Resolving Conflicts in Educational Game Design through Playtesting

Brian Winn, Assistant Professor, Dept. of Telecommunication, Information Studies, and Media, Michigan State University, winnb@msu.edu

Carrie Heeter, Professor, Dept. of Telecommunication, Information Studies, and Media, Michigan State University, heeter@msu.edu

Introduction

The multidisciplinary nature of educational game design requires multidisciplinary teams. Mishra and Koehler (in press) define "technological pedagogical content knowledge" (TPCK) as an emergent form of knowledge that goes beyond all three individual components (content, pedagogy and technology). An award winning game designer is unlikely to also have a doctorate in education and be a leading paleobiologist. Serious game design teams must resolve fascinating ideological disagreements resulting from disparate disciplinary values, vocabulary, and culture. This essay describes a sequence of conflicts encountered during the game design process and how each was resolved through iterative playtesting to arrive at convergent, amicable game design features that balanced pedagogy, content, and gameplay. For clarity, it will be written as if each area was represented by a single individual.

Tensions between TPCK perspectives are natural and important. Game designers hope to create a highly interactive, compelling experience that is fun to play. Pedagogy experts insist that the game must be a great teacher. Content experts expect the game to include accurate, richly detailed content; ideally inspiring at least some of the passion they feel for their field of study.

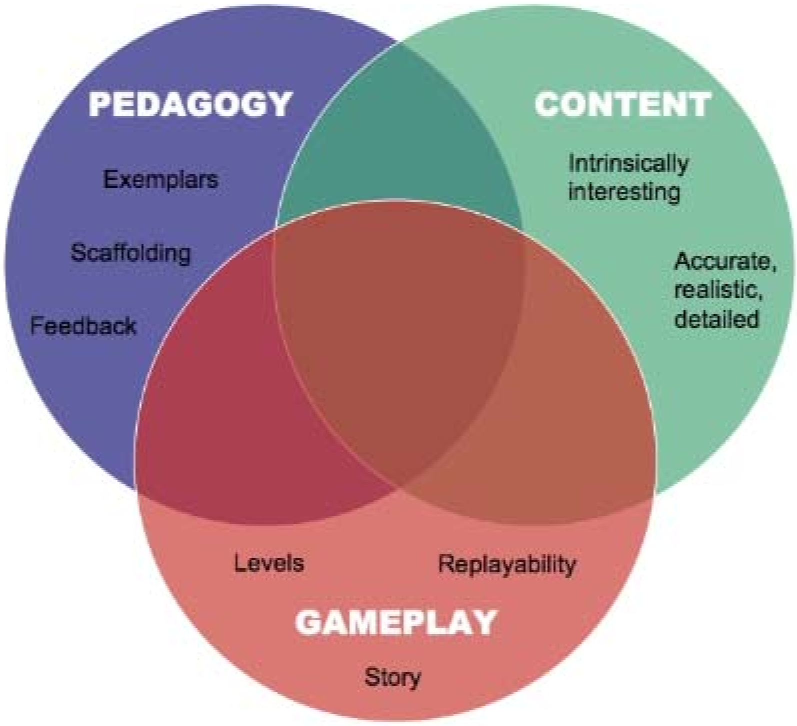
Pedagogy, content, and gameplay conflicts rarely arise in the design of exogenous educational games when the game mechanics and pedagogy are already defined. Exogenous games separate learning content and game mechanics (Halverson 2006; Malone and Lepper 1987). They typically re-use successful game mechanics, such as hangman, a jeopardy-style game show or a visual matching game, inserting content to be learned to the pre-existing game structure and rules. Content is often the only new input.

Endogenous educational games integrate learning content into the structure of the game (Halverson 2006). Exploring the game space and meeting game challenges contribute to learning goals. An endogenous game need not involve immersion or role play. The defining characteristic is that learning content is embodied in the game itself. Endogenous educational game designers begin with a more or less blank slate. Game designers seek an idealized convergence of content, pedagogy and game mechanics that achieve the hypothetical potential of games to promote advanced forms of learning. (For discussion of the games and learning, see Prensky 2001; Gee 2003, 2005; Van Eck 2006; and Kirriemuir and McFarlane 2004.) This ill-specified design problem has infinite possible solutions.

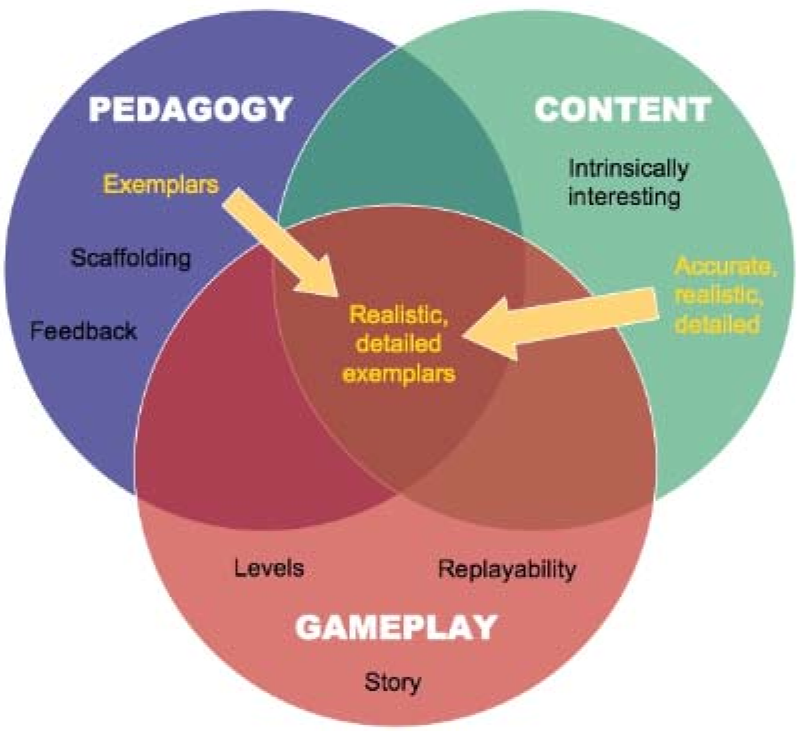
Case Study

The Michigan State University Games for Entertainment and Learning (GEL) Lab created a web-based endogenous learning game designed to teach a subset of the National Science standards on adaptation and evolution (AAAS 1994; Center for Science, Mathematics, and Engineering Education 1996). Funded by the National Science Foundation, the game targeted 7th through 9th grade students and was intended to be playable within a 45 minute class period. A case study of this game design process is used to illustrate how playtesting resulted in progressive revisions that brought pedagogy, content, and gameplay goals into ever-closer alignment.

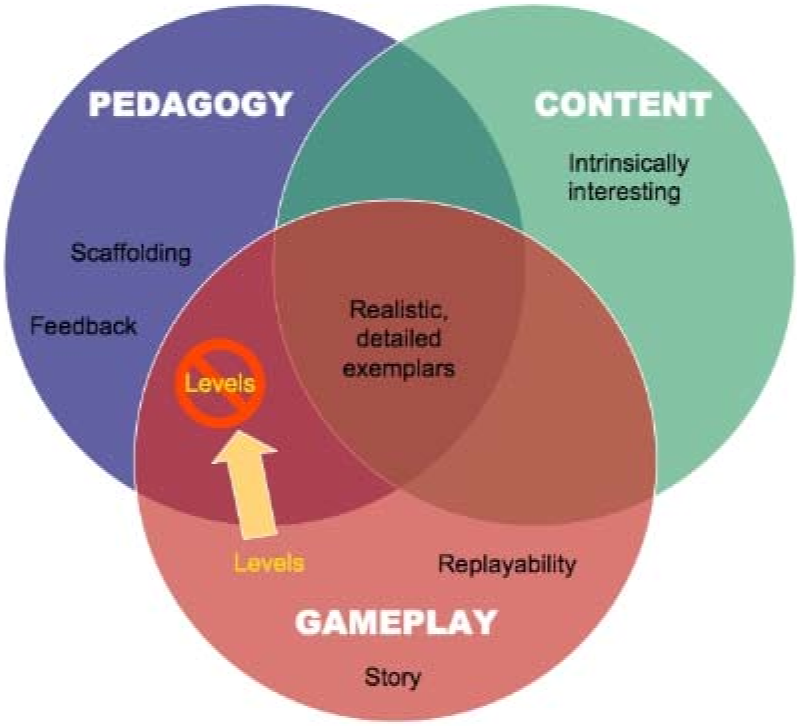
A detailed discussion of the game design and prototype-playtesting progression, was presented at the 2005 Designing User eXperience conference and is available online. In this discussion we will highlight some of the key moments in resolving pedagogy-content-gameplay conflicts. Figure 1 is an interactive diagram showing the 9 step progression through three prototypes.



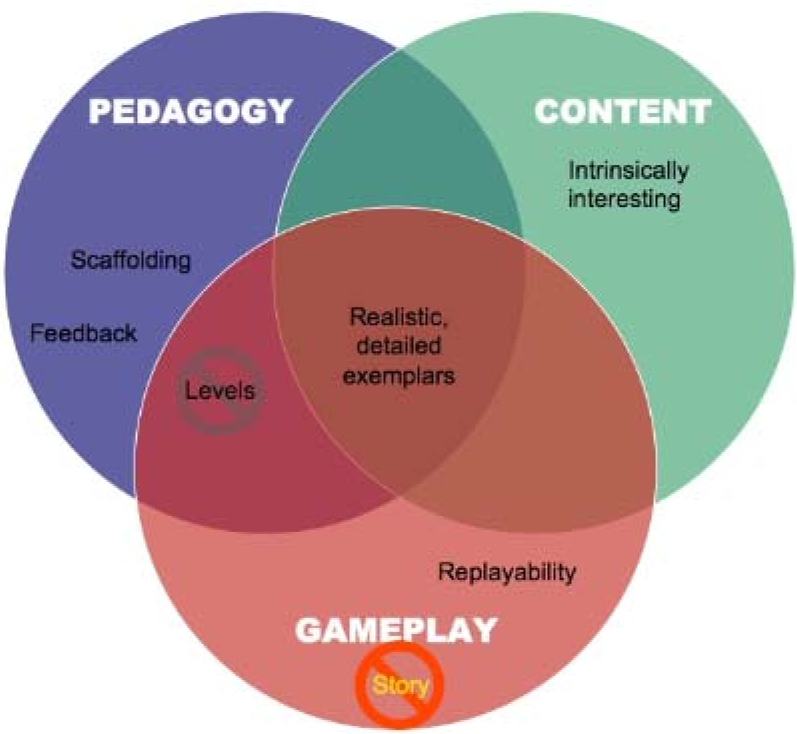
Project Onset



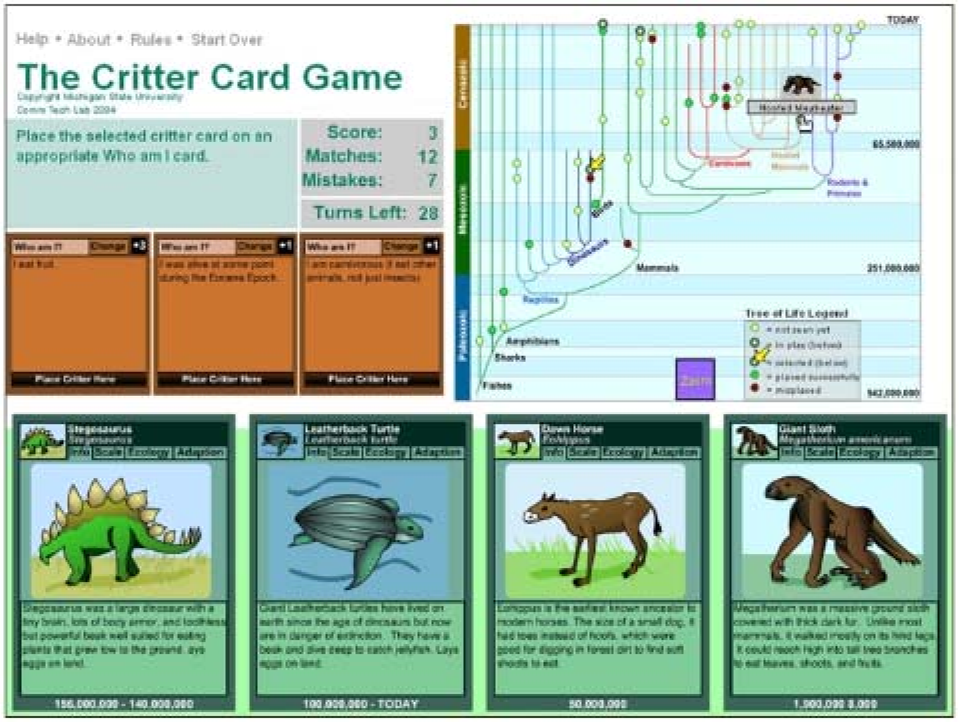
Step 1


PEDAGOGY
CONTENT
Intrinsically interesting
Scaffolding
Feedback
Realistic, detailed exemplars
Levels
Levels
Replayability
GAMEPLAY
Story

Step 2

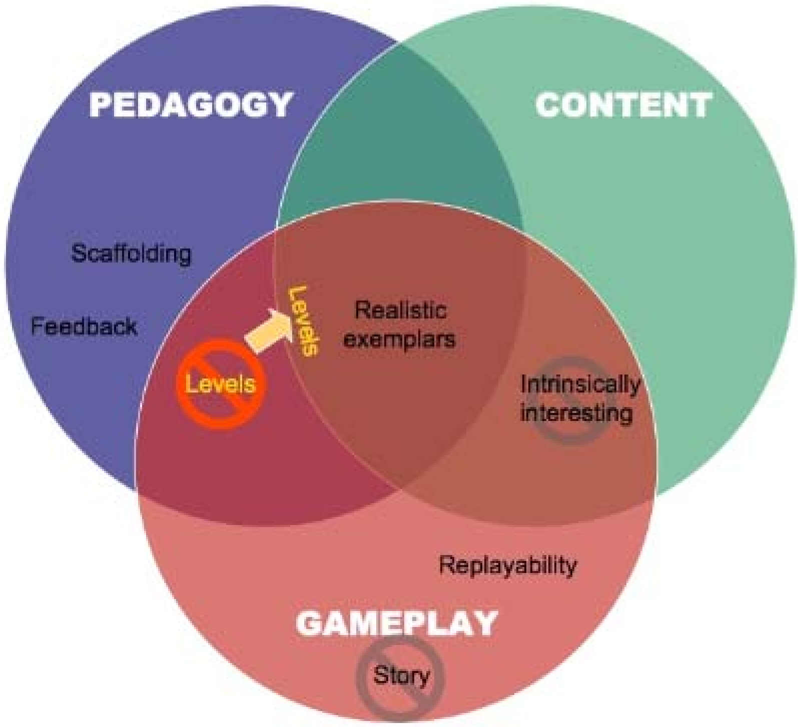

PEDAGOGY
CONTENT
Intrinsically interesting
Scaffolding
Feedback
Realistic, detailed exemplars
Levels
Replayability
GAMEPLAY
Story

Step 3



Prototype One - The Critter Card Game

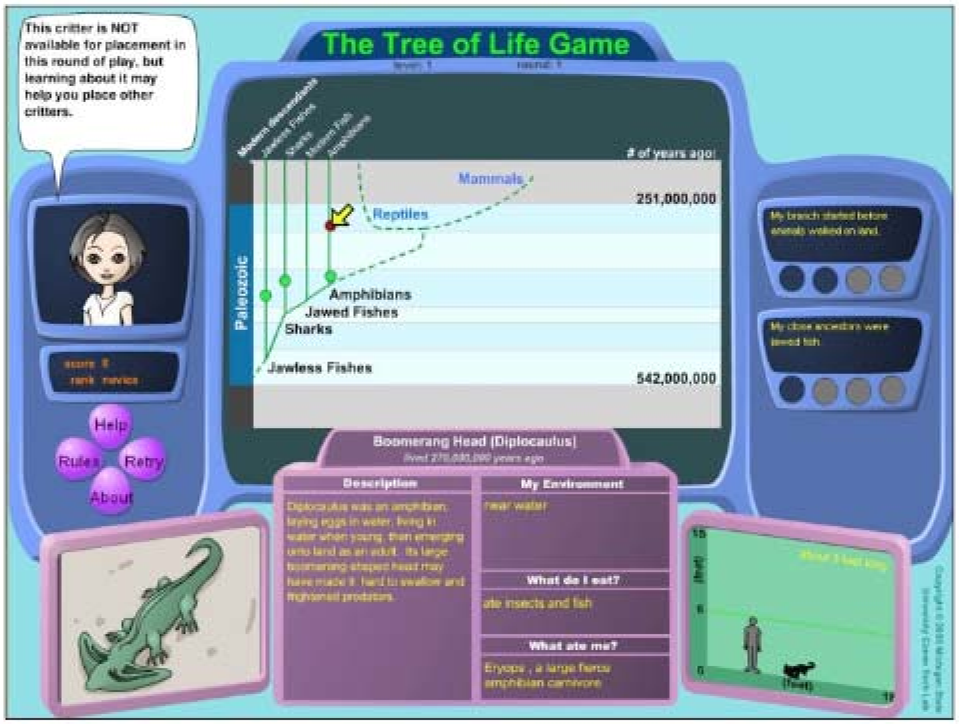
Step 4


PEDAGOGY
CONTENT
Scaffolding
Feedback
Levels
Levels
Realistic exemplars
Intrinsically interesting
Replayability
GAMEPLAY
Story

Step 5

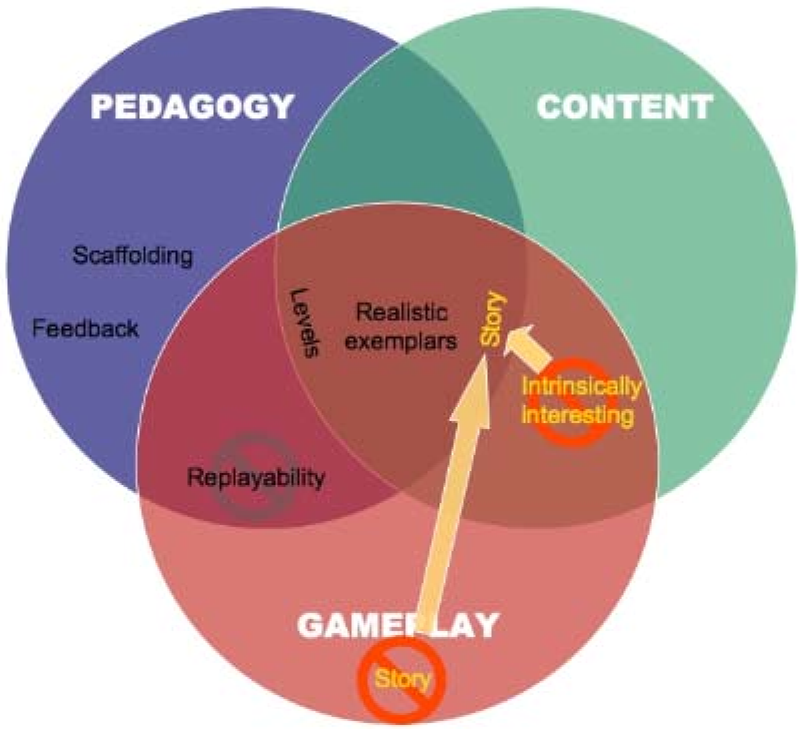

PEDAGOGY
CONTENT
Scaffolding
Feedback
Levels
Realistic exemplars
Intrinsically interesting
Replayability
Replayability
GAMEPLAY
Story

Step 6

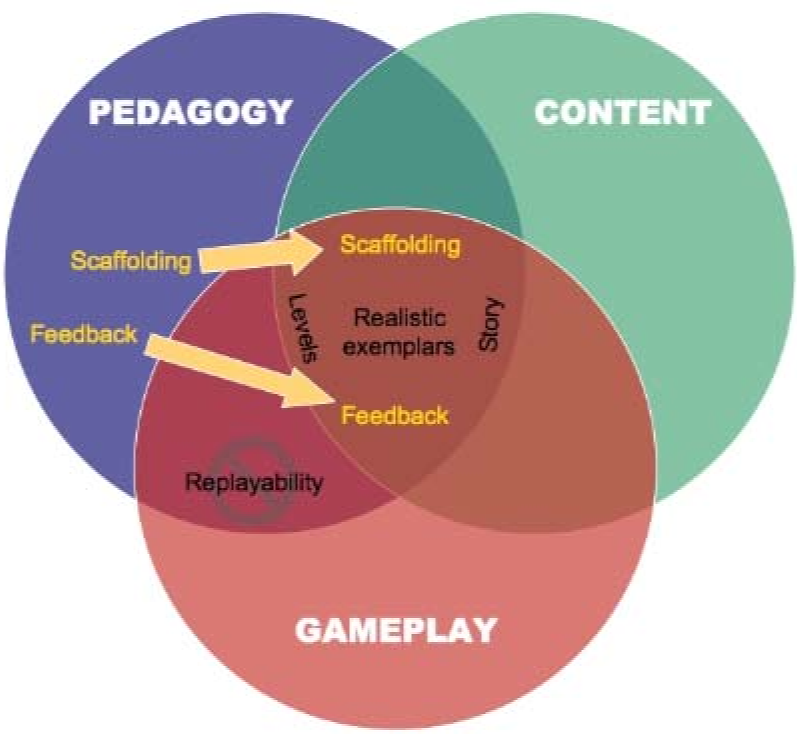


Prototype Two - The Tree of Life Game

Step 7

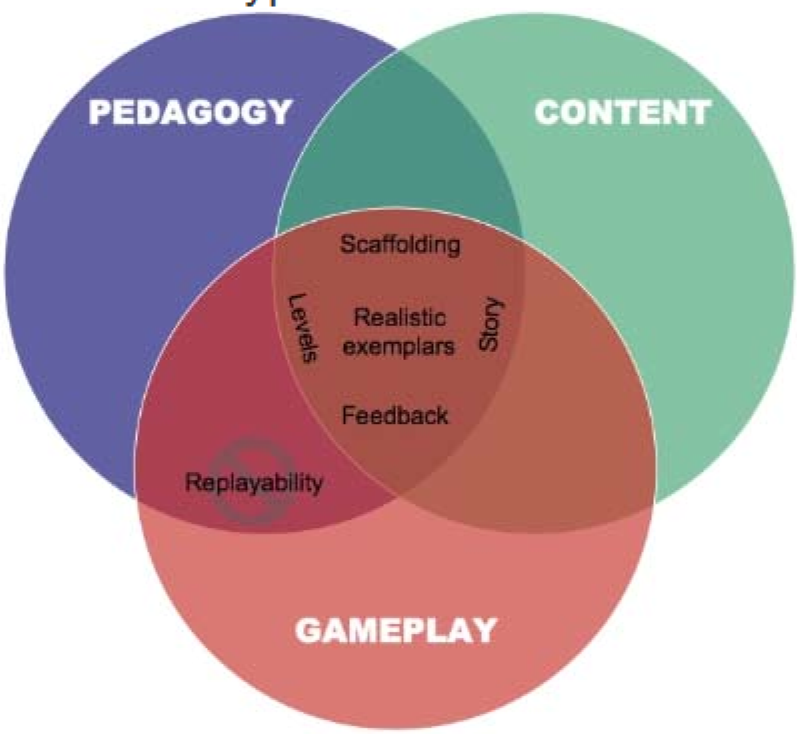

PEDAGOGY
CONTENT
Scaffolding
Feedback
Levels
Realistic exemplars
Story
Intrinsically interesting
Replayability
GAMEPLAY
Story

Step 8


PEDAGOGY
CONTENT
Scaffolding
Scaffolding
Feedback
Levels
Realistic exemplars
Story
Feedback
Replayability
GAMEPLAY

Step 9

Prototype Three - Life Preservers



Project Conclusion

To illustrate the dilemma (which can also be enthusiastically described as "the opportunity"), consider an overly simplistic goal of creating a game to teach only one of the 12 national science standards related to evolution: "The basic idea of biological evolution is that present day species developed from earlier distantly different species" (Center for Science, Mathematics, and Engineering Education 1996). Like the other national science standards, this standard is a highly abstract concept. It does little to narrow the scope of game design, pedagogy or specific science content possibilities.

The content and pedagogy experts easily converged on the desire to base the game on realistic, detailed exemplars of now-extinct creatures from earth's past as a mechanism for teaching evolution concepts. This was accepted without objection by the gameplay expert as a central design specification. See *Step 1 (Expert Convergence: Content and Pedagogy Meet).*

> Levels
>
> Levels are a design feature of many digital games today. Levels are often used in games to break up the storyline and/or gameplay into logical chunks (Rouse, 2001). For example, a level may contain a distinct part of an overall storyline, akin to a chapter in a book. Further, a level may introduce a new gameplay feature that the player must learn and master. Game designers often design a level so the difficulty and tension ramp up toward the end of the level. This creates a barrier the player must pass to continue the game. To play level 2, the player must first "beat" level 1. In this way, levels provide a subgoal or milestone in the overall game. They also provide the player a means to gage his or her progress through the game.

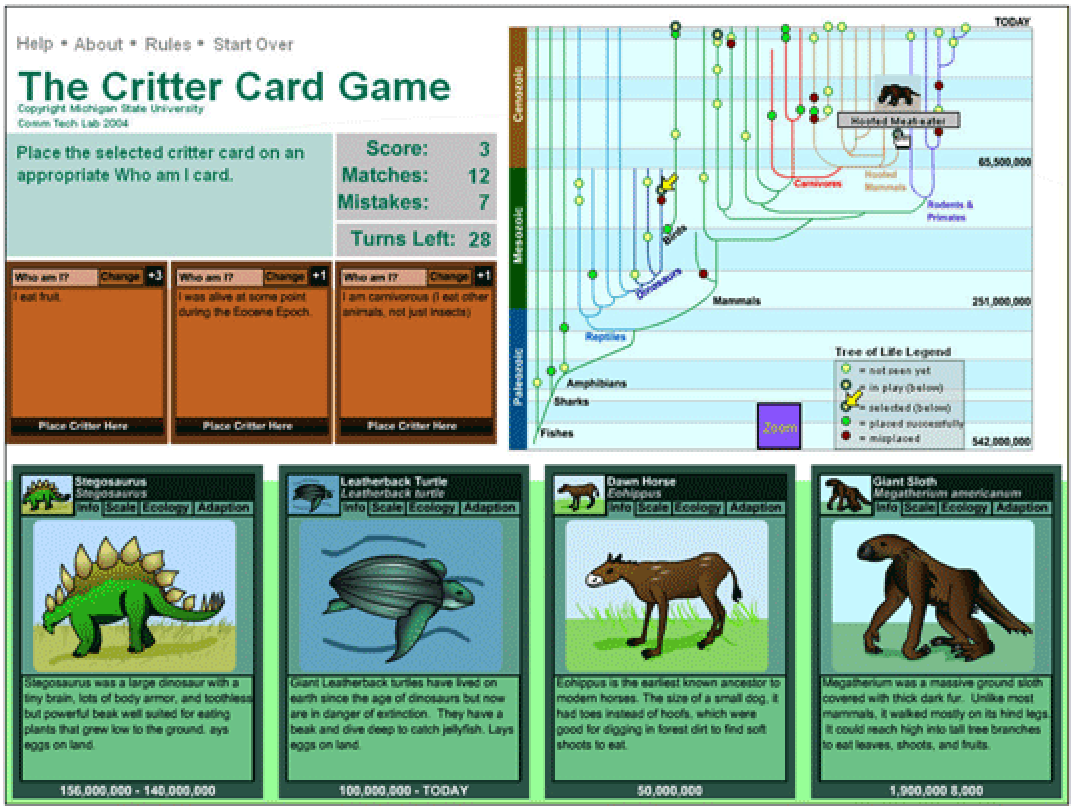
The game designer initially included levels (see levels insert above) as a design feature because levels provide players with subgoals and a means to gage their progress through the game. However, levels also create barriers to advancing in the game. That is, if players are unable to beat level 1, they will not be allowed to play level 2. The pedagogy expert deemed this unacceptable. From her point of view, she did not want the game to prevent any learners from accessing learning content. Therefore, levels conflicted with the pedagogy perspective and the feature was eliminated. See *Step 2 (Expert Veto: Levels eliminated because they are barriers to learning).*

Before the team converged enough on any design concept to be able to create a prototype, different team members' game ideas were categorically vetoed by one or more of the other experts. The pedagogy expert and scientist proposed game ideas that were immediately dismissed by the game designer because it was "not a game." Many of the game designer's game ideas were immediately rejected by the scientist because the science was incorrect.

Most initial conflicts centered on the game's story. There are two perspectives on storytelling in games, the "designer's story" and the "player's story" (Rouse 2001). The designer's story is the storytelling built into the game. Story can be used to set the stage (often called backstory), provide purpose and engagement, and convey content, among other things. The player's story is the story of the interactions and choices the player makes as they play the game. Some games have stronger designer's stories, such as typically adventure and role-playing games, while others have little to no designer's story, such as classic arcade games like Pacman and puzzle games like Tetris. However, all games have a player's story. Because story ideas were so difficult to agree on, consideration of story was temporarily shelved. See *Step 3 (Expert Veto: Backstory placed on backburner).*

Scientists who did not understand games collaborating with game designers who did not understand evolution wasn't working. The entire team delved into the science concepts, gaining a more sophisticated understanding of the content domain. The game designer also taught team members basic concepts about games and game design. Although mutual education helped it did not always yield agreement. The most hotly disputed design features were implemented into a playable prototype and the first playtest was conducted.

First Prototype (The Critter Card Game)



Iterative Design

Game design is an iterative process, including a repetitive cycle of designing, prototyping, playtesting, and modifying the design based on the playtest results (Salen and Zimmerman 2006, 21-22). An underlying premise is that one cannot create a great game by simply envisioning it, writing the specifications, and then building it, precisely matching the initial specifications. Reality checks and refinements are critical to tuning the play experience. Prototyping and playtesting start as soon as possible, often with a paper or software prototype of some small aspect of the overall game. Each prototype acts as a hypothesis to be tested, revised, and testing again. Zimmerman (2003) describes his GameLab company's design-test-analyze playtesting cycle as a means of "discovering the answers to questions you didn't even know were there."

Methodologically, playtesting resembles usability testing, but the focus is much broader, centered on the player experience rather than usability of an interface (Ermi and Mäyrä 2005). Like usability testing, playtesting typically involves small numbers of human subjects, designers are usually not physically present with the human subjects to avoid influencing their experience though they may observe through a two-way mirror, and one player is watched at a time. In-game data may be collected about play choices, particularly as the game matures. Player actions and reactions are carefully observed and systematically recorded, looking for signs of engagement, signs of boredom, in game choices and play patterns. Players are usually interviewed after playing. In educational game playtesting, general questions about what was learned may be supplemented by more formal knowledge tests.

Playtesting is progressive. The iterative cycle continues ad infinitum (or perhaps, ad nauseam), progressively refining the play experience. Fullerton, Swain, and Hoffman (2004, Chapter 8) describe the progression of playtesting, from self-testing to playtesting with confidants, then playtesting with people you don't know, and finally, playtesting with representatives of the target audience. First the design team playtests amongst themselves. (At this stage, playtesting simply means playing the game over and over again and noticing how it feels.) As the idea improves, the team feels ready to test with colleagues or friends not directly involved with the design process. The protocol may be formal or informal, mostly depending on the budget and scope of the game.

Each prototype is essentially a hypothesis to be tested, revised, and tested again (see "iterative design" insert above). The first playable prototype (The Critter Card Game or CCG – Figure 2) was arrived at through a mix of compromise and inspiration. The gameplay designer, pedagogy expert, and scientist agreed on a set of design concepts and features. Players earned points by matching the "critter cards" they were dealt with adaptation challenges. As is typical of early playtests, a small sample of relatively convenient strangers was recruited. Five male and five female college students playtested the CCG prototype, one at a time. Researchers observed player engagement as well as how they progressed through the game and then interviewed players about what they learned. The CCG instructions and prototype are playable online at http://gel.msu.edu/prototypes/ccg/.

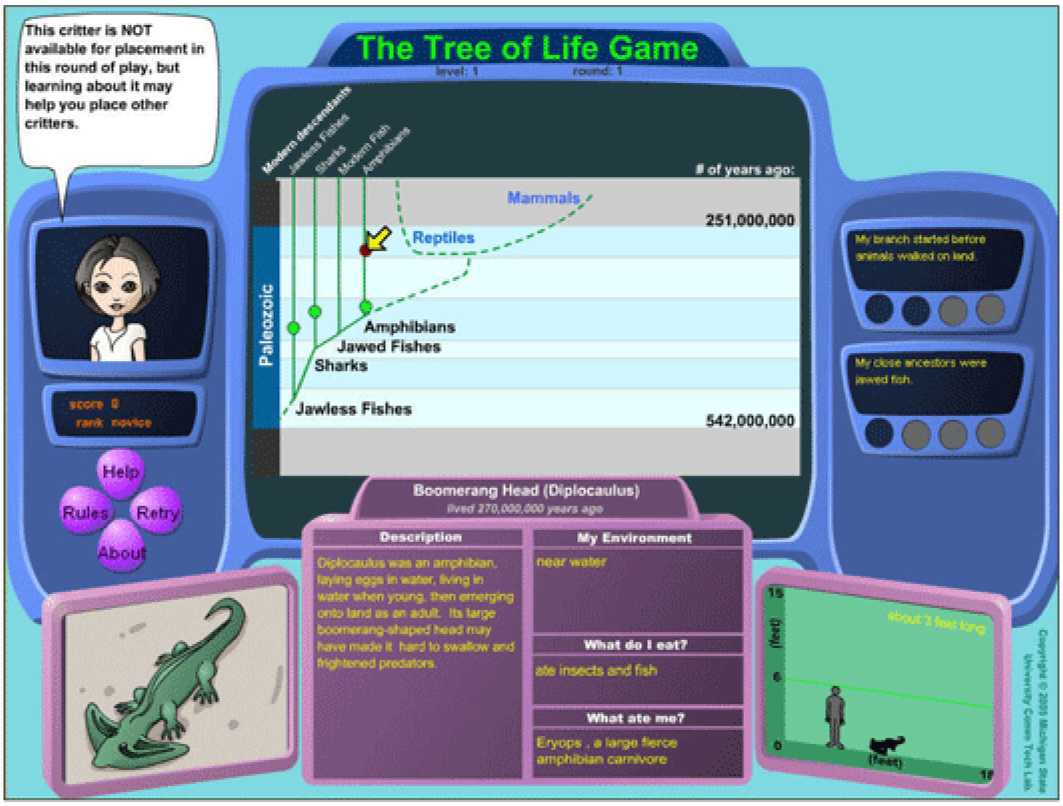
Three assumptions were disproved by playtesting the first prototype. The first was the scientist's expectation that extinct creatures from Earth's past would be so intrinsically interesting that the players would be motivated to read and explore as much as possible. In fact, the playtesters mostly guessed and only one of the ten playtesters clicked to reveal any of the detailed critter content. The entire design team concluded that including less detail would actually increase the amount of detail players noticed. Furthermore, detail should be immediately visible without relying on optional player actions to reveal extra content. See *Step 4 (Evidence-driven Convergence: Greatly reduce level of detail).*

Male playtesters easily but without enthusiasm traversed from beginning to end of the first prototype. All five female playtesters appeared to be lost and unmotivated to play, making little headway. No female playtester reached the midpoint of the game and all female scores were far below all male scores.

The design team reconsidered the initially vetoed concept of levels, to use as a possible means to increase player motivation. Levels were reintroduced accommodating both the pedagogy and science perspectives. Levels allowed us to logically organize the science learning by breaking the game into three ages: Paleozoic (age of amphibians), Mesozoic (age of dinosaurs), and Cenozoic (age of mammals). Players would hopefully be motivated to get to the next level. To mitigate pedagogical concerns, any player could eventually succeed in beating a level. Levels were created to entice players, not to lock learners out. See *Step 5 (Evidence-driven Convergence: Levels are needed to motivate play).*

Replayability is a characteristic of a good game (Prensky 2001, 179-180). Replayability means that a player can and wants to play the game many times, ideally endlessly, and have a different experience each time. Replayability is achieved in games by providing many ways to accomplish the goals and subgoals of the game and/or providing an element of randomness in the game. Initially the pedagogy expert accepted gameplay expert's advocacy of replayability as a design goal. In playtesting the CCG prototype sometimes cards were randomly drawn in a sequence that happened to be better for learning than other random sequences. Given playtesters' failure to learn and the intent to create a game for one time classroom use, the pedagogy expert changed her mind about replayability, deciding that the best pedagogy could be achieved through purposive ordering of carefully selected challenges. The card metaphor was abandoned and play was situated inside of the Tree of Life. See *Step 6 (Evidence-driven Veto: Replayability harmed pedagogy).*

Second Prototype: The Tree of Life Game



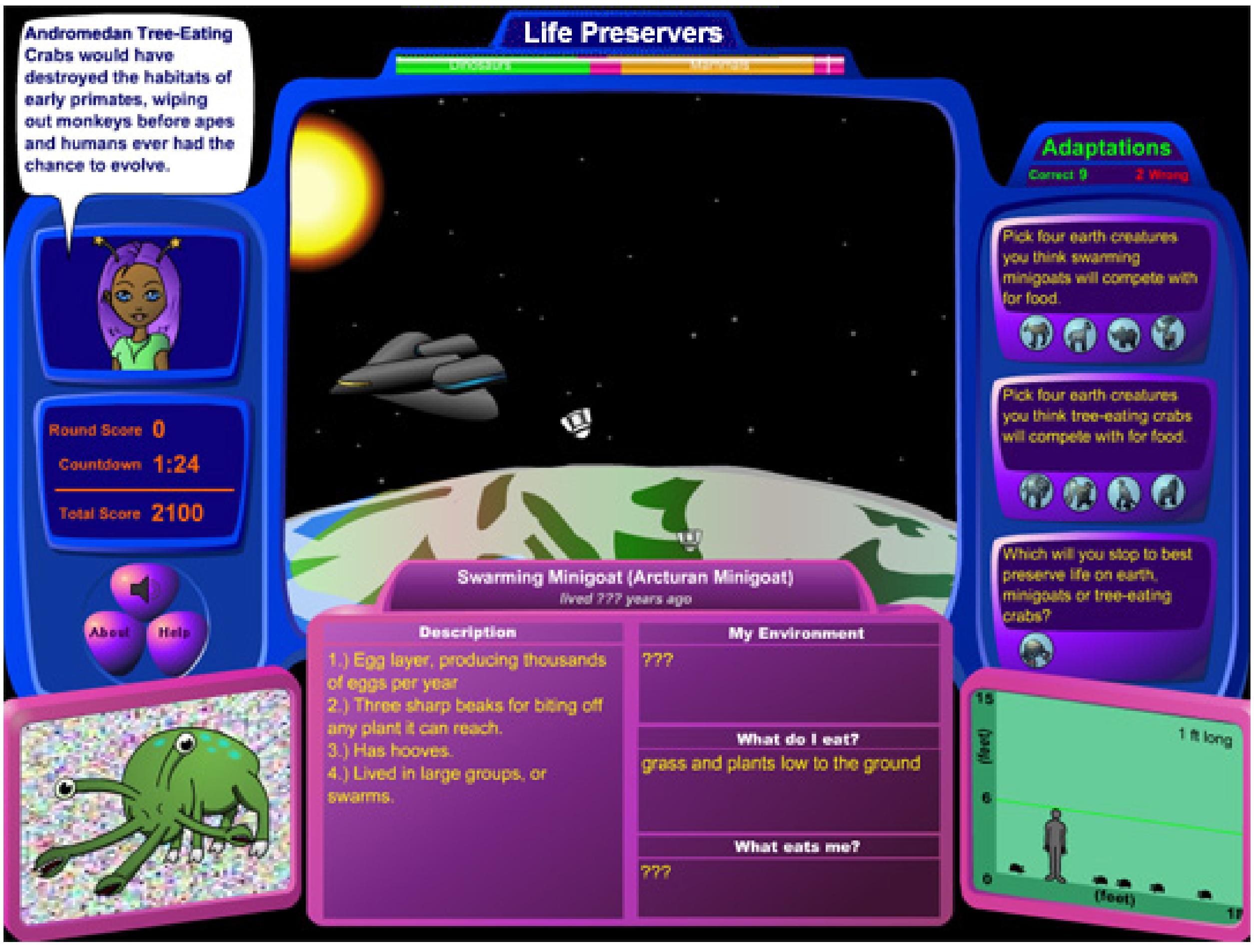
The second playtest was conducted on the modified prototype, now called The Tree of Life Game or TOLG (Figure 3), which embodied the design convergences described above. Five male and five female college students playtested this prototype. The TOLG prototype is playable online at http://gel.msu.edu/prototypes/tolg/.

The TOLG prototype was on the right track. Player engagement was better, and players appeared to be learning some of the intended concepts. However, female players in particular were uninterested in the amphibians of level 1, and all five of them ran out of time during the playtest without getting past the first round of the age of dinosaurs. We eliminated the first level and further reduced the amount of irrelevant (though to some of us, fascinating) content in the game. See *Step 7 (Evidence-driven Refinement: Reduce complexity and detail even more).*

With basic game mechanics in place, a backstory, one that was consistent with and ideally even contributed to the science and learning goals, could now be invented to increase engagement and motivate both play and learning. Content, pedagogy, and gameplay enthusiastically embraced adding an alien invasion round at the end of level 1 (the age of dinosaurs) and level 2 (the age of mammals). Each alien invasion round challenged the player to identify which creatures from Earth's past would be most affected by each of two (imaginary) invading alien species. See *Step 8 (Evidence-driven Refinement: Add more engaging story to enhance motivation and reinforce science concepts).*

In the TOLG prototype a guide provided basic feedback (right-wrong) about the player's actions. For the third prototype, feedback was customized and expanded, offering longer explanations for each right and wrong choice to better scaffold learning. "Points" and a "score" replaced a simple count of number right and wrong to re-introduce the feeling of competitive play. Cut scenes between rounds reinforced learning and contributed to storytelling. See *Step 9 (Evidence-driven Refinement: Add cut scenes and feedback to scaffold and reinforce learning).*

Third Prototype: Life Preservers



The third prototype, now called Life Preservers or LP (Figure 4), was the first version of the game the design team felt was good enough to test with target users. Local 7th and 8th grade girl scouts were recruited. The LP playtest validated design decisions. All six players appeared to be strongly engaged throughout the game. They were disappointed when they made mistakes, and spent time carefully considering the games' challenges. After playing the game, when asked what they had learned, playtesters volunteered national science standards, almost verbatim.

LP acts as a good teacher, guiding learners to think about questions of evolution and adaptation in a carefully constructed order. Game play takes place inside of the Tree of Life. Although gameplay involves answering questions, the pedagogy is more complex than a trivia game. Learners explore a carefully selected content domain (part of the Tree of Life containing only selected critters). Questions in each round guide players to explore real scientific examples from Earth's past which illustrate key concepts in evolution. Inferring from the structure of the Tree of Life that the first bird adapted from therapod dinosaurs around 150 million years ago is interesting, but the science standard to be learned is that "the basic idea of biological evolution is that present day species developed from earlier distantly different species." Each round is reinforced by a narrated, animated cut scene that reiterates the key learning concepts. The final version of LP can be played online at http://lifepreservers.msu.edu/

The third prototype validated convergence on the basic game mechanics and pedagogy. After the playtest, two months of subsequent polishing and refinement ensued, improving the game's visuals and sounds, science content, wording of questions and feedback, and so on. Significant ideological conflicts did not arise during these refinements because the team was in agreement on the overall plan for what the game was and how it would be played. At times pedagogy and content experts offered opinions about game design refinements to optimize learning. Quantitative surveys and in-game data collection with 150 players helped show where fine-tuning of the science content was needed. An external scientist validated the learning content when the game was complete. Nearly one thousand 6th through 9th graders have now played LP at school. Teachers have observed high levels of engagement.

Conclusion

Iterative prototypes and playtesting are critical to the design of a great commercial game (Salen and Zimmerman 2006, 21-22). They are even more necessary for designing great learning games. Playtesting not only helps designers refine the game mechanics but also can help resolve conflicts between pedagogy, content, and gameplay by moving disagreements from theoretical stances to demonstrated success or failure of design concepts.

The holy grail of educational game design is finding an optimal convergence of these perspectives for a particular content domain and set of learning goals. As Gee (2006) pointed out at the Serious Games conference, game designers working with instructional designers can often be overheard complaining about "those damn academics" while academics complain about "those damn designers." Although uncomfortable, these conflicts draw attention to critical game design decisions. Gee concludes, "If we do not learn that game design and learning design are compatible, serious games will remain a small space." Great educational game design teams give voice to all sides of TPCK and make good choices about which perspective prevails for each issue based on playtesting and overarching goals and constraints of the game.

[This article was modified from a presentation, called Using Player Research to Mediate Battles Between Pedagogy, Learning Content, and Fun, given at the Serious Games Summit during the Game Developers Annual Conference in San Jose, CA, March 2006. The conference audience was gameplay designers. The presentation was modified to address pedagogy and content experts who may or may not have personal experience with game design.]

References

American Association for the Advancement of Science. 1994. Benchmarks for science literacy. Cary, NC: Oxford University Press.

Center for Science, Mathematics, and Engineering Education. 1996. National science education standards. Washington, DC: The National Academies Press.

Ermi, L. and F. Mäyrä. 2005. Player-centred game design: experiences in using scenario study to inform mobile game design. *Game Studies* 5 (1). http://www.gamestudies.org/0501/ermi_mayra/ (accessed July 12, 2006).

Fullerton, T, C. Swain, and S. Hoffman. 2004. *Game design workshop: designing, prototyping, and playtesting games.* CMP Books, San Francisco.

Gee, J. 2003. *What video games have to teach us about learning and literacy*. New York: Palgrave Macmillan.

Gee, J. 2006. *What's wrong with serious games*. Panel Session at the Serious Games Summit 2006, Game Developers Annual Conference, San Jose, CA, March.

Gee, J. 2005. *Why video games are good for your soul*. Common Ground Publishing Pty Ltd, Australia, www.theLearner.com.

Heeter, C., B. Winn, and D. Greene. 2005. Theories meet realities: Designing a learning game for girls. Paper presented at DUX (Designing the User eXperience) conference, San Francisco, November. http://commtechlab.msu.edu/publications/files/dux2005.htm (accessed July 20, 2006).

Halverson, R. 2005. What can K-12 school leaders learn from video games and gaming? *Innovate* 1 (6). http://www.innovateonline.info/index.php?view=article&id=81 (accessed July 24, 2006).

Kierrimuir, J. and A. McFarlane. 2004. Literature review in games and learning. NESTA FutureLab Series, Report #8. http://www.futurelab.org.uk/research/reviews/08_01.htm (accessed July 25, 2006).

Malone, T. and M. Lepper. 1987. Making learning fun: A taxonomic model of intrinsic motivations for learning. In R. E. Snow and M. J. Farr, eds., *Aptitude, Learning, and Instruction*, Volume 3: *Cognitive and Affective Process Analysis* (Hillsdale, N.J.: Erlbaum): 223–53.

Mishra, P. and M. Koehler. In press. Technological Pedagogical Content Knowledge: A new framework for teacher knowledge. To appear in *Teachers College Record*.

Prensky, M. 2001. *Digital game-based learning*. McGraw-Hill, New York.

Rouse, R. 2001. *Game design theory and practice*. Wordware Publishing: Plano, TX, 216-218, 406-433.

Salen, K. and E. Zimmerman. 2006. *The game design reader*. MIT Press: Cambridge, MA.

Van Eck, R. 2006. Digital game-based learning: It's not just the digital natives who are restless. *Educause*, March/April, 16-30.

Zimmerman, E. 2003. Play as research. In Brenda Laurel, Ed., *Design Research*. The MIT Press: Cambridge, MA, 176-184.