Assessment-based Gaming Environments

Diego Zapata-Rivera

ETS, Princeton, NJ

Paper presented at the annual meeting of the American Educational Research Association (AERA) and the National Council on Measurement in Education (NCME)

April 13-17, 2009, San Diego, CA.
Abstract

This paper describes an approach to creating assessment-based gaming environments that builds on Evidence-Centered Design (ECD) principles (Mislevy, Steinberg, & Almond, 2003). BELLA, an assessment-based gaming environment for English vocabulary and math is used to illustrate this approach. Results from a small usability study are also presented.
Introduction

Educators have long recognized the ability of games to capture the attention and imagination of students. Many authors (i.e., Gee, 2003; Prensky, 2001; Shaffer, 2006; Squire 2002) posit that online games represent an educational breakthrough that could transform how people learn. However, integrating valid assessment activities and feedback into gaming environments without disrupting the “flow” experience (Csikszentmihalyi, 1990) is a big challenge.

We have addressed this issue by following an evidence-based methodology that reconciles the need for obtaining valid assessment information and creating an engaging environment that children want to use. Evidence-Centered Design (ECD) (Mislevy, Steinberg, & Almond, 2003) is a methodology for assessment design that emphasizes a logical and explicit representation of an evidence-based chain of reasoning from proficiencies to tasks.

We have applied ECD principles to the design and development of assessment-based gaming environments. These environments are composed of various interactive activities (assessment tasks) that are developed to explicitly gather evidence of behavior that demonstrates proficiency (or the lack of it) on previously defined constructs. BELLA, a game-inspired assessment and learning environment aimed at supporting vocabulary learning in academic subject areas (i.e., math), is used to illustrate this approach. BELLA integrates assessment and learning into an interactive gaming system that includes written conversations, math activities, oral and written feedback in both English and Spanish, and a visible psychometric model that is used to adaptively select activities, and feedback levels. This paper describes how ECD principles were used to create internal assessment
models in BELLA. We also report on a small usability study carried out in a public middle school in New York City.

**The Approach**

The main goal of this process is to design assessment-based gaming scenarios that can be used to help students learn the content while at the same time help the system capture valid assessment information to adapt its behavior. An assessment-based gaming environment should be designed so it can respond to the following questions posed by Samuel Messick (1994): (a) *What complex of knowledge, skills, or other attributes should be assessed?* (b) *What behaviors or performances should reveal those constructs?* And (c) *What tasks or situations should elicit those behaviors?*

**Figure 1.** The three main ECD models

ECD is a methodology for assessment design that emphasizes a logical and explicit representation of an evidence-based chain of reasoning from proficiencies to tasks and vice versa. ECD helps us design assessments that can respond to the questions above. We followed an ECD-based approach in designing and developing assessment-based gaming environments. ECD three central models include (see Figure 1): (a) *Competency models* (CM) that define the knowledge, skills, and abilities (KSAs) that need to be measured in the game. These KSAs are clearly linked to claims about the student (i.e., assessment
claims), and are defined to support the scores and sub-scores in the game; (b) *Evidence models* (EM) that describe how evidence about student performance on particular tasks or game activities is used to support assessment claims. Evidence models are clearly linked to the KSAs in the CM. The evidence model is implemented using evidence rules (i.e., a scoring component or rubric) and a measurement or statistical model; and (c) *Task models* (TM) that describe the assessment tasks or game activities, including a specification of the observables (e.g., records of mouse clicks, actions selected, etc.) to be produced by the student. The range of permissible values or settings for a game activity feature may be specified, thus constituting a model of a class or family of activities.

Logically, these assessment models would be defined before individual activities (tasks) are designed. Appropriate task models capture all salient work product features that have been specified in the EM. Task models are linked to the CM via evidence models. These task model designs must take into consideration the characteristics that will affect validity and difficulty of the activities, operational constraints (e.g., scoring capacity and time), and game related aspects that will keep students engaged in the game (e.g., immediate feedback and progress indicators).

Evidence-based game scenarios are composed of various interactive activities (i.e., assessment tasks). Each scenario has an underlying storyline defining: (a) the behavior that we want to observe based on the CM and EM, and (b) the interactive activities that we need to elicit such behaviors (TM). Figure 2 outlines this design and development process. This process requires the input from an interdisciplinary team including users (i.e., students or players), domain experts (e.g., teachers and researchers), assessment specialists and interactive design experts.
Steps for designing evidence-based, gaming scenarios are as follows:

- **Gathering domain knowledge information (Step 1).** Gather domain knowledge information from a variety of sources, including an alignment document (i.e., curriculum standards and competencies), and text books, which provide information regarding how the topic is introduced, taught, and assessed.

- **Designing initial competency and evidence models (Steps 2a and 2b).** Define the competencies of interest and relationships among them (see Figure 2, CM) and
describe initial evidence. That is, what do you want the students to do in order to demonstrate their competency on the above mentioned competencies?

- **Selecting competencies and required evidence (Step 3).** Select a competency or a group of competencies and corresponding pieces of required evidence from initial CM and EM models.

- **Generating new ideas (brainstorming sessions) (Step 4).** Brainstorm about scenarios and activities that can be used to elicit desired behavior (i.e., these activities/tasks will be used to gather evidence that will inform the EM). Users (i.e., players) are welcome to participate in some of these brainstorming sessions.

- **Describing scenarios and activities (Step 5).** This step involves describing each task/activity and defining how work products will be assessed (e.g., using scoring rubrics and the number of points). It includes the following actions: describe each task/activity (i.e., what students need to read and do, the type of response desired, and how target words will be integrated into the activity); define the role of the student (e.g., how students interact with pedagogical agents in the game); describe the role of the teacher (what kind of supporting role the teacher plays); define the role of the pedagogical agents (e.g., what level of feedback [or scaffolding], coordination, or assessment activities will be provided to the learner [depending upon his/her knowledge level] after each activity is completed); establish work products for each activity; and describe the evidence rules for the task/activity (e.g., task rubrics and number of points). Each activity needs to be carefully defined in order to seamlessly capture the assessment information that is needed, without sacrificing the flow of the game. In some cases, new scenarios need to be
designed in order to create an engaging sequence of actions. This tradeoff between creating engagement, learning interactions and gathering valid assessment information is an interesting research challenge that we have just started exploring.

- *Updating task models (Step 6).* Once an activity is described, information about its characteristics (e.g., nature of the interaction, observables, expected work product, rubric, and number of points) is kept as part of a task model.

- *Updating competency and evidence models (Step 7).* Once information regarding activities for a particular scenario is available, changes to the EM and the CM need to be made to reflect the most recent version of each scenario. Changes could include the addition or deletion of pieces of evidence linked to one or more competencies.

This is an iterative process which results in a blueprint of how assessment and interactive learning components are integrated in the assessment-based gaming environment. Once a scenario is described, interactive design experts and system developers can implement a prototype of it, pilot test it with users, and make changes based on the feedback that is gathered. It is possible that some of these changes would require updating the TM, EM or CM. BELLA is an example of applying this evidence-based, scenario design process.

**BELLA**

The game starts when the student chooses and customizes a character with which to play the game. The student also selects a friend who will accompany the character while
playing the game. BELLA takes place in a virtual city. The student’s mission is to help
his/her student character interact in the virtual city. In this virtual environment, the
student character is invited to visit friends and to participate in various activities (e.g., a
pizza party, a birthday party, a visit to a museum). Each activity provides an integrated
learning and assessment scenario for the student. As part of each activity, the student
character interacts with virtual people (e.g., the mother of one of his/her friends) who
provide guidance, feedback, and, at the same time, administer embedded assessment
tasks to the learner related to predefined vocabulary and math proficiencies. Evidence of
student knowledge is obtained through the student’s interaction with these characters and
his/her performance on various math and vocabulary activities. As the learner advances
in the game, s/he accumulates points for his/her student character. In addition, knowledge
level estimates (i.e., power levels) are continuously updated based upon performance in
the game and are visible to the learner through his/her virtual cell phone. These
knowledge levels are externalized as progress bars (one for vocabulary and one for math)
and are referred to as the student character’s power levels, implementing an indirectly
visible student modeling approach (Zapata-Rivera, 2007).

**Competency Model (CM)**

While ESL teachers usually teach everyday vocabulary (Tier I) to their students,
mainstream teachers focus their attention on domain specific vocabulary (Tier III).
Teaching of Tier II words is often neglected by teachers. Tier II words and expressions
have been classified by Calderón (2007) into importance and utility words or those high-
frequency words that appear in many domains (e.g., represent, investigate, expression),
polysemous words or those words that change meaning depending on context (e.g., trunk,
table, root), cause-effect words (e.g., as a result, due to, so that, therefore, thus), contrast words (e.g., although, in contrast, nevertheless, on the other hand), and addition and comparison words (e.g., as well as, likewise, moreover). The target student/player is expected to have basic conversational and reading knowledge of Tier I, everyday vocabulary. Because the scenarios are set in the context of mathematics, specialized Tier III words are also used. The focus of BELLA, however, is to teach and reinforce Tier II words. The BELLA CM was developed to represent the cognitive processes students must use to access information, solve problems, and communicate results, both within the English and Mathematics domains and between the domains. It has two main branches (math and language) that are linked through common cognitive processes (e.g., communication). The language branch shows vocabulary broken into Tier II and III words, which are the types of vocabulary used during the BELLA scenarios. The math branch shows various math skills related to fractions (e.g., representing fractions, equivalent fractions, simplifying, adding and subtracting fractions).

**Evidence Model (EM)**

Bayesian networks (Pearl, 1988) were used to implement BELLA’s internal statistical model. The Bayesian network expresses probability distributions among competencies and between competencies and tasks/activities. Thus, variables in this Bayesian model include English vocabulary and mathematics competencies, common cognitive processes, and tasks/activities. Tasks/activities are linked to competencies in the CM. This model takes into account task difficulty estimates and prior information about the target population elicited from ESL teachers and from word frequency analysis, to create initial
prior and conditional probabilities. Student data will be used to refine these initial conditional probability values.

Power levels (i.e., student’s knowledge estimates) are computed based on the Bayesian student model. These values are shown as progress bars in the game (one for math and one for vocabulary). Power levels are calculated based on the expected a posteriori (EAP) value of a particular node/competency (e.g., math or language) (Zapata-Rivera, 2007). Power levels and points earned act as two different indicators of progress in the game. Power levels are estimates of the student’s proficiency that fluctuate according to the student’s performance, as defined by the rules used to integrate new evidence into the Bayesian student model. Whereas, the points earned always increase in the game. These two indicators of progress can have different uses (e.g., teachers could be interested in knowing how much progress a student has made on a particular skill, as opposed to the number of points the student has accumulated).

**Task Model (TM)**

BELLA makes use of written interactive discourse completion tasks (IDCTs; Kuha, 1999), mathematics activities and pop-up questions. These embedded assessment activities have been integrated into a game scenario. IDCTs, the text that appears on the virtual cell phone screen, consists of messages from other characters in the scenario. New dialogues are developed by writing a script for one character to introduce a new math activity and its vocabulary. Other characters in the scenario use vocabulary that has already been introduced. Students can ask for additional feedback by clicking on highlighted Tier II and Tier III words. A student can choose to initiate a pop-up question by clicking on a character, or the system may initiate a pop-up activity to collect
additional evidence of student knowledge of a particular Tier II word or mathematics concept.

Mathematics activities were designed to assess knowledge of various concepts related to fractions that typically arise in real life. Each activity is aligned to the New York Mathematics Core Curriculum and linked to the concepts in the CM. Scenario activities were organized to show increasing cognitive skills related to students’ understanding of fractions, from basic understanding to advanced applications. Mathematics pop-up questions aligned to each activity were developed to provide additional evidence of students’ mastery of mathematics concepts. Each activity has a rubric that is used to implement the evidence rules that connect them to the EM.

**Usability Study**

A small usability study of BELLA was carried out in 2007. Participants were 9 seventh graders from a public middle school in New York City. Students answered a Tier II word familiarity survey (self-assessment, 10 minutes), played BELLA (1 hour), and answered a short usability survey (10 minutes) and a demographic questionnaire (5 minutes). Each student had his/her own laptop and headphones to use during the session. Results of the usability study showed that, overall, students found the game engaging and challenging. Eight of the students agreed that they liked the story (i.e., Pizza Party) as well as the pop-ups (mathematics and vocabulary). Seven of the nine students liked how the BELLA game looked. All nine of the students liked selecting their character’s face and being able to give their character a name. Eight of the students agreed that they liked helping their character and talking to Mom. All nine students thought that they had learned vocabulary and eight thought that they had learned mathematics using BELLA. When asked if they
would be willing to play the game again, all nine students agreed. When asked what they would add to the game, students suggested adding additional characters and more interactive activities. Teachers also expressed their support for the tool and provided valuable feedback on how to integrate it into their practice.

Conclusion and Future Work

This paper described an approach for building an assessment-based game that combines assessment activities and instructional feedback. This gaming environment is aimed at helping students learn mathematics at the same time that it supports the development of English language skills. Our evidence-based, scenario design process proved to be instrumental in helping us combine game and assessment requirements. This process resulted in assessment models, storyboards, and the final Bayesian student model that acts as the inference engine of the application. We are planning to start exploring student learning effects in controlled contexts. Future work also includes the development of additional ECD-based gaming scenarios in various domain areas.
Acknowledgments

We would like to thank the members of our team: Waverely VanWinkle, Alyssa Buteux, Bryan Doyle, Malcolm Bauer, Thomas Florek, Janet Stumper and Kristine Jacobson-Wiley. We would also like to acknowledge our external expert: Professor Margarita Calderón, John Hopkins University, and John Young and Marjorie Biddle for supporting this work. We would like to acknowledge the teachers and students who participated in our study.
References


Retrieved from the Web on January 27, 2008, at:

http://www.gamestudies.org/0102/squire/