2009 Games for Change Conference
Assessment Workshop Summary
Jim Diamond, Katie McMillan Culp
EDC/Center for Children & Technology

Acknowledgement: EDC/CCT would like to thank Rafi Santo of Global Kids and Jae Osenbach of Fordham University for their invaluable notes and their contributions to the write-up of these proceedings

Introduction

On May 27, 2009, a diverse group of experts met to consider how to assess the impact of social issue digital games on players' knowledge, attitudes, and actions. The group was convened as part of the annual Games for Change Festival in New York City, and included game designers, psychometricians, educational researchers, civic educators, and funders. Participants shared an interest in building a framework for assessing the impact and quality of digital games that focus on issues of social change and civic engagement.

With funding from the MacArthur Foundation, Games for Change organized the evaluation day as a response to calls for more focused, ongoing discussions of social issue game assessment and evaluation, which originally arose during a games and evaluation panel at the 2008 Festival. Assessment is one of the key factors influencing the future development and deployment of games for education and civic engagement. Digital games are a young medium, and we are just beginning to understand how they can educate, influence, and engage citizens young and old.

The day included a mix of conversations and presentations and focused on four broad topics:

1. *Potentially generalizable assessment models that could be applied across genres and settings for serious games*. Models discussed included evidence centered assessment design (ECD) and "stealth assessment"; using player choices, as opposed to knowledge outcomes, as the essential unit of analysis for game-based learning; and games as preparation for future learning**.** Participants from a variety of fields and disciplines also discussed how games might be used to promote social change and to teach skills associated with civic engagement, with a focus on how game designers and developers might share best practices for educational and social change games.
2. *Exemplars of educational games designed to foster content knowledge acquisition and skill development.* Both *Quest Atlantis* and the in-development game *Our Courts: Guardian of Law* were presented. Both of these games have specific educational goals, and in both cases assessment was an explicit consideration during game design and development. Representatives from both projects discussed some of the challenges involved in designing games whose activities map to assessable learning outcomes.
3. *The importance of establishing instructional design principles for game-based learning.* Throughout the day the group discussed the importance of

developing common frameworks and strategies for developing and implementing assessment strategies. This will both support the accrual of knowledge across games and studies, and generate a common language and set of practices for naming, supporting, and assessing specific forms of game-based learning.

4. *The relationship between social, participatory game play and broader civic engagement, and the relevance of those connections to an assessment agenda*. The group discussed the challenges involved in moving beyond the assessment of the impact of an individual game on an individual player's knowledge or skills, to assessing impact on more distal constructs such as changes in individuals' or groups' actions over time, or changes in the beliefs and values of individuals or groups. Empirical investigation of these outcomes will require multiple, coordinated strategies for assessment program evaluation, and other forms of research.

Serious gaming is a large and crowded boat

The day's participants represented the diverse interests and objectives within the serious gaming community. This group of game designers, scholars, and researchers from fields such as civics, science and social studies education; communications; developmental and cognitive psychology; and law all recognized the difficulty of addressing the needs of such a diverse set of players, but many also agreed on the opportunities for social and civic learning afforded by games, because of the opportunities they offer to explore and play in authentic environments that can prepare people for other experiences in the real world. For example, many games, including *Quest Atlantis*, enable players to take on identities through role-play in order to solve real world problems in a virtual world. As students play *Quest Atlantis,* they are faced with representations of challenges encountered by scientists in the real world. That is, within the narrative world of the game, problems are practical and solutions have consequences. In addition to influencing the development of scientific reasoning skills, this kind of play may also influence the growth of students' social understanding and ethical reasoning.

Throughout the day, participants referenced the complexity of the task. They raised issues related to the *diversity of the audiences* they seek to reach with evaluation findings and discussed the *difficulty of defining measurable outcomes* that fully capture the ambitious goals associated with these kinds of games. They also sought to disentangle the goals of *assessment* from the broader need for evaluation that also arises as both individual games, and the field of social issue games as a whole, mature.[1]

Specific issues discussed in the course of the day included:

[1] This day focused specifically on *assessment*—the measurement of the impact of game play on specific outcomes, whether for individual players or for groups. *Evaluation* is the process of judging aspects of the quality, effectiveness, or impact of a program, product, or experience. Assessment data can serve many purposes, including providing evidence for program or product evaluation.

1. The many different *motivations* for assessment. Assessment data can be used to inform further design and development efforts, to provide evidence of impact to existing funders or excite interest among new funders, or to raise awareness and interest among potential stakeholders such as educators, leaders in informal education settings, and others in the game development community. Different goals have different implications for assessment design.
2. The varying needs of different *audiences* for assessment data. Practitioners, designers, funders, and policymakers are all likely to have very different ideas of what aspects of player experience are most valuable to assess. Additionally, different audiences may be in need of different kinds of data (scores, portfolios, case studies). The diverse settings social issue games may be played in further complicates the picture of stakeholder needs, as teachers, after-school program staff, and youth organizers may also have different perspectives on the purpose and goals of assessment.
3. The many different levels and types of *outcomes* that can be assessed, including dispositions, attitudes, creativity, content knowledge, values, or skills, any of which would require the development of distinct measures and methods for assessment. Additionally, the challenge of *defining desirable outcomes* influences how assessments should be crafted in terms of game play.
4. The many different *contexts* in which social change games are played, ranging from casual games played at kiosks, to at-home play, to play at special events, to play in traditional school settings. People's engagement with games is likely to vary widely across these settings, as is their opportunity to follow up on what they have learned or experienced through the game. All of these factors complicate the assessment process. Though not always the case, schools, non-profits, and NGO's may not be interested in the same measurable outcomes. Commercial game developers, non-profit actors who work outside of education, and developers who are engaged with the education system all bring very different priorities and perspectives to bear on this process, and they may have very different notions about what aspects of player experiences should be assessed.
5. The difficulty of extracting meaningful assessments of an individual player from game play that may focus on *group experiences*. Assessments typically focus on demonstrating the impact an intervention has on the individual, which may not match up well with games that privilege the group.
6. The challenges of designing games that are engaging, fun to play, and *based on instructional design models that promote the intended (and, in some cases, unintended) learning outcomes*. The process of creating a deep integration of the practices of instructional design and the practices of good game design is still in its early stages.

This summary document highlights themes and commonalities that emerged from the day's presentations and conversations. It also reviews some of the challenges to rigorous and systematic game evaluation, a category that encompasses the assessment of individual player experiences that participants identified.

We hope that this document will serve as a reference as this conversation about assessment and evaluation continues, and will inform discussion about future research and development efforts.

Theme 1: We need to build innovative assessments that educators and policymakers will recognize, but that are well suited to the goals of social issue games.

Throughout the day, the group debated whether and how to associate the evaluation of social issue games with school-based assessment practices. Traditional assessment frameworks, such as standardized tests of content mastery, produce evidence that educators and policymakers will easily recognize, but are limiting and poorly suited to the goals of social issue games. Given a wide range of informal learning frameworks, they are also poorly suited to assessment of out-of-school play experiences, which need to be well accommodated in any shared framework for assessment. While some prominent games (such as *Quest Atlantis*) are designed for in-school use, many other social games are intended for casual play, for use at special events, or for use in alternative, informal learning environments. The variety of individual play experiences and the lack of instructional or authorial control complicate the notion of standardized assessment of learning outcomes.

The group acknowledged that aligning assessment efforts for games for social change with school-based assessment paradigms involved a number of tradeoffs. First, attempting to connect games for social change to narrow measures of student achievement, such as scores on state accountability tests, could potentially bring significant recognition from some school-level practitioners or from state-level administrators and policymakers. But these kinds of outcome measures are very distant from the goals of most games for social change, which can include long-term attitude and behavioral changes related to specific social issues, making them unappealing and unlikely to capture real change or the value of playing such games. While some participants felt external pressure to produce these kinds of outcomes to gain legitimacy with some stakeholders, few participants felt these measures were actually well suited to the nature of games for social change, or to the purposes of most evaluation efforts for games for social change. Indeed, several participants acknowledged that assessment-driven games might be counterproductive when game feedback is useless or even harmful.

Developers interested in connecting with schools and teachers can also assess the impact of games for social change using other assessment strategies that can be more difficult to create and implement, but that can capture meaningful outcomes that are relevant to the goals and intentions of the games being evaluated. Presentations by Dan Hickey, Valerie Shute, and Dan Schwartz illustrated important, innovative work being done in this area and are discussed in more detail below.

As some participants pointed out, many developers of games for social change are not invested in traditional school structures. They need other approaches to assessment, that respond to the different play experiences their audience is likely to have, and the narrow opportunities they often have for data collection. The group agreed that in some ways these settings are the most challenging of all for meaningful assessment. School-based games may be easier to assess in terms of learning outcomes, as the players are a captive audience, learning objectives can be made discrete and explicit, and data from game play and other classroom-based activities can be collected longitudinally through a variety of methods. In informal environments (including casual, Web-based games), however, where many serious games produced by social change advocates are played, it is more challenging to collect data and to assess relationships between game play experience and changes in attitude or behavior in other arenas. Many in the group agreed that managing expectations for how much change small games can effect is important, for those in the social issue gaming community and for institutions from whom they hope to acquire support. But several participants also suggested that game play tasks should be authentic and related to the outcomes the game has been developed to address in order to provide players with opportunities to reflect on the experience.

Assessment needs to be fully and deeply integrated into game design.

The group discussed video games as potentially strong examples of performance-based assessments. This type of assessment approach, which has a long history in school reform efforts, would involve players accomplishing goals in ways that reflect their understanding of embedded content and concepts, as well as their ability to engage in problem solving and other skills associated with learning content. In order to describe relationships between game tasks, player inputs, and learning outcomes confidently, however, explicit instructional design and theories of game-based learning would need to inform the design of games from their inception.

The group also agreed that, in order for games to remain gamelike, game-based assessment should be invisible to the player—that is, directly and meaningfully related to game play. Further, it should not take players out of the game, and should not shut down exploratory play behaviors. One criticism of school-based, "sequestered" assessments voiced by several participants was their inauthentic and abstract nature and their frequent misuse as tools for summative assessment, rather than as formative tools with which students can continue to learn as they practice.

Three presentations by Dan Hickey (Indiana University), Valerie Shute (Florida State University), and Dan Schwartz (Stanford University) provided specific examples of promising approaches to making games more authentic forms of assessment.

1. Dan Hickey (Indiana University): Assessing for design in *Quest Atlantis*

Hickey focused on the value of formative assessment as feedback for both players and game designers, suggesting a general design principle of "provide feedback that is useful and used," noting that feedback can sometimes do more harm than good in

educational settings. Critically, formative feedback in game play should provide players with opportunities to reflect on learning and provide evidence for how they might improve. Similarly, within the framework of design-based research in which game-based learning environments such as *Quest Atlantis* are studied, useful formative feedback provides designers with opportunities to continue to refine aspects of the game that support player learning.

Hickey also addressed the role of feedback in motivating learners to engage in participatory discourse practices associated with learning domains. That is, formative feedback should provide learners with additional opportunities to improve their ability to engage publicly in argumentation and problem solving related to specific content areas through authentic tasks, rather than with disincentives through isolated tests that evidence "what they don't know." In short, feedback should provide learners with opportunities to improve as they engage in meaningful activities. To that end, levels of formative assessment (in games and in learning environments in general) ought be aligned such that the relationships among them are clear to learners: "immediate, close, and proximal" levels of assessment should be related in order to enable learners to engage in domain-related discourse across a continuum of assessment practices.

2. Valerie Shute (Florida State University): Evidence centered design and stealth assessment

Valerie Shute argued that well-designed games are effective environments for enabling students to develop systems thinking skills, a core "21st century competency." While several participants noted there are multiple definitions of systems thinking, Shute described it as the ability to reason about and understand complex, dynamic interactions in natural, artificial, and social systems.

Applying the Evidence Centered Design framework to games as performance-based assessments, Shute suggested that games can be effective "stealth assessments" of student learning when their designs are based on models of effective domain practices. Briefly, evidence centered frameworks provide assessment designers with frameworks for explicating competency models (for example, what does successful scientific argumentation look like and how is it developed?), describing the practices that provide evidence of competency, and mapping the features of tasks that will provide the evidence. The model is an effective strategy for tying each element of game play to the desired learning outcomes. As domain competency models are tied to player action models, the game is a valid measure of player learning and obviates the need for external, isolated tests.

When assessing learning, player decisions about tasks are evaluated in light of the game's underlying model and theory of learning. Over time, evaluations of game play provide portraits of players' learning and their reasoning (as evidenced through game play related to problems) about content. Thus, assessing game play (in addition to discourse and other non-game related activities) provides evidence of how students engage in problem solving. Shute noted that a challenge for this

form of assessment is balancing attention to expert models of problem solving with novel behaviors and solutions during game play.

3. Dan Schwartz (Stanford University): Assessing choice patterns in game play

Schwartz argued that games are effective as tools for preparation for future learning, as game play can prime students to think about content they will encounter during direct instruction. Games might be used as advance organizers that make subsequent learning more meaningful by providing learners with concrete opportunities to explore the content. From this perspective, video games do not bear the entire burden for instructional design and supporting learning outcomes and can retain more gamelike qualities, rather than engage in direct instruction and risk being perceived as overly pedantic. Most significantly, Schwartz noted, games provide learners with opportunities to make adaptive choices, which is (or ought to be) the defining goal of education.

Video game interactivity is defined by player action, and actions are predicated by choices about what to do in a game environment. Unlike earlier, more static, instructional technologies, well-designed games provide players with artificial worlds that respond to player choices with dynamic, meaningful feedback that highlights the consequences of those choices. Because society, Schwartz argued, is ultimately concerned with the choices that individuals make about how to learn and how to engage in problem solving, the focus of game-based learning in educational settings should be the choices that players make as they encounter problems.

Making choices, rather than knowledge outcomes, the focus of assessment, however, raises the challenge of distinguishing favorable learning pathways from those that might be less productive, while ensuring that design does not prevent players from engaging in exploratory play behaviors. Schwartz suggested the solution is to identify productive choice patterns for individuals and for groups over time. Data collection of game play over time allows for portraits of games choices to emerge; choice patterns can be analyzed to predict future decision-making and to anticipate forms of scaffolding those along different pathways might require. Analysis of paths will also reveal where students are "just clicking through" in order to succeed.

We need to distinguish between effective evaluation of individual games and the larger process of building a case for the potential of games for social change in general.

The day's discussion made clear that while this community cares about creating tools for assessing the impact of specific games on students, there is also a strong, shared desire to use evaluation, more broadly defined, to advocate for a larger set of ideas about learning and its relation to civic engagement. As the group discussed outcomes they hope to target through assessments, words such as "perspectives," "values," and "orientations" came up repeatedly, as participants sought to describe the larger societal goals they are seeking to address through their games.

As the presentations described above make clear, specific assessments of the outcomes from specific games can be innovative and sophisticated in their design, and can capture a wide range of aspects of young people's experiences. But the complexity of the work presented reminded the group that no one game should be expected to demonstrate the kinds of broad, sustained impact on young people's beliefs and actions that this community is seeking to achieve over time.

Rather, the day's discussion suggests that at least two distinct projects need to be pursued. First, designers and researchers need to work together to carefully define and design for a set of specific, desired outcomes for particular games, and ensure that the game play experience young people will have is well aligned with those desired outcomes. They need to also invest in developing and deploying innovative, minimally intrusive measures that allow them to document the impact of playing that particular game on those targeted outcomes. This is a resource-intensive process and requires close collaboration among multiple players.

Second, the broader community of stakeholders in the larger project of developing and disseminating games to support social learning and civic engagement must continue to work together to define a set of broader goals that can be described and shared across a wide range of projects. These might include those longer-term outcomes such as attitudes and levels of civic engagement, which as Joe Kahne described, can be measured among large populations through survey research. Over time, it may be possible to draw on that line of research or other similar strategies to explore whether measurable associations are appearing between, for example, young people's experience with digital games for social change and their levels of civic engagement.

Theme 2: We also need assessment strategies for games that facilitate civic engagement and social participation in a democratic society.

As one of the day's central themes, several presenters spoke about games as tools for promoting civic engagement. While methods for measuring how games influence players' long-term behaviors, attitudes toward, and types of participation in a democratic society are not yet well established, many agreed that games could provide people with experiences that influence the development of participatory dispositions. As Joe Kahne (Mills College) commented, games might facilitate player experiences that socialize them for future engagement. Defining engagement, then, is part of the challenge of designing games that seed future civic participation.

Tying the discussion of civic engagement to the earlier discussions of assessment are issues of instructional design and modeling. Presenters discussing civic participation were also interested in how games might be based on experiential models that allow for comparisons of player activity to norms of civic engagement. Simulation games can provide players with opportunities to contemplate the outcomes of their decisions and activities. These are just two of many possible ways that players' engagement with ideas or images of civic engagement could be captured and potentially assessed within a game environment.

While a number of speakers addressed this theme throughout the day, we summarize three presentations below as examples of the issues being considered around games and civic engagement.

1. Melissa Gresalfi (Indiana University): Designing assessment for *Quest Atlantis*

Gresalfi discussed assessment and games for change within the context of *Quest Atlantis (QA),* a MUVE (multi-user virtual environment) created by researchers at Indiana University to immerse adolescents in virtual problem-solving environments based on educational content. To date, *QA* has over 15,000 participants in classrooms around the world who are exploring science-, math-, literature-, and social studies-related topics.

While playing *QA,* students engage in problem-solving exercises, or quests, by becoming immersed in a virtual world and engaging in narratives that explore dilemmas that require the practice of disciplinary skills to solve. Quests are based on educational content and the scenarios are designed to help students develop discipline-based problem solving skills. Gresalfi also noted that *QA* is based on socio-scientific issues that require players to practice the skills of scientific thinking and to make decisions that have real-world significance. Interacting with other questers and non-player characters to complete the quests, players face competing demands and must make choices about their solutions to problems based on how they think those solutions might affect the rest of the environment.

Gresalfi commented that playing *QA* is tied to student identity development. Within the game, they must practice skills associated with specific communities of practice. But the problems they encounter in games are also tied to issues students might encounter in the real world. Thus, problem solving in *Quest Atlantis* provides students with experiences that might prepare them to think about those issues in the real world.

2. Chad Raphael (Santa Clara University): A conceptual framework for exploring ethics in games

Chad Raphael proposed an instructional design framework to encourage game designers concerned with issues of ethics and citizenship to design games more intentionally and with an eye toward learning based on desired learning outcomes. The framework is constructed from two axes: agency ↔ structure and expediency ↕ ethics. Changes in any of these factors (in terms of game narratives and mechanics) will influence players' experiences around ethics in a game. For example, a game based in the *Ethics & Structure* quadrant (ethical development over a longer game and an inability to make substantive changes in the game) of the framework might provide players with a normative (restricted in terms of game play choices) environment in which to reflect on the consequences of their activities. Such a game might be effective for citizenship development if it provides players with experiences they might encounter through civic engagement in other areas of their lives as citizens. In another scenario, a game located in the *Agency & Expediency*

quadrant (ability to make choices in the game and learning lessons quickly) might be suited for management and leadership training. Raphael suggested that practice in this kind of environment could transfer into the real world. To assess their efficacy, however, the games should be based on models such that player activities can be compared to models of behavior.

3. Joe Kahne (Mills College): Youth civic engagement

Kahne discussed the potential links between game play and civic engagement, specifically as it relates to participation in a social democracy. He noted that not all social activities pursued by individuals in a democratic society fall under the civic umbrella—and that youth are not always clear on the distinctions between "civic" and "social." He defined "civic" activities as those in which some legitimate public issue is at stake or where the public takes a role in the decision-making.

He discussed surveys and questionnaires as proxies for civic engagement and suggested that longitudinal survey data related to civic engagement and game play might allow for future exploration of relationships between game play and changes in civic activity. But surveys are not related to skill building toward civic engagement, whereas games might be. To create games that facilitate the practice of those skills, Kahne suggested, a normative model for the most effective types of participation is required against which to compare student play. Successful games should not only be measured by a rise in participation, but by their ability to help players develop skills that are relevant to civic action.

Kahne argued that dispositions to engage in democratic participation stem from experience and knowledge, and games provide venues to acquire both of these. Kahne argued that games could prepare students for future experiences in the civic sphere by providing opportunities to investigate social models and simulations, to engage in deliberation, and to network with a wide body of people. Further, games can also engage players by addressing controversial issues and topics of public import, increasing the likelihood of engagement.

Conclusions: Ongoing considerations for assessing game-based learning

The day's discussions made clear how much work is already underway to define and assess the impact of social issue games on their players. It also made clear what an ambitious project this is, how diverse the audience for results will be, and how complex the process of defining and operationalizing specific outcomes can be.

Several concrete suggestions for next steps that could support assessment efforts across multiple games or programs included the following:

1. *Distinguish "critical" from "less important" game design choices:* Continuing in the vein of choice analysis and evidence-centered design, game-based instructional design might focus on "critical choice points" (as determined by the competency model) in games as indicators of successful learning, or where learners might need to improve with formative feedback, while "less important" choices might be areas that allow more latitude for creative problem solving and exploratory behaviors.

2. *Creating a database of "worked examples":* Jim Gee suggested that as serious games become more widespread, people could begin to share strong examples of game design that support problem solving and knowledge acquisition. These designs might be repurposed in games to support learning (or attitude changes) in school-based and non-school-based contexts. This suggestion led to a discussion of the role of instructional design, per se, in social issue games that are not necessarily intended to support students' learning of traditional content, but instead address attitudes, beliefs, or social orientations. While these games undoubtedly need to be guided by design principles and clear target outcomes, "instruction" in the traditional sense may or may not be the right metaphor for the structures the games provide to their players.

3. *Developmental concerns:* Social issue game designers need to consider developmental differences among their audiences, particularly when considering the role of game feedback and how players make use of it in order to modify choices about problem solving. Particularly when assessing outcomes from casual games that may be played by a wide range of people, designing for developmental appropriateness will need to be accompanied by the creation of assessments that accommodate and adapt to players' ages and literacy levels.