The PBS KIDS Iterative Design Process For Building a Successful Augmented Reality (AR) Game

Elizabeth McCarthy, Ph.D., WestEd
Yvonne Kao, Ph.D., WestEd
Iulian Radu, Georgia Institute of Technology
Sara Atienza, WestEd
Michelle Tiu, WestEd

Overview of the Iterative Design Process and the Use of Formative Evaluation

This paper focuses on the iterative design process in the development of a successful PBS KIDS' AR tablet app, *Cyberchase Shape Quest*, targeted at children ages 6-8. *Cyberchase Shape Quest* uses a tablet camera and a printable board game to interact with shapes in five different virtual environments. Students apply spatial memory, visualization, and modeling skills to complete levels. In the first week of release, *Cyberchase Shape Quest* rose to be the #1 free iPad kids app in the iTunes store. The app is also available for Android devices. Development of *Cyberchase Shape Quest* involved an iterative design process. Incorporating formal and external formative evaluation into several of the design cycles proved to add significant value during game development because it helped reduce the uncertainty around a large number of unknowns—AR is a relatively new technology and unfamiliar to young children. This paper focuses on the utility of conducting formal formative evaluation at specific points of game development: 1) early game concept testing and 2) user testing on an alpha build.

Formative Evaluation, Cycle 1: Early Game Concept Testing

PBS KIDS initially proposed several game concepts that used AR. These game concepts addressed several mathematics topics, including tessellations, measurement, and spatial reasoning. PBS KIDS and its producers worked with researchers at WestEd to iterate and refine the initial game concepts into the idea behind *Cyberchase Shape Quest*, chosen because it best used the affordances of AR to address an established need in math education—spatial reasoning. At this point the PBS KIDS team was faced with a large number of game design decisions, such as level design and the design of game controls. Rather than make decisions based on limited knowledge of how children would react to the game concept, the team turned to a round of formative evaluation to guide the design.

Methods: Paper Prototype Testing and Game Mechanics Testing

The formative evaluation used a two-part approach in order to address the large number of questions regarding appropriate game design, focusing on questions related to level design and game controls. First the team, which included PBS KIDS developers, math education experts, and WestEd researchers, sketched a series of possible level designs that used a variety of spatial skills (e.g., 2D rotation vs. 3D rotation), were developmentally appropriate, and which progressed in difficulty. Researchers then constructed 13 physical block-and-tile versions of these levels based on the sketches. These physical versions of the puzzles were used in a variation on paper prototype testing. The goal of paper prototype testing was to collect data on the difficulty of the different types of spatial reasoning problems, document the problem-solving strategies students used to solve the problems, and document the scaffolding techniques used by an in-person adult facilitator. The second part of the formative evaluation was to collect data on the usability of different types of game controls—at the time the game was being developed, there was little published data on the usability of AR games for children. Rather than spending resources developing parts of the game, the team opted to test a variety of controls (e.g., pressing buttons with thumbs, shaking the tablet, ability to maintain AR tracking) from a collection of commercially-available AR and non-AR tablet games. Seven children ages 6-9 (3 boys and 4 girls) participated in the formative evaluation. The paper prototype testing and the game mechanics testing each took 30 minutes, with the order of tests counterbalanced across participants. All sessions were video recorded and analyzed.

Findings and Contributions to Later Iterations

Researchers analyzed the video data in order to establish the relative difficulty of the different puzzle levels based on the number and age of students who completed each level, the time needed to complete each level, and the amount of scaffolding needed. The video data was further analyzed to generate more detailed data on children's problem-solving strategies, the amount and type of scaffolding required for students to complete each level, children's ability to use different games' controls and to perform the actions required to maintain AR tracking, and how long children could hold the tablet up to maintain AR tracking before experiencing fatigue. As a result of this

first round of formative evaluation, game developers were able to determine which styles of puzzle levels were too easy for the target age group and which styles were too difficult. Game developers were also able to use the scaffolding interactions from the paper prototype testing as models for the scaffolding algorithms into the game. Specifically, the paper testing informed the timing of scaffolding, the progression of scaffolding (e.g., moving from general to specific hints), the wording of scaffolding (sometimes using exact words captured during the study), as well as the modality of scaffolding (e.g., what types of verbal and visual hints to include and when to deploy each type of hint). In addition, the formative evaluation determined that the most usable approach to game control was to have children hold the tablet with both hands and have UI buttons on the edge of the screen. It was also determined that children lose the AR tracking frequently, so providing adequate support for avoiding and recovering from tracking loss would be a critical element for game usability.

Formative Evaluation, Cycle 2: User Testing on an Alpha Build

Following the first round of formative evaluation, the PBS KIDS game developers and producers continued iterating on the game design and produced an alpha version of the tablet game that underwent another cycle of formal formative evaluation. This cycle of formative evaluation was guided by these research questions: 1) What usability problems did the participants encounter when playing the game?, 2) Is the difficulty of the levels appropriate for the participants and do the levels build in difficulty?, 3) What would teachers change about the game to improve the instructional content?. Seven children ages 6-10 (3 boys and 4 girls) tested the game. Additionally, 3 teachers who currently work with this age group tested the game and provided feedback around the game's usefulness as an instructional tool. Each participant played *Cyberchase Shape Quest* for 30-40 minutes (with a break in the middle). Game play was followed by a post-interview to obtain participants' feedback. Observers took note of usability and pedagogical issues that arose during game play. Researchers debriefed after each session to synthesize what they observed in context of the research questions. Observations, interview responses, and researcher debriefs were transcribed and qualitatively coded.

Findings and Contributions to the Final Game

The evaluation confirmed that the game is generally usable and fun, that kids can persist through multiple levels of the game at a time, and that the game is interesting and valuable to teachers. We also uncovered several usability issues that were particularly critical to address in the final game. Users had major problems with understanding the mechanics of AR and perceiving the 3D space—both children and adults tended to sit in one position rather than moving around to view the 3D puzzles from different orientations. This led the developers to create new "tutorial" levels that required kids to physically move, and made them sensitive to the 3D depth of the game. This also led the creative team to amplify perceptual cues in the game (adding shadows, making shapes bounce around). In addition, children struggled to understand the scaffolding provided by the game. As a result, the development team adjusted the scaffolding wording and visuals to be simpler and more direct, as well as adjusting the actions that triggered the scaffolding. Finally, having teachers participate in the evaluation enabled the team to identify more teaching opportunities in the game.

Significance and Recommendations for Future Game Development

Given our experience with *Cyberchase Shape Quest*, PBS KIDS and WestEd have the following recommendations for designing high-quality AR-based educational games: 1) Expect to have many unanswered questions early in the design process, when dealing with bringing technology into a new domain, 2) Explore unknowns by incorporating formal formative evaluation into an iterative design process, 3) Involve different stakeholders and domain experts (game developers, researchers, content area experts) early in the development process, and keep them involved throughout, and 4) Conduct testing early and often, as it is not necessary to have a playable game for testing to add value, and be sure to include teacher participants in the testing of educational games. In conclusion, using formative evaluation was very useful to develop a successful game in a design space with many unknowns. Cyberchase Shape Quest launched in January 2014, and was downloaded nearly 250,000 times in its first month on the App Store. It currently averages 12,000 daily active users. Formative evaluation helped the PBS KIDS team understand how children think about spatial problems, and identified new opportunities for teaching and scaffolding. It also helped us design a user interface and game mechanics that are appropriate for the intended age group. The project is funded by a Ready To Learn grant (PR/AWARD No. U295A100025, CFDA No. 84.295A) provided by the Department of Education to the Corporation for Public Broadcasting.