Analyzing the educational potential of existing games using Evidence-Centered Design and Cognitive Task Analysis

Diego Zapata-Rivera

Eric G. Hansen

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 analyzing the educational potential of existing games and adapting existing games for educational purposes that makes use of Evidence Centered Design (ECD) principles and Cognitive Task Analysis (CTA) techniques. The main outcome of this approach is the development of arguments for the quality of educationally enhanced games. A game activity based on a commercial game is used to illustrate this approach. Results from a small-scale usability study are also presented.
Introduction

Many researchers agree that those who play are learning something, yet there is far less agreement on what players are learning. There are still relatively few studies that rigorously investigate claims about learning benefits of game play (Goodman, Powers, & Hansen, in preparation). The most consistent learning outcomes seem to be related to the improvement of visual-motor and visual-spatial skills in specialized domains rather than general academic skills (Tobias & Fletcher, 2008). Games can be used for educational or entertainment purposes. Games might prove useful as assessments, though again for specialized skills. For example, Rosser, Lynch, Cuddihy, Gentile, Klonsky, and Merrell (2007) found that video game skill accounted for the greatest amount of variance in laparoscopic proficiency and accumulated past video game experience accounted for additional variance (cited in Tobias & Fletcher, 2008). These findings underscore the importance of focusing on specific skills or attributes and the importance of explicating whether the intended educational benefit is to be derived from the game functioning as assessment, learning-intervention, or perhaps both.

It seems clear that a research-based process that can be used to identify the learning potential of existing games would provide information relevant to both educators and game developers. This paper describes a process that relies on Evidence-Centered Design (ECD; Mislevy, Steinberg, & Almond, 2003) and Cognitive Task Analysis (CTA; e.g., Jonassen, Tessmer, and Hannum, 1999; Crandall, Klein, and Hoffman, 2006) to analyze the educational potential of existing games in terms of learning effectiveness, learning efficiency, engagement, assessment validity, and accessibility. One game (i.e., Roller Coaster Tycoon 3) was used to illustrate this approach. This process results in a list of
argument-based assessment tasks (or game activities) that can be used to improve the validity of the game and the opportunities for successful deployment in educational settings.

**Evidence Centered Design**

Evidence-Centered Design (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. ECD helps design assessments that can respond to the 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? ECD has recently been extended to include work that seeks to ensure accessibility to special populations without undermining the validity of assessment results (Hansen & Mislevy, 2006). The work on ECD for accessibility of assessments was described extensively in a National Research Council report on accommodations in large-scale assessments (National Research Council, 2004, chapter 6).

ECD principles were used in this study to inform the expert knowledge elicitation process that resulted in argument structures for a selected game scenario representing (a) the targeted competency to foster or measure, (b) the person’s characteristics (both focal [targeted] and nonfocal [nontargeted]), (c) the intent for growth or learning, and (d) the cognitive and other demands imposed on the person by the task situation.

**Cognitive Task Analysis**

CTA has been used to elicit information about cognitive processes from individuals, represent it, analyze it and make use of it. CTA techniques have been used successfully in
many areas, including: the design of new devices or interfaces, gathering software requirements from users, doing market analysis, integrating technologies into organizations, and designing instructional systems.

In this approach, CTA was used to gather information from experts regarding the educational potential of various games, to analyze the knowledge, skills, and abilities (KSAs) required by the games, and to elicit potential enhancements/modifications that could be used to better align them with academic curriculum standards.

**Analyzing existing games and selecting a game scenario**

A team of five experts was assembled that, in aggregate, had expertise in the areas of cognitive science, instructional psychology, mathematics, usability studies, assessment development, ECD principles, games, and teaching. The full team met for an hour almost weekly for several months.

Experts were asked to review various games (about 40 including entertainment and educational games) and report their findings regarding: relevance of the game to valued academic constructs, ease of access to usage data, popularity and other evidence of high student engagement, age appropriateness (middle school students) and novelty.

Based on the criteria above, ten out of the forty games were selected for further examination. Experts were assigned two games each and asked to prepare a presentation/demo of the game describing how the game could be used for educational purposes. In particular, experts were asked to answer questions such as: Is there an explicit educational intent in design? If yes, what is the purpose (learning, assessment or both)? Also, if yes, what constructs are measured or fostered? What key observables are currently available (e.g., scores currently reported)? What other KSAs must the student
possess in order to perform well under typical conditions? How does the game rely on parents, teachers, and mentors (if at all)? How could the game's educational potential be enhanced? What would be the nature of the enhancements? What observables (beyond those currently available) might plausibly be collected (e.g., timing data, keystrokes)? How readily could the enhancements be made and validated? For example, does the game provide scripting or modding (modification) capabilities? What issues might arise in enhancing the game to improve qualities such as engagement, accessibility, assessment validity, learning effectiveness and efficiency?

Experts answered the questions and presented the games. Each session was followed by a discussion comparing the strengths and weaknesses of the games regarding their educational potential. Information was collected, analyzed and used to select a game to illustrate the enhancement process (Roller Coaster Tycoon 3).

**Enhancing a game scenario**

A game scenario can be seen as a problem solving activity. CTA and ECD were used to analyze the KSAs required to successfully accomplish the goal of the scenario. This process is iterative. It requires experts to clearly understand the game scenario and the KSAs that will be the focus of the enhanced game activity; identify situations (tasks) in the game that can be used to gather evidence of these KSAs; and define a sequence of steps, additional materials and scaffolding that will be provided to the player when interacting with enhanced game activity. This information is represented in the form of an argument structure (see Figure 1). This argument structure is modified during the elicitation process. Figure 1 represents the status of the argument before the usability
study had been carried out. Data gathered from the usability study will be used to refine the argument to make it more accurate and defensible.

Selecting KSAs

Given the nature of the game (i.e., a business [amusement park] tycoon simulation game), several ideas about how to use the affordances provided by the simulation to address particular curriculum standards were elicited from experts. From these ideas we selected an activity aimed at exploring the relationship between ticket price and weekly profit.

This scenario offered various opportunities for designing activities directly linked to middle school math standards. Target (focal) KSAs included: (a) collecting data, (b) plotting data, and (b) describing and explaining relationships among variables. These KSA’s are aggregated into a single targeted proficiency (i.e., Analyze data).

Developing an argument structure

Among the key parts of the argument structure is the “Focal Value” column (see Figure 1), which defines the targeted proficiency being fostered. Other key parts of the argument are broken into sub-arguments, one for each of three main activities (Activities 1, 2, and 3). The primary purpose of Activity 1 is to ensure that the player can satisfy the nonfocal requirements (e.g., requirements [cognitive or other demands] for knowing the RCT3 environment, knowing about the scenario topic, and knowing about the format of the paper-based templates used in this activity). The primary purpose of Activity 2 is to foster growth in the three focal KSAs (collect data, plot data, and describe and explain the relationship between price and profit). Finally, Activity 3 has the primary purpose of
assessing the overall targeted proficiency (“Analyze data”). Additional information (not shown in Figure 1) includes mechanisms for gathering initial profile information (e.g., a screener questionnaire or a pre-assessment), and a post-assessment. Please note that for a given KSA, generically, the 0 stands for low (or absent) and the 1 stands for high (or present).

Having defined the values for the matrix, rules are applied to determine if the matrix meets technical quality criteria. Some of these rules follow: (1) “If for each focal KSA, the pre-experience profile (initial profile) value is at least as high as the focal value, then the person already has a high targeted competency value, which is an indication that the person does not need the intervention.” (2) “If for any nonfocal KSA, the pre-experience profile value is less than the requirement value, then this is an unsatisfied nonfocal requirement.” This rule provides a critical link to issues not only of accessibility but to a wide range of other issues where the learner may lack some critical prerequisite. (3) "If the purpose of an experience is to measure a targeted competency, then, for each focal KSA, the requirement value must be equal to the focal value." This rule provides a critical link to the domain of educational assessment. And (4) “If there is an intended growth outcome, then there must be a matching requirement value.” This rule is based on the assumption that in order for a person to grow a skill, it is necessary to impose a requirement (cognitive or other demand) for that skill. The argument for each activity in Figure 1 is coherent and of high quality in the sense that it meets key operational quality criteria related to accessibility, learning effectiveness, and validity.
Figure 1. Argument structure for an enhanced game activity based on Roller Coaster Tycoon (fragment)

Developing a protocol for the enhanced game activity

In order to demonstrate this approach, a research protocol was developed. This protocol included a screener questionnaire, an intervention (activity 1 for addressing “access
skills” [nonfocal KSAs] and activity 2 for addressing the skills targeted for growth [focal KSAs]), and a post-assessment (activity 3). A script of each of the activities was produced based on its sub-argument structure (see Figure 1). Thus, for example, activity 1 provided practice opportunities aimed at helping players get familiar with the game about how to use the game. Each activity was based on a storyline that required players to work on several tasks linked to one or more KSAs. Activity 2, for example, asked the player to help his/her parents (the owners of the amusement park) make the amusement park as profitable as possible by finding the most profitable ticket price. The script for Activity 2 guides the player through a process that involves changing ride ticket prices, running the simulation to see the reaction of the riders, inspecting weekly profit values, tabulating and plotting information and writing about his/her findings.

A small-scale usability study

The main goal of the study was to explore the feasibility of the methodology (i.e., how well the “resulting argument” was actually achieved) and to get initial usability feedback about the materials produced from participants. A screener questionnaire was used to identify possible participants. This questionnaire was instrumental in ensuring that some of the nonfocal requirements were met before using the game (e.g., participants were able to read and write at an appropriate level to understand and work on the tasks developed for the game).

Participants consisted of five middle-school students. Participants played the game for about 75 minutes. The game consisted of 3 tasks that were designed according to the parameters shown in the argument structure described in Figure 1. At the end of the
study, participants answered a final usability survey that took about 10 minutes. The proctor provided guidance and took notes during the study. The nature of feedback provided by the proctor was also carefully planned based on the goals of each section as outlined in the argument structure.

In general participants liked the enhanced game activity and valued the educational purpose of it. Participants mentioned that they would also like to play the original game and explore more choices. Additional issues that will be used to refine the argument structure include: (a) although most of the participants seemed to have accomplished the tasks, a participant who spoke English as a second language required additional support. Changes to the argument structure might include using simpler language for English language learners or possibly revising upward the minimum levels for language related KSAs; (b) Players could have solved problems without proctor/teacher assistance (Activity 2) if additional written materials (e.g., a glossary of terms) were provided. Based on the argument, players were expected to satisfy nonfocal requirements during Activity 1, which included a glossary of terms. However, we noticed that some players made good use of these materials during Activity 2 as well. The argument structure could be revised accordingly; and (c) the protocol did not handle every particular misconception that could hinder a player’s performance. The argument could be revised to specify instructional activities or educational content linked to particular misconceptions, or changes could be made to the delivery mechanism to warn the teacher about particular misconceptions.
Conclusions

This approach shows how ECD principles and CTA techniques can be used explore the educational potential of games and to create enhanced game activities that clearly link game performance to valued KSAs. The approach facilitates design decisions by providing a principled way for designing rich and authentic games activities that meet specific educational criteria. It results in a better understanding of the strengths and weaknesses of the game and helps guide decisions about how to improve its effectiveness.
Acknowledgments

We would like to thank the members of our team: Russell Almond, Irvin Katz, Heather Nadelman, Sarah Ohls and Liz Marquez. We also like to acknowledge the students who participated in our study.
References


