.000
Level1	Level1	90.000	10.000
Level1	Level2	20.000	80.000
Level1	Level3	20.000	80.000
Level2	Level0	90.000	10.000
Level2	Level1	90.000	10.000
Level2	Level2	20.000	80.000
Level2	Level3	20.000	80.000
Level3	Level0	90.000	10.000
Level3	Level1	90.000	10.000
Level3	Level2	20.000	80.000
Level3	Level3	20.000	80.000


0.46
0.44
0.42
0.4
0.38
0.36
0.34
0.32
0.3
average correlations
1
2
3
4
level
level 1
level 2
level 3
level 4

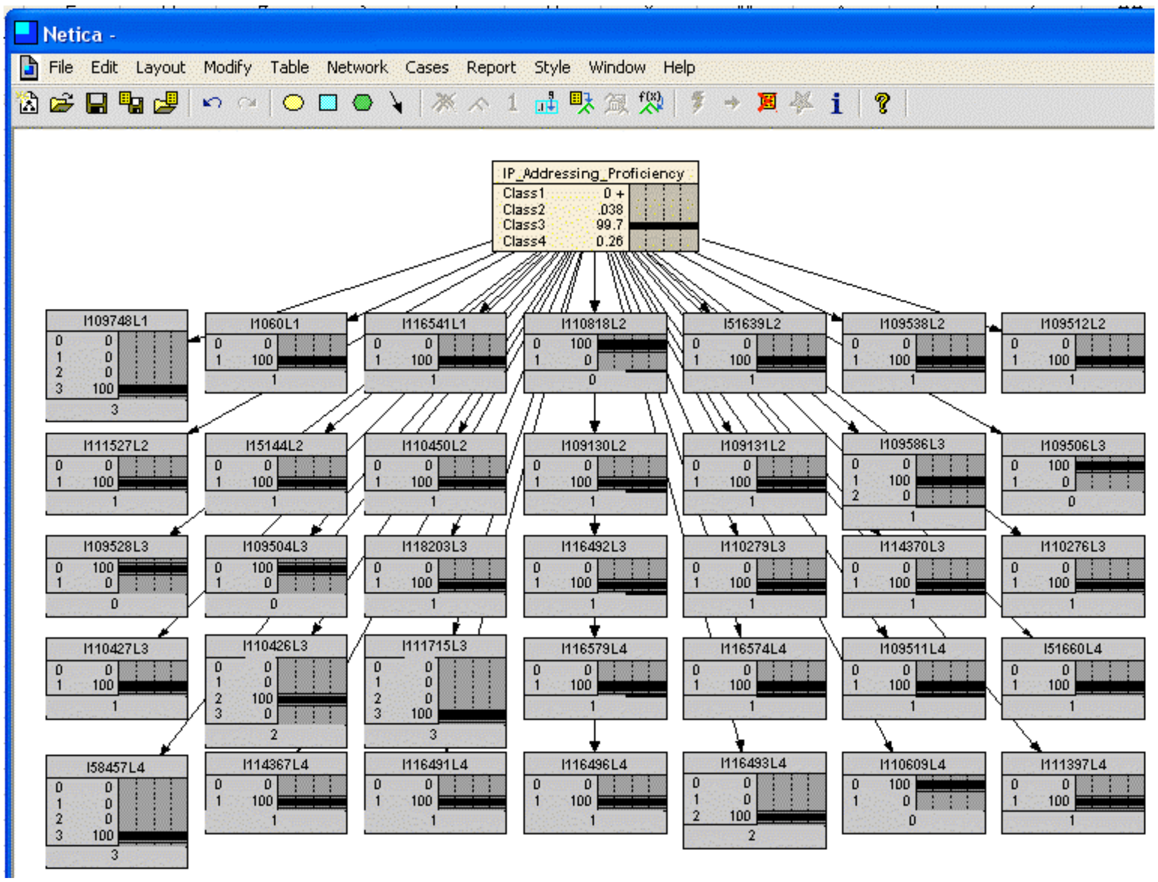

0.55
0.5
0.45
0.4
0.35
0.3
average correlations
3
5
6
9
chapter
chp 3
chp 5
chp 6
chp 9

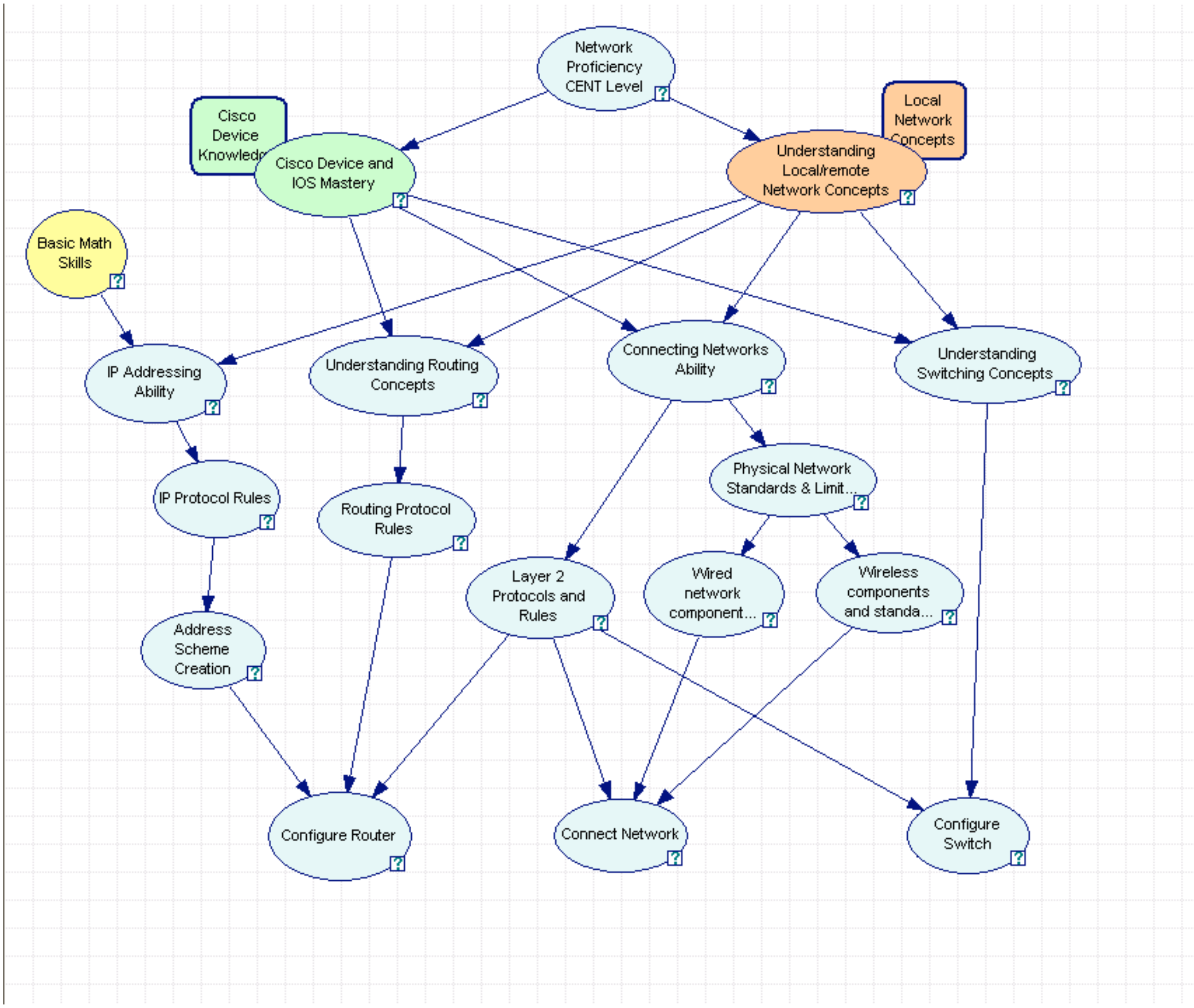

Eigenvalues
18
16
14
12
10
8
6
4
2
0
-2
0
5
10
15
20
25
30
35
40
Component


Netica
File Edit Layout Modify Table Network Cases Report Style Window Help
IP_Addressing_Proficiency
Class1 47.6
Class2 33.2
Class3 17.2
Class4 2.06
I109748L1
I1060L1
I116541L1
I110818L2
I51639L2
I109538L2
I109512L2
I111527L2
I15144L2
I110450L2
I109130L2
I109131L2
I109586L3
I109506L3
I109528L3
I109504L3
I118203L3
I116492L3
I110279L3
I114370L3
I110276L3
I110427L3
I110426L3
I111715L3
I116579L4
I116574L4
I109511L4
I51660L4
I58457L4
I114367L4
I116491L4
I116496L4
I116493L4
I110609L4
I111397L4

IP_Addressing_Proficiency	Score0	Score1
Class1	84.200	15.800
Class2	74.420	25.580
Class3	62.440	37.560
Class4	12.480	87.520

IP_Addressing_Proficiency	Score0	Score1	Score2
Class1	40.300	49.000	10.700
Class2	11.870	36.280	51.850
Class3	2.350	12.050	85.600
Class4	0.000	5.110	94.890


Netica -
File Edit Layout Modify Table Network Cases Report Style Window Help
IP_Addressing_Proficiency
Class1 0 +
Class2 .038
Class3 99.7
Class4 0.26
I109748L1
I1060L1
I116541L1
I110818L2
I51639L2
I109538L2
I109512L2
I111527L2
I15144L2
I110450L2
I109130L2
I109131L2
I109586L3
I109506L3
I109528L3
I109504L3
I118203L3
I116492L3
I110279L3
I114370L3
I110276L3
I110427L3
I110426L3
I111715L3
I116579L4
I116574L4
I109511L4
I51660L4
I58457L4
I114367L4
I116491L4
I116496L4
I116493L4
I110609L4
I111397L4


Network Proficiency CENT Level
Cisco Device Knowledge
Local Network Concepts
Cisco Device and IOS Mastery
Understanding Local/remote Network Concepts
Basic Math Skills
IP Addressing Ability
Understanding Routing Concepts
Connecting Networks Ability
Understanding Switching Concepts
Physical Network Standards & Limit...
IP Protocol Rules
Routing Protocol Rules
Layer 2 Protocols and Rules
Wired network component...
Wireless components and standa...
Address Scheme Creation
Configure Router
Connect Network
Configure Switch

Appendix A
IP Addressing Skills Progression

Skills Progression:

Level 1 – Novice -Knowledge/Skill (possibly pre-course knowledge and skills)

1. Student can navigate the operating system to get to the appropriate screen to configure the address.
2. Student knows that four things need to be configured: IP address, subnet mask, default gateway and DNS server.
3. Student can enter and save information.
4. Student can use a web browser to test whether or not network is working.
5. Student can verify that the correct information was entered and correct any errors.
6. Student knows that DNS translates names to IP addresses.
7. Student understands why a DNS server IP address must be configured.

Level 2 – Basic – Knows Fundamental Concepts

1. Student understands that an IP address corresponds to a source or destination host on the network.
2. Student understands that an IP address has two parts, one indicating the individual unique host and one indicating the network that the host resides on.
3. Student understands how the subnet mask indicates the network and host portions of the address.
4. Student understands the concept of local –vs- remote networks.
5. Student understands the purpose of a default gateway and why it must be specified.
6. Student knows that IP address information can be assigned dynamically.
7. Student can explain the difference between a broadcast traffic pattern and a unicast traffic pattern.

Level 3 – Intermediate – Knows More Advanced Concepts:

1. Student understands the difference between physical and logical connectivity.
2. Student can explain the process of encapsulation.
3. Student understands the difference between Layer 2 and Layer 3 networks and addressing.
4. Student understands that a local IP network corresponds to a local IP broadcast domain. (both the terms and the functionality)
5. Student knows how a device uses the subnet mask to determine which addresses are on the local

Layer 3 broadcast domain and which addresses are not.

6. Student understands the concept of subnets and how the subnet mask determines the network address.

7. Student understands why the default gateway IP address must be on the same local broadcast domain as the host.

8. Student understands ARP and how Layer 3 to Layer 2 address translation is accomplished.

9. Student knows how to interpret a network diagram in order to determine the local and remote networks.

10. Student understands how DHCP dynamically assigns IP addresses.

Level 4 –Advanced – Can Apply Knowledge and Skills in Context

1. Student can use the subnet mask to determine what other devices are on the same local network as the configured host.

2. Student can use a network diagram to find the local network where the configured host is located.

3. Student can use a network diagram to find the other networks attached to the local default gateway.

4. Student can use the PING utility to test connectivity to the gateway and to remote devices.

5. Student can recognize the symptoms that occur when the IP address or subnet mask is incorrect.

6. Student can recognize the symptoms that occur if an incorrect default gateway is configured.

7. Student can recognize the symptoms that occur if an incorrect DNS server (or no DNS server) is specified.

8. Student knows why DNS affects the operation of other applications and protocols, like email or file sharing.

9. Student can use NSlookup output to determine if DNS is functioning correctly.

10. Student can configure a DHCP pool to give out a range of IP addresses.

11. Student knows the purpose of private and public IP address spaces and when to use either one.

12. Student understands what NAT is and why it is needed.

Level 5 – Expert – Can Readily Apply Advanced Skills

1. Student can recognize a non-functional configuration by just looking at the configuration information, no testing of functionality required.

2. Student can interpret a network diagram to determine an appropriate IP address/subnet mask/default gateway for a host device.

3. Student can recognize the symptoms that occur if an incorrect subnet mask is configured on the intermediate routers or destination host.

4. Student can interpret a network diagram in order to determine the best router to use as a default gateway when more than one router is on the local network.

5. Student can evaluate a connectivity problem to determine if it could possibly be caused by an

incorrect setting configured on the host.

6. Student can propose changes to a host configuration to solve a connectivity problem.

7. Student can make and test proposed changes to a host configuration to solve an identified connectivity problem.

8. Student can implement NAT to translate private to public addresses.