PATTI WEST, DAISY WISE RUTSTEIN, ROBERT J. MISLEVY, JUNHUI LIU, ROY LEVY, KRISTEN E. DICERBO, AARON CRAWFORD, YOUNYOUNG CHOI, KRISTINA CHAPPLE AND JOHN T. BEHRENS

A BAYESIAN NETWORK APPROACH TO MODELING LEARNING PROGRESSIONS

A central challenge in using learning progressions (LPs) in practice is modeling the relationships that link student performance on assessment tasks to students' levels on the LP. On the one hand, there is a progression of theoretically defined levels, each defined by a configuration of knowledge, skills, and/or abilities (KSAs). On the other hand, there are observed performances on assessment tasks, associated with levels but only imperfectly and subject to inconsistencies. What is needed is a methodology that can be used to map assessment performance onto the levels, to combine information across multiple tasks measuring similar and related KSAs, to support inferences about students, and to study how well actual data exhibit the relationships posited by the LP. In terms of the "assessment triangle" proposed by the National Research Council's Committee on the Foundations of Assessment (National Research Council [NRC], 2001), coherent theoretical and empirical connections are needed among the theory embodied in a progression (cognition), the tasks that provide observable evidence about a student's understanding relative to that progression (observation), and the analytic models that characterize the relationship between them (interpretation).

This chapter discusses the use of Bayesian inference networks, or Bayes nets for short, to model LPs. Bayes nets are a class of statistical models that have been adapted for use in educational measurement. At present, the use of Bayes nets in LP contexts is in its relative infancy. We describe the fundamentals of the approach and the challenges we faced applying it in an application involving a LP in beginning computer network engineering.

The first section of the chapter reviews our framework of model-based reasoning. Subsequent sections map the development of LPs and associated assessments onto this framework and show how Bayes nets are used to manage the problems of evidence and uncertainty in the relationship between LPs and assessment task performances. We then explain in more detail what Bayes nets are, how they can be used to model task performance in the context of LPs, and the challenges that we face in this work.

MODEL-BASED REASONING

The lens of model-based reasoning helps clarify the role Bayes nets can play in modeling LPs. A model is a simplified representation focused on certain aspects of

Alicia C. Alonzo, Amelia Wenk Gotwals (eds.), Learning Progressions in Science: Current Challenges and Future Directions, 257–292.

a system (Ingham & Gilbert, 1991). The entities, relationships, and processes of a model provide a framework for reasoning about any number of real world situations, in each instance abstracting salient aspects of those situations and going beyond them in terms of mechanisms, causal relationships, and/or implications that are not apparent on the surface.

The lower left plane of Figure 1 shows phenomena in a particular real world situation. In the case of LP research, the situation is students' task performances. A mapping is established between this situation and, in the center of Figure 1, the semantic plane of the model; that is, structures expressed in terms of the entities, relationships, and properties of the model. The lines connecting the entities in the model represent causes, influences, mechanisms, and other relationships. The analyst reasons in these terms. In modeling LPs, this layer concerns progressions and their levels, relationships among different progressions, and expected performance on assessment tasks based on the features of tasks (what students are asked to do) and the features of their performances (what they actually do).

The real world situation is depicted in Figure 1 as fuzzy, whereas the model is well-defined. This suggests that the correspondence between real world entities and the idealizations in the model is never exact. The reconceived situation in the lower right plane of Figure 1 is a blend of selected aspects of the real world situation and elements of the model (shown in dotted form). The match between the real world and the data is not perfect, but a framework of meaning that the situation does not possess in and of itself can enhance our understanding of it (Suarez, 2004; Swoyer, 1991). It is here that descriptions, explanations, and implications for real world phenomena are formed. In the case of LPs, it is here that patterns of students' performance are interpreted in terms of their status or development with respect to the LP levels.

Symbol systems that are associated with some models further support reasoning, such as the algebraic and graphical representations of regression models shown above the semantic plane in Figure 1 as Representational Forms A and B. Similarly, Bayes nets provide mathematical and graphical representations to support reasoning about LPs, students' status on them, and evaluations of their performances across tasks.

DEVELOPMENT, ASSESSMENT, AND MODELING OF LEARNING PROGRESSIONS

When we speak of modeling a LP, we refer to a coherent set of elements: a progression defined in terms of the psychology and the substance of the domain under consideration, a specification of how real-world situations can be set up to evoke evidence about a student's status on the LP, and a measurement model (in our case, a Bayes net) that articulates the probabilistic relationship between student performances and status on the LP. These are the vertices of an "assessment triangle" (NRC, 2001): cognition, observation, and interpretation. Cognition refers to a theory about what students know and how they know it (the learning progression). Observation relates to the tasks we ask students to perform to gather evidence about what they know. Interpretation is the meaning we assign to these observations. Specifying and validating a probability model—Bayes nets in this case—helps analysts develop coherence among these elements in order to reason

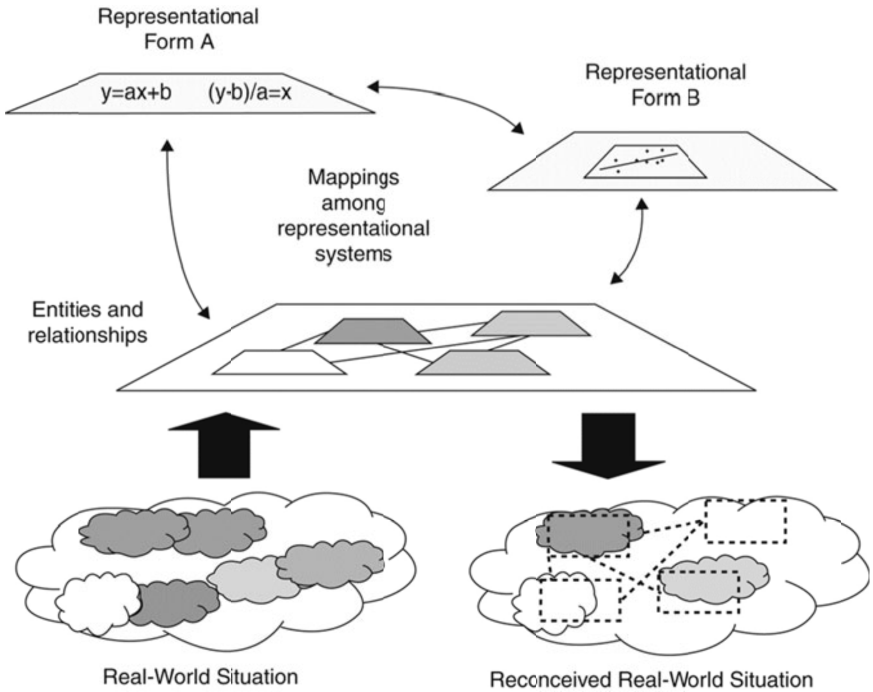

Figure 1. Reconceiving a real world situation through a model. From "Validity from the Perspective of Model-Based Reasoning," by R. J. Mislevy in The Concept of Validity: Revisions, New Directions, and Applications by R. L. Lissitz (Ed.) (p. 95), 2009, Charlotte, NC: Information Age Publishing. Copyright 2009 by Information Age Publishing. Reproduced with permission of Information Age Publishing.

about students' levels in the progression from their task performances. This section provides an overview of our work with LPs in the Cisco Networking Academy using this framework.

Learning Progression Development (Cognition)

LPs are empirically grounded and testable hypotheses about how a student's understanding and ability to use KSAs in a targeted area develop over time (Corcoran, Mosher, & Rogat, 2009). LPs are the pathways that bridge the gap between some starting point, such as the inability to connect two routers in computer networking, and a desired endpoint, such as the successful configuration of routers (National Research Council [NRC], 2007). There are five components of LPs (Corcoran et al., 2009; NRC, 2007): (1) learning targets or clear end points that are defined by social aspirations, (2) progress variables that identify the critical dimensions of KSAs that are developed over time, (3) levels of achievement or stages that define significant intermediate steps, (4) learning performances that are the operational definitions of what KSAs would look like at each stage, and (5) assessments that measure performances with respect to each key KSA over time. Although students may progress along different pathways, common paths can be tested and legitimated. It should also be noted that student learning and thinking progress in the context of instruction and experiences. This progression must be considered in creating, assessing, modeling, and interpreting LPs.

The following discussion describes the development of LPs in a specific content area: beginning computer networking. The context is the Cisco Networking Academy (CNA), a global program in which information technology is taught through a blended program of face-to-face classroom instruction, an online curriculum, and online assessments. These courses are offered in high schools, 2- and 3-year community college and technical schools, and 4-year colleges and universities. Since its inception in 1997, CNA has grown to reach a diverse population of approximately 900,000 students annually, in more than 160 countries (Levy & Murnane, 2004; Murnane, Sharkey, & Levy, 2004). Behrens, Collison, and DeMark (2005) discuss the framework that drives the ongoing assessment activity that, in turn, provides the data for this work.

In 2007, CNA updated and redesigned the curriculum for its primary network course offerings. A group of subject matter experts, working with psychometricians and educational psychologists, sketched out provisional LPs based upon several lines of work integral to the design of the curriculum (for details, see West et al., 2010). First, they conducted statistical analyses of student exams from the previous four-course curriculum. Classical and item response theory (IRT) analyses of end-of-chapter and final exam data revealed patterns in the difficulty of certain assessment tasks based upon their placement in the curriculum. For example, the same item used to assess IP addressing had different difficulty depending on whether the item was used before or after students learned basic routing concepts. Second, these patterns were considered in combination with the results of cognitive task analysis research into novice and expert performance in the domain (Behrens, Frezzo, Mislevy, Kroopnick, & Wise, 2007; DeMark & Behrens, 2004). Finally, external research highlighting the real-world KSAs necessary for various job levels was used to validate the subject matter expert opinion and statistical analyses. Thus the initial LP framework was developed through the interaction of various experts using both theory and data.

To make this discussion more concrete, Table 1 presents an example of a LP in Internet Protocol (IP) Addressing, a key area taught in the four-semester Cisco Certified Network Associate (CCNA) course sequence. IP addressing is the mechanism by which all pieces of equipment in a network (PCs, routers, etc.) are given unique "addresses" so information sent to them knows where to go and information sent from them is properly labeled for return if necessary. An analogy is the street address of a house. A five-level progression is defined based on clusters of interrelated, assessable elements that describe a student's capabilities at each level. The levels reflect increasingly sophisticated understandings of IP Addressing. Table 1 presents an abridged version of the KSAs at each level.

Table 1. Sample of Knowledge, Skills, and Abilities in the IP_Addressing Progression.

Level 1 – Novice – Knows Pre-requisite Concepts: Can recall factual information and perform highly scripted activities • Student can navigate the operating system to get to the appropriate screen to configure the address. • Student can use a web browser to check whether or not a network is working.
Level 2 – Basic – Knows Fundamental Concepts: Able to understand reasoning behind actions, but can't apply in unknown situations • Student understands that an IP address corresponds to a source or destination host on the network. • 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. • Student understands the default gateway is the address that data is sent to if the data is leaving the local network and why it must be specified. • Student understands how the subnet mask indicates the network and host portions of the address. • Student can create subnet masks based on octet boundaries.
Level 3 – Intermediate – Knows More Advanced Concepts: Able to apply concepts to actions • Student understands the difference between physical and logical connectivity. • Student can explain the process of encapsulation. • Student understands how Dynamic Host Control Protocol (DHCP) dynamically assigns IP addresses.
Level 4 –Advanced – Applies Knowledge and Skills: Able to apply concepts in context in an unscripted manner • Student can use the subnet mask to determine what other devices are on the same local network as the configured host. • Student can use a network diagram to find the local network where the configured host is located. • Student can recognize the symptoms that occur when the IP address or subnet mask is incorrect.
Level 5 – Expert – Applies Advanced Knowledge and Skills: Able to apply concepts in new contexts in an unscripted manner and predict consequences of actions • Student can recognize a non-functional configuration by just looking at the configuration information; no testing of functionality is required. • Student can interpret a network diagram to determine an appropriate IP address/subnet mask/default gateway for a host device. • 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.

Task Design (Observation)

The CNA assessment development process follows an Evidence Centered Design (ECD; Mislevy, Steinberg, & Almond, 2003) approach. ECD guides the assessment design process by addressing a series of questions: “What claims or inferences do we want to make about students?” “What evidence is necessary to support such inferences?” “What features of observable behavior facilitate the collection of that evidence?” At each level of the LP, a subject matter expert created multiple claims based on the set of related KSAs that define the level. In order to assess student performance with respect to these claims, the curriculum contains end-of-chapter tests and end-of-course final exams consisting of multiple-choice questions. Each chapter or course typically addresses multiple LPs. Our current focus is the case in which each item in an assessment is designed to measure one LP level. Work on modeling more complex assessment tasks that address multiple LPs is discussed later in the chapter.

In this example, IP_Addressing is called a student model variable (SMV) because it represents an aspect of a student’s proficiency. SMVs, like IP_Addressing, are latent variables, which means their values cannot be observed directly. However, students’ task performances provide evidence about them. Two items that provide evidence about a student’s level on IP_Addressing are shown in Figure 2. They both concern knowledge of the syntax of a router command. These two seemingly similar items provide evidence to distinguish between different levels of a LP due to a small but conceptually important difference in task features: Changing the stem from /24 to /28 requires students to have a more advanced IP Addressing skill, namely the skill to subdivide one of the octets. Item A distinguishes between Level 1 and Level 2 (students can create subnet masks based on octet boundaries), while Item B distinguishes between Level 3 and Level 4 (students can use the subnet mask to determine what other devices are on the same local network as the configured host).

Modeling Responses (Interpretation)

We can represent the different patterns of evidence provided by the sample items in Figure 2 with a Bayes net. First, a student’s level on the IP_Addressing LP can be represented with a variable called IP_Addressing. The variable has five possible values, one for each level of the LP. For each level, there is a probability that a student is at that level. Figure 3 represents ignorance about a student’s level, expressed as probabilities of .2 at each level. (The Netica program, Norsys Software Corp., 2007, shows these probabilities as percentages, hence 20 rather than .20.) This is called a *prior* probability distribution, reflecting the belief about a student before observing any of the student’s responses. We will see how observation of student responses allows us to update our beliefs and express them in a *posterior* probability distribution.

Item A	**Item 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 which the host 192.168.16.43/**24** resides?	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?
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

Figure 2. *Sample items assessing levels of IP_Addressing. From A Bayesian Network Approach to Modeling Learning Progressions and Task Performances (CRESST Report 776) (p. 6), by P. West, D. W. Rutstein, R. J. Mislevy, J. Liu, Y. Choi, R. Levy, A. Crawford, K. E. DiCerbo, K. Chappel, and J. T. Behrens, 2010, Los Angeles, CA: University of California, CRESST. Copyright 2010 The Regents of the University of California. Reprinted with permission of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST).*

To this end, our simple Bayes net also contains variables for responses to Item A and Item B. They are called observable variables (OVs) because we learn their values, usually with certainty, when we observe a student's responses. Their possible values are 1 and 0, scores for right or wrong responses, respectively. In our example, Item A distinguishes between Level 1 and Level 2, and Item B distinguishes between Level 3 and Level 4. The Bayes net indicates these relationships through conditional probability distributions for OVs, such as scores on Item A and Item B given a student's level on IP_Addressing, the latent SMV. The distributions indicate the conditional probability of getting Item A right or wrong at each level of the IP_Addressing LP.

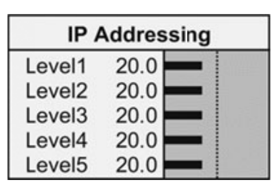

Figure 3. *A depiction of a student model variable that represents a five-level learning progression, indicating equal probabilities that a student is at each level. (Note: This figure and the Bayes net graphics that follow were produced with the Netica computer program. Probabilities are displayed in the percent metric and thus sum to 100 rather than to 1.)*

Table 2a specifies the relationship between IP_Addressing and Item A. Each row in the table is the conditional probability distribution for the values of the Item A OV, given the value of IP_Addressing. The row for Level 1, for example, says that a student at Level 1 has a probability of .8 of answering incorrectly and only .2 of answering correctly. (We discuss the source of these probabilities later in the chapter.) A student at Level 2 has a probability of .7 of answering correctly. Table 2 reflects the LP structure since students at Level 1 will probably get Item A wrong, but students at or above Level 2 will probably get Item A right. These expectations are probabilities rather than certainties because a student at Level 4 might miss Item A owing to carelessness, an arithmetic error, or a gap in knowledge. Although the capabilities at a given level are interrelated for both concepts and curriculum, students may be stronger on some elements at one level than on others. Table 2b, which gives conditional probabilities for Item B, shows a jump in conditional probabilities between Level 3 and Level 4.

Table 2. Conditional Probabilities for Item Responses Given the Level of IP_Addressing.

a) Item A

	Item A	
IP_Addressing	*Score 0*	*Score 1*
Level 1	80	20
Level 2	30	70
Level 3	20	80
Level 4	10	90
Level 5	10	90

b) Item B

	Item B	
IP_Addressing	*Score 0*	*Score 1*
Level 1	90	10
Level 2	80	20
Level 3	70	30
Level 4	20	80
Level 5	10	90

To summarize this section, theoretically defined levels of the learning progression provide information about what students know and how they know it. This is the cognition vertex in the assessment triangle. The theory and research underlying the LPs suggest how we might design tasks to elicit student performances that depend on their status in the progressions. This is the observation vertex. The interpretation vertex of the triangle addresses analytic models that connect assessment performances with the cognitive structure of the LP; these models are used to validate and improve the LP and task framework and to reason about individual students in that framework. The following sections explore this vertex using Bayes nets.

MODELING LPS USING BAYES NETS

Even the simplest LP structure poses issues of evidence and uncertainty since a student at a given level of a progression may provide responses that vary across levels from one task to the next. What degree of regularity should we expect in performance? How do we infer back from students' performances to the levels at which they likely work? How much evidence about a level does one task, or several tasks at different levels, provide us? Does a particular task operate differently than others when our theory says they should operate similarly? Additional complexities arise when we consider multifaceted clusters of concepts. Are there multiple typical paths students take that extend beyond the usual variability in performance? Are there identifiable strands of concepts that display their own regularities within a larger, less tightly structured progression? What hard prerequisites, or soft tendencies, seem to influence students' paths? How do we discover these patterns in noisy data?

Measurement models posit a relationship between (a) a student's status on inherently unobservable SMVs that characterize some aspects of their capabilities and (b) OVs that provide evidence about these SMVs. Specifying a measurement model becomes a matter of specifying these relationships and, in so doing, specifying how assessment data should be interpreted to yield inferences about students. In addition to Bayes nets, other modern measurement models with this same essential structure include latent class models (Dayton & Macready, 2007), cognitive diagnosis models (Rupp & Templin, 2008), and structured IRT models such as the Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM; Adams, Wilson, & Wang, 1997). In the context of LPs, the SMVs correspond to LPs since students are presumed to be at a given but unobservable level on each LP. We obtain evidence about their level from performance on tasks built, based on the theory underlying the LP, for this purpose.

More specifically, modern measurement models facilitate inferences about students using two key features. The first is latent variables. These variables recognize that what we would ultimately like to know about students (i.e., their levels on the LP) is unobservable, and must be inferred from what we can observe, namely, performance on tasks. The relationship between a student's level on the LP—the latent SMV—and performance on the tasks—captured in OVs—is at the heart of the inference. By specifying which values of the OVs (i.e., the performances on tasks) are expected based on the value of the SMV (i.e., the level of the LP), the measurement model allows us to make inferences about the SMV from observed values of the OVs.

The second feature of modern measurement models is the use of probability models to express these relationships. Student performances (OVs) are modeled as probabilistically dependent on the student's level on the LP (SMV). A student may exhibit task performances that do not exactly agree with the expectations based on the model. For example, a student who has reached a given level of the LP might demonstrate a higher or lower level of performance on a particular task; task performance could be the result of chance, of inconsistency in applying concepts, or of the influence of factors not encoded in the model. This is why the conditional probabilities in the introductory example (Table 2) are not all ones and zeros.

Combining these two features produces a modern measurement model—performances on tasks (OVs) are modeled as probabilistically dependent on the unobservable level of the LP (SMV). Box 1 provides a formal definition of a measurement model formulation in the LP paradigm. Given such a model, we can characterize tasks' effectiveness at distinguishing between levels (through the patterns in the conditional probabilities as estimated from data), and we can draw inferences about the status of students on a LP (as we will see shortly, through posterior probability distributions once we observe students' performances). Further, probability theory helps a researcher explore the fit and misfit of a model to data and iteratively fine-tune both tasks and theories.

==

Box 1: Formal definition of a measurement model

To more formally define the measurement model structure used in the modeling of LPs, let θ denote an unobservable SMV. Further, let X_1, X_2,…, X_J represent some number J of OVs, the values of which summarize performance on tasks (e.g., scored item responses). A measurement model then specifies the *conditional probability* for each OV, denoted $P(X_j \mid \theta)$. The conditional probability expression yields different probabilities of values of the OV depending on the value of the SMV, capturing how a student's performance depends on his/her level of proficiency. Each OV is permitted to have its own conditional probability distribution given the SMV, as tasks may differentially measure the KSAs.

==

Bayesian Inference Networks

Bayes nets combine probability theory and graph theory to represent probabilistic relationships among variables. Bayes nets are so named because they support reasoning from any set of observations to any other variables (either latent or observable but not yet observed) in a network using algorithms that incorporate Bayes' theorem (Lauritzen & Spiegelhalter, 1988; Pearl, 1988). As a general modeling approach, Bayes nets focus on conditional probabilities in which the probability of one event is conditional on the probability of other events: in forecasting, for example, probabilities of tomorrow's weather given today's weather and climate patterns; in animal breeding, characteristics of offspring given characteristics of ancestors; in medical diagnosis, probabilities of syndromes given disease states and of test results given syndromes. In assessment, interest lies in item responses or features of performances given students' KSAs. Bayes nets can be used to structure relationships across large multivariate systems, allowing us, for example, to synthesize the results of many observed responses to support inferences about student thinking (Mislevy & Gitomer, 1996).

One way to represent these networks of variables and the resulting computations is with a graphical model (such as Figure 4) consisting of the following elements (Jensen, 1996):

- A set of variables, represented by ellipses or boxes and referred to as nodes. All the variables in the most widely used Bayes nets have a finite number of possible values, corresponding to a set of exhaustive and mutually exclusive states (e.g., the IP addressing example has five mutually exclusive LP levels that comprise all possible states of IP_Addressing as it is being modeled).
- A set of directed edges (represented by arrows) between nodes, indicating 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. In our example, IP_Addressing is the parent of Item A and Item B. The direction of edges is often determined by theory, such as disease states as the parents of symptoms or the status of some indicator at Time *k* as a parent of status at Time *k*+1.
- For each variable without parents (such as IP_Addressing), there is an initial probability distribution for that variable. This could be uninformative, as in Figure 2, or based on background knowledge about a group or individual.
- For each variable with parents (such as Item A in Figure 4), there is a set of conditional probability distributions corresponding to all possible combinations of the values of the parent variables (as in the rows of Table 2a).

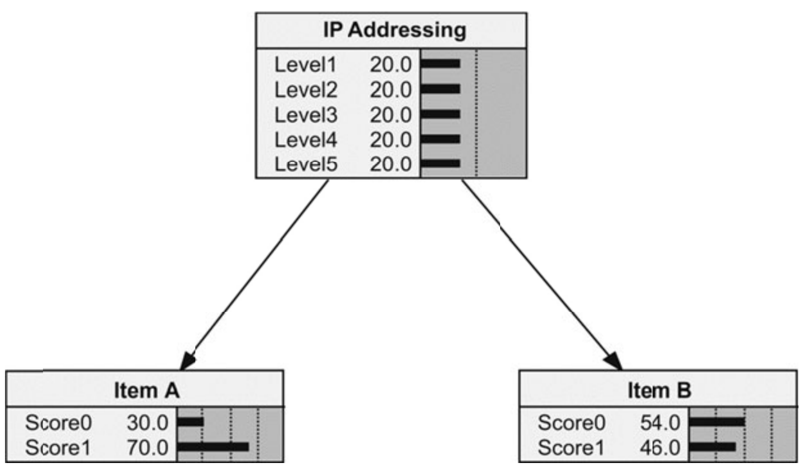

Figure 4. A Bayes net showing observable variables for two tasks. The performance is dependent on a student's level on the learning progression represented in the student model variable IP_Addressing.

Box 2 more formally describes Bayes nets as probability models. Because Bayes nets are framed in terms of the standard theory of probability and statistics, general approaches to model construction, parameter estimation, and model criticism are available to researchers seeking to model LPs or any other substantive situation.

==

Box 2: Bayes nets as probability models

In general, Bayes nets can be described as a probability model for the joint distribution of a set of finite-valued variables, say $(Y_1,\ldots Y_N)$, represented recursively in terms of the product of conditional distributions:

$$P(Y_1,\ldots,Y_N)=\prod_j P\left(Y_j \middle| Pa\left(Y_j\right)\right), \tag{1}$$

where $Pa(Y_j)$ refers to the subset of variables with indices lower than j upon which Y_j depends. These are the variables that have edges pointing from them to Y_j in the graphical representation of the network. Theory and experience suggest which variables should be considered parents of others. For example, in weather forecasting, variables for today's conditions are parents of variables for tomorrow's conditions. In genetics, variables representing genotypes of individuals are parents of variables representing the phenotypes of the same individuals, and variables for genotypes of literal parents are Bayes net parents of variables for the genotypes of their literal children. Theory and experience also provide information for determining if one set of variables should be modeled as independent from another set of variables, given the values of a third set of variables ("conditional independence"). For example, in Table 2, the probabilities of the Item A responses are independent of the probabilities of Item B responses given information about the levels of IP_Addressing; if we knew the value of a student's IP_Addressing variable, observing the value of the Item A response would not change our expectations for her response to Item B. When theory and experience suggest many conditional independence relationships, the variables in a Bayes net will have relatively few parents, and the diagram and the recursive expression simplify. The relationships among variables in the network can then be expressed in terms of interactions among relatively small clusters of variables.

Once such a representation has been built, one can update belief about any subset of the variables given information about any other subset using Bayes theorem. The rapid increase of the use of Bayes nets is due to efficient algorithms that allow these computations to take place in real time when the dependency structure is favorable.

In the context of Bayesian networks for LPs, θ is a latent discrete SMV with states that correspond to the levels of the progression. Formally, the values can be ordered, partially ordered, or unordered. A single LP would typically be represented by ordered levels. The OVs are discrete variables with states corresponding to the different possible scored performances on items or other tasks (e.g., a correct or incorrect response on an item, or levels or types of performance qualities in more complex tasks). The Bayes net specifies $P(X_j \mid \theta)$, a table of the conditional probabilities of observing different performances on tasks given the student's level on the LP. Multiple tasks yield OVs that may have different associated conditional probability tables. For example, one item may require a student to be at least at a low level of the progression in order to have a high probability of performing well, whereas another item requires the student to be at a higher level to have a high probability of performing well. In the case of the two items for IP_Addressing (Table 2), Item A requires a student to be at level 2 or above on IP_Addressing in order to have a high probability of getting a correct score, while Item B requires a student to be at level 4 or level 5 to have a high probability of answering correctly. The specification of the model is completed by defining an initial probability distribution for θ—i.e., a prior distribution—capturing how likely it is that a student is at each level of the progression. The prior may be uninformative, as in the introductory example (Figures 2 and 4), or based on other information such as student background data or instructors' expectations.

When a Bayes net is specified in this way to model assessment of a single discrete SMV, it can be viewed as a latent class model (Dayton & Macready, 2007; Lazarsfeld & Henry, 1968). A traditional formulation for a *C*-class latent class model (i.e., a model with *C* levels of a learning progression) specifies the probability that examinee *i* responds to item *j* yielding an OV value of *r* as

$$P(X_{ij} = r) = \sum_{c=1}^{C} P(\theta_i = c) P(X_{ij} = r \mid \theta_i = c), \tag{2}$$

where $P(\theta_i = c)$ is the prior probability that examinee *i* is in class *c* (i.e., level *c* of the progression) and $P(X_{ij} = r \mid \theta_i = c)$ is the conditional probability that an examinee in class *c* responds to item *j* in response category *r*. The usual restriction, $\sum_{c=1}^{C} P(\theta_i = c) = 1$, is imposed. Similarly, within latent classes the conditional probabilities over response categories are restricted such that $\sum_{r=1}^{R_j} P(X_{ij} = r \mid \theta_i = c) = 1$, where R_j is the number of distinct response categories for item *j*. The graphical representation contains edges from θ to each *X* (e.g., the edges from IP_Addressing to both items in Figure 4). The recursive representation is

$$P(X_1, \ldots, X_N, \theta) = \prod_j P\left(X_j \middle| Pa\left(X_j\right)\right) = \prod_j P\left(X_j \middle| \theta\right) P(\theta). \tag{3}$$

More complex cases can include multiple LPs as well as progressions that allow for different pathways so that the Bayes net must address a finer grain-size of KSAs to distinguish points along different pathways. In these cases, θ is vector-valued. Performance on a given observable X from a task can depend on more than one component of θ; that is, conditional probabilities for such an observable are estimated for possible combinations of its entire set of parent SMVs. In networking, for example, doing well on a certain troubleshooting task may require a student to be at Level 3 or higher in the IP_Addressing progression and at Level 2 or higher in the Connectivity (also called "Connect Networks") progression.

==

To continue with our example, the graphical representation of a Bayes net depicts the structure of relationships among variables—in our case, how performance on tasks depends on LP levels—and probabilities that represent the analyst's knowledge of a student at a given point in time. Probabilities that arise from (1) knowing nothing about a particular student's level in IP_Addressing and (2) knowing the conditional probabilities of item responses from Table 2 are shown in Figure 4. The direction of the arrows reflects the direction of the conditional probabilities in the tables, namely, that item performance depends on the student's level in IP_Addressing.

Once this probability structure has been built, we can reason in the other direction as well. We work back through an arrow to obtain a better estimate about an individual student's level on the LP, given the response the student makes to a given item, and then revise what we would expect to see on other items as a result. Figure 5 shows that if a student answers Item A incorrectly, he or she is probably at Level 1. The probabilities for IP_Addressing are obtained by applying Bayes' theorem as follows: Multiply the initial probabilities for each level of the LP (in this case, .2) by the corresponding conditional probabilities in the column for Score 0 of Table 2a; then normalize the result (i.e., rescale the results of the multiplications so they add to 100%). The result gives the posterior probabilities (IP_Addressing in Figure 5). These updated probabilities can then be used to obtain the probabilities for Item B. This updating is a simple example of Bayes' theorem with just two variables. In more complicated networks, algorithms are used that build on Bayes' theorem but take advantage of conditional independence structures to update many variables efficiently (Jensen, 1996).

Figure 6 shows that if the student answers item A correctly, s/he is probably at Level 2 or higher. Figure 7 shows that if the student who answers Item A correctly also answers Item B incorrectly, belief shifts to Levels 2 and Level 3. (The probabilities for IP_Addressing in Figure 4 have now been combined with the column for Score 0 in the Item B conditional probability table, Table 2b). If we wanted to sort out these possibilities, we would administer an item that focuses on capabilities that emerge in Level 3. Finally, if the student had answered B correctly, then our belief would shift to Level 4 and Level 5 (Figure 8).

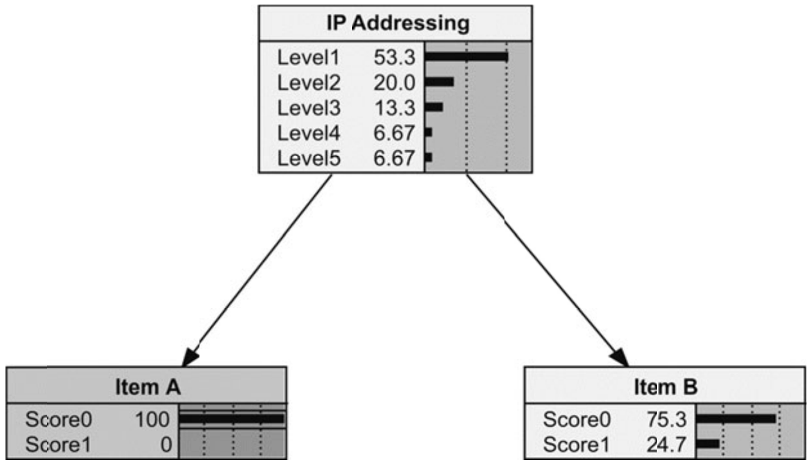

Figure 5. A Bayes net showing updated belief about IP_Addressing and expectations for Item B, after having observed an incorrect response to Item A.

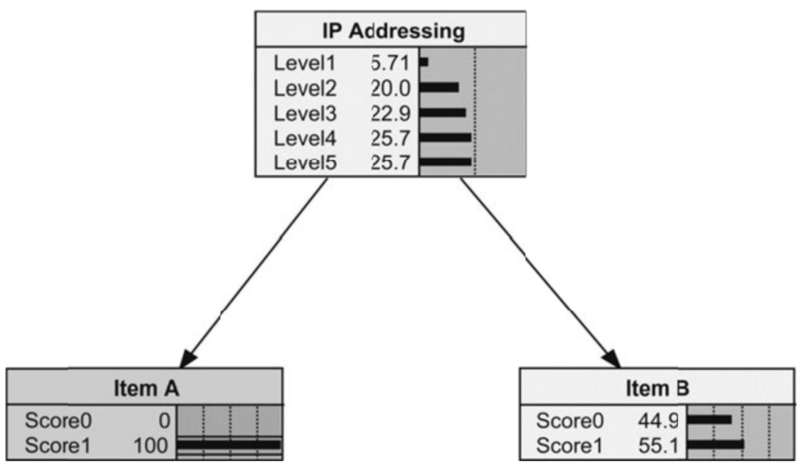

Figure 6. A Bayes net showing updated belief about IP_Addressing and expectations for Item B, after having observed a correct response to Item A.

We stated earlier that we would say more about the source of the numbers in the conditional probability matrices. At this point, we note the essential idea: We posit the basic structure of relationships (e.g., how LPs relate to each other and how LPs relate to tasks) and of the probabilities (e.g., where the jumps are) from our theory about learning and the way we construct tasks. We collect student responses to see how well theory fits data. If the model fits, we use the data to estimate the

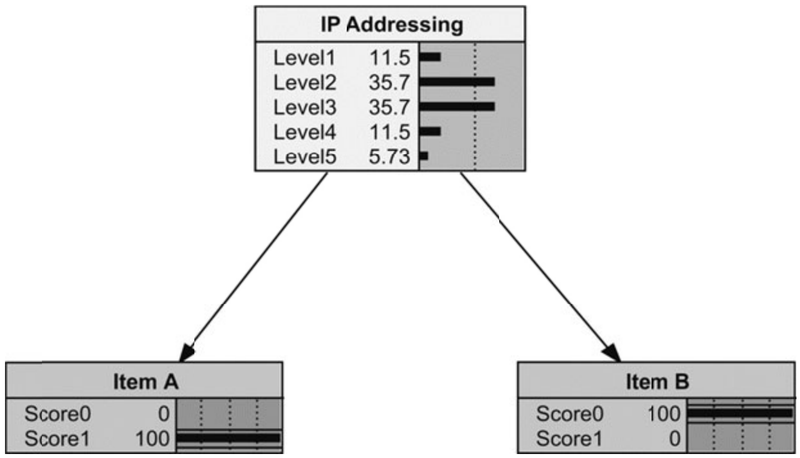

Figure 7. A Bayes net showing updated belief about IP_Addressing after having observed a correct response to Item A and an incorrect response to Item B.

probabilities. Bayesian model fitting and model-checking approaches such as described by Gelman, Carlin, Stern, and Rubin (1995) can be used. See Mislevy, Almond, Yan, and Steinberg (1999) and Levy and Mislevy (2004) for details on modeling and estimation approaches for fitting large conditional probability matrices for more complex Bayes nets. If the model doesn't fit, we use the data to revise the model, the theory, or the way we collect data (e.g., revision of tasks).

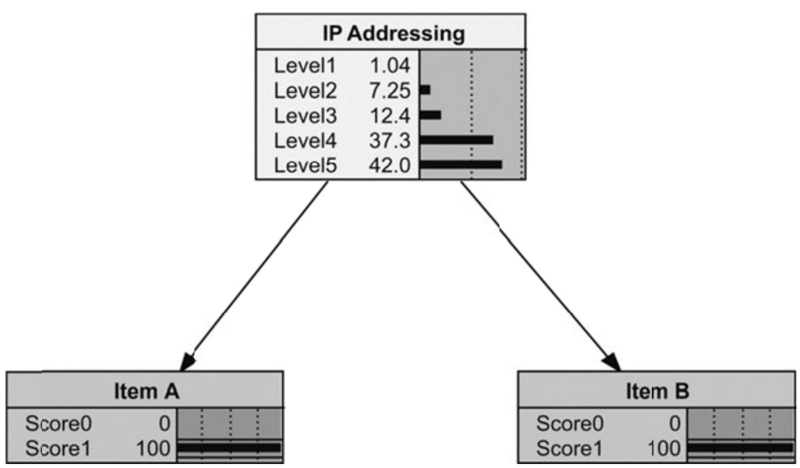

Figure 8. A Bayes net showing updated belief about IP_Addressing after having observed correct responses to both Item A and Item B.

CHALLENGES IN USING BAYES NETS TO MODEL LEARNING PROGRESSIONS

A number of challenges exist when using Bayes nets in modeling LPs and associated data from assessments targeting LPs. One challenge concerns the development of a toolkit of Bayes net techniques tuned for modeling assessment in the context of LPs. Bayes nets support probability-based inferences in complex networks of interdependent variables and are used in such diverse areas as forecasting, pedigree analysis, troubleshooting, expert systems, jurisprudence, intelligence analysis, and medical diagnosis (e.g., Pearl, 1988). When using Bayes nets in any new domain (like LPs), it is a challenge to develop an experiential base and modeling strategies using the general model-building, model-fitting, and model criticism tools of Bayes nets to address the relationships that are particular to that domain. For example, Bayes net fragments for properties of weather patterns and meteorological instruments can be assembled and tailored for weather forecasting applications (e.g., Edwards, 1998). Fragments concerning witness credibility and lines of argumentation also recur in Bayes nets in legal evidentiary arguments (e.g., Kadane & Schum, 1996). The unique features of LPs dictate that certain recurring structures will likely be present in applications of Bayes nets used to model LPs. We discuss three of these applications—(1) interrelationships among LPs; (2) KSA acquisition over time; and (3) evidence from complex tasks—which are likely to be part of sophisticated applications of modeling LPs.

A second, more local, challenge arises when one applies Bayes nets to model any specific substantive LP. In every application there are challenges in defining the LP, creating tasks, and iteratively fitting and improving the model and theory. In the NRC's (2001) terms, the cognition, observation, and interpretation components of an assessment must cohere. Bayes nets instantiate the last of these components. In and of themselves Bayes nets do not dictate the choices faced by researchers in any application, including the grain-size and number of levels in LPs. The definition and modeling of the middle levels of a LP present specific challenges in connecting cognition, observation, and interpretation.

There is a continual interplay between these two kinds of challenges. A Bayes net toolkit for LP research, at any stage of development, aids the analyst in all projects. Every project has its unique wrinkles, offers the possibility of insights about model structures or modeling strategies that may be more broadly useful for successive projects, and, as such, motivates expressing these new understandings in resources for the toolkit. Since Bayes net analysis of LPs is relatively new, we note in the following discussion the local challenges we faced. These challenges highlight recurring patterns that the field may expect to encounter more broadly in modeling LPs. For the interested reader, the Appendix gives details of an application of Bayes nets to modeling a learning progression in the CNA context.

Interrelationships Among LPs

The IP_Addressing example we discuss in this chapter concerns a single LP. In any complex domain, however, multiple KSAs must be developed not only with respect to sophistication in and of themselves but also in terms of their connections to other KSAs, and jointly as the basis for more integrated understandings. The knowledge maps in the *Atlas of Science Literacy* (American Association for the Advancement of Science [AAAS], 2001, 2007) suggest that such relationships are common and may become the object of study in LP research. This phenomenon occurs in the CNA curriculum. Therefore, we can use our experience to illustrate the broader challenge of modeling the interrelationships among LPs. The relationships we build in Bayes nets can, when schematized, be starting points for future researchers who tackle LP modeling challenges that resemble ours.

In learning computer network skills, the student goes beyond understanding isolated concepts to a synthesis of related KSAs. Figure 9 is a graphical representation of the KSAs required for computer networking in the CNA curriculum. It was created from discussions with subject matter experts and instructors in the curriculum. This is not a Bayes net. Rather, it is a kind of concept map that is similar to the maps in the *Atlas of Science Literacy* (AAAS, 2001, 2007). The map is one source of information we use in building LPs in the CNA domain and in building the Bayes nets for modeling them. The map suggests that a student's capability is directly related to some KSAs that are specific to particular networking devices and to other KSAs that are more general. For example, in Figure 9, IP_Addressing depends on Basic Math Skills and Understanding Local & Remote Network Concepts.

In a model of a domain consisting of multiple LPs, the structure and strength of the relationships among different LPs can be incorporated in a Bayes network. These relationships can be straightforward, such as when two LPs are correlated or when mastery of one LP is a prerequisite for mastery of a more advanced LP. Other relationships can be more complicated. For example, exploratory analysis in the CNA curriculum suggests that to master certain levels of the IP_Protocol_Rules progression, learners must be at a certain level of understanding in the IP_Addressing progression. It can be challenging to determine how to model the relationships between the LPs. While there are methods to learn the structure of a Bayesian network just from data, it is often useful to hypothesize the structure first and then use data to verify or to revise this model.

Using Bayes nets to model the hypothesized structure of multiple LPs, we structure the joint distribution among a set of LPs by constructing relationships among latent variables in a multivariate system. As previously discussed, under a Bayes net approach, each LP is represented as a discrete latent variable (node) with categories corresponding to different levels of KSAs in the LP. In the graphical representation, directed edges connect latent variables according to a model structure suggested by subject matter experts or exploratory analyses; for example, Figure 10 indicates that there is a dependence between IP_Addressing and IP_Protocol_Rules as discussed above. That is, the arrow from IP_Addressing to IP_Protocol_Rules indicates that the probabilities of the levels of the IP_Protocol_Rules SMV are different, depending on the level of IP_Addressing.

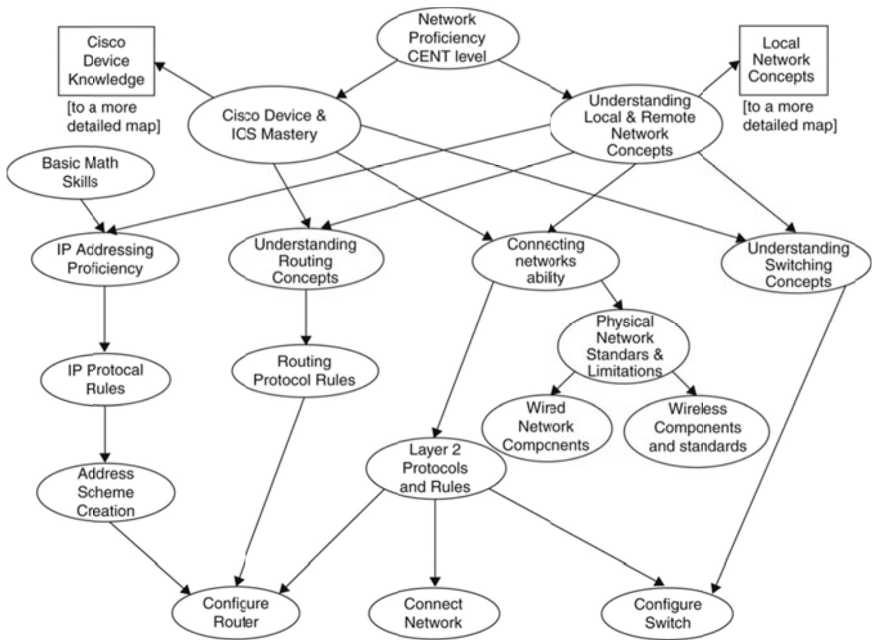

Figure 9. *Concept map of the CNA curriculum. This model displays the relationship between different networking KSAs of which IP Addressing Proficiency is a part. A Bayes Net could be constructed with a student model variable corresponding to each node. From A Bayesian Network Approach to Modeling Learning Progressions and Task Performances (CRESST Report 776) (p. 23), by P. West, D. W. Rutstein, R. J. Mislevy, J. Liu, Y. Choi, R. Levy, A. Crawford, K. E. DiCerbo, K. Chappel, and J. T. Behrens, 2010, Los Angeles, CA: University of California, CRESST. Copyright 2010 The Regents of the University of California. Reprinted with permission of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST).*

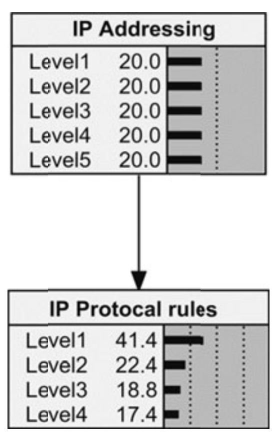

Figure 10. *Graphical representation indicating that there is a relationship between two LPs. The nature of the relationship is specified by the conditional probability matrix in Table 3.*

The edge in this graph only indicates that there is a relationship, not its nature or strength. This information is contained in the conditional probabilities. Table 3 shows one possible relationship. Reading conditional probability distributions across rows, we see that if a student is at Level 1 or Level 2 of IP_Addressing, there is a high probability the student will be at Level 1 of IP_Protocol_Rules. However, a student at Level 3, Level 4, or Level 5 of IP_Addressing has very similar probabilities of being at any level of IP_Protocol_Rules. The interpretation of this structure is that students at Level 1 or Level 2 of IP_Addressing are usually at Level 1 of IP_Protocol_Rules. For students at or above Level 3 in IP_Addressing, there is only a mild positive association between the two variables.

Table 3. A Conditional Probability Table with a Dependency Estimated from Data.

	IP_Protocol_Rules			
IP_Addressing	*Level 1*	*Level 2*	*Level 3*	*Level 4*
Level 1	0.72	0.14	0.08	0.06
Level 2	0.59	0.21	0.11	0.09
Level 3	0.32	0.28	0.22	0.18
Level 4	0.23	0.25	0.26	0.26
Level 5	0.21	0.24	0.27	0.28

Conditional probabilities can be determined based on data alone or in conjunction with constraints suggested by subject matter experts. Table 4 is very similar to Table 3, but Table 4 has theory-based constraints. The zeros in Table 4 imply that a person who is at Level 1 or Level 2 on IP_Addressing cannot be at a high level of IP_Protocol_Rules. That is, Level 3 of the IP_Addressing LP is a prerequisite for being at Level 3 of the IP_Protocol_Rules LP. Such a structure could be suggested by the substantive relationship between the KSAs at the levels of the two LPs. With data, a statistical test could be applied to test whether constraining the probabilities at the upper right of Table 4 to zero provides acceptable fit.

Table 4. A Conditional Probability Table with Constraints on Conditional Probabilities that Affect a Prerequisite Relationship.

	IP Protocol Rules			
IP_Addressing	*Level 1*	*Level 2*	*Level 3*	*Level 4*
Level 1	0.8	0.2	0*	0*
Level 2	0.7	0.3	0*	0*
Level 3	0.25	0.25	0.25	0.25
Level 4	0.25	0.25	0.25	0.25
Level 5	0.25	0.25	0.25	0.25

* Constrained value.

Multiple Time Points

Another challenge related to the Bayes net implementation of LPs is modeling change over time. As mentioned above, LPs can be characterized as measurable pathways that students may follow in building their knowledge and in gaining expertise over time. The Bayes nets we have discussed in this chapter have only addressed student status at a single point in time. Dynamic Bayes nets can be used to model student LPs over multiple time points where a student's level may change from one set of observations to another.[1] At each time point there are (a) one or more SMVs representing the LP(s) and (b) OVs with probabilistic dependence upon the SMVs. In addition, there is a copy of the SMVs for each time point. We model the relationship between unobservable LPs over time with conditional probability distributions that reflect transition probabilities. Transition probabilities indicate the probability of moving from a particular level at one measurement occasion to the other levels at the next measurement occasion.

Figure 11 shows an example of modeling LPs with a dynamic Bayes net. The Bayes net contains two parts: (1) four SMVs, which are actually the same LP but assessed at four successive time points where each measurement occasion modeled is dependent on the previous one, and (2) four OVs at each time point that are dependent on the SMV for that time point. Different patterns of transition matrices can be considered that depend on the developmental theory that grounds the LPs and on the students' experiences between measurement occasions. For example, the effectiveness of instructional treatments can be compared in terms of the transition probabilities they produce. Figure 11 depicts a situation in which observations have been made at all four time points. At each occasion, the results of four tasks were observed. This student was most likely at Level 1 of the SMV on the first occasion, at Level 2 on the second occasion, at Level 3 on the third occasion, and at Level 4 on the fourth occasion.

Complex Tasks

In the examples discussed thus far, each observable variable depends on only one SMV (i.e., LP). More complex tasks, however, may require jointly employing the KSAs that are modeled to reflect levels in more than one LP. Conducting an investigation in Mendelian inheritance, for example, may require KSAs from both a LP for the concepts in Mendelian genetics and the skills in a LP for proficiency in scientific inquiry.

In computer networking, students solving real-world network design and troubleshooting problems often encounter tasks that require them to draw upon multiple KSAs. Figure 9 suggests that assessing a student's capabilities in configuring a router involves the student's understanding of IP addressing and router concepts plus the student's ability to connect networks. While tasks can be defined to measure just one skill (and most of the multiple choice questions in end-of-chapter tests are so designed), in order to determine whether students can solve problems in real-world environments we must design tasks that require KSAs from multiple LPs.

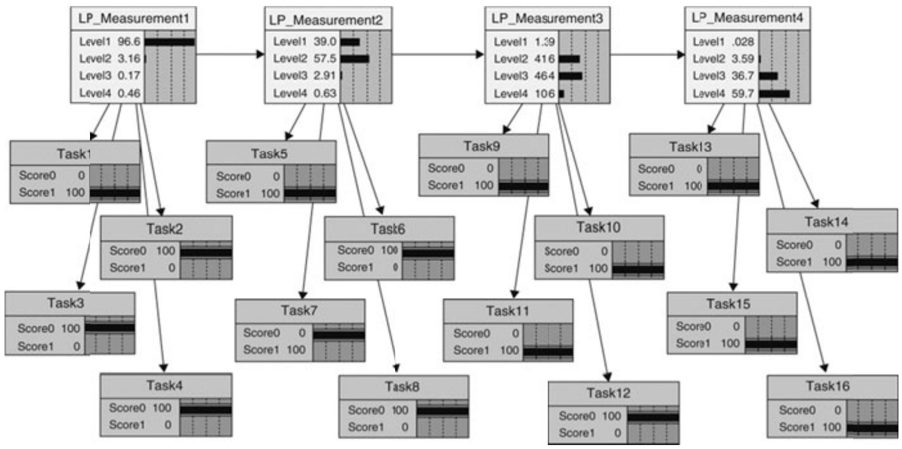

Figure 11. A dynamic Bayes net.

Cisco has developed a simulation environment for creating complex assessment tasks using a tool called Packet Tracer (Frezzo, Behrens, & Mislevy, 2009). Students interact with these tasks as they would with a real network environment. As in a real troubleshooting or network design problem, students must draw on KSAs from multiple LPs to complete these tasks. These relationships are incorporated in a Bayes net by having multiple SMV parents of an observable variable. Figure 12 shows a hypothetical example. Some tasks depend only on capabilities in the Connectivity LP, some depend only on capabilities in the IP_Addressing LP, and others depend on capabilities in both. Table 5 shows the conditional probability table for an observable variable for such a task, identified as ConAddTask1. Each row gives the probabilities of a right and wrong response (Score1 or Score0) assuming that a student is at a given combination of values for the SMV parents, Connectivity and IP_Addressing. The probabilities represent students' status on two LPs, namely, Connectivity with three levels and IP_Addressing with five. By choosing task features and performance requirements in accordance with the definitions of the LPs, the author of ConAddTask1 designed the task so that, to be likely to succeed, students need to be at Level 2 or higher on Connectivity and at Level 4 or higher on IP_Addressing. The pattern of conditional probabilities reflects this structure.

The challenges in modeling complex performance tasks involve determining not only how the relationships from the model are to be structured but also what type of observations are most appropriate for incorporation as evidence in the model. In Packet Tracer, the students' log files include a record of all commands that they have performed as well as a final network configuration that would resolve the problem in a troubleshooting task or meet the client's requirements in a configuration task. Currently the final configuration is evaluated by comparing it to

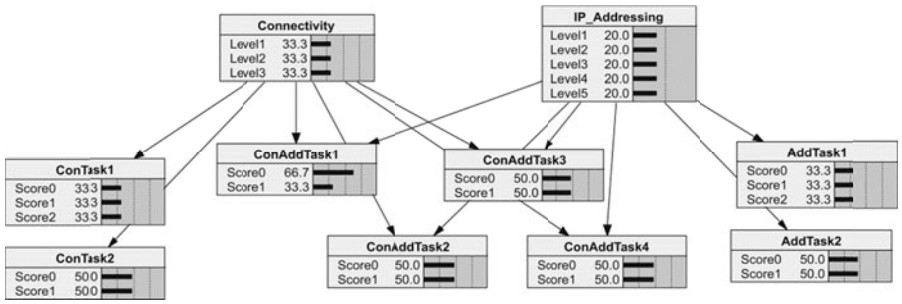

Figure 12. A graphical representation of a task dependent on multiple learning progressions. Note that some tasks depend strictly on Connectivity (ConTasks), some on IP_Addressing (AddTasks), and some on both (ConAddTasks). Adapted from A Bayesian Network Approach to Modeling Learning Progressions and Task Performances (CRESST Report 776) (p. 13), by P. West, D. W. Rutstein, R. J. Mislevy, J. Liu, Y. Choi, R. Levy, A. Crawford, K. E. DiCerbo, K. Chappel, and J. T. Behrens, 2010, Los Angeles, CA: University of California, CRESST. Copyright 2010 The Regents of the University of California. Reprinted with permission of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

a target configuration on the quality and features of a number of network aspects. Additional information is being collected, however, that can be used for evaluation. Earlier research with the NetPASS prototype (Williamson et al., 2004) used sequences and timing of actions recorded in the log files as evidence of student capabilities. We are currently studying whether other types of evidence can be used operationally.

Grain-size of a LP

The grain-size of a LP refers to how broadly or finely focused the progression of the described KSAs is. There are tradeoffs between modeling LPs at coarser or finer levels of detail; the choice depends on the purpose of the research or the assessment project at hand. Fine-grained LPs may be appropriate for studying changes in a narrowly defined concept, but a more broadly cast LP might be preferable for summarizing students' general levels of reasoning in a large-scale educational survey. The LPs we have worked with in CNA are built to help instructional designers, task developers, and teachers coordinate their work to define the curriculum and guide instruction at the level of lessons and exercises. The IP_Addressing LP shown in Table 1, for example, is at a medium grain-size. Each level is associated with a set of KSAs that are related to IP_Addressing. Students operating at a given level in this LP are likely to display most of the KSAs at lower levels and few of the KSAs at higher levels. However, more finely-grained LPs that focus more closely on one KSA in the set are possible—that is, IP_Addressing may be broken into a number of finer-grained LPs. On the other hand, a coarser LP may be defined by combining IP_Addressing with, say, Connectivity, Routing, and Switching for a broader description of students' KSAs. In this example, the grain-size was chosen for reasons related to subject matter and data.

Table 5. Conditional Probability Table for the Observable Variable ConAddTask1, Which Has Two SMV Parents (IP_Addressing and Connectivity).

Connectivity	*IP_Addressing*	*ConAddTask1*	
		Score 0	*Score 1*
Level 1	Level 1	90	10
Level 1	Level 2	90	10
Level 1	Level 3	90	10
Level 1	Level 4	90	10
Level 1	Level 5	90	10
Level 2	Level 1	90	10
Level 2	Level 2	90	10
Level 2	Level 3	90	10
Level 2	Level 4	20	80
Level 2	Level 5	20	80
Level 3	Level 1	90	10
Level 3	Level 2	90	10
Level 3	Level 3	90	10
Level 3	Level 4	20	80
Level 3	Level 5	20	80

In terms of subject matter, the groupings reflected in the IP_Addressing LP are based on clusters of related concepts that are taught and practiced together as variations on a "key idea" that is addressed in instruction and built on in subsequent levels. Two data-driven lines of reasoning influenced our choice of grain-size: (1) analyses of existing test data and subsequent identification of patterns of stability in that data and (2) variation in performance across two different organizations of the CNA curriculum.

Analysis of end-of-chapter test data revealed items with similar difficulties in terms of statistics and clustering of students in accordance with latent classes that represented those who "got the idea" and those who did not—usually one central concept, sometimes two, in a chapter. We conducted exploratory analyses using unconstrained latent class models (see Haertel, 1989) to identify structures that may suggest portions of LPs. These exploratory analyses and additional latent class analyses revealed dependencies across chapters that reflect curriculum developers' beliefs that certain concepts build on others. Tracking these dependencies revealed linear progressions of concepts across chapters that formed a LP, such as IP_Addressing. There was instructional value in defining a LP at this grain-size because the central theme in a given LP level (as discussed in connection with the IP_Addressing example) could account empirically for a cluster of related KSAs addressed in the chapter and the associated learning exercises. We also found cases in which knowledge at a given level of one LP was necessary for advancing to a given level of a different LP.

We gained further insights by comparing results across different presentations of material. Different classes present information in different sequences. Our analyses are still underway, but it appears that the patterns of performance in different courses can be understood in terms of the different orders in which the LP levels are addressed. In other words, modeling at a coarser grain-size would produce a very "messy middle" because after, say, two courses, students would have very different performance profiles. Modeling at the medium grain-size allows us to understand the middle in terms of different profiles across the same set of LPs.

The "Messy Middle"

Another challenge related to the definition of LPs may arise when modeling middle levels of proficiency. It is typically easiest to define the endpoints of a LP, where the lowest level refers to a novice state and the highest level refers to an expert state. In the simple LP described in the beginning of this chapter, learning generally proceeds as the successive attainment of KSAs in a single order, as shown in Table 1. In such cases, it is possible to define a LP in terms of ordered levels of a single SMV (as shown in Figure 13a).

It is more challenging to model the intermediate levels in the LP, however, when there are multiple pathways a student may follow in acquiring the KSAs associated with the various levels. To illustrate some of these possibilities, Figure 13b depicts alternative structures for the sequencing of KSA acquisition. In each sequence, the nodes represent different KSAs associated with the LP. Note that KSA 1 and KSA 5, at the beginning and the end of the sequences, represent the lowest and highest endpoints of the LP, respectively. The sequence on the left represents the acquisition of KSAs 1–5 in a particular order: students acquire KSA 1, followed by KSA 2, followed by KSA 3, followed by KSA 4, and finally KSA 5. The sequence in the middle offers a similar structure in which KSAs are acquired in an ordered fashion although the order differs. The sequence on the right depicts a different structure in which students acquire KSA 1 and then KSA 2. They then may acquire either KSA 3 or KSA 4, both of which must be acquired before KSA 5.

As Figure 13 illustrates, numerous patterns of KSA acquisition are possible. It is often unclear which sequence or pattern holds, or, as may be possible, if students experience different sequences of KSA acquisition. The difficulty in defining a single sequence that applies to all students, or of enumerating all the sequences that students experience—to say nothing of identifying which sequence students progress along—is what is referred to as the "messy middle" (Gotwals & Songer, 2010, p. 277). Approaches for modeling multiple sequences in the "messy middle" can be found in the psychometric literature on diagnostic and classification models (e.g., Haertel & Wiley, 1993; Leighton & Gierl, 2007; Rupp & Templin, 2008; Tatsuoka, 2002). These approaches can be expressed in Bayes net structures by extending the ideas discussed in the previous section.

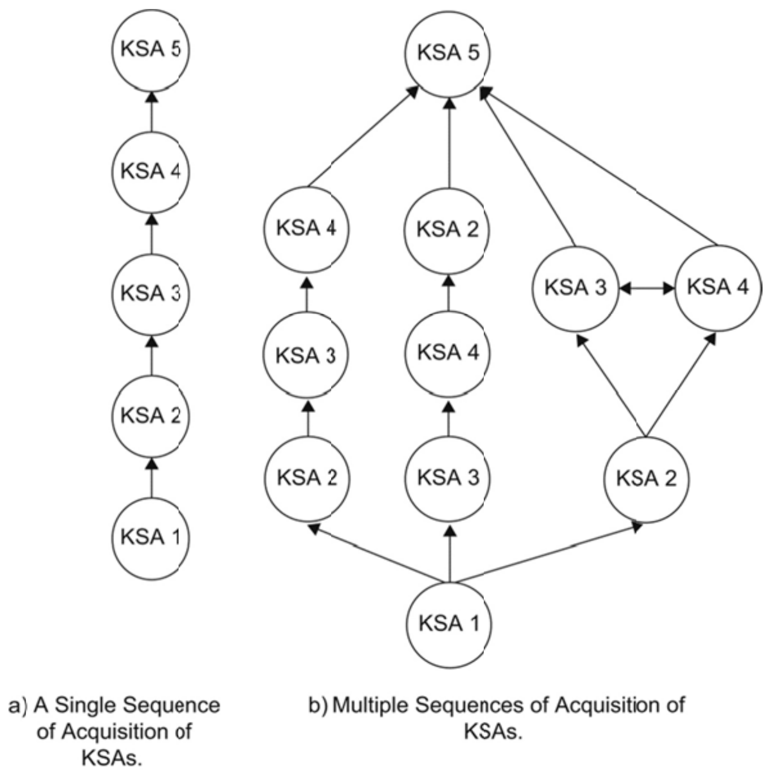

Figure 13. Single and multiple sequences of acquisition of KSAs.

In particular, a Bayes net for the multiple path scenario can be written in terms of multiple nodes. Figure 14 shows a Bayes net structure that accommodates all the paths in Figure 13b, using SMVs with two values for each KSA. This structure posits that, in students' development, the acquisition of KSA 1 tends to occur first and the acquisition of KSA 5 last. However, different orderings of KSAs 2, 3, and 4 may occur along the way. The conditional probability matrices corresponding to the arrows indicate if paths are common. If a single ordering is required, say, for reporting purposes, it is possible that the state with KSA 1 only is at a low level of a LP; any state with KSA 1 and KSA 2, 3, or 4 is at a second level; any state with KSA 1 and two of KSA 2, 3, and 4 is at a third level; a state with KSAs 1–4 is at a fourth level; and a state with five KSAs is at the highest level. According to this definition, students at the same broad level of the LP could have different profiles of probabilities for tasks, depending on their states within levels.

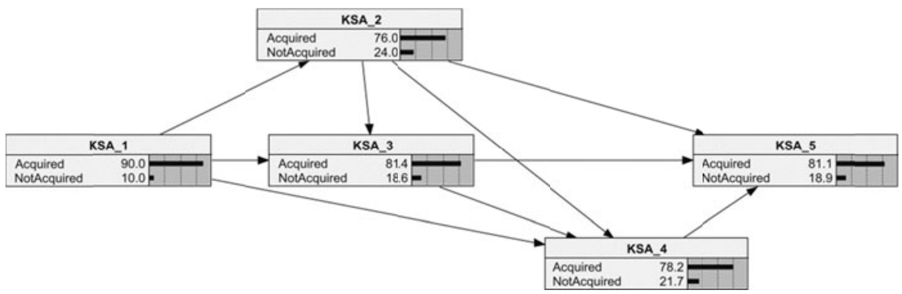

Figure 14. Bayes net for LP with multiple sequences of acquisition of KSAs. A student could progress from KSA_1 to KSA_2, 3, or 4 and then through any of the various paths indicated by the arrows.

DISCUSSION

Benefits of Using Bayes Nets to Model LPs

Why use Bayes nets to model LPs? One strength of the Bayes net approach is that it builds on a structure that can be based on theory and expert opinion; initially, conditional probabilities can be based on expert opinion and fragmentary data. This means a network can be used immediately for low-stakes inferences. As data accumulate, however, the probability-based reasoning foundations allow for coherent improvement of the estimates and more powerful investigations of model fit and subsequent model improvement. A particular advantage of Bayes nets is the flexibility in modeling complex and often subtle evidentiary relationships (Schum, 1994) among LP levels and between LPs and task performances. This is especially important if multiple interrelated progressions are considered and if KSAs from multiple progressions are involved in complex tasks or multiple aspects of a performance are evaluated (Mislevy, 1994; Mislevy & Gitomer, 1996). Estimates for students' LP levels can be updated in real time. Fragments of Bayes nets can be pre-built and assembled on the fly in response to an evolving situation, which is important in adaptive testing and in games- and simulation-based assessment (Almond & Mislevy, 1999; Conati, Gertner, & VanLehn, 2002; Martin & VanLehn, 1993; Shute, Ventura, Bauer, & Zapata-Rivera, 2009).

Although the work of Bayes nets is focused on modeling student performance in light of LPs, Bayes nets also make LPs more useful in some practical and important ways. Using Bayes nets to model LPs can lead to efficient and valid task design. The process of identifying initial LPs focuses test developers on the theory of cognition in the domain and defines the characteristics of individuals at various levels of the LP. Bayes nets confirm these levels and progressions, allowing designers to specify the levels of KSAs at which they aim assessment items. For example, it is possible to state that a task is designed to target Level 3 of one progression and requires knowledge at Level 2 of another progression. This helps make task design more principled, more planful, and ultimately more valid. Bayes nets also help connect curriculum to assessment. For example, curriculum designers can use information from a Bayes net structure to make decisions about which content areas to emphasize so that students have a greater probability of mastering future KSAs (DiCerbo & Behrens, 2008).

An area we continue to explore is how Bayes nets can provide feedback to students and instructors (DiCerbo, 2009). Such feedback could be achieved in at least two ways. First, students could receive reports that update their estimated levels on various KSAs, given their assessment performance. Based on these reports, students could be directed to other activities. This is the idea behind intelligent tutors or, when wrapped in a “fun” scenario, behind games (Shute et al., 2009). Teachers can use the structure of Bayes nets the same way that the curriculum designers (mentioned above) do when making decisions about content emphasis. In addition, teachers can diagnose student problems. For example, if a student is struggling in one area, teachers can look backwards to a network of variables to see what prerequisite KSAs the student probably lacks.

Challenges for the Community

Model-building is, by nature, iterative. Progressions are hypothesized, and models are built, based on understandings of the substantive area and data at a given point in time. As discussed above, building a Bayes net for a particular application involves encoding the relationships and making hypotheses of interest from the domain in the Bayes net. The Bayes net is fit to the data, and data-model fit and related model-checking tools are used to identify strengths and weaknesses of the Bayes net in terms of overall features, subsets of variables, or subsets of examinees (Levy, 2006; Sinharay, 2006; Sinharay & Almond, 2007; Williamson, Mislevy, & Almond, 2001). The results of these analyses have several interpretations. In a statistical sense, adequate data-model fit indicates that the probabilistic relationships in the Bayes net account for what actually takes place in terms of the data at hand, whereas data-model misfit indicates the relationships in the Bayes net do not reflect what actually takes place. More substantively, because the Bayes net is explicitly built to reflect domain-specific hypotheses, adequate data-model fit constitutes support for those hypotheses, whereas data-model misfit constitutes evidence against the hypotheses. Data-model misfit might indicate that some approximations or choices made in the model are not precise enough, or that certain relationships are poorly understood, or that the hypotheses and relationships hold for certain students but not for others.

As noted above, Bayes nets are flexible statistical models applicable to a wide variety of problems. Assessment and assessment in the context of LPs constitute just a few of these problems. The unique features of assessment of LPs, however, dictate that certain recurring features of the model are likely present in applications of Bayes nets to LP assessments. At present, the development of Bayes nets to accommodate these aspects is in its relative infancy. Similarly, related aspects of modeling need to be tuned to the particular features of assessment in the context of

LPs. For example, efficient data-model fit procedures to evaluate hypotheses about sequences of KSA acquisition through the "messy middle" need to be developed.

As discussed above, new challenges arise in every application in which a researcher models a LP in a particular substantive area. A comprehensive approach to assessment for LPs develops Bayes nets in concert with the specification of the desired inferences and tasks. This process, which is often iterative, is always localized to the specific situation as defined by the purpose of the assessment. Any serious application of Bayes nets involves the interplay between the methodological tools of Bayes nets and the substantive expertise required to build appropriate model approximations for the domain.

NOTE

[1] Because the LP variables are unobservable, the resulting Bayes net is formally a hidden Markov model (Cappé, Moulines, & Rydén, 2005; Langeheine &Van de Pol, 2002).

REFERENCES

Adams, R., Wilson, M., & Wang, W.-C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21,* 1–23.

Almond, R. G., & Mislevy, R. J. (1999). Graphical models and computerized adaptive testing. *Applied Psychological Measurement, 23,* 223–237.

American Association for the Advancement of Science. (2001, 2007). *Atlas of science literacy* (Vols. 1–2). Washington, DC: Author.

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 & 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. Baker, J. Dickieson, W. Wulfeck, & H. F. O'Neil (Eds.), *Assessment of problem solving using simulations* (pp. 59–80). New York, NY: Routledge.

Cappé, O., Moulines, E., & Rydén, T. (2005). *Inference in hidden Markov models*. New York, NY: Springer.

Conati, C., Gertner, A., & VanLehn, K. (2002). Using Bayesian networks to manage uncertainty in student modeling. *User modeling and user-adapted interaction, 12*, 371–417.

Corcoran, T., Mosher, F. A., & Rogat, A. (2009, May). *Learning progressions in science: An evidence-based approach to reform* (CPRE Research Report #RR-63). Philadelphia, PA: Consortium for Policy Research in Education.

Dayton, C. M., & Macready, G. B. (2007). Latent class analysis in psychometrics. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics* (pp. 421–446). Cambridge, MA: Elsevier.

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.

DiCerbo, K. E. (2009, April). *Communicating with instructors about complex data analysis*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

DiCerbo, K. E., & Behrens, J. T. (2008, March). *Identifying keystone concepts: Big ideas in computer networking*. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Edwards, W. (1998). Hailfinder: Tools for and experiences with Bayesian normative modeling. *American Psychologist, 53,* 416–428.

Frezzo, D. C., Behrens, J. T., & Mislevy, R. J. (2009). Design patterns for learning and assessment: Facilitating the introduction of a complex simulation-based learning environment into a community of instructors. *The Journal of Science Education and Technology, 19*, 105–114.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. London, UK: Chapman & Hall.

Gotwals, A. W., & Songer, N. B. (2010). Reasoning up and down a food chain: Using an assessment framework to investigate students' middle knowledge. *Science Education, 94*, 259–281.

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement test items. *Journal of Educational Measurement, 26,* 301–321.

Haertel, E. H., & Wiley, D. E. (1993). Representations of ability structures: Implications for testing. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 359–384). Hillsdale, NJ: Erlbaum.

Ingham, A. M., & Gilbert, J. K. (1991). The use of analogue models by students of chemistry at higher education level. *International Journal of Science Education, 13*, 193–202.

Jensen, F. V. (1996). *Introduction to Bayesian networks*. New York, NY: Springer.

Kadane, J. B., & Schum, D. A. (1996). *A probabilistic analysis of the Sacco and Vanzetti evidence.* New York, NY: Wiley.

Langeheine, R., & Van de Pol, F. (2002). Latent Markov chains. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 304–341). Cambridge, UK: Cambridge University Press.

Lauritzen, S., & Spiegelhalter, D. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, 50*, 157–224.

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

Leighton, J. P., & Gierl, M. J. (Eds.). (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge, UK: Cambridge University Press.

Levy, R. (2006). *Posterior predictive model checking for multidimensionality in item response theory and Bayesian networks.* (Doctoral dissertation, University of Maryland). Retrieved from http://drum.lib.umd.edu/bitstream/1903/3483/1/umi-umd-3311.pdf

Levy, R., & Mislevy, R. J. (2004). *Specifying and refining a measurement model for a simulation-based assessment* (CSE Technical Report 619). Retrieved from University of California Los Angeles, National Center for Research on Evaluation, Standards, & Student Testing website: http://www.cse.ucla.edu/products/reports/r619.pdf

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

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

Martin, J. D., & VanLehn, K. (1993). OLEA: Progress toward a multi-activity, Bayesian student modeler. In S. P. Brna, S. Ohlsson, & H. Pain (Eds.), *Artificial intelligence in education: Proceedings of AI-ED 93* (pp. 410–417). Charlottesville, VA: Association for the Advancement of Computing in Education.

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

Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika, 59,* 439–483.

Mislevy, R. J., Almond, R. G., Yan, D., & Steinberg, L. S. (1999). Bayes nets in educational assessment: Where do the numbers come from? In K. B. Laskey & H. Prade (Eds.), *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence* (pp. 437–446). San Francisco, CA: Morgan Kaufmann.

Mislevy, R. J., & Gitomer, D. (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., & 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). Washington, D.C.: National Academy Press.

Muthén, L. K., & Muthén, B. O. (1998–2006). *Mplus User's Guide* (4th ed.). Los Angeles, CA: Author.

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K-8.* Washington, DC: The National Academies Press.

Norsys Software Corp. (2007). *Netica manual*. Retrieved from 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*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Rupp, A. A., & Templin, J. L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement: Interdisciplinary Research and Perspectives, 6*, 219–262.

Schum, D.A. (1994). *The evidential foundations of probabilistic reasoning.* New York, NY: Wiley.

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

Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 295–321). Philadelphia, PA: Routledge.

Sinharay, S. (2006). Model diagnostics for Bayesian networks. *Journal of Educational and Behavioral Statistics, 31*, 1–34.

Sinharay, S., & Almond, R. G. (2007). Assessing fit of cognitive diagnostic models: A case study. *Educational and Psychological Measurement, 67,* 239–257.

Suarez, M. (2004). An inferential conception of scientific representation. *Philosophy of Science, 71*, S767–779.

Swoyer, C. (1991). Structural representation and surrogative reasoning. *Synthese, 87*, 449–508.

Tatsuoka, C. (2002). Data analytic methods for latent partially ordered classification models. *Applied Statistics, 51,* 337–350.

West, P., Rutstein, D. W., Mislevy, R. J., Liu, J., Levy, R., . . . Behrens, J. T. (2010, August). *A Bayesian network approach to modeling learning progressions.* (CRESST Research Report 776). Retrieved from University of California Los Angeles, National Center for Research on Evaluation, Standards, & Student Testing website: http://www.cse.ucla.edu/products/reports/R776.pdf

Williamson, D. M., Bauer, M., Steinberg, L. S., Mislevy, R. J., Behrens, J. T., & DeMark, S. (2004). Design rationale for a complex performance assessment. *International Journal of Measurement, 4,* 303–332.

Williamson, D. M., Mislevy, R. J., & Almond, R. A. (2001, April). *Model criticism of Bayesian networks with latent variables*. Paper presented at the annual meeting of the National Council on Measurement in Education, Seattle, WA.

APPENDIX

An Example of Building and Fitting a Bayesian Network for a Learning Progression

This example demonstrates how assessment data can be used to help validate a learning progression (LP) using statistical modeling in the form of a Bayes net. The data are from scored responses to 35 items written to target specified levels of the IP_Addressing progression. The hypothesized structure is a Bayes net with these 35 items as conditionally independent observable variables, dependent on a single discrete latent variable with values that indicate LP levels.

Owing to the connection between a Bayes net of this description and latent class analysis (Box 2), a series of latent class analyses were conducted using the poLCA package (Linzer & Lewis, 2007) in R (R Development Core Team, 2008). These were exploratory analyses that did not constrain the solution to finding the theoretical levels that motivated the item writers or to yielding conditional probabilities that reflected jumps that accorded to levels. Rather, they were unconstrained latent class analyses for 2-Class, 3-Class, 4-Class, 5-Class, and 6-Class solutions. Furthermore, it was not required that classes be ordered or that the conditional probability matrices for items would show jumps at levels targeted by the item writers. The structure that emerged would be driven by patterns in the data. As the ensuing discussion shows, the latent class structure that emerged empirically closely reflected the theoretical structure of LP levels and conditional probabilities for items with jumps at intended levels.

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. The four classes identified in the analysis corresponded to increasing levels of performance on the items and were interpretable as increasing levels of KSAs. A hypothesized further distinction at the high end of the LP was not realized due to the small number of items targeted at this level. In other words, there was insufficient information in the data set to differentiate students at the two highest theorized levels. A Bayes net representation of a model with a single SMV containing four levels (classes) was then constructed in Netica (Norsys Software Corp., 2007), represented in Figure A1.

INFERENCES REGARDING ASSESSMENT ITEMS

An item was classified as "at the level" of a certain class if it supported an interpretation that students at that level would be able to solve or complete the task,

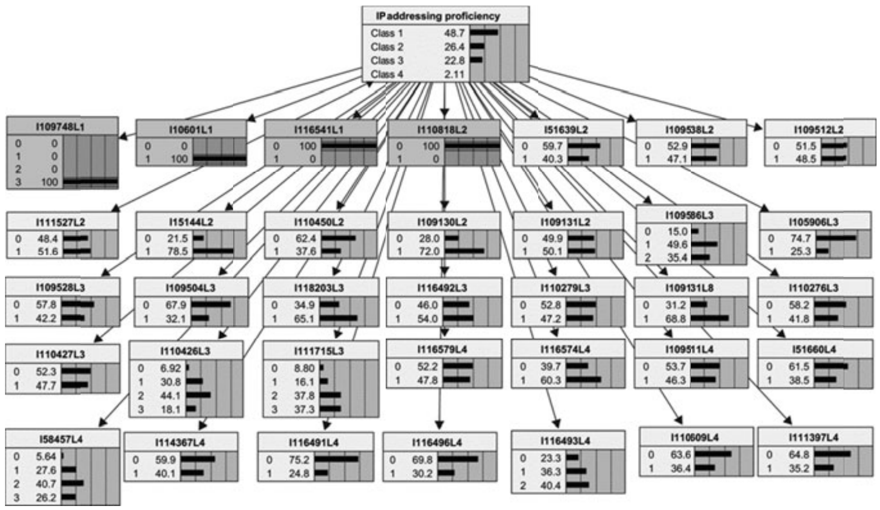

Figure A1. Graphical representation of the Bayes net. Adapted from A Bayesian Network Approach to Modeling Learning Progressions and Task Performances (CRESST Report 776) (p. 17), by P. West, D. W. Rutstein, R. J. Mislevy, J. Liu, Y. Choi, R. Levy, A. Crawford, K. E. DiCerbo, K. Chappel, and J. T. Behrens, 2010, Los Angeles, CA: University of California, CRESST. Copyright 2010 The Regents of the University of California. Reprinted with permission of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

whereas students at lower levels would likely be unsuccessful. 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 (our interpretation of the lowest class emerging from the analysis—a class with low conditional probabilities of correct responses to all items) 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 indicated that many items discriminated strongly between classes. For example, Figure A2 contains the conditional probability table for an item that only students in the fourth (highest) class are likely to solve successfully. Statistically, this item aids in distinguishing students in the fourth latent class from those in the remaining classes. Substantively, the item exemplifies one aspect of what it means to be at the fourth level of the LP. Students at the fourth level have acquired the KSAs necessary to have a strong chance of correctly answering this item; students at the lower levels have not.

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

Figure A2. Conditional probability table for an item that discriminates well between the fourth and the other classes. From A Bayesian Network Approach to Modeling Learning Progressions and Task Performances (CRESST Report 776) (p. 19), by P. West, D. W. Rutstein, R. J. Mislevy, J. Liu, Y. Choi, R. Levy, A. Crawford, K. E. DiCerbo, K. Chappel, and J. T. Behrens, 2010, Los Angeles, CA: University of California, CRESST. Copyright 2010 The Regents of the University of California. Reprinted with permission of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Other items were more ambiguous in terms of their levels. For example, Figure A3 shows the conditional probability table for an item on which students in the second class have a .88 probability of 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.

Across all items, the results were largely consistent with the experts' expectations. Ten items exhibited clear and distinct patterns that distinguished between classes exactly as predicted by experts. That is, these items were "located" at the expected level. Figure A2 shows an example of one such item; the expert prediction of this item as a level 4 item is strongly supported by the empirical results.

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

Figure A3. Conditional probability table for an item that discriminates differentially at multiple points. From A Bayesian Network Approach to Modeling Learning Progressions and Task Performances (CRESST Report 776) (p. 19), by P. West, D. W. Rutstein, R. J. Mislevy, J. Liu, Y. Choi, R. Levy, A. Crawford, K. E. DiCerbo, K. Chappel, and J. T. Behrens, 2010, Los Angeles, CA: University of California, CRESST. Copyright 2010 The Regents of the University of California. Reprinted with permission of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Five items were scored polytomously; these distinguished roughly well at the level predicted by experts. This is seen in terms of differential probabilities between the targeted LP levels at one score level and the two other LP levels at another score level. This phenomenon is illustrated for the item whose conditional probabilities appear in Figure A3. This item was expected to be a level 4 item. This item is located at class 2 with respect to being able to obtain a score of 1 as opposed to 0; it is also is located at class 3 in terms of being able to obtain a score of 2 as opposed to 1.

Overall, eighteen items were located at a level adjacent to the predicted level (e.g., an item expected at level 4 was located at class 3). One item was located adjacent to the predicted class and was also located at another class not adjacent. Only one item was clearly located at a class that was not equal to or adjacent to the predicted level. Initial reviews of these results indicated revisions that would help the items more sharply target the concepts at their intended levels.

INFERENCES REGARDING STUDENTS

The conditional probability tables also reveal how inferences regarding students are conducted in the Bayes net. For example, observing a correct response for the item in Figure A2 is strong evidence that the student is in class 4; observing an incorrect response for the item in Figure A2 is relatively strong evidence that the student is not in class 4. The use of a Bayes net 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 algorithms based on Bayes' theorem to yield posterior distributions for the remaining unknown variables (Pearl, 1988), including the SMV corresponding to the LP. For example, Figure A1 contains the Bayes net for a student who has completed four items. The student correctly answered the first two items and incorrectly answered the next two items. On the basis of this evidence, the posterior distribution for his/her latent skill variable indicates that this student has a probability of being in classes 1–4 of .487, .264, .228, and .021, respectively. On this basis we may infer that the student is almost certainly in one of the first three classes (i.e., is at one of the first three levels of the progression) and is more likely in the first class than either the second or third. Yet there still remains considerable uncertainty. The collection and inclusion of more data would lead to a more refined inference.

COMMENT ON THE EXAMPLE

The results of the modeling offer a data-based interpretation of the development of KSAs that constitute the LP. In some cases, the results for items confirm the experts' expectations. For other items, the results are more ambiguous or offer an alternative to the experts' expectations. To take a more comprehensive perspective on assessment of LPs, the results of the statistical analyses will be submitted to the subject matter experts for consultation and possible refinements in terms of the definition of the LP, the items that assess the aspects of the LP, and the utility of additional items for modeling students' progression.