DOCUMENT RESUME

ED 477 928		TM 035 024
AUTHOR	Williamson, David M.; Bauer, Malcolm; Mislevy, Robert J.; Behrens, John T.	
TITLE	An ECD Approach to Designing for Reusability in Innovative Assessment.	
PUB DATE	2003-04-00	
NOTE	52p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 21-25, 2003).	
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)	
EDRS PRICE	EDRS Price MF01/PC03 Plus Postage.	
DESCRIPTORS	*Educational Innovation; *Research Design; *Test Construction	
IDENTIFIERS	*Evidence; *Usability	

ABSTRACT

The current pace of technological advance has provided an unprecedented opportunity to use innovative simulated tasks in computerized assessment. A primary challenge for the successful use of innovation in assessment rests with the application of sound principles of design to produce a valid assessment. An additional challenge is to maximize the utility from the investment in innovative design through leveraging successful innovation to new assessment tasks and new educational tools. This paper describes the Evidence Centered Design (ECD)approach to design of an innovative simulation-based assessment of computer networking ability. The paper emphasizes the design components and how these components may be leveraged for reusability in a variety of ways, including the generation of new assessment tasks, applications to alternative purposes within the domain of computer networking, or as a basis for extending knowledge of proficiencies needed for performance in the domain, and the extension of these assessment design components to related domains of interest.(Contains 2 tables, 9 figures, and 28 references.) (Author/SLD)

Running Head: ECD DESIGN FOR REUSABILITY

An ECD Approach to Designing for Reusability in Innovative Assessment

David M. Williamson

Malcolm Bauer

Educational Testing Service

Robert J. Mislevy

University of Maryland

John T. Behrens

Cisco Systems

Correspondence should be addressed to:

David M. Williamson, Educational Testing Service, Rosedale Road, Princeton, NJ 08541

Abstract

The current pace of technological advance has provided an unprecedented opportunity to utilize innovative simulated tasks in computerized assessment. A primary challenge for the successful use of innovation in assessment rests with the application of sound principles of design to produce a valid assessment. An additional challenge is to maximize the utility from the investment in innovative design through leveraging successful innovation to new assessment tasks and new educational tools. This paper describes the Evidence Centered Design (ECD) approach to design of an innovative simulation-based assessment of computer networking ability. The paper emphasizes the design components and how these components may be leveraged for reusability in a variety of ways, including for the generation of new assessment tasks, applications to alternative purposes with the domain of computer networking, as a basis for extending knowledge of proficiencies needed for performance in the domain, and the extension of these assessment design components to related domains of interest.

An ECD Approach to Designing for Reusability in Innovative Assessment

INTRODUCTION

The study and development of innovative assessment tasks is an ongoing endeavor for educational measurement professionals. This desire is motivated by a commonly held belief that traditional assessment techniques, particularly those based solely on multiple-choice items, are sub-optimal for the measurement of certain complex domains (e.g. medical diagnosis), some integrated knowledge, skills and abilities (KSAs) of interest, complex problem solving ability, highly performance-oriented abilities (e.g. troubleshooting mechanical systems), and other "higher-order" skills and abilities. To address these needs there is a history of leveraging emerging technologies for educational assessment. The technological advances of the past decade have brought an unprecedented capacity for using video, sound, and other multimedia technologies in large scale assessment through the internet. While recent technological advances have enabled a vast new range of potential innovations in assessment tasks, this is a necessary but not sufficient condition for effective implementation of innovative assessments. Innovation is accompanied by associated risks that must also be considered in the design process. These risks include the risk to psychometric principles of valid and fair measurement from innovation and, equally important, the financial risks of implementing and maintaining an innovative assessment design given the complexities and development costs of such endeavors. This paper describes an application of Evidence Centered Assessment Design (ECD) to the design and implementation of an innovative assessment for computer networking ability with the goals of mitigating the risks of innovation through explicit evidential reasoning to maintain psychometric standards and designing for reusability.

The technology for innovation in large-scale assessment currently exists, but the risks of such innovation remain a concern. The potential for innovation in the kinds of activities that people perform as part of computerized assessment is largely self-evident given the rapid advances in computer processing speed, graphics capabilities, video presentation, and data transmission capabilities with modern desktop computers. The potential for large scale internet administration of tasks with these multimedia components, and for capturing a plethora of actions, processes and outcomes from user interactions, including audio and image data, from remote users is a potentially powerful tool for innovative design. Technology now exists for assessments to incorporate sophisticated and realistic simulations, which require examinees to draw upon a wide range of relevant knowledge, skills, and abilities as they solve meaningful tasks from the domain. A good simulation system, however, is not the same as a good assessment system (Melnick, 1996). To be effective as an assessment, a simulation system must be able to evoke, record and interpret observable evidence about targeted knowledge in a way that is principled, defensible, and suited to the purpose at hand (e.g., licensure, achievement testing, coached practice). It is not an effective strategy to design a sophisticated simulation system and interesting tasks, and only then ask "How do you score it?" In order to meet professional measurement standards when developing innovative assessment tasks the foundation for sound assessment should be laid at the beginning and the principles of assessment design should guide every decision throughout the development process—tasks, scoring, psychometrics, simulators—so that the many elements perform in harmony to best serve the assessment purpose. The initial risk in innovative assessment concerns this ability to develop a strong validity rationale regardless of the complexities introduced by the innovative approach to assessment.

Risks of innovation in assessment design extend beyond the threats to measurement standards from complexities inherent in innovative design rationale. The cost of developing many types of innovative assessment tasks, particularly complex and realistic simulations, can be substantial. While the benefits can be equally substantial, there is risk in such an investment of resources for a static assessment tool, used for a single purpose, that may become outdated or over-exposed. To reinvest an additional, perhaps substantial, amount in redevelopment is undesirable when an assessment design incorporating intent for long-term model-based assessment as well as leveraging the modeled components for other educational tools can be a preferable, and more cost effective, alternative. Therefore, the need for cost-effective design of innovative assessment is a critical feature for the long-term success of such assessments and reusability of design components is a key part of this strategy.

Until recently, the grounding in sound design principles and corresponding methods for developing, administering, scoring, and maintaining innovative assessments was not sufficiently developed to assure that resultant assessments would meet professional standards. Evidence Centered Assessment Design (Mislevy, Steinberg, & Almond, 2002; in press) leverages knowledge of cognition in the domain and sound design principles to provide a robust framework for designing assessments, be they simple and familiar or complex and innovative. The application of ECD to simple designs provides a rigorous and re-usable assessment blueprint, which facilitates developing assessments that meet professional measurement standards. For complex and innovative measurement needs, the ECD framework provides for assessment design that maintains an evidentiary focus, to guide the professional through the complexities of innovative design. This framework permits professionals to devote their efforts to the central design issues while avoiding unnecessary complexity and extraneous content (e.g. complete real-world fidelity, glitzy technology, etc.). Through this framework ECD approaches mitigate the two primary risks of innovation in assessment design: maintaining sound measurement practice (i.e. construct validity); and maximizing the utility of innovative design through reusability.

REUSABILITY

The decision to take an innovative approach to assessment is a choice of the assessment designer, but it is a choice based on the needs of the assessment and the nature of the domain that the assessment is designed to serve. In most cases, innovative approaches to assessment are more expensive to develop, implement and maintain than traditional multiple-choice counterparts. The increased expense of such assessment and the corresponding risks that accompany innovation demand a degree of effort from test designers to maximize their probability of a successful, valid assessment and to make the most of successful innovation. A key component to the long-term success of such innovations is the ability to design for reuse.

When discussing reusability it is useful to distinguish among a variety of potential types of reuse, each with it's own characteristics and implications for the assessment design. These different targets of reusability include:

- *Extensibility of Task.* This target of reusability focuses on the ability to use the design elements from innovative tasks in new tasks, either to refresh the task pool, develop alternative forms, or to modify the design of tasks to improve their ability to provide evidence of examinee proficiencies. Such reusable characteristics for innovative tasks are dependent on the nature of the tasks themselves but include such examples as the interface with which the examinee interacts, the nature of the tools available to the examinee to resolve the task, and the task characteristics that determine level of complexity and difficulty and the necessary KSAs to complete it successfully. In many ways this effort for reusability of innovative task design components is parallel to automatic item generation efforts (e.g. Bejar, 2002; Irvine & Kyllonen, 2002; Mislevy, Steinberg & Almond, 2002, in press), sometimes referred to as item cloning, being undertaken on the basis of explicit task modeling at targeted levels of abstraction.
- *Extensibility of Purpose.* This form of reusability targets the reuse of the understanding of proficiency in a domain to serve new purposes different from those initially targeted in the assessment design. A readily accessible example of such a repurposing might be to begin with an assessment initially targeted for placement of students along a unidemnsional continuum of 5th

grade mathematical ability. Let us assume that there is an existing documented multidimensional structure of proficiency underlying the model (e.g. the documented understanding that the summary of ability on a unidimensional continuum is based on explicit representation of a variety of mathematical sub-abilities, such as addition, subtraction, multiplication, tables, graphs, etc.). An example of this reuse would be the repurposing this existing model of proficiency to develop an educationally diagnostic assessment that targets the reliable reporting of ability in the various areas of mathematical work that make up the 5th grade curriculum. A less familiar example of extensibility of purpose might be the reuse of a proficiency model underlying an intelligent tutoring system to serve an assessment used for certification or licensure. The key component of extensibility of purpose is that the understanding of what makes up proficiency in the domain remains unmodified and is reused in an assessment that is then used for a different purpose than the original model was designed to serve, even if some of the characteristics of the original purpose persist in the new application.

- *Extensibility of Proficiency*. The nature of reusability implied by this type of focus is that the reuse is directed at the same domain of interest as the original assessment, but the understanding of proficiency in the domain is altered. This situation is very similar to extensibility of purpose with the exception that the model of proficiency in the domain is extended and/or modified, usually through supplementary analysis of the domain, in order to address a different areas of interest in the domain. An example of such a situation might be the converse of the previously described scenario; when a certification assessment for a domain is used as the basis for an intelligent tutoring system. If we assume that the understanding of proficiency in the domain must be revisited and extended in order to serve this new purpose then the original understanding of domain proficiency serves as the extensible foundation for this effort. As a result of these additional efforts, which might include such activities as cognitive task analyses and curriculum analyses, a new understanding of proficiency emerges that is based on the one originally serving the certification assessment.
- *Extensibility of Domain*. This form of reusability emphasizes leveraging the existing understanding of proficiency and task construction for one domain to be applied in a new domain. Of course, the extent of generalizability of these elements will depend in part on the degree of

similarity between domains (e.g. we might expect an assessment structure from a calculus assessment to have greater generalizability to an engineering assessment than to an accounting assessment). However, there can often be a greater degree of generalizability among domain proficiencies than might be initial suspected, particularly in cases where the emphasis of the initial assessment is on cognitive and strategic elements of performance in addition to specific domain knowledge. An example of such a similarity between apparently incompatible domains (in this case, between computer networking and aircraft hydraulics maintenance) is provided later in this paper.

As the above discussion suggests, there is something of a hierarchy among these levels of targeted extensibility, with each subsequent type of extensibility tending to subsume the types that come before it. This paper will describe some of the basic components of the ECD process with an emphasis on the reusability of design components at these various levels of extensibility. This extensibility will be illustrated using examples from the design of NetPass, an assessment of computer networking proficiency designed for Cisco Learning Institute's (CLI) Cisco Networking Academy Program (CNAP), that incorporates innovative computer simulations. We begin with a brief overview of ECD followed by a general description of the ECD process applied for NetPass. Once these ECD design models are laid out we explore the potential for reusability of these elements and the computerized tools that facilitate the documentation, sharing and reuse of design elements.

EVIDENCE CENTERED ASSESSMENT DESIGN

The inherent complexity of innovative assessment for complex domains demands meticulous care in assessment, particularly in the design of simulations to evoke behavior that provides targeted evidence about key skills and knowledge. Furthermore, the design must provide for principled interpretations of elicited behavior and how it constitutes evidence that suit the purpose of the assessment. Psychometric models for such complex assessments are feasible only after substantial integrative design work prior to the implementation of the statistical model. The ECD framework ensures that appropriate evidence is gathered for assessment inferences, thus supporting the validity of those inferences by requiring explicit connections between the claims to be made about students, the evidence supporting these claims, and tasks that elicit such evidence from students. Evidence Centered Assessment Design (Mislevy, Steinberg, & Almond, 2002, in press) emphasizes the explicit definition and full explication of design elements, including their relation to assessment purpose and other design components, in each step of the design process. As such, ECD is a systematic approach to design that requires consideration and explicit definition of measurement constructs and evidence that would support the types of inferences to be made from the assessment (Mislevy, 1994). Through this systematic approach to design the resultant design structures lend themselves to reuse and redesign as needed for the ongoing life of the assessment and for adaptation in other related educational assessment tools. It is this capability for reuse of the assessment design structures, in part or in whole, that provide an advantage in mitigating the costs and other associated risks of innovative assessment design. As is true with most systematic design methodologies, ECD requires greater initial effort, but yields greater ability to understand the constructs being measured and the impact that changes in assessment requirements and designs will have on subsequent design and development stages (see Steinberg, Mislevy, Almond, Baird, Cahallan, Chernick, DiBello, Kindfield, Senturk, & Yan, 2000) as well as producing a re-usable blueprint for assessment design.

In simplified form, the ECD process consists of an interdependent series of assessment design steps, with subsequent steps continuing to influence the outcomes of previous work and vice versa. As in any complex problem of design under constraints, the process is iterative and cyclical, with results and insights from any part of the work able to prompt us to rethink or to revise other parts. While we can present the issues being addressed in a linear order, we certainly can't expect to address them in such strict linear sequence! These key questions for assessment design exploration and elaboration can be expressed as:

- Why are we assessing? (**Purpose Definition**)
- What will be said, done, or predicted on the basis of the assessment results? (**Claims** and **Prospective Score Report**)
- What portions of a field of study or practice does the assessment serve? (**Domain Analysis**)
- Which knowledge and proficiencies are relevant to the field of study or practice? (**Domain Model**)
- Which knowledge or proficiencies will be assessed? (**Student Model**)
- What behaviors would indicate levels of proficiency? (**Evidence Models**)
- How can assessment tasks be contrived to elicit behavior that discriminates among levels of knowledge and proficiency? (**Task Models**)
- How will the assessment be conducted and at what point will sufficient evidence be obtained? (**Assembly Models**)
- What will the assessment look like? (**Presentation Models**)

- How will the assessment be implemented? **(Delivery System Model)**

An evolving assessment design iteratively cycles through these questions to provide a set of answers in terms of design elements, consistent with one another, coherent as an expression of an assessment argument, and able to address an assessment's purpose within its constraints. Designing assessments in concert with this conceptual approach and implementing them in a common infrastructure provides several practical benefits. It provides a roadmap for innovation by systematically addressing unfamiliar constructs, data, purposes, task types, and content areas. It makes explicit the choices and the accompanying rationale of good test development practices and in the process lays out the validity argument that underlies assessment or other educational tools. Finally, it promotes the re-use of assessment elements and processes for a multitude of potential variations on the initial design. The following sections walk through some of the decisions and examples of the outcomes from the application of these ECD processes in the development of NetPass, and discussion of the potential for reusability of these design elements.

DOMAIN ANALYSIS

Under ECD, the design process begins by specifying the purpose of the assessment to specify why the design process is being undertaken. The design team uses this understanding as the origin as the first stages of *domain analysis*, a systematic study of the domain to determine the critical knowledge, skills and abilities (KSAs) needed to perform domain tasks successfully and the key portions of a domain of practice that are targeted by the assessment. This first stage of assessment design fulfills several purposes: defining the scope and intent of the assessment in the domain context, documenting and delineating the specific claim-based utility of assessment results, and specifying the rationale and assumptions underlying these decisions. While these design components are repeatedly revisited and revised during the design and development process as a result of domain analysis activities, these decisions set the stage for documenting the claims to be made on the basis of assessment results. The explicit relationship between the claims to be made, the proficiencies that students must possess to meet these claims, the nature of data that provides sufficient evidence of these proficiencies, and the critical features of situations that can elicit such evidence, provides the evidential framework for the development of an assessment capable of producing valid and relevant results.

PURPOSE DEFINITION

The CLI collaborates with high schools, community colleges, and vocational schools to provide education on the fundamentals of computer networking. The CNAP is four-semester curriculum teaching the principles and practice of designing, implementing, and maintaining computer networks capable of supporting local, national, and global organizations. Instruction is provided in classrooms as well as through on-line curriculum and activities and assessments are likewise conducted through classroom exercises and on-line testing. The CNAP uses the World Wide Web for both instruction and assessment administration and data maintenance. World Wide Web usage facilitates global access to educational resources, with approximately 150,000 students in 60 countries participating in the CNAP and an average of 10,000 administrations of online assessments per day. This high-volume global access presents both opportunities and challenges for educational research. Computer networking demands considerable technical knowledge as well as strategic and procedural expertise to become accomplished at common networking tasks. Despite the importance of these domain abilities current web-administered assessments consist of multiple-choice items primarily assessing declarative knowledge. This fact prompted the lead CNAP curriculum designer, Dennis Frezzo (extracted with permission from *Assessment Pains*, an essay from a Cisco Learning Academy internal communication, May 2, 2000), to caution instructors not to depend on these assessments alone to evaluate students:

> *"[The current] online assessments ... are limited checks for understanding that will help the students get ready for that [certification] exam. But to produce students who can make real networks run, your assessment must be MUCH broader and deeper than any online assessment. ... Recall a primary goal of the program -- designing, installing, and maintaining networks. Quite frankly, the Assessment Server tests and [certification] test do not adequately test the complex problem-solving and manual set of skills required to maintain actual school networks. That is why the Instructor's Guide and Training model emphasize project-based, hands-on, lab-based, troubleshooting, "authentic", journal-and-portfolio-based assessments -- making cables, configuring routers and switches, building networks, wiring schools, all graded by rubrics."*

These comments reflect the widely held belief that the then-current on-line assessments were inadequate for determining student understanding of some of the most important aspects of networking ability. The absence of standardized assessment of critical elements of ability force a reliance on local evaluation efforts, which may be prone to substantial variability in curriculum emphasis and evaluation standards.

The limited scope of standardized assessment and the potential for substantial variability in student capabilities in critical areas of computer networking ability is being addressed by CLI, in partnership with ETS, through a redesign of the CNAP networking assessment program. This redesign leverages current simulation technology and remote connection capabilities to produce an innovative assessment administered online and exercising the *cognitive* aspects of network design, implementation, and troubleshooting. The initial outcome is a prototype assessment, called NetPass, using network simulations with realistic interactive tasks to measure students' abilities and provide targeted educational feedback. This feedback includes reporting on the students' knowledge of networking, their mastery of various networking skills, their ability to carry out procedures and strategies for networking tasks, and their misconceptions about network functionality and operational procedures.

The primary purpose of the NetPass prototype assessment is to provide a means for CNAP students to obtain a standardized evaluation of their networking skills, to complement current multiple-choice tests assessing declarative networking knowledge. This purpose implies assessment of strategy, efficiency and other aspects of operational computer networking performance that are difficult to capture with multiple-choice questions. The goal is to both assess networking proficiency and to provide educational benefits to the student through targeted diagnostic feedback from their performance. A primary goal of the assessment is to provide access to evaluation on a single CNAP standard for networking ability regardless of physical location of the student as well as instructionally relevant diagnostic feedback from highly monitored performance. This service is expected to be a substantial improvement over reliance on potentially varied standards provided by regional instruction alone.

As part of the specification of assessment purpose it was decided to limit the scope of NetPass to the most critical aspects of computer networking—network design, network implementation, and network troubleshooting—and to therefore exclude routine planning and maintenance from the prototype assessment. This explicit exclusion of portions of the domain from the targeted purpose of assessment is the first opportunity to explore implications for reusability of assessment design. The decision to eliminate certain aspects of the domain from the purpose of the assessment extended to two explicitly omitted types of domain activity: routine network maintenance activities and interactions with the physical layer of a network (e.g. physical checking and manipulation of cable connections, computer cards, etc.). By documenting the relationship of these excluded activities to other domain activities and the purpose of the assessment there exists a basis for revisiting and repurposing future tools by reviewing the existing documentation rather than being forced to repeat an analysis of the domain of practice to determine what was omitted from the scope of the current assessment design. This could be important both for new extensions of the assessment and for educational tools, for example, if the purpose was changed to include targeting the ability of students to understand and remediate faults in physical components of computer hardware. This stage also provides an opportunity to consider multiple purposes for the design process and to prioritize these purposes so that the design can be systematically targeted toward the initial priorities, yet be extensible to additional purposes as resources permit.

JOB ANALYSIS/TASK ANALYSIS

As part of the domain analysis process an investigation is undertaken to determine what the critical knowledge areas are and the key domain activities are that discriminate among the domain ability levels targeted by the purpose of the assessment. In this process extensive use is made of all domain relevant resources, including domain documentation such as curriculum materials, training manuals, job performance evaluation criteria, prior research in the domain and related domains, manuals, subject matter experts (SMEs) both from academic and practitioner settings, domain task performance observations, and original investigations of domain performance. In many ways this process of collecting domain relevant information and extracting from these sources the critical KSAs and domain tasks to be performed is parallel to many activities commonly used in job analysis and task analysis activities that are a routine part of development of test specifications in certification and licensure applications. Indeed, common approaches to job and task analyses are a means of building information that contribute to the domain analysis process.

For the NetPass project this process required multiple sources of information, including CNAP curricular material, subject matter experts (practitioners and instructors), existing assessments, computer networking documentation, educational texts and courseware, and prior research in related fields. As the four-semester CNAP curriculum was developed from a similar process focusing on success as a computer networking practitioner it was not surprising to find that the resultant characteristics of the critical domain KSAs and key tasks to be performed in computer networking overlapped substantially with the existing CNAP curriculum.

ASSESSMENT CLAIMS

On the basis of a full specification of the assessment purpose and an ongoing investigation of the domain the design team can begin specifying domain-relevant *claims* to be made about students on the basis of the assessment, which specify what will be said, done, or predicted on the basis of assessment results. Claims are specific statements about the examinee's knowledge, skills, or abilities to be made on the basis of observed evidence from the assessment. These explicit purpose-related claims guide development of the assessment to ensure that the observable evidence provided by the assessment tasks is both relevant and sufficient to support such claims. In particular, the claims provide the foundation for the identification and specification of proficiencies students must possess, and that the assessment must elicit evidence of, in order to make valid and relevant inferences about the student. The full list of domain-based claims is developed through an iterative process of claim addition, culling, and hierarchical reorganization. The hierarchical structure can reflect choices in degree of specificity with which claims are established and ultimately, can be reflected in the hierarchy of the student model for the assessment.

In service to NetPass specific claims were developed through domain research about the knowledge, skills and abilities necessary for computer networking performance at the ability levels targeted (on the basis of the previous delineation of assessment purpose) for the assessment. These were organized in a claim hierarchy and used to drive the rest of the assessment design process. For example, three claims that emerged from the NetPass analysis were:

1. Students can use a systematic approach to identify and solve network problems
2. Students can identify the cause of connectivity problems at the physical, data link, and network layers of the OSI model
3. Students can use TCP/IP utilities to troubleshoot network connectivity problems.

The first claim is a general statement about troubleshooting ability while the second claim is a subclaim specifying a component ability needed for the first claim, in this case the ability to troubleshoot a common network symptom, namely, loss of connectivity. The third claim is in turn a subclaim of the second claim and is one of the subskills needed to have the ability described in the second claim. The full set of claims in the NetPass design is a hierarchical tree with four tiers, starting with a high level statement about students' networking abilities and branching down to more specific abilities and subskills. Ultimately, these hierarchical claims are partially represented in the final form of the student model for the assessment and provided explicit targets guiding the design process delineating a sufficient body of supporting evidence.

The resultant purpose statements, domain claims, and documented domain analysis can be reused in a variety of ways, including expansions or reductions in the breadth of these areas (e.g. rank-ordering or placement based on subsets of the domain, such as for classroom workgroups), that is, in the *nature* of claims to be made, as well as in the depth of representation (e.g. a more fine-grained representation of claims for diagnostic feedback or for guiding interventions of an intelligent tutoring system), that is, the *grain size* of the claims being made. By leveraging the work that is required for these initial ECD activities the potential exists to have a representation of the purpose hierarchy, claim structure, and domain analysis that is extensible to such additional products as diagnostic tools, intelligent tutoring systems, or for providing links to educational interventions through references to training materials and related self-directed educational exercises.

DOMAIN MODEL

On the basis of, and typically overlapping with some portions of, the initial domain analysis a *domain model* is constructed that specifies the knowledge and proficiencies that are relevant to the field of study or practice. The domain model represents the next step in the chain of evidential reasoning for the assessment, with the purpose of assessment guiding the selection and organization of domain-based claims and these claims, in turn, suggesting the KSAs that learners or practitioners in the domain must possess in order to perform as various levels. This initial structure of characteristics critical for performance in a domain constitute the domain model and is constructed without impingement from known constraints about what is feasible or possible to reliably measure in the assessment. The domain model identifies a catalog of KSAs as well as *knowledge representations* in the domain, which are observable characteristics of domain performance that can contain indicators of domain ability. With a draft model of proficiency and an indication of the tasks that are performed in the domain, and the observable characteristics of these tasks that may contain evidence of domain ability, the assessment designer is prepared to continue on to the next stage of development.

While for many domains and purposes of assessment it may be sufficient to build the understanding of domain performance purely as a model of knowledge acquisition for the computer networking domain there are important cognitive components of performance, particularly with regard to strategies and efficiencies of effort, that are important parts of proficiency in the domain. On such occasions when constructing a model of proficiency in the domain extends into cognitive and/or information processing components the domain modeling process can add a cognitive task analysis (CTA) to the series of domain investigations upon which the domain model is based.

COGNITIVE TASK ANALYSIS

A CTA is a disciplined process of investigating the knowledge structures and strategies that individuals at targeted levels of ability use to solve specific types of tasks, and observable evidence of those structures and strategies (e.g. Steinberg & Gitomer, 1993). While a job analysis is typically concerned with the frequency and importance of domain task performance, a CTA focuses on identifying knowledge and strategies people use to address those tasks. A CTA seeks to expose (a) essential features of task situations for eliciting certain behaviors; (b) internal representations of task situations; (c) the relationship between problem-solving behavior and internal representation; (d) processes used to solve problems; and (e) task characteristics that impact problem-solving processes and task difficulty (Newell & Simon, 1972).

With the ultimate objective of designing optimal assessment tasks, CTA methods were adapted from expertise literature (Ericsson & Smith, 1991) to capture and to analyze the performance of CNA students at different known levels of expertise, under standard conditions, across a range of tasks. CTA conducted in the service of assessment design will focus on skills, behaviors, knowledge structures and cognitive strategies that are directly related to the assessment purpose, rather than on the general understanding of cognition in a domain that characterizes CTA in cognitive science. The preceding steps in domain analysis for computer networking establish a fundamental understanding of the domain and the intent of the assessment, thereby directing the focus of the CTA on specific areas of task performance most critical to the assessment claims, and thus improving the efficiency of the CTA in support of the assessment purpose.

In order to conduct the CTA the information from earlier steps in the Domain Analysis were used to draft a representation of the general understanding of the domain proficiencies, the nature of observable work products that are produced when solving domain tasks, and preliminary hypotheses about how characteristics of this observable behavior might constitute evidence of domain proficiency. Each of these steps is driven by the goals of the assessment and the objective of being able to support the claims targeted by the assessment results. A goal of the CTA is to further facilitate the ability to document the relationship between the evidence observed from examinee work and the claims to be made about the examinee in order to strengthen the validity and utility of the assessment for its intended purpose. As assessment design continues both the claims and the evidential requirements are revisited and revised to reflect the dynamic elements of the design process (e.g. subsequent task development, etc.) and to ensure that the evidence being provided by the assessment is sufficient to support each claim.

In preparation for the CTA for the NetPass assessment, on the basis of established claims, the design team worked with SMEs to develop descriptions of ways knowledge is represented in the domain and the features of these representations that can provide evidence about the knowledge and skills possessed. An example of the types of representations corresponding to Claim 2 (discussed above) and a table expressing how features of those representations provide evidence for Claim 2 is provided as Table 1. In Table 1 a *network diagram* is a common way for networking professionals and students to describe the structure of a network and it displays network devices, connections between network devices, and network address information. A *log file* consists of the sequence of network commands used to configure or troubleshoot a network. A *configuration file* describes the state of variables associated with a network device (typically a router, for which IP addresses and other characteristics may be specified).

The design team, again with SMEs, identified the features of work products that provide evidence regarding Claim 2, which is provided as Table 2. Scoring features, identifiable characteristics of an examinee solution that can be used as evidence to support claims, are provided in the first column. The second column provides the general category containing the feature and the third column provides the link to representations in Table 1 that are capable of representing the feature. For example, the second feature, "identification of network problems", provides critical evidence for the claim because identifying the specific network problem (a bad network address, wrong communication protocol, etc.) is essentially identifying the cause of the loss of network connectivity. This evidence would be contained in multiple representations: the log file; configuration file; and worksheet. The log file would record the commands the student entered while conducting troubleshooting to investigate and fix the connectivity problem. Similarly, a configuration file would provide the subsequent configuration of the network after the student's interventions during troubleshooting. Finally, a worksheet could record faults identified by the student. As such, Tables 1 and 2 describe the evidence that can be gleaned from different representations in support of specific claims.

Explicating the relationship between claims and different types of evidence contained in various representations also involved specifying the necessary characteristics of domain task situations. Documented task characteristics describe the necessary task elements that enable a student to employ targeted knowledge and skill, and, given these situations, the ways he or she could act to display that knowledge. The components of the computer network, their associated representational systems, and the beginning and ending network states were particularly important in the specification of task performance situations.

Consideration of the necessary characteristics of domain tasks includes addressing the method of administering such tasks, since this can impact the nature of evidence collected. Particular administration requirements were thus derived from the nature of claims and evidence for NetPass and the goal of assessing student ability to interact effectively with a computer network while addressing complex networking tasks. NetPass required networking tasks that, if not performed on an actual computer network, must have an interactive computer interface which functions with a high degree of fidelity to an interface for a real computer network. Additionally, on-line web-based delivery was chosen as the administration vehicle in order to provide a universal assessment standard in a practical and readily accessible manner for remote learning locations.

Design decisions such as these have the effect of both targeting the primary goals of the assessment and simultaneously precluding other potential design features. For example, by meeting the purpose of assessing computer network interaction skills through widely accessible internet administration, the assessment must exclude a direct assessment of the student's ability to examine the physical layer of the network (e.g. to check that network cables are connected during troubleshooting activities, to physically install network cards in computers, etc.). As proficiency in physical layer operations was not a high priority goal of the assessment, this was an acceptable constraint on the assessment design. Such tradeoffs illustrate how the relationship among claims, the evidence that support those claims, and the characteristics of tasks and administration characteristics must be repeatedly revisited and reexamined during the design process. This work was a precursor for the cognitive task analysis, which provided the level of detail needed to fully define the remaining elements of the assessment design.

The CTA for this project was designed to (a) tap the knowledge and strategies used by CNAP students of various ability when designing, implementing, and troubleshooting networks, and (b) identify observable behaviors that manifest this knowledge and strategy at various levels of proficiency. The CTA was designed to flesh out the assessment design through further identification of knowledge and skills required for network design, implementation, and troubleshooting, with respect to the assessment purpose of low-stakes (learning) assessment with supplementary feedback. The assessment claims most relevant to the third semester of CLI curriculum, the primary target of NetPass, represented the focus of the CTA. These claims cover three skill areas: network troubleshooting, network configuration, and virtual local area network (VLAN) design (for a more complete description of the CTA for this project see Bauer, Williamson, Steinberg, Mislevy & Behrens, 2001; Williamson, Bauer, Mislevy, & Behrens, 2003). In this targeting of skill areas four primary computer networking activities were targeted for the task analysis:

- *Network troubleshooting*. The student is introduced to an existing network, with specified properties and meant to perform in predetermined ways. User reports of certain failures are provided. It is the student's task to determine the fault(s) and fix them.

- *Network configuration*. The student is presented with the design specification of a network. The student configures a provided pod of routers to reflect the design specification.
- *VLAN design*. The student is presented with user requirements and constraints for designing a local or wide area network involving a VLAN. The student develops a design to satisfy these requirements and constraints.
- *Network Diagram*. In association with each of the above three activities, the student produces a diagram of the structure of the network, the relationships and connections between network elements, and the network functionality for the network they designed, implemented, or troubleshot. This activity was required for each scenario based on the domain analysis, which indicated that the understanding of physical systems and their relationships in a computer network is a critical component of networking ability.

The CTA data consisted of transcripts of talk-aloud protocols during task performance solutions, researcher-assisted retrospective protocols of students recalling the rationales for their problem-solving processes, log files of all computer workstation commands during task completion, and diagrams and calculations produced as participants solved the tasks. This data was analyzed and discussed by a team of ten researchers, computer networking instructors, and subject matter experts with the intent of identifying recurring patterns that distinguished performance scenarios and ability levels. In this process prior design work was again called upon to suggest hypotheses for investigation in talk-aloud protocols and resultant log files and diagrams. These identified patterns served as the foundation for defining re-usable observed variables from assessment tasks for use in the evidence models.

CTA RESULTS

Troubleshooting Tasks

Examination and discussion of student troubleshooting protocols and associated log files ultimately resulted in the classification of various commands and sequences of commands into four major categories:

1) *Actions associated with gathering information about router configuration.* These commands provide information about the state of a router and network connectivity (e.g. the "show" command and its variants).

2) *Actions associated with changes to router configuration to fix faults.* These commands institute changes to a router (e.g. adding a clock signal, setting router protocol, setting IP addresses on interfaces, etc.).

3) *Actions associated with testing network behavior after fixes.* These commands also provide information about the state of the network but for post-fix testing these commands are typically directed at network connectivity (e.g. "ping" and "telnet") rather than provision of router configuration information (e.g. "show"). However, some of these commands can overlap with those described for information gathering above. Therefore, in establishing whether actions are information gathering or post-fix testing the prior fault-fixing actions must also be referenced.

4) *Actions associated with getting information about commands.* These commands do not address the network but instead access available information about commands or command targets (e.g. uses of the help system, "?" command).

Through this command classification structure it was possible to analyze troubleshooting activities with respect to the frequency, character, and appropriateness of actions based on these command classifications.

These characterizations and classification of commands facilitated subsequent SME examination of protocols and command patterns, ultimately resulting in identification of three patterns of solutions. It must be emphasized that particular students did not necessarily exhibit consistent patterns on all tasks, and that the patterns were correlated with, but not identical to, the instructor designated ability levels. That is, the following are descriptions of *behaviors*, not of *students*. The characteristics that differentiated the three performance patterns could be summarized by means of two main categories, which themselves are composed of a hierarchy of contributing components according to the following structure:

a) Correctness of Procedure
   i) Procedural Sequence Logic
      (1) Sequence of targets
      (2) Sequence of actions
   ii) Efficiency of Procedure
      (1) Help usage
      (2) IOS syntax

(3) Volume of actions

b) Correctness of Outcome

i) Error Identification

ii) Error Over-Identification

The typical patterns in troubleshooting solutions were the following:

- **Troubleshooting Pattern A.** The student followed unsystematic and inefficient troubleshooting procedures, rarely found and fixed faults, and used the help system extensively to guide command use.
- **Troubleshooting Pattern B.** The student found and fixed faults correctly and efficiently.
- **Troubleshooting Pattern C.** The Student followed standardized troubleshooting procedures rather than a sequence of procedures tailored to the network in question. These patterns of solutions fell between Patterns A and B in their overall success on the troubleshooting tasks. These patterns follow a more prescribed set of procedures and usually take more steps to isolate each problem than a Pattern B solution. The student attempted to be systematic in their testing of the network, but sometimes flailed to identify the fault(s) in the network, or did so through a long and circuitous series of actions. They correctly fixed some of the network faults, but they also fixed faults that were not actually present.

Design Tasks

Based on the examination of CTA results, several patterns of performance emerged which distinguish varying degrees of quality in Design solutions. The basic distinction among these design tasks was with the *Correctness of Outcome*, which focuses on whether the resulting network design is functional, correct, and complete.

In professional practice there are two aspects to Correctness of Outcome: the Functionality of Design, which is a measure of the extent to which the network serves its intended purpose, and the Efficiency of Design, which considers aspects which affect network performance, such as the number of components used, the cost of components, and the maintenance and performance implications of the selected components. Both aspects are important in professional practice, but since the vast majority of professional designs are functional they differ mainly in their efficiency. In contrast, efficiency is not a major factor discriminating among the student-generated designs in the relatively simple problems addressed in third-semester CNA curriculum; for the most part, they either satisfy the requirements or they don't. Therefore, only the Functionality of Design was found to be relevant in analyzing the student design solutions.

Functionality of Design can be decomposed into two parts: Core Requirements and Peripheral Requirements. The Core Requirements represent the critical elements of the function of the network design that must be in place for the network to meet even the rudimentary intent of the network function (for example, in a VLAN design a student must have indicated the right broadcast domains). The Peripheral Requirements are elements that are required for network performance, but which are not central to the purpose of the network or the intent of the task (for example, having an appropriate number of signal repeaters to account for distance). The characteristics distinguishing patterns of performance for Design tasks are:

a) Correctness of Outcome
   i) Functionality of Design
      (1) Core requirements
      (2) Peripheral requirements

On this basis three patterns of performance were distinguished among the solutions to the design problems in the CTA:

- **Design Pattern A.** This pattern was characterized by designs that result in a network that do not meet the core operational requirements for functionality of design.
- **Design Pattern B.** This pattern was characterized by designs that result in a network that meets all the core operational requirements and many of the peripheral operational requirements for functionality of design.
- **Design Pattern C.** Designs that resulted in a network that meets all, or very nearly all, of the core operational requirements for functionality of design.

Implementation Tasks

Review of the CTA results and discussion with SMEs established that the overall framework of observations appropriate for implementation is similar to that for troubleshooting. Both the way in which students perform the configuration (correctness of procedure), and the quality of the resulting configuration (correctness of outcome) are important high-level behavioral characteristics. The same types of actions (gathering information, making changes or adding to the configuration, testing, and getting help) that apply to troubleshooting activities apply to implementation as well. Compared to troubleshooting, however, the implementation tasks demand greater emphasis on the network configuration changes than on gathering information and testing the network function. The characteristics that differentiated the patterns could be summarized by means of these two main categories, which themselves are composed of contributing components according to the following structure:

a) Correctness of Procedure
  i) Procedural Sequence Logic
  ii) Efficiency of Procedure
    (1) Help usage
    (2) IOS syntax
    (3) Volume of actions
b) Correctness of Outcome

Correctness of Procedure has the same two components as for troubleshooting: efficiency of procedure and procedural sequence logic. Efficiency is largely the same as in troubleshooting, focusing on the overall number of actions, use of help, and correct syntax of IOS commands. There are some optimal patterns of behaviors relating to sequence of procedure, but they are less constrained than in troubleshooting; often, the order in which commands are used does not matter in configuration. It does matter in some cases, however, including (1) the order of commands in constructing an access control list, and (2) the sequencing of actions that are used to configure and test connectivity (e.g. it is necessary to set up the routing protocol and interfaces on the routers before testing for connectivity).

Correctness of outcome consists of getting the configuration right – however, some components of the configuration are more difficult than others (e.g. access control lists). In addition, there are some aspects of the configuration (e.g. setting passwords) that students should perform even if the scenario only explicitly mentioned higher-level requirements (e.g. security issues, in the case of passwords). The categories of behavior that reflect these observed differences in patterns of performance are:

- **Implementation Pattern A.** This pattern of performance was indicative of solutions in which students experienced substantial difficulties in configuring network devices.
- **Implementation Pattern B.** This pattern represented those solutions in which students implemented correct solutions in a straightforward manner.

- **Implementation Pattern C.** This pattern was characteristic of solutions in which the student experienced some difficulty in properly configuring network devices.

Network Diagram Construction

For all three types of tasks (design, implementation and troubleshooting), students were asked to sketch a diagram of the network described in the text of each scenario before beginning the actual task required for the scenario. While this diagram construction task was done in the context of a larger scenario, the critical characteristics of performance were similar across scenario types, and the results are grouped together here. Based on the design determination that the sequence of procedures in constructing the diagram is not important evidence for the inferences to be made from the diagram task only the final diagram was collected during the CTA and no data was collected on sequence of diagramming processes. Therefore, this discussion does not address the correctness of procedure and instead focuses on the outcome characteristics of the paper and pencil diagrams students constructed during the CTA.

Overall, the diagrams students constructed varied by level of detail and accuracy. Many students were able to represent the network at the correct level of abstraction for the purpose of the scenario, including essential devices, connections between devices, and appropriate labels. Other students, however, failed to represent important network connections and/or devices. In addition, some students represented unnecessary detail by focusing on irrelevant physical characteristics of devices and the buildings which housed them. The patterns of behavior for the diagram activity are described in terms of a single observable feature: Correctness of Outcome. This Correctness of Outcome is composed of:

a) Extraneous Components
b) Necessary components

On this basis three patterns of performance were distinguished among the solutions to the network diagram problems in the CTA:

- **Network Diagram Pattern A.** This pattern of performance was characterized by the omission of essential features of the network.
- **Network Diagram Pattern B.** This pattern of performance was characterized by the inclusion of all essential characteristics without the addition of erroneous diagram elements. Furthermore, this

pattern was also characterized by a restriction of the detail of the network representation to the necessary level of abstraction for the required tasks.

- **Network Diagram Pattern C.** This pattern of performance was characterized by the inclusion of all essential characteristics without the addition of erroneous diagram elements.

Summary

The results revealed a degree of similarity in the overall classification structure across tasks. For example, the Correctness of Outcome is a universal characteristic of evaluation across the task types and Correctness of Procedure is common, though not universal. However, while the general classifications (e.g. Correctness of Outcome) can be consistent across task types, the characteristics of solutions that contribute to those classifications are specific to the tasks, as is evident in the more fine-grained characteristics for each task type. The observed patterns of performance for all task types were consistent with the three primary classifications of performance patterns, with one (pattern A) indicative of an inadequate pattern of performance, one (pattern B) indicative of a high level of performance, and one (pattern C) indicative of a moderately successful performance. These patterns of performance may be generally summarized by saying that Pattern A is ineffective; Pattern B is both effective and efficient, that is, efficacious—a common feature of expert representations (Kindfield, 1999)—and Pattern C is somewhat effective but not efficient.

The Domain Model

The domain analysis, including the CTA, culminated in a Domain Model (sometimes called an unrestricted student model) representing the constellation of proficiencies and related claims important for success as a CNAP student. The resultant Domain Model is provided as Figure 1 (as are a few associated claims, indicated by stars), representing not only the knowledge, skill and ability necessary, but also the dependencies among them. For example, a student's Network Proficiency (which consists of five interrelated skills) is modeled to be dependent on their Domain Disciplinary Knowledge (which also consists of multiple components). In addition, Design and Troubleshooting are modeled to be conditionally independent given Network Proficiency. Further explication of Design is documented through its associated claims(s), two of which (128a and 400) are provided as examples. In the latter stages of assessment design, the development of reporting rules for a student model further specify, through functions of the posterior distributions for student-model variables, how the values of one or more of these variables relate to particular claims. This Domain Model will later comprise the foundation for the Student Model for the NetPass assessment.

The Domain Model is composed of a number of variables representing aspects of knowledge, skill and ability. The Domain Disciplinary Knowledge variable represents the declarative knowledge of network components and operation, and therefore is the type of knowledge typically assessed through tasks requiring the recall of specific elements of network knowledge (e.g. multiple-choice). There are a number of elements of declarative knowledge represented as part of the Domain Disciplinary Knowledge, including: the OSI network model; addressing schemes; hardware components of a network; media; IOS; protocols; and security.

The Network Proficiency variable represents the skills and procedural knowledge (as opposed to declarative knowledge) necessary to perform a variety of tasks critical to successful network operations. These network operations include several skills: Implementing (configuring) a network; Troubleshooting; Designing; Operating; and Planning. (Only Troubleshooting, Implementation and Design tasks were utilized in the CTA and the prototype assessment). As each of these network activities requires declarative knowledge in order to conduct the procedures required to perform these tasks, there is a modeled relationship between the declarative knowledge represented in Domain Disciplinary Knowledge and the procedural knowledge required for Network Proficiency.

The Network Modeling variable is the ability of the student to represent a network structure that may facilitate their Network Proficiency in various types of tasks. The ability to produce a model of a network requires, in part, Domain Disciplinary Knowledge, which is therefore represented as a prerequisite of Network Modeling ability. Since the ability to produce a model of the network in question is thought to facilitate such tasks as troubleshooting and design, there is a modeled relationship between this ability and the Network Proficiency ability.

<u>Implications for Evidence Models</u>

By utilizing tasks in the CTA that were explicitly designed to yield evidence relevant to anticipated assessment claims the CTA results yielded specific examples of evidence supporting these claims. From the CTA, we have observable examples of errors and misconceptions of students at various levels of expertise, as well as the elements they successfully implement. This evidence will inform development of the Evidence Models, which relate students' actions to the states of the Student Model variables. The Evidence Model is the vehicle by which the raw data from the student's solution is transformed into evidence about the specific claims to be made about the student, and absorbed into the Student Model to update our understanding of the student knowledge, skills and abilities.

These Evidence Models consist of two subcomponents: the evidence rules and the statistical model. The evidence rules provide for the representation of work product elements as variables (called *observables*) that are used as evidence in the evidence accumulation process. Their values are summaries of the lower level features of the assessment data compiled in a manner similar to Clauser Subhiyah, Nungester, Ripkey, Clyman, & McKinley (1995). The statistical model for the assessment takes the values of these fine-grained observables from student performances and uses them to update the values of student model variables. The structure of the Domain Model has implications for the required characteristics and capabilities of the statistical model used to link the values of observables to the estimates of Student Model variables.

Implications for Task Models

The results of the CTA also serve to suggest required task characteristics to maximize their tendency to elicit evidence distinguishing among students, thus fulfilling the summary reporting and diagnostic feedback requirements and ensuring that the evidential requirements for the assessment claims can be fully met. As the CTA provides an opportunity to observe actual student behavior and strategies as they work through the complex tasks this provides an ideal opportunity to determine which aspects of the tasks are serving well and which task elements might be altered to yield better evidence of student ability, thereby better supporting the inferences made about student model variables. As such, the findings of the CTA have several implications for task design in honing the tasks into a better tool for updating the student model variables in the prototype implementation.

One aspect of the Domain Model with implications for the tasks used in assessment is Domain Disciplinary Knowledge. This proficiency correlates with the Network Proficiency student model variable. As our estimate of students' Network Proficiency increases, so does our estimate of their Domain Disciplinary Knowledge. There is, however, a potential ambiguity in the model. If students encounter difficulties with tasks we cannot be certain, based on the network simulations alone, whether those difficulties are due to deficiencies in Domain Disciplinary Knowledge or a lack of Network Proficiency. To disambiguate estimates of these proficiencies direct evidence of Domain Disciplinary Knowledge could be obtained through the inclusion of items designed to specifically target this declarative knowledge in addition to the skills and abilities targeted by the design of simulation tasks.

Similarly, Error Identification (part of the Correctness of Outcome from the troubleshooting tasks in the CTA, see above) in troubleshooting tasks is based on the observation that the student successfully remedies an existing network error. By observing that the student remedies an existing error this observation is used as evidence that the student has both identified the presence of the error and understands the procedures for remedying the network error. However, the possibility exists that a student has sufficient skills to identify an error in a network but insufficient skill to actually intervene and correct the deficiency. The initial evidence identification process did not attempt to disambiguate these two possible factors (error identification and error correction) contributing to the Error Identification variable. As a result there would be implications of this potential ambiguity for assessment design, possibly requiring the addition of a task allowing for this ability to disambiguate deficiencies in ability.

As a result of the Domain Analysis and the Domain Modeling processes a substantial body of information exists to continue with the development of the assessment. Specifically, we have gained an understanding of the tasks that practitioners in the domain must undertake, framed these tasks within the purposes of the assessment, investigated the characteristics of practitioners that influence successful performance of domain tasks, provided a domain model specifying the relationship among key domain abilities, and most importantly made the evidential connections between key features of task performance (and by implication, task design for assessment) and the inference of individual ability. On the basis of these efforts we have the raw materials and fundamental structures necessary to complete the assessment design.

CONCEPTUAL ASSESSMENT FRAMEWORK

The Conceptual Assessment Framework (CAF) is the final design of the full assessment that builds specific models for use in a particular assessment product (taking into account the specific purposes and requirements of that product). The Conceptual Assessment Framework (CAF) for ECD is depicted in Figure 2. It consists of four fundamental models: the *student model*, representing student KSA structure and forming the foundation for claims that will be made about the student on the basis of assessment evidence; *task models*, specifying content and construction of tasks in order to provide the necessary evidence to support claims about the student; *evidence models*, specifying how evidence from assessment tasks inform student model variables in support of specific claims about the student; and an *assembly model*, specifying the strategy used to select and present tasks to a student (Almond, Steinberg, & Mislevy, 2001). The CAF addresses a sequence of questions posed by Sam Messick (1994) that represent the foundation of assessment design:

- "What complex of knowledge, skills, or other attribute should be assessed?"
- "What behaviors or performances should reveal those constructs?"
- "What tasks or situations should elicit those behaviors?"

"What complex of knowledge, skills, or other attribute should be assessed?" The student model is comprised of variables representing characteristics of the examinees—knowledge, skills and abilities—that are the focus of the inference process and that determine the outcomes (e.g. licensure, placement, diagnostic feedback, or some combination) consistent with assessment purpose. The student model represents knowledge about an examinee's possession of these KSAs based on assessment evidence. The extent and functionality of a student model can vary widely by assessment purpose, and include multiple student models to support multiple purposes from a common assessment (e.g. one to provide for outcomes such as an overall score and another to provide for educationally relevant feedback). It is the values of some combination of student model variables in the CAF that are referenced by *reporting rules* that transform these statistical values into text and graphics used in the score reports for different purposes (although for some purposes these reporting rules might also reference portions of the evidence model).

For the NetPass assessment the Domain Model (see Figure 1) becomes the Student Model provided as Figure 3. The student model for NetPass represents the ability variables targeted by the assessment. Since the NetPass assessment is specifically targeting the performance components of computer networking through innovative design in order to supplement the pre-existing knowledge tests the representation of various domain knowledge variables are dropped from the Domain Model to the Student Model. In addition, due to the relative importance of some computer networking tasks it was decided to focus on the three most critical computer networking tasks and to omit routine operation and planning from the NetPass assessment. The Student Model for NetPass reflects these design decisions.

"What behaviors or performances should reveal those constructs?" An evidence model expresses how the observable features of examinee work products from a task constitute evidence about student-model variables. The evidence model is made up of two components: (a) the evaluation component, which consists of the *rules of evidence* describing the evidence extraction process, in which features of the work product are identified as observable variables that constitute evidence for inferences about students, and (b) the *statistical, or measurement, model*, which specifies how information provided by these observables should influence belief about the values of student model variables.

In the case of NetPass the rules of evidence for the evidence model follow the framework outlined above from the results of the task analysis. For example, in the case of troubleshooting tasks the rules of evidence are a number of if-then rules that assess the work products (log file, network diagram, a list of identified faults, and the final network configuration file) according to the hierarchical structure outlined above for troubleshooting the if-then rules produce summary values for the observables described above as the highest variables in the hierarchy (correctness of outcome for the faultlist, correctness of outcome for the configuration file, correctness of procedure for the log file, efficiency of procedure for the log file, and correctness of outcome for the diagram). Once the values of these observables are obtained from the evidence rules then the statistical model is applied to produce the summary values of the student model variables. Figure 4 provides a graphical example of the relationship between the observables for a troubleshooting task and the student model variables informed by these observables. For NetPass, the statistical updating for these evidence models is accomplished through the application of Bayesian Inference Networks (Jensen, 1996; Mislevy, 1994; Pearl, 1988), a probabilistic model for the drawing of inferences on student model variables based on the evidence provided by the observable variables.

"What tasks or situations should elicit those behaviors?" Task models describe characteristics of the assessment tasks intended to elicit particular types of examinee performance. A task model provides a framework for characterizing and constructing situations with which a candidate interacts to provide evidence about targeted aspects of knowledge. The task model specifies key elements of 1) performance situations, 2) material presented to the student, and 3) student work produced in responding to the task (see Bennett & Bejar, 1998 for a discussion of the importance of this and other considerations). A typical assessment uses not one but many task models, each capable of generating many tasks according to model specifications. Tasks generated from such models share similarities in task requirements and the student work produced in response. Both task models and instantiations of items from them play a vital role in automatic item generation efforts (Bejar, 2002; Irvine & Kyllonen, 2002; Mislevy, Steinberg & Almond, 2002, in press) and there is a degree of commonality between the task models constructed from an ECD perspective and the item modeling efforts that underlie efforts to create automatic item generation systems (Singley & Bennett, 2002; Newstead, Bradon, Handley, Evans & Dennis, 2002; Dennis, Handley, Bradon, Evans & Newstead, 2002; Sheehan & Deane, 2003) such that the efforts of each endeavor facilitates the development of the other.

In the case of NetPass the assessment requires evidence from a variety of sources and the task models used to generate the simulations must specify a large number of design variables, both describing the key components of the scenario that tend to discriminate levels of ability as well as to control peripheral elements of the simulation to prevent the parameters on these elements from extending beyond the controls established to provide for consistent evidential value from the design tasks. An example of a small portion of a task model for NetPass troubleshooting scenarios is provided as Figure 5.

In an implemented assessment, the student model(s) accumulate and represent current beliefs about the targeted aspects of examinee proficiency that are expressed as student-model variables. The evidence models identify the key features of examinee behaviors and work products that provide evidence about aspects of proficiency in the student model (Steinberg & Gitomer, 1996), and use a psychometric model to express how the evidence relates to student-model variables (Mislevy, 1994). The task models guide the design of situations engineered to evoke responses that provide appropriate evidence for the claims relevant to the purpose of the assessment (Almond, Steinberg, & Mislevy, 2001). An example of the use of these models to develop a design rationale for simulation-based problem-solving in the domain of dental hygiene is provided by Mislevy, Steinberg, Breyer, Almond, and Johnson (1999; in press). A more complete description of the ECD process applied to produce NetPass and emphasizing the early stages of design can be found in Bauer, Williamson, Steinberg, Mislevy, & Behrens (2001) and Williamson, Bauer, Mislevy & Behrens (2003). By virtue of emphasizing systematic consideration of targeted inferences, the nature of evidence required for such inferences, and the explicit relationship between design components and the delivery processes, the final assessment design from ECD also provides specifications for implementation and delivery.

REUSABILITY OF ASSESSMENT DESIGN COMPONENTS

As previously discussed, one of the significant advantages of the ECD approach to innovative assessment design is that each stage of the design process provides a documented design structure that can be utilized for other purposes, thereby improving the feasibility of ongoing assessment and maximizing the utility of the successes of innovation. This discussion will illustrate some anticipated reusability of design components for NetPass, with an example of such anticipated reusability for each of the four major types of extensibility: task, purpose, proficiency and domain.

The extensibility of task is perhaps the most obvious potential area of reusability and relies on leveraging the CAF task models to produce new tasks in a controlled manner while maintaining a constant structure for the other assessment design elements (e.g. student model, purpose, etc.). In part, the current efforts to develop approaches for automatic item generation represent a recognition of this potential for reusability of task design. While to date, most efforts toward automatic item generation are concerned with the generation of multiple-choice items the principles remain the same when considering innovative item types. This underlying principle is the belief that given some hierarchy of task structure (from the vaguely defined to the tightly constrained) and specifications of conditions under which the components of task structure are free to vary, this abstract task structure can be readily employed to produce a number of items with a controlled degree of similarity based on the modeled structure. While the opportunity for an automatic approach to task generation can be expected to be limited for many innovative task types, the use of ECD task models to produce the innovative items provides the immediate framework for continued manual production of more innovative items from a common, controlled structure. There are two general goals typically targeted for the reuse of task models to produce new innovative assessment tasks: (a) the continuation of an assessment with new, unexposed, but measurement-equivalent assessment tasks, referred to as *isomorphic* equivalent tasks and (b) the extension of a successful innovative task framework to target a systematically varying level of ability or interaction of abilities in the domain.

When the goal is to produce new, unexposed tasks that are isomorphic equivalent for the purposes of the assessment the task models serve as a means of controlling the variation in both evidential relevance and in difficulty of the resultant tasks. Since the task model identifies the variables used to produce a task the key variables relating to the evidential relevance can be held constant to ensure that the same *nature* of evidence is targeted by the task (e.g. that the same student model variables are the target of inference). In addition, in order to control the *difficulty* of the task both the construct-relevant and construct-irrelevant variables in the task model can be specified to be within identified parameters to exercise control. The degree of control required and the number of variables that must be controlled for this approach will vary by the complexity of the innovative task design, standards of exchangeability (e.g. higher standards for high-stakes assessment than for medium or low stakes assessment), the number of decision points in the reporting scale and the relationship between modeled task parameter variation and measurement error. A similar approach to the model-based production of isomorphic complex simulation tasks in high-stakes assessment has been used by the National Council of Architectural Registration Boards for their architectural design simulations employed in the Architect Registration Examination (Williamson, Johnson, & Sinharay, 2002).

The fundamental approach to the extensibility of task remains the same when leveraging the task models for the purpose of targeting varying levels of ability or for shifting the target of the nature of evidence provided to different student model variables. The procedure still relies on the identification of the task model variables that must remain constant and those that can be varied within certain parameters but it is no longer necessary for the resultant assessment task be isomorphically equivalent to previously generated tasks. An example of two operationalized task models for NetPass are provided as Figure 6, for a troubleshooting task of medium difficulty, and Figure 7, for a troubleshooting task of higher difficulty. These illustrate the operationalized decisions from the general task model in Figure 5, illustrating the extension of a common design framework to produce multiple tasks.

An additional area in which the extensibility of task for NetPass has been particularly fruitful is in the interface design for the design tasks. The interface for the design tool used in specifying the design of a network, or the existing structure of a preexisting network (such as during the troubleshooting work) is presented as Figure 8, with a completed design presented as Figure 9. This interface has proved so successful as a means for representing the key parts of network structure that CLI and others associated with the networking academy program have started using the interface independently of the assessment system as a means for communicating about network design over long distances (Behrens, 2003). In this case the extensibility of task has been by providing an interactive tools for representing network structure that also fills a professional need for the domain of practice itself.

When leveraging the assessment design for extensibility of purpose the student model remains the same as in the original CAF and the key alteration to the assessment design rests with the *reporting rules* used in accessing and reporting on configurations of the values of assessment design variables of interest. An example of such reusability is when the purpose shifts from placement to diagnostic assessment. If we were using the NetPass model to determine the placement of examinees in the most appropriate level of the CNAP curriculum the reporting rules would be based on the summary values of the student model variables in the CAF. If, however, the purpose shifted to diagnostic assessment with educationally relevant feedback the reporting rules would need to be altered to address key elements of the evidence model. For example, in diagnostic assessment for troubleshooting the reporting rules would be expanded to reference the log file correctness of procedure in the evidence model. The values of the evidence rules would be leveraged to provide a summary of the examinee performance with respect to the characteristics of good troubleshooting procedure as described above (e.g. sequence of targeting to begin with the router most likely to be the source of the fault, following the proper sequence of identification, remediation, and testing, etc.). These would be the basis for reporting on a comparison of the examinee procedural performance with the recommended procedure for troubleshooting.

A key component of extensibility of purpose that distinguishes it from other forms of extensibility is that it requires no changes to the student model driving the assessment. Depending on the nature of the shift in purpose this shift may require associated changes in the task models used for task generation and the associated assembly model.

For example, if the purpose shifted from relatively low stakes assessment to high stakes assessment there may be associated changes in the need for isomorphic equivalence among tasks generated from the task model and in the assembly model to assure sufficient reliability of reported results. This form of extensibility also allows for dual purpose from a single assessment. That is, there is always the potential to use a single assessment to both place students at the appropriate level of CNAP curriculum AND to provide educationally diagnostic assessment feedback.

The extensibility of proficiency is primarily focused on the reuse of elements of the student model in modified form for the same domain of interest. A straightforward example of such reuse of a prior student model in a new student model for NetPass would be by extending the student model to include the two additional areas of practice represented in the domain model (see Figure 1). This would constitute an example of extending the breadth of the student model but the extension of a student model could also be directed at the depth of the model as well. An application in an intelligent tutoring system, for example, would require extension to a substantially finer grain-size than the current student model attempts to represent. These extensions of the proficiency model can include explicitly modeled misunderstanding of the domain that may be common in certain populations, much like the efforts that test developers apply at the item level in developing good distracters for multiple-choice items. These same principles can be used to leverage the proficiency model into a model of educational curriculum designed to provide a model of ability acquisition in the domain. Of course, the reusability of student models are not restricted to extension as a reduction of scope of a student model for a new assessment with a precise, targeted purpose (e.g. an assessment of domain modeling ability alone) would also constitute an example of extensibility of proficiency.

The final type of reuse is extensibility of domain. Initially, it would seem that there would seldom be an opportunity to leverage design components for different domains. However, experience has suggested that the opportunity for such inter-domain sharing of assessment design elements may be more prevalent than previously assumed. As an example of the potential for possibilities of interdomain reusability the authors have observed a degree of similarity between the strategic elements of troubleshooting for computer networks and troubleshooting for military aircraft hydraulic systems (Gitomer, Steinberg & Mislevy, 1995). Specifically, mechanics tended to use one of two strategies for troubleshooting the aircraft hydraulic systems: serial elimination or space-splitting. In serial elimination, the less efficient approach, mechanics would sequentially test parts in the hydraulic system in the sequence of their connective relationship until they reached the part that was causing the malfunction. By contrast, in space-splitting, the more efficient approach, mechanics would select a key component toward the middle of the series of hydraulic components and if it showed functionality they would know that all of the components prior to that one in the series were functioning appropriately as well. Through this process those who used space-splitting techniques could accomplish with a single diagnostic test what would take those using serial elimination a sequence of tests to accomplish. Similarly, on the basis of the CTA for NetPass it became evident that students high in troubleshooting ability for computer networks would strategically select the router and diagnostic tests that would provide information allowing for the quick elimination of large sections of the network as candidates for the networking fault.

While this domains of aircraft hydraulics and computer networking are quite dissimilar in content, the similarity in strategic approach to troubleshooting these two systems suggests an unanticipated degree of potential reusability in the understanding of effective strategies in troubleshooting. In this case the common element between domains rests with the act of troubleshooting complex systems for operational faults. A formalized structure for troubleshooting strategy in the assessment design models might therefore find application in other content domains that benefit from such strategies. It is possible that with a formalized structure of design from ECD and an emphasis on considering the cognitive components of domain performance rather than limiting investigations to domain content that there could be a greater degree of generalizability of assessment design elements across domains initially thought to be incompatible.

TOOLS FOR EFFECTIVE REUSE

If we accept the potential benefits of designing for reusability in innovative assessment as a goal of the deign process it would be beneficial to examine ways that we can facilitate our ability to maximize the potential reusability of design components. Currently at Educational Testing Service (ETS) some divisions are applying some of these design methods in a paper-based format and leveraging the efficiencies of reuse for traditional assessments. Some other work areas at ETS have been using a computerized prototype tool built to explore these approaches to innovative assessment design which includes features that support the reusability of design components. On the basis of such interest in reusability, both for innovative assessments and for traditional measures, we are currently developing a production tool currently referred to as the Assessment Designer Tool (ADT). The purpose of the ADT is to support the design of assessments via an Evidence-Centered Design approach and the review of the assessments by the various stakeholders that contribute to the design. The ADT facilitates the creation of an assessment design and associated tasks through the use of various forms, each of which provides a different view of the assessment design being constructed. The ADT does not prescribe a particular design process; different pieces of the assessment design are completed on an as-needed basis. Core functionality to support reuse includes abilities to a) save and retrieve design objects, b) generalize objects that were designed for a specific assessment, and c) search and browse libraries of design objects, the ability to generalize objects that were designed for a specific assessment.

CONCLUSION

This paper has outlined the development of Cisco Learning Institute's NetPass assessment as an illustration of an ECD approach to development of an innovative assessment. By virtue of applying the ECD process a the resultant design provides for a transparent construct validity rationale for the use of the assessment and for elements thereof. This approach to assessment design also provides a substantial and modular set of models that lend themselves to future reuse and adaptation to other purposes. As such the NetPass assessment design has provided an example of an innovative assessment that has both an underlying, transparent validity argument as well as the ability to leverage the successful elements of this innovation in future work. The challenge for the future is to further explore the degree of reusability of such assessment design models through the reuse of NetPass design components in future work and the evaluation of the success of these reusable components. As technological advances continue the challenges of educational measurement will extend beyond a primary concern with technical capabilities and become more concerned with the basis of a sound validity argument underlying the use of such innovative approaches and the means for maximizing the success and economic efficiencies of such successful applications of innovative design. Towards that end ECD provides a means for valid design targeting opportunities for innovation, and efficiencies through reusability of design.

References

Almond, R.G., Steinberg, L.S., & Mislevy, R.J. (2001). A sample assessment using the four process framework. *CSE Technical Report #543.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Bauer, M., Williamson, D. M., Steinberg, L., Mislevy, R., & Behrens, J. T. (2001, April). *How to create complex measurement models: A case study of principled assessment design.* Paper presented at the meeting of the National Council on Measurement in Education, Seattle, WA.

Behrens, J. T. (2003, January). *How ETS's evidence centered design is changing my life (and that of 500,000 others).* Paper presented at the Research Colloquium series at Educational Testing Service, Princeton, NJ.

Bejar, I.I. (2002). Generative testing: From conception to implementation. In S.H. Irvine & P.C. Kyllonen (Eds.), *Item generation for test development* (pp. 199-217). Hillsdale, NJ: Erlbaum.

Bennett, R. E. & Bejar I. I. (1998). Validity and automated scoring: It's not only the scoring. Educational Measurement: Issues and Practice, 17(4), 9-17.

Clauser, B.E., Subhiyah, R., Nungester, R.J., Ripkey, D., Clyman, S.G., & McKinley, D. (1995). Scoring a performance-based assessment by modeling the judgments of experts. *Journal of Educational Measurement, 32,* 397-415.

Dennis, I., Handley, S., Bradon, P., Evan, J., Newstead, S. (2002). Approaches to modeling item-generative tests. In S. H. Irvine & P. C. Kyllonen (Eds.), *Item Generation for Test Development* (pp. 53-72). Mahwah, NJ: Erlbaum.

Ericsson, K. A., & Smith, J., (1991). Prospects and limits of the empirical study of expertise: An introduction. In K.A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits.* Cambridge: Cambridge University Press.

Gitomer, D.H., Steinberg, L.S., & Mislevy, R.J. (1995). Diagnostic assessment of trouble-shooting skill in an intelligent tutoring system. In P. Nichols, S. Chipman, & R. Brennan (Eds.), Cognitively diagnostic assessment (pp. 73-101). Hillsdale, NJ: Erlbaum.

Irvine, S.H., & Kyllonen, P.C. (Eds.) (2002). *Item generation for test development.* Hillsdale, NJ: Erlbaum.

Jensen, F.V. (1996). An Introduction to Bayesian Networks. New York, NY: Springer-Verlag New York, Inc.

Kindfield, A.C.H. (1999). Generating and using diagrams to learn and reason about biological processes. *Journal of the Structure of Learning and Intelligent Systems, 14(2),* 81-124.

Melnick, D. (1996). The experience of the National Board of Medical Examiners. In E.L. Mancall, P.G. Vashook, & J.L. Dockery (Eds.), *Computer-based examinations for board certification* (pp. 111-120). Evanston, IL: American Board of Medical Specialties.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13-23.

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

Mislevy, R.J., Steinberg, L.S., & Almond, R.G. (2002). On the roles of task model variables in assessment design. In S. Irvine & P. Kyllonen (Eds.), *Generating items for cognitive tests: Theory and practice* (pp. 97-128). Hillsdale, NJ: Erlbaum.

Mislevy, R.J., Steinberg, L.S., & Almond, R.G. (in press). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Discussion.*

Mislevy, R.J., Steinberg, L.S., Breyer, F.J., Almond, R.G., & Johnson, L. (1999). A cognitive task analysis, with implications for designing a simulation-based assessment system. *Computers and Human Behavior, 15,* 335-374.

Newell, A., & Simon, H.A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Newstead, S., Bradon, P., Handley, S., Evans, J., Dennis, I. (2002). Using the psychology of reasoning to predict the difficulty of analytical reasoning problems. In S..H. Irvine & P. C. Kyllonen (Eds.), *Item Generation for Test Development* (pp. 35-52). Mahwah, NJ: Erlbaum.

Pearl, J. (1988). Probabilistic reasoning in intelligent systems: Networks of plausible inference (2^{nd} ed.). San Francisco, CA: Morgan Kaufmann.

Sheehan, K, & Deane, P. (2003, April). *A partially automated system for generating passage-based multiple-choice verbal reasoning items*. Paper presented at the meeting of the National Council on Measurement in Education, Chicago, IL.

Singley, M. K., & Bennett, R. E. (2002). Item generation and beyond: Applications of schema theory to mathematics assessment. In S. H. Irvine & P. C. Kyllonen (Eds.), *Item Generation for Test Development* (pp. 362-384). Mahwah, NJ: Erlbaum.

Steinberg, L.S., & Gitomer, D.H., (1993). Cognitive task analysis and interface design in a technical troubleshooting domain. *Knowledge-Based Systems,6,* 249-257.

Steinberg, L.S., & Gitomer, D.G. (1996). Intelligent tutoring and assessment built on an understanding of a technical problem-solving task. *Instructional Science, 24,* 223-258.

Steinberg, L., Mislevy,R., Almond, R., Baird, A., Cahallan, C., Chernick, H., Dibello, L., Kindfield, A., Senturk, D., Yan, Duanli (2000). Using evidence-centered design methodology to design a standards-based learning assessment. Research Report. Educational Testing Service. Princeton, NJ.

Williamson, D. M., Bauer, M., Mislevy, R., & Behrens, J. T. (2003, April). *Creating a complex measurement model using evidence centered design*. Paper presented at the meeting of the National Council on Measurement in Education, Chicago, IL.

Williamson, D. M., Johnson, M. & Sinharay, S. (2002, April). *Hierarchical IRT examination of isomorphic equivalence of complex constructed response tasks*. Paper presented at the meeting of the National Council on Measurement in Education, New Orleans, LA.

Acknowledgements

This paper is based on research conducted jointly by the Cisco Learning Institute (CLI) and Educational Testing Service (ETS). The principal investigators of the project are John Behrens, Cisco Systems, and Malcolm Bauer, ETS. Additional members of the research team whose contributions we gratefully acknowledge include, from Cisco, Tara Jennings, Sarah Demark, Ken Stanley, Bill Frey, Michael Faron; from ETS, Peggy Redman; and from Unicon, Perry Reinhert. We would like to express our gratitude to the 27 students who volunteered to work through the cases of the cognitive task analysis, and their instructors, Mike Hogan, Dara West, K Kirkendall, and Tenette Petelinkar. We especially wish to acknowledge and thank the Cisco Learning Institute for funding this research.

Table 1

Example Claim and Evidence Representations

Claims	Representations to capture information from student
Identify the cause of connectivity problems at the physical, data link, and network layers of the OSI model	1. Log file of IOS commands 2. Configuration files for routers (state of network) 3. Worksheet (set of faults) 4. Network Diagram 5. Essay

Table 2

Observable Features (Evidence) from Specific Representations

Observable Feature	*General category*	Representations
1. Steps taken to identify problem(s)	*Correctness of procedure*	1
2. Identification of network problem(s)	*Correctness of outcome*	*1,2,3*
3. Connection between steps and problem(s)	*Connections between procedure and outcome*	*1,3,5*
4. Problem solving logic used for determining cause	*Rationale for procedure*	*5*

Figure Captions

Figure 1. Networking domain model with selected example claims.

Figure 2. The primary models of the Conceptual Assessment Framework.

Figure 3. The operational student model for the NetPass assessment.

Figure 4. An example evidence model for summary score reporting from a NetPass troubleshooting task.

Figure 5. A Portion of a task model for a NetPass troubleshooting task.

Figure 6. A Portion of a task model specified for a medium difficulty NetPass troubleshooting task.

Figure 7. A Portion of a task model specified for a high difficulty NetPass troubleshooting task.

Figure 8. The initial layout of the network design interface for a NetPass design task.

Figure 9. A completed layout in the network design interface for a NetPass design task.

Figure 1

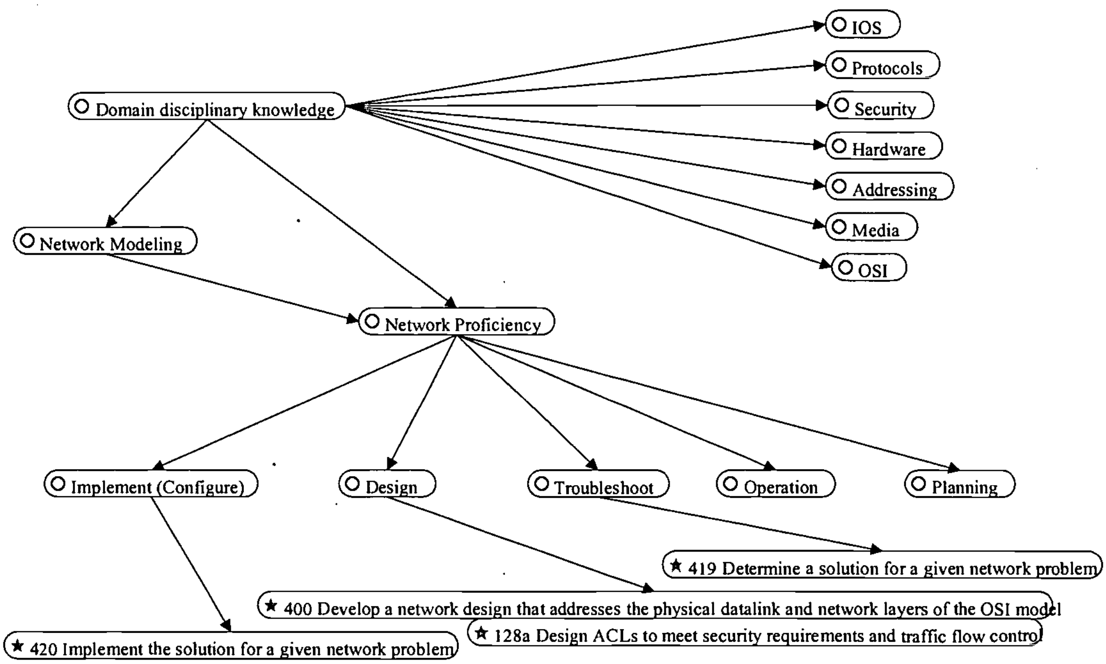

Domain disciplinary knowledge
IOS
Protocols
Security
Hardware
Addressing
Media
OSI
Network Modeling
Network Proficiency
Implement (Configure)
Design
Troubleshoot
Operation
Planning
419 Determine a solution for a given network problem
400 Develop a network design that addresses the physical datalink and network layers of the OSI model
128a Design ACLs to meet security requirements and traffic flow control
420 Implement the solution for a given network problem

Figure 2

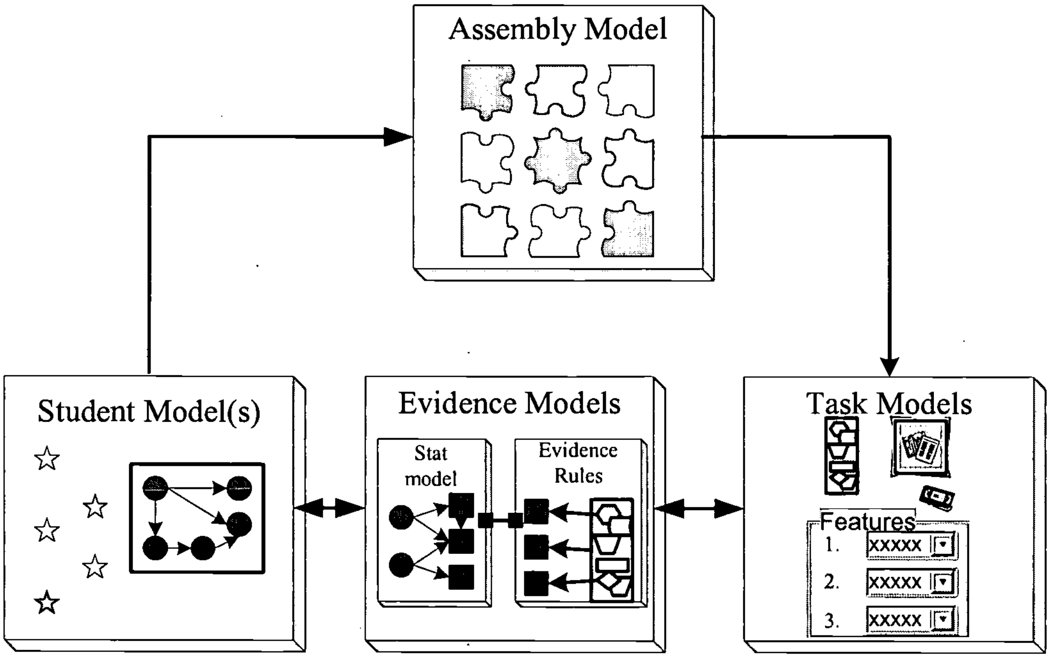

Assembly Model
Student Model(s)
Evidence Models
Stat model
Evidence Rules
Task Models
Features
1. XXXXX
2. XXXXX
3. XXXXX

Figure 3

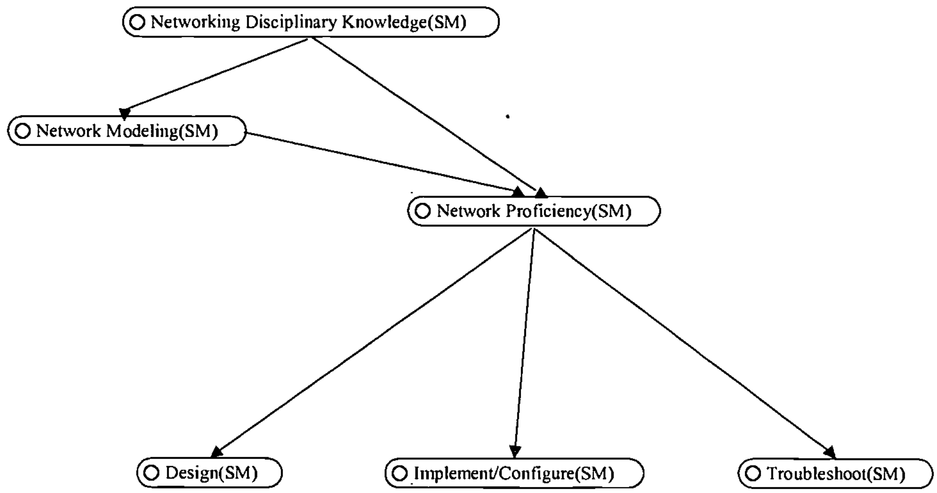

Networking Disciplinary Knowledge(SM)
Network Modeling(SM)
Network Proficiency(SM)
Design(SM)
Implement/Configure(SM)
Troubleshoot(SM)

Figure 4

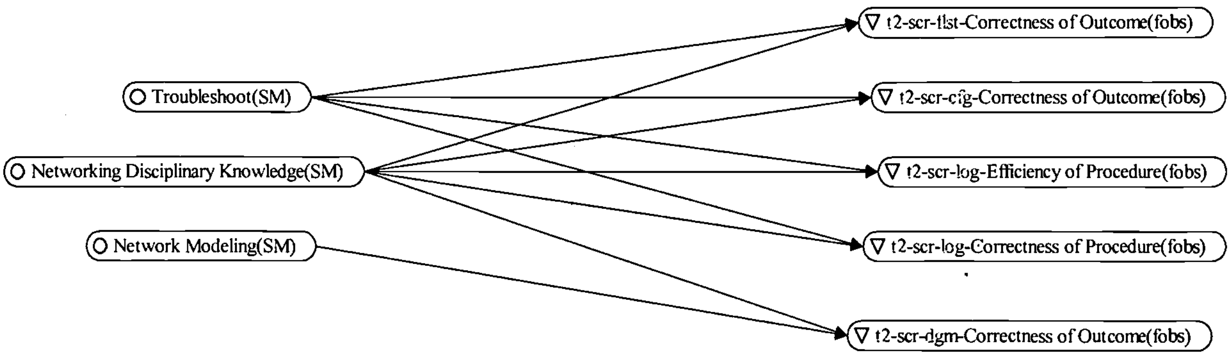

Troubleshoot(SM)
Networking Disciplinary Knowledge(SM)
Network Modeling(SM)
t2-scr-flst-Correctness of Outcome(fobs)
t2-scr-cfg-Correctness of Outcome(fobs)
t2-scr-log-Efficiency of Procedure(fobs)
t2-scr-log-Correctness of Procedure(fobs)
t2-scr-dgm-Correctness of Outcome(fobs)

Figure 5.

Task Model Variable
..... Fault Symptoms
.... Diagnostic specificity and power
.... Fault Symptom Overlap
..... Item-level feedback required
..... Nature of Claim * Troubleshooting
..... Network Complexity
.... Network Connectivity
.... Number of routers
..... Network Testing Characteristics
.... Number of tests required
.... Test Types
..... Number of faults
.... Data Link Fault
.... Frame type fault
.... NIC Card fault
.... Fault distribution across OSI Layers 1 and 2 and 3
.... Fault relationships
.... Fault-Symptoms Matrix Requirement
.... Interface Fault
.... Interface Status
.... Interface Type
.... Network layer fault
.... Access List fault
.... Access List direction fault
.... Access List numbers
.... Access List port fault
.... IP addressing fault
.... Subnet Mask Fault
.... Routing Protocol fault
.... Physical layer fault
.... Cable fault
.... Cable location
.... Cable status
.... Cable termination
.... Clock rate on serial port fault
.... Interference fault
..... Response type
.... Characteristics of distractors
.... Misconceptions targeted
.... Number of distractors
.... Similarity of responses
..... Simulator usage
.... Simulator Configuration
.... Simulator Integration
..... Troubleshooting Story
..... Type of Requirements Given
.... Fault Symptom Presentation
.... Setting
.... Subgroup name
.... User class)
.... Test requirement presentation
.... Test result presentation

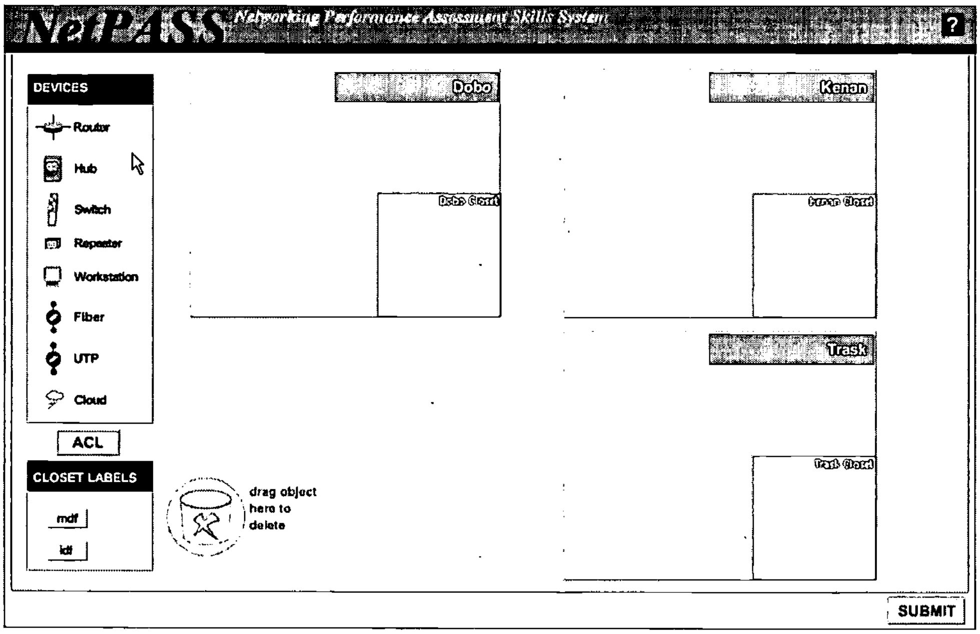
Figure 6

Task Model Variable
..... Fault Symptoms * Significant
.... Diagnostic specificity and power * Medium
.... Fault Symptom Overlap *
..... Item-level feedback required * Yes
..... Nature of Claim * Troubleshooting
..... Network Complexity * Complex
.... Network Connectivity * Partial connectivity
.... Number of routers * 3
..... Network Testing Characteristics * Applicable
.... Number of tests required * 3
.... Test Types * Show run, Ping,
..... Number of faults * 3
.... Data Link Fault * No
.... Frame type fault * NA
.... NIC Card fault * NA
.... Fault distribution across OSI Layers 1 and 2 and 3 * 1 layer only
.... Fault relationships * Inter-related
.... Fault-Symptoms Matrix Requirement * Yes
.... Interface Fault * No
.... Interface Status * Functioning
.... Interface Type * NA
.... Network layer fault * Yes
.... Access List fault * No
.... Access List direction fault * NA
.... Access List numbers * NA
.... Access List port fault * NA
.... IP addressing fault * Yes
.... Subnet Mask Fault * No
.... Routing Protocol fault * Yes
.... Physical layer fault * No
.... Cable fault * No
.... Cable location * NA
.... Cable status * NA
.... Cable termination * NA
.... Clock rate on serial port fault * Present
.... Interference fault * No
..... Response type * Constructed Response
.... Characteristics of distractors * NA
.... Misconceptions targeted * NA
.... Number of distractors * NA
.... Similarity of responses * NA
..... Simulator usage * a simulator is required for task performance
.... Simulator Configuration * not dynamically configurable
.... Simulator Integration * NA
..... Troubleshooting Story * e-mail connectivity interupted in 1 building
..... Type of Requirements Given * Network problem indicators
.... Fault Symptom Presentation *
.... Setting * College campus
.... Subgroup name * students
.... User class * NA
.... Test requirement presentation * Not given
.... Test result presentation * Not given

Figure 7

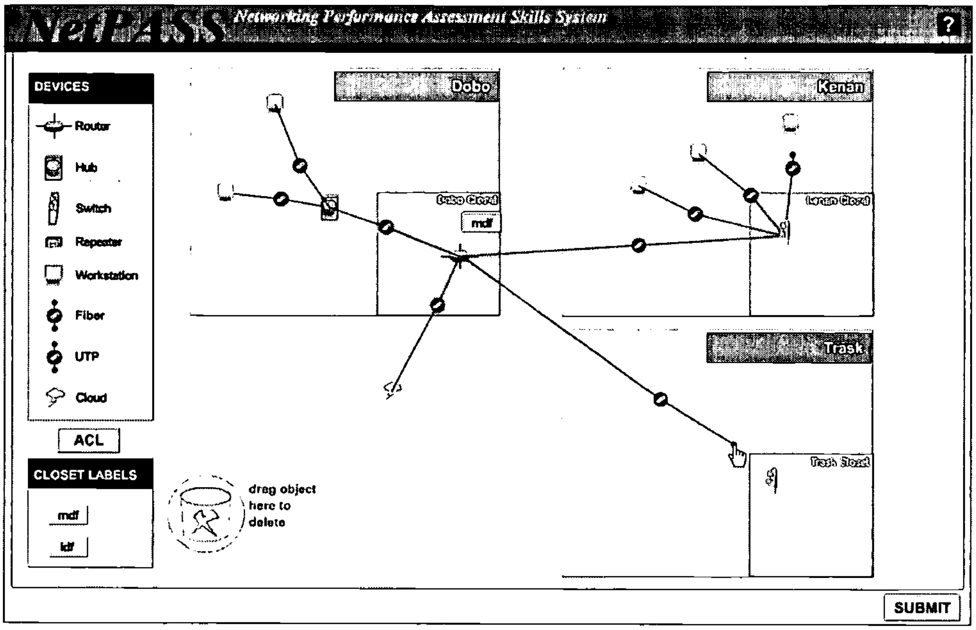
Task Model Variable
..... Fault Symptoms * Significant
.... Diagnostic specificity and power * High
.... Fault Symptom Overlap *
..... Item-level feedback required * Yes
..... Nature of Claim * Troubleshooting
..... Network Complexity * Complex
.... Network Connectivity * Partial connectivity
.... Number of routers * 3
..... Network Testing Characteristics * Applicable
.... Number of tests required * 4
.... Test Types * Show run, Ping, Telnet
..... Number of faults * 3
.... Data Link Fault * No
.... Frame type fault * No
.... NIC Card fault * No
.... Fault distribution across OSI Layers 1 and 2 and 3
.... Fault relationships * Inter-related
.... Fault-Symptoms Matrix Requirement * Yes
.... Interface Fault * No
.... Interface Status * Functioning
.... Interface Type * NA
.... Network layer fault * Yes
.... Access List fault * Yes
.... Access List direction fault * Yes
.... Access List numbers * No
.... Access List port fault * No
.... IP addressing fault * Yes
.... Subnet Mask Fault * No
.... Routing Protocol fault * No
.... Physical layer fault * No
.... Cable fault * No
.... Cable location * NA
.... Cable status * NA
.... Cable termination * NA
.... Clock rate on serial port fault * present
.... Interference fault * No
..... Response type * Constructed Response
.... Characteristics of distractors * NA
.... Misconceptions targeted * NA
.... Number of distractors * NA
.... Similarity of responses * NA
..... Simulator usage * a simulator is required for task performance
.... Simulator Configuration * not dynamically configurable
.... Simulator Integration * NA
..... Troubleshooting Story * internet connectivity interrupted in 1 building
..... Type of Requirements Given * Network problem indicators
.... Fault Symptom Presentation *
.... Setting * Business
.... Subgroup name * branches
.... User class * NA
.... Test requirement presentation * Not given
.... Test result presentation * Not given

Figure 8



BEST COPY AVAILABLE

Figure 9



BEST COPY AVAILABLE

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC
Educational Resources Information Center

REPRODUCTION RELEASE

(Specific Document)

TM035024

I. DOCUMENT IDENTIFICATION:

Title: An ECD Approach to Designing for Reusability in Innovative ~~Asst~~ Assessment	
Author(s): David M. Williamson; Malcolm Bauer; Robert J. Mislevy; John T. Behrens	
Corporate Source: ETS	Publication Date: April 2003

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ✓	Level 2A ☐	Level 2B ☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please

Signature: [signature]	Printed Name/Position/Title: David M. Williamson; Research Scientist	
Organization/Address: ETS Rosedale Road Princeton, NJ 08541	Telephone: 609-734-1303	FAX:
	E-Mail Address: dmwilliamson@ets.org	Date: 6/4/03

(Over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20742-5701
ATTN: ACQUISITIONS

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfacility.org