



USING EVIDENCE-CENTERED DESIGN FOR DEVELOPING VALID ASSESSMENTS OF 21ST CENTURY SKILLS

Rim Razzouk, Ph.D.

August 2011

Abstract

Today's digital world requires that students have opportunities to develop critical 21st Century skills such as communication, collaboration, creativity, innovation, and higher-order thinking. It is imperative that educational systems provide students with relevant opportunities to develop these skills through incorporating learner-centered methods (e.g., problem-based learning) of learning. However, assessing those skills is beyond the ability of the traditional assessments (e.g., multiple-choice tests). As a result, innovative educational assessments that aim to measure those skills are needed. One of these innovations is the Evidence-centered design (ECD) framework. ECD is a powerful process for designing and developing excellent (i.e., valid and reliable) assessments of 21st Century skills, and is especially suited for assessments that involve complex competency models and dynamic, interactive environments. ECD provides a means of reasoning about assessment design, and about learner performance. ECD has many advantages; however, it can only lead to valid and reliable assessments if thoughtfully applied.

Introduction

Today, teaching and assessing our students are taking place in a rapidly changing world. The world is shrinking and the technological revolution has created a boundless space of interconnected information and communication. Thus, to effectively function in this environment requires our students to acquire [21st Century skills](#) such as critical-thinking, problem-solving, systems-thinking, communication skills and others. Unfortunately, traditional assessments such as multiple choice tests are inappropriate for assessing or guiding learning in today's world. Educators need to be able to accurately and reliably measure these skills, systematically evaluate them, infer students' success, and continuously improve the development of these 21st Century skills through designing right tasks based on real data (Behrens, Mislevy, DiCerbo, & Levy, 2010; Shute, Hansen, & Almond, 2008).

This short paper is based on a working example designed to explain, apply, and determine the advantages of the evidence-centered design process (See Shute, Kim, & Razzouk, 2010 <http://myweb.fsu.edu/vshute/ECD%20for%20Dummies/ECD%20for%20Dummies.swf>). The evidence-

centered design (ECD) approach to assessment design was introduced by Mislevy, Steinberg, and Almond (2003). ECD is an integrated approach to constructing educational assessments in terms of evidentiary arguments that can be used to improve the validity and maximize efficiencies in the design, development and delivery processes. The following section provides an overview of the ECD approach while describing the assessment design process that includes competency, evidence, and task models.

Evidence Centered Design (ECD)

ECD builds on the vision of Samuel Messick (1994), “the nature of the construct being assessed should guide the selection or construction of relevant tasks, as well as the rational development of construct-based scoring criteria and rubrics.” It is a principled framework for designing developing, and delivering valid assessments. ECD is a systematic approach to the design of assessments that focuses on the evidence (student performance and products) of proficiencies as the basis for constructing assessment tasks. ECD has two main functions. It provides a way to reason about assessment design, and a way to reason about learner performance. It can be used to design assessments of all kinds, and is especially suited for assessments that involve complex competency models and dynamic, interactive environments that lie beyond the analytic capabilities of simpler assessments.

The key idea of ECD is to specify the structures and supporting rationales for the evidentiary argument of an assessment. By making the evidentiary argument explicit, it becomes easier to examine, share, and refine. Argument structures encompass, among other things, the claims (inferences) one wishes to make about a student, the observables (performance data) that provide support for those claims, the task performance situations that elicit the observables from the students, and rationales for linking it all together (Mislevy, Steinberg, & Almond, 2003; Shute et al., 2008). The three main models used in ECD for student assessment include: the competency model, the evidence model, and the task model.

- **Competency/Student Model: What collection of knowledge, skills, and other attributes should be assessed?**

A given assessment is meant to support inferences for a specific purpose, such as grading and providing diagnostic feedback. The competency model describes the knowledge, skills, and other attributes about which inferences are intended, the context under which the ability is demonstrated, and the range and relations of competencies in the knowledge domain (Shute & Torres, 2011). This model shows the variables needed to measure and reflect aspects of students' proficiencies (Snow, Fulkerson, Feng, Nichols, Mislevy & Haertel, 2010).

- **Evidence Model: what behaviors should reveal different levels of the targeted competencies?**

The evidence model bridges the competency model and the task model. This model defines the evidence that is needed to support the above-mentioned claims. Evidence models describe what is to be scored, how to score it, and how to combine scores into claims. It consists of two parts: evidence rules, which include rubrics or scoring models and statistical models, which range from a simple number-right scoring to the use of complex probabilities calculations. These models establish the boundaries of performance and identify observable actions that are within those boundaries (Mislevy et al., 2003).

- **Task/Activities Model: What tasks/situations can elicit the behaviors that make up the evidence?**

The task model describes the environment in which students say, do or make something to provide evidence of their proficiencies (Snow et al., 2010). The purpose of this model is to identify tasks, features of tasks, and/or activities that are able to elicit the evidence defined in the evidence model. Task models specify the inputs required to perform the observable actions as well as the work products that result from performing the observable actions (Shute et al., 2008).

Figure 1 demonstrates the flow between the differing ECD models.

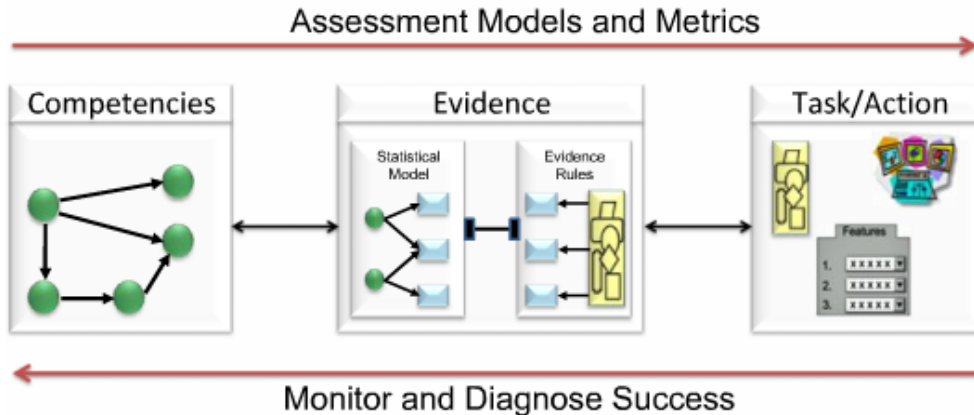


Figure 1. Evidence centered design model (adopted from Mislevy et al., 2003; Shute & Torres, 2011).

Why ECD?

Collecting the right information from assessments that help to make accurate inferences about students' competencies is critical because accurate inferences that support smart, right decisions can promote learning. The ECD approach yields accurate inferences since it focuses on assessing learning processes, authentic knowledge and skills, and what learners understand and can do; and focuses less on learning outcomes, declarative knowledge, and what learners do not know (Shute, Kim, & Razzouk, 2010). ECD is an effective framework for new assessments because it:

- Can yield valid assessments for different purposes
- Can aggregate information from various sources (e.g., qualitative and quantitative data)
- Provides accurate estimates of complex competencies and other unobservable variables and performances
- Allows linking specific performance data, which is observable, to theoretical constructs that are unobservable, thus, affording transparency and accountability to stakeholders.

Conclusion

The last two decades have seen innovations and changes in the world that require our students to acquire critical 21st Century skills to solve complex problems and to survive in an interconnected, digital environment. Assessing those skills, however, is beyond the scope of traditional and multiple-choice tests. As a result, innovative educational assessments that aim to measure those skills are beginning to evolve. ECD has proven to be a powerful framework for developing valid and reliable tasks and assessments of 21st Century skills (e.g., Behrens et al., 2010; Shute & Torres, 2011; Snow et al., 2010). However, it is a non-prescriptive and can only lead to valid and reliable assessments if thoughtfully applied.

References

- Behrens, J. T., Mislevy, R. J., Dicerbo, K. E., & Levy, R. (2010). *An evidence centered design for learning and assessment in the digital world* (CSE Technical Report 778). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Center for the Study of Evaluation, UCLA. Retrieved from <http://www.cse.ucla.edu/products/reports/R778.pdf>
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Education Researcher*, 23(2), 13-23.
- Mislevy, R.J., Steinberg, L.S., & Almond, R.G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1), 3-62.
- Shute, V. J. & Torres, R. (2011). Where streams converge: Using evidence-centered design to assess Quest to Learn. In M. Mayrath, J. Clarke-Midura, & D. H. Robinson (Eds.), *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research*. Charlotte, NC: Information Age Publishing.
- Shute, V. J., Hansen, E. G., & Almond, R. G. (2008). You can't fatten a hog by weighing it—Or can you? Evaluating an assessment for learning system called ACED. *International Journal of Artificial Intelligence and Education*, 18(4), 289-316.
- Shute, V. J., Kim, Y. J., & Razzouk, R. (2010). ECD for Dummies. In J. Gee & D. Davidson (Eds.) *Working Examples*, <http://myweb.fsu.edu/vshute/ECD%20for%20Dummies/ECD%20for%20Dummies.swf>
- Snow, E., Fulkerson, D., Feng, M., Nichols, P., Mislevy, R., & Haertel, G. (2010). *Leveraging evidence-centered design in large-scale test development* (Technical Report 4). Menlo Park: SRI International, Center for Technology in Learning. Retrieved from http://ecd.sri.com/downloads/ECD_TR4_Leveraging_ECD_FL.pdf