43.

Designing Games with Assessment in Mind

Diverse Processes of Integrating Design Thinking with Evidence-Centered Design

Abstract

Each year many papers are presented at GLS discussing the role of assessment in game design and the methods by which different games intend to support and measure educationally valuable knowledge and skills. While many of them claim to use ECD, it is often unclear how they leverage assessment to conceptualize game design around the competency of interest. This symposium aims to discuss and synthesize different processes and methods beyond ECD that are currently being applied by different groups.

Description of the Symposium

For the past several years, educational game researchers and practitioners have increasingly used evidence-centered design (ECD) as a framework to align what we value in students' learning with what students do in games. While many of them claim to use ECD, it is often unclear how they leverage assessment to conceptualize game design around the competency of interest. Some of the process-related issues that we rarely discuss include: How do you formalize assessment models? How do you translate formalized assessment models to game design elements? When in the game design process does this translation occur most effectively? How do you transform competency models into interesting, engaging game mechanics? How do you ensure psychometric qualities without being too prescriptive? This symposium aims to discuss and synthesize different processes and methods beyond ECD that are currently being applied by different groups. More specifically, we aim to unveil diverse methods and processes related to how design teams, often including learning scientists, subject-matter experts, and game designers, can seamlessly integrate design thinking and the formalization of assessment models. Each presenter will ground their talk around: (a) description of game development project(s) that each group is involved, (b) specific competencies/learning objectives aligned in the game, (c) the process and methods you employed to go from conducting domain analysis to creating engaging game elements and how rigorously or loosely ECD was applied in the process, and (d) encountered challenges and gained insights.

Designing Games for Competency-Based Teacher Education Program

Yoon Jeon “YJ” Kim, MIT Teaching Systems Lab, yjk7@mit.edu

At the MIT Teaching Systems Lab, we are using educational games to develop and assess pre-service teachers’ competencies in partnership with the Woodrow Wilson Academy of Teaching and Learning. The WW Academy will be purely competency-based, therefore, it is crucial to align learning and assessment tools with the learning outcomes. We focus on two games we are designing for this purpose: *Unrubric Rubrics* and *Eliciting Learner Knowledge (ELK)*.

Both games were inspired by the design team’s understanding of ambitious teachers’ competencies, yet the employed process for each game is somewhat different. Because *Unrubric Rubrics* has a heavier focus on assessment capacity than *ELK*, how “rigorously” ECD is applied varies between the two games. For *Unrubric Rubrics*, we conducted extensive domain analysis to fully understand futuristic and ambitious teaching practices, and developed a progression a new teacher might take to build this expertise. The analysis included literature review, subject-matter experts interview, and teacher interviews. The outcomes of this analysis include a set of formal assessment models that describe (a) operationalized competency model variables and (b) features of actions that we want to observe (therefore, designed interactions in the game) in relation to those competency variables. These formal models, then will be translated into core mechanics of the game, which will be refined via paper prototyping. For *ELK*, our domain analysis work began with considering the literature on student misconceptions in science, particularly in natural selection. We developed profiles for the “student players” reflecting common misconceptions about natural selection, adaptation, and evolution (Anderson et al, 2002).

A competency, by definition, is a combination of skills, abilities, and knowledge needed to perform a specific task that is objectively measurable (Klein-Collins, 2012). Therefore, capturing the demonstration of a competency in authentic contexts and designing those contexts that elicit evidence for the competencies requires careful design. The biggest challenge that we have been facing is the translation of our understanding of complex competencies (e.g., formal assessment models) to creative and engaging game elements. For both the *ELK Game* and *Unrubric Rubrics,* translating our understanding of competencies to game elements will require collaborative thinking among learning scientists and game designers.

Designing Cross-content Assessment Games on a Digital Learning Platform

Kevin Miklasz, BrainPOP, kevinm@brainpop.com

At BrainPOP, we are designing a series of “assessment games” or “playful assessments,” the most recent of which is called TimeZoneX (BrainPOP 2105). TimeZoneX is a game about sequencing historical events, similar to how the card game “Timeline” works. Each of BrainPOP’s assessment games are focused on a skill (in this case, sequencing) that can incorporate many different kinds of content (in this case, historical events from BrainPOP’s 800 animated videos).

For these games, our process generally starts with identifying a mechanic that will cut across content areas and is related to a general learning goal. Next we simultaneously design the game mechanics and content, while having conversations about assessment design. Once the design starts to become settled (but not finalized and immune to change), the role of assessment becomes stronger- we undergo a ECD-like discussion, and the results of that discussion leads to an initial exploratory analysis of the data to test assumptions. We then start to do some assessment-focused user-tests that get at issues of validity of our constructs. Next the design becomes pretty finalized, the game goes live, and we do additional assessment testing on the live game using thousands of anonymous playthroughs.

We use ECD mainly as a communication tool between the game designers and the assessment specialists. In the claims-evidence-tasks framework, the claims are more of the designers territory (what higher level actions the game design is trying to achieve), the tasks are more the assessment data territory (what are the individual events being captured during gameplay), and the evidence is where the two groups meet in the middle and reach understanding. The ECD process (which in practice is a 2-3 hour meeting with all staff on the project) often highlights hidden assumptions that must be tested about the game, and also highlights what data or clickstream events are expected to be most crucial and deserve the most attention in analysis- thus it guides the initial exploratory analysis, and the design of the assessment user testing.

Designing Games to Teach Middle School Science

David Gagnon, University of Wisconsin, david.gagnon@wisc.edu

I was the producer for a project to develop nine 10-20 minute “mini-games” to teach middle school science topics, called “The Yard Games.” The project rose out of a challenge to live in a “game jam” mode for a few months, finding new ways to develop educational games quickly and inexpensively. We were curious if the slow adoption of games in the classroom could be improved if we would: (a) develop small games that could be easily used and completed in one class period, (b) involve teachers as co-designers from the very beginning, (c) ensure that the games were free and open source so that others could remix or modify them at will, (d) tightly align each game with 1 or 2 Next Generation Science Standards.

The design team loosely used ECD as a way to describe and critique our level design, the ways that a player would progress through the game challenges. Levels in these simple games serve two purposes: 1. Incrementally introduce user interface elements and the high-level player role and goal. 2. Provide evidence for player competency of a topic so that the game can increase in complexity. By creating a tight coupling between competency demonstration with game progression, the hope is that we would be able to say that a player who had completed level X had demonstrated competency with concept Y. In theory, this should lead toward very playable games because the conceptual trajectory grows in complexity at a pace that the player can handle. It should also lead to games that would act as an assessment of player competency.

During the rapid playtest/iteration cycles each game undertook, this approach certainly helped us design better games, ones that were more playable as well as formatively demonstrate conceptual growth. At of the time of writing this proposal, we are preparing for a much more rigorous study that will measure correlation between traditional assessments of each concept with success rates for completing each related game level. Because the games are online and we are able to test several versions in parallel, our hope is that we will be able to rapidly iterate on level designs while the games are in production with a goal to increase correlation between game progress and traditional assessment. By doing so we will not only be able to inform a teacher that one of their students are struggling with a part of the game, but with the concept itself. Formative results from this process will be available by the time of the conference.

Formative Assessment of Implicit STEM Learning in Games

Jodi Asbell-Clarke, TERC, jodi_asbell-clarke@terc.edu

EdGE at TERC has studied Scientific Inquiry in an MMO game called Martian Boneyards, Physics learning in Impulse and Quantum Spectre, Biomechanics and relations to the Natural World in Ravenous, and we are currently studying computational thinking in Zoombinis. We design our games with mechanics that are grounded in STEM phenomena such as Newton's Laws of Motion, Optics, or Algorithmic thinking. We bake the data collection into the game as so that the game mechanics, learning mechanic, and assessment mechanics are inherently aligned.

We observe players in the game, using a combination of screen (and mouse) capture, audio, and video recordings so that we see and hear the strategies that players define for themselves as they play the game. We have multiple researchers code the videos (establishing reliability) for pertinent strategies (moves and behaviors that align with the STEM learning goals) that emerge from players own gameplay and explanations. We then build learning analytics (such as data mining detectors) to detect those same strategies in the game logs. In this way we find evidence of implicit STEM learning in the game. This model does not use a prescribed task model, as often found in ECD, but rather focuses on emergent strategies that come from the learner and demonstrate implicit learning that may not yet be formally expressed by the learner.

Because we are studying implicit game-based learning, we are not relying on formalisms that are easily translatable to school based summative assessments. Our focus is on *how* players build knowledge in the game, and our studies show that how they play makes a difference in their learning. Our game-based learning model relies on a teacher to bridge the implicit learning to explicit (more easily measured) learning in the classroom. We have gained insights on measuring implicit learning that is not yet expressed by learners in school's formalized language, yet is foundational for their learning process (Polanyi, 1966). We exploit game-based learning analytics to dig deep into the learning process and reveal insights that are beyond what can be seen by traditional assessments.

References

Anderson, D. L., Fisher, K. M., & Norman, G. J. (2002). Development and evaluation of the conceptual inventory of natural selection. *Journal of research in science teaching*, *39*(10), 952-978.

Klein-Collins, R. (2012). Competency-based degree programs in the US: Postsecondary credentials for measurable student learning and performance.