

Running Head: INTRODUCTION TO ECD, NETPASS, CISCO

Introduction to Evidence Centered Design and Lessons Learned From Its Application in a Global  
E-learning Program

John T. Behrens  
Cisco Systems, Inc, San Jose, California

Robert J. Mislevy  
University of Maryland

Malcolm Bauer

David M. Williamson  
Educational Testing Service, Princeton, New Jersey

Roy Levy  
University of Maryland

To appear in *The International Journal of Testing*

### Abstract

This paper introduces the assessment and deployment contexts of the Networking Performance Skill System (NetPASS) project and the papers in this section that report on findings from this endeavor. First, the educational context of the Cisco Networking Academy Program is described. Second, the basic outline of Evidence Centered Design is described. In the third section, the intersection of these two activities in the NetPASS project is described and the subsequent papers introduced.

## Introduction to Evidence Centered Design and Lessons Learned From Its Application in a Global E-learning Program

The following series of papers describe the application of Evidence Centered Design (ECD; Mislevy, Steinberg, & Almond, 2003) to the construction of an on-line performance based assessment activity as well as lessons learned in this endeavor. The undertaking described here is one of the few comprehensive applications of ECD reported to date. The context of this application was the Cisco Networking Academy Program, which is a philanthropic educational outreach program with the goal of supporting educational institutions in their efforts to teach students about computer technologies and computer networking. Between its inception in 1998 and April of 2004, the program grew to participation by over 10,000 schools in more than 150 countries. Because curricular and assessment materials are presented to students on-line as computer based networked aids, the research described here reports on the intersection of psychological, inferential, statistical, and computing technologies. ECD proved a valuable aid in the integration of these many aspects of assessment design and implementation.

### The Context

Cisco Systems, Inc. and the Cisco Learning Institute fund the development and deployment of the Cisco Networking Academy Program. This program provides on-line curriculum and on-line assessments for free to participating secondary and higher education public schools throughout the world. Typically, students read and interact with on-line curricula out of class and participate in instructional and assessment activities in class. For their part, participating schools send their teachers to training on the content and pedagogy of the program and must fulfill certain quality control agreements including the use of real networking hardware for hands on instruction, ongoing use of the Academy Connection World Wide Web portal for

students and instructors, and a regular cycle of professional development for instructors.

Demonstrated proficiency of the instructors is expected on both the multiple-choice and hands-on skill-focused final exams. Chapter exams, mid-term exams (for some curricula), and Final exams are provided in an on-line format for students that allows immediate automated scoring, storage of results into a grade book and computerization of other common administrative tasks. These are designed to promote interactivity and provide the needed depth to pass an industry standard certification that is delivered independently of the Networking Academy Program. The curriculum and assessments are typically presented in nine languages, but the number varies slightly over time and curricular update cycle. As of April 2004, over 32 million on-line exams have been delivered into classrooms to support student instruction using the World Wide Web in the Cisco Networking Academy Program. The philanthropic context of the Cisco Networking Academy Program is described by Porter & Kramer (2002) while the larger training and educational context is described in Murnane, Sharkey, & Levy (2002) and Levy & Murnane (2004). Interested readers can find out more at <http://cisco.netacad.net/public/index.html>.

Having made significant inroads into the mechanics of the world-wide assessment delivery of fixed-response multiple-choice exams, the Cisco Networking Academy assessment personnel decided a comprehensive investigation of possibilities for on-line assessment design and deployment was in order. More specifically, it was decided that explorations into the application of ECD (Mislevy, Steinberg, & Almond, 2003) would be most useful because ECD is one of only a few frameworks that provides comprehensive guidance across design, analysis, deployment and measurement aspects of assessment.

#### The Goal

While the Networking Academy assessments program was widely considered to be successful (cf Murnane et al., 2002; Levy et al., 2004), it was believed that a number of areas for continued improvement remained. First, there was general sense that the web-based delivery model and all the associated technologies were not being used as completely as possible. More importantly, however, it seemed the old language of classical assessment texts did not provide the richness the assessment team was looking for to envision another generation of assessment. Accordingly a language and structure was needed that could free the assessment developers to move forward in an agile way. Second, the assessment system had become bifurcated into on-line fixed response tasks, and off-line complex “skills-based” tasks that required the completion of integrated tasks with numerous steps. While the skills-based exams were widely believed to have high content validity, the logistics of scoring students on exams that required dedicated networking hardware for each student were formidable. In addition, the variability in instructor proficiency throughout the world led to variability in the detail of the feedback given to students and possibly the quality of scoring.

Given these influences, it was determined that the goal of the research project would be to build an on-line performance based assessment prototype that provided rich diagnostic feedback to the student along with summative information as well. A secondary goal was to “field test” the language and process of ECD and explore strengths and weakness of this framework which had previously been described primarily in theoretical terms.

### The ECD Framework

ECD is a relatively comprehensive descriptive and proscriptive framework for assessment activity (Mislevy, Steinberg, & Almond, 2003). It describes end-to-end processes in a number of conceptual layers. The top layers concern cataloging and describing the domain of

interest. This includes domain analysis, which addresses questions concerning the primary representations, categories, and features of a domain. Next, domain modeling concerns setting scope around objects, ideas, and representations to address specific assessment goals. This is frequently done in the context of ordering and prioritizing claims an assessor wants to make about examinees. Aspects of domain analysis and domain modeling described in the subsequent paper by Williamson et al. (in submission).

The middle layer of assessment activity concerns the Conceptual Assessment Framework (CAF) which specifies submodels associated with task features (the task model), scoring rules and their accumulation (evidence models), and specification of the summarization of evidence to support characteristics of evidence about claims (the student model). The paper by Williamson et al. (in submission) walks the reader through the construction of many of these models based on the domain analysis and modeling activity that was conducted. DeMark and Behrens (in submission) discuss explorations in the use of statistical techniques to analyze information that may help guide the development of scoring rules for the evidence model.

While the first layer considers assessment from the perspective of the domain, and the second from the perspective of an evidentiary argument about claims, the third layer of the ECD takes assessment from the perspective through a four process delivery model. The four process model conceptualizes assessment delivery as having components for task selections (what do I show the examinee next?), presentation (what do I show, how do I show it, and what do I get back?), response process or evidence identification (what did they do, and how do I summarize it) along with a statistical summary which is also called evidence accumulation (how do I incorporate the evidence). Levy and Mislevy (in submission) discuss advances in the application

and computation of Bayesian inference networks to support sophisticated evidence accumulation.

### Highlights of Lesson's Learned

While the papers that follow report on specific technical aspects of lesson's learned in this endeavor, we want to begin by highlighting two lessons that are most salient from this experience: the value of the language of ECD and the impact of working in an exclusively computer environment.

#### *Language Matters*

Learning Evidence Centered Design can be an intimidating experience, in part, because it proposes a detailed language to describe many of the aspects that are common to standard assessment practice. Accordingly, it is easy to try to “translate” the ECD concepts into older more familiar concepts. Such a fall back to earlier language, however, is likely to lead to a fall back to earlier conceptualizations. During the year and one-half of the bulk of the NetPASS project, we frequently found the consistent use of the terminology of ECD gave us a language to conceptualize and articulate possibilities that could not have been brokered with older language. For example, the four process model described above, and illustrated in Figure 1, specifies the output of the presentation process is a work product and the output of response processing is an observation, and that multiple observations can be made on the work product by different rules.

While this language subsumes the traditional multiple-choice “item” which has the chosen option for a work product, matching for the response processing process, and correctness for the observation, it can do much more. Consider what happens when the presentation process becomes more complex. For example, a task with the goal of troubleshooting a network may take anywhere from 15 minutes to 1 hour to complete. What should be “scored”? The complex

system designer will reply that “that depends”. The work product can be the list of all the commands the examinee entered, the final functioning of the network, or the final configuration of the network devices (independent of their functioning). Having started with a rich and complex presentation process, the designer will likely seek a rich and complex work product (or set of work products), and likewise like to apply a rich set of rules to the work products. Lastly, after all that effort, the information out of the evidence identification/response processing activity should be much richer than a simple binary bit of information (0/1). Rather, we would like to see numerous characterizations of the work with numerous levels of detail in the characterization.

In this world view, evidence identification is not simply a fancy expression for item scoring, but a replacement that allows a broader conceptualization of possibilities. If assessment designers consistently employ the older language they are likely to work under, its assumptions and fail to see the full range of possibilities suggested by ECD. If we think of activities as items, we are likely to think of their results as indications of correctness, missing the opportunity to see tasks as complex activities that can be scored along multiple dimensions of goodness (or badness!), with multiple pieces of information coming out of them. Indeed, numerous assessment computing systems have greatly limited themselves by considering “items” the smallest unit of analysis, thereby failing to take advantage of more fine grained architectures that can score and store numerous pieces of information from a single complex activity.

### *Computers Change Everything*

While many assessment activities are faced with the transformation from paper-based to computer-based, the NetPASS project benefited from the constraint of being entirely computer based. This meant that the capture of complex digital information was a foregone conclusion

and careful consideration of the representations of the world using digital information would run throughout the project. Such a setting allowed the complex tasks described by Williamson et al. (in submission), the complex analysis described by DeMark & Behrens (in submission), and the complex evidence accumulation described by Levy and Mislevy (in submission). Here again, the language of ECD suggested for modularity computer architecture that matched the distributed computing allowed over the internet. In some cases, presentation was occurring over the internet at the examinee's desk top based on servers on the East Coast of the U.S. while scoring and reporting was occurring on servers on the West Coast of the U.S.. This isomorphism with the components of the four process model was not trivial, insofar as it allowed the designers to match computing expertise in different parts of the country with the assessment components. Other aspects of computer impact on NetPASS and the application of ECD have been discussed by Behrens, Mislevy & Bauer (2002) and Behrens, Reinert, Mislevy & Levy (in submission).

### Summary

ECD has been described thoroughly in a number of papers (e.g., Almond, Steinberg, & Mislevy, 2002; Mislevy, Steinberg, & Almond, 2003) with a theoretical emphasis or limited application. The papers reported subsequently describe the application of that framework to the development of a research prototype to a global context of in-class and on-line instruction. As will be seen in the following papers, the complexity of ECD matched the complexity of the issues surrounding such an ambitious project and numerous lessons were learned in the transition from theory to practice.

The current age is an exciting time assessment, but requires assessment professionals to integrate the many aspects of the stories told here: knowledge of the domain, a framework for describing it usefully, a model for task features, the accumulation of information and the

deployment of the assessment to the world. The world is complex, quality assessment is relatively complex, and ECD provides a structure and language to deal with the complexity when no other framework can provide such clarity and structure.

## References

- Almond, R. G., Steinberg, L. S., & Mislevy, R. J. (2002). Enhancing the design and delivery of assessment systems: A four-process architecture. *Journal of Technology, Learning, and Assessment*, 1(5). Available from <http://www.jtla.org>.
- Behrens, J. T., Mislevy, R. J., Bauer, M. (2001, August). Future Prospects for Assessment in the On-line World. Paper presented at the University of Maryland Conference on Assessment and Cognition, College Park, MD.
- Behrens, J. T., Reinert, P., Mislevy, R. J., Levy, R. (in submission). *Assessment design in an age of digital information and computing*.
- DeMark, S., & Behrens, J. T. (in submission). Using statistical natural language processing for understanding complex responses to free-response tasks. *International Journal of Testing*.
- Levy, F., & Murnane, R.J. (2004). *The new division of labor: How computers are creating the next job market*. Princeton, NJ: Princeton University Press.
- Levy, R., & Mislevy, R. J. (in submission). Specifying and refining a measurement model for a computer based interactive assessment. *International Journal of Testing*.
- Mislevy, R.J., Steinberg, L.S., & Almond, R.G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1, 3-62.
- Murnane, R.J. , Sharkey, N.S. & Levy, F. (2004). A role for the internet in American education? Lessons from the Cisco Networking Academies. In P. A. Graham & N. G. Stacey (Eds.), *The Knowledge Economy and Postsecondary Education: Report of a Workshop*. (pp. 127-158). Committee on the Impact of the Changing Economy on the Education System, National Research Council. Washington, D.C.: National Academies Press.

Porter, M. E., & Kramer, M. R. (2002). The competitive advantage of corporate philanthropy.

*Harvard Business Review*, 80, 12, 56-68.

Williamson, D. M., Bauer, M., Steinberg, L. S., Mislevy, R. J., & Behrens, J. T. (in submission).

Design rationale for a complex performance assessment. *International Journal of Testing*.

Figure 1.

