SGSEEM: Evaluating Serious Game Frameworks from a Stakeholder Experience Perspective

ABSTRACT

Evaluation of serious game frameworks is emerging as an important area of research. This paper describes an evaluation mechanism called the Serious Game Stakeholder Experience Evaluation Method (SGSEEM). SGSEEM is designed to provide detailed insights into the strengths and weaknesses of serious game frameworks through a stakeholder perspective based approach.

In this paper, we report on the use of SGSEEM to evaluate Makahiki, an open source serious game framework for sustainability. Makahiki facilitates the development of serious games for the purpose of education and behavioral change regarding energy and water consumption. Our results provide useful insights into both Makahiki as a serious game framework and SGEEM as an evaluation method.

Author Keywords

serious games; framework evaluation; sustainability

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g., HCI): MiscellaneousK.8.0. Personal Computing Games

General Terms

Serious Game; Evaluation; Game Design; Case study.

INTRODUCTION

Serious games (games with additional goals beyond just entertainment) have been a topic of academic research for decades [15]. Such games show great potential as interactive media that provide engaging interfaces in various serious contexts [10, 12]. The recent phenomenon of gamification [4] also calls for game-related research in areas beyond traditional entertainment purposes.

One fundamental question in assessing a serious game is the extent to which the game achieves its "serious" purpose. This is quite different from traditional entertainment games, in which assessment focuses on usability or playability [14]. In the field of serious games, there is an increasing focus on the methodology of research and evaluation [9]. De Freitas and Oliver describe a four dimensional framework [3] for evaluating an educational game, consisting of: the context, the pedagogy, the representation, and the learner (or player). Harteveld proposes an alternative approach called "Triadic Game Evaluation" [5], consisting of three perspectives: Reality, Meaning, and Play.

The above approaches focus on evaluation of a single game, as opposed to a game *framework*. Game frameworks (also known as game engines) are "comprised of a collection of different tools, utilities, and interfaces that hide the low-level details of the various tasks that make up a game" [13]. One of the benefits of using a serious game framework is that, if correctly designed, it will provide useful and reusable "building blocks" with which to develop a variety of serious games. These building blocks enable the serious game developer to focus more time and thought on content and results instead of on infrastructure. Yet how are we to know if a serious game framework has been "correctly designed"?

To help answer this question, this paper proposes a method for evaluating serious game frameworks, called the Serious Game Stakeholder Experience Evaluation Method (SGSEEM). In a nutshell, SGSEEM identifies the most important stakeholders of a serious game framework and provides a method for gaining insight into whether the framework is effective and efficient with respect to each stakeholders' needs.

To understand SGSEEM, we will start by briefly introducing Makahiki, our serious game framework for sustainability, and how its development motivated us to create the SGSEEM method. We then describe our preliminary results from the application of SGSEEM to Makahiki. We conclude with the insights this evaluation process provides for our own work on Makahiki as well as for serious game design in general.

MOTIVATION FOR SGSEEM

Sustainability education and conservation have become an international imperative due to the rising cost of energy, increasing scarcity of natural resources, and irresponsible environmental practices. Over the past decade, energy and water challenges have become focal points for sustainability efforts at both university and industry campuses. For example, college residence hall energy competitions have been a widespread mechanism for engaging students in energy issues, with more than 160 taking place or being planned for the 2010–2011 academic year in North America [6].

Designers of such challenges typically have three choices for information technology support: (a) build their own custom in-house solution (as was done at Oberlin College in 2006 [11]); (b) out-source to a commercial provider (as was done at the University of British Columbia in 2011); or (c) use a minimal tech solution such as a web page and manual posting of data and results (as was done at Harvard in 2012).

Gamification'13, October 2-4, 2013, Stratford, ON, Canada.

None of these choices are ideal: the custom in-house solution requires sophisticated design and implementation skills; out-sourcing can be financially expensive and impedes evolution; and the minimal tech solution does not fully leverage the possibilities of advanced information technology.

To provide a new alternative to these three choices, we designed and implemented an open source serious game framework for sustainability called Makahiki [8]. Makahiki is an extensible software system with a variety of common services for developing sustainability games including: authentication; game mechanics such as leaderboards, points, and badges; a variety of built-in games and content focused on sustainability; a responsive user interface; cloud-based deployment; and the ability to customize the game to the needs of individual organizations. Figure 1 illustrates a home page implemented using Makahiki.

Figure 1. Makahiki Game Instance

To explore the ability of the Makahiki framework to support sustainability games in different environments, we ran challenges at three different organizations in Fall 2012: The University of Hawaii, Hawaii Pacific University, and the East-West Center. While these experiences provided anecdotal evidence that Makahiki constitutes a framework, we realized that a more rigorous evaluation would yield better insight into its current quality and requirements for future enhancement.

Upon review of the literature, we found little prior work concerning formal evaluation for the particular needs of serious game frameworks. As a result, we designed SGSEEM, and began applying it to gain better insight into Makahiki and serious game evaluation in general.

SERIOUS GAME STAKEHOLDER EXPERIENCE EVALUATION METHOD (SGSEEM)

The goal of SGSEEM is to determine to what extent the serious game framework under evaluation, as an Information Technology (IT) infrastructure, can effectively and efficiently support the development and play of a serious game.

An *effective* serious game framework can produce a game with the desired outcome with respect to its "serious" goals for the players. For example, an effective serious game framework for energy education and conservation produces a game that increases players' energy literacy and reduces their energy consumption during (and, hopefully, after) the game. Because the goals of serious games are always subject specific, the desired effect of a serious game for sustainability is, for example, different than the desired effect of a serious game for language learning. In this paper, we will refer to subject-specific goals relevant to sustainability, but users of SGSEEM in other domains will substitute goals for their area.

An *efficient* serious game framework can support the full life cycle of game development, execution, and wrap-up of the serious game, including design, management, administration, development, and improvement of the game.

Methodology

Creswell [2] categorizes research methods into three approaches: quantitative, qualitative, and mixed-methods, according to what constitutes knowledge and how knowledge is best acquired. SGSEEM is a mixed-methods approach with both qualitative and quantitative data collection and analysis.

In SGSEEM, qualitative analysis involves structured interviews and questionnaires designed to gain insight about stakeholder experiences with the framework under study. In addition to qualitative analysis, SGSEEM also employs quantitative analysis involving the analytical data recorded by the system, including website logs, player interaction logs, feedback, resource usage, etc.

Stakeholders

There are a great variety of potential stakeholders in serious games. The following are the stakeholders we identified for Makahiki:

- *Players*: those who participate in the game play.
- *System Admins*: those who install and maintain the technological game infrastructure.
- *Game Designers*: those who design the content and game mechanics.
- *Game Managers*: those who manage the game during the period of game play.
- *Developers*: those who extend, enhance and debug the game framework.
- *Researchers*: those who are conducting research using the game framework.
- *Spectators*: those who do not participate in the game play but are interested in the game and the results of game play.
- *Community partners*: those who partner with the game organizers to help run the game (such as coordinating real-world events as part of the game).
- *Facilities*: those who are responsible for the resources (energy, water, etc) associated with the game.
- *Funding organizations*: the organizations who provide funding to the project.

The overall success of a serious game framework for sustainability depends on the individual success of all of these

stakeholders. However, as SGSEEM focuses on the software infrastructure, it does not address the spectator, community partner, facilities, and funding organization stakeholders. These are important stakeholders but outside the scope of our evaluation method.

Our case study evaluation of Makahiki using SGSEEM evaluates: (1) aspects of effectiveness of the system for Players, and (2) aspects of efficiency for System Admins, Game Designers, Game Managers, Developers, and Researchers.

Figure 2 provides an overview of the evaluation framework. The following sections describe in detail the evaluation mechanism for each stakeholder.

Stakeholder	Evaluation Goal
Players	Effectiveness of the game to players in terms of literacy and behavior change in sustainability, player engagement
System admins	Efficiency in administrating the system
Game designers	Efficiency in designing a game
Game managers	Efficiency in managing a game
Developers	Efficiency in developing a game or enhancing the system
Researchers	Efficiency in performing research

Figure 2. Overview of SGSEEM

1. Player Effectiveness

SGSEEM assesses the "effectiveness" of a serious game framework for the Player stakeholder by addressing three questions: (a) To what extent does the game increase player's literacy in sustainability? (b) To what extent does the game produce positive player behavior change in sustainability? (c) To what extent does the game engage players?

Figure 3 illustrates the process for player effectiveness evaluation, which involves a pre-game and post-game measurement for literacy and behavior change, as well as the in-game data logging to measure the level of player engagement.

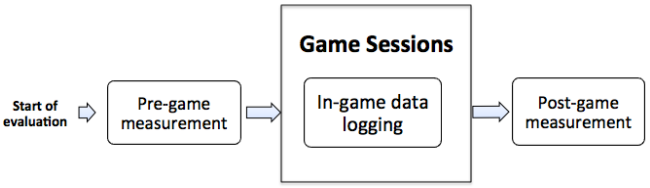

Figure 3. Player Effectiveness Evaluation Process

(a) Literacy assessment: One important goal of a serious game for sustainability is to produce a change in knowledge in the players. A literacy assessment can indicate whether this goal is being met.

SGSEEM uses an approach similar to that described in [1] to assess the impact of the game on player literacy. In general, a set of literacy survey questionnaires are presented to a random selection of the players before the game. After the game ends, the same survey (post-game) is presented to the players who responded the pre-game survey. These two set of survey response data are compared to understand if the game has had an impact on literacy.

The extent of players' sustainability literacy change will indicate the degree of educational effectiveness of the serious game for sustainability.

(b) Behavior change assessment: Positive behavior change is another main goal of a serious game for sustainability. A serious game for sustainability may include some degree of resource consumption measurement. SGSEEM uses resource consumption data before and after the game as part of the assessment of the players' sustainability behavior change. A resource consumption baseline can be established based on historical consumption data. During and after the game, we can compare the resource consumption with the baseline for a particular day to understand to what extent the resource consumption has changed.

The problems with using a baseline to assess the energy reduction in the case of dormitory energy challenges is discussed in detail by Johnson et al. [7]. As a method for evaluating the effectiveness of serious game for sustainability in a broader context beyond the dormitory challenge, we continue to use the baseline method as one way to assess changes in resource consumption.

In addition to resource consumption, SGSEEM can include the use of a behavior survey to measure self-reported behavior change. A pre-game survey is presented to the players to ask about their current sustainability behavior, then after the game, a post-game survey is presented to ask about the players' behavior again. These two sets of survey response data can be compared to understand if there is any changes.

The combination of resource consumption changes and self-reported behavior changes can be used to assess the degree of behavior effectiveness of the serious game for sustainability.

(c) Engagement assessment: Player engagement is an important measure for understanding the effectiveness of a serious game. By investigating the degree of engagement, we can determine to what extent individuals are participating in the game, as well as to what extent the community population is participating in the game.

Engagement has a subtle relationship to the overall effectiveness of a serious game. It is possible for the game to be played by only a subset of the target population, but have an impact on those not playing by virtue of their contacts with players. Gaining better insight into this effect is an area of active study for us.

To obtain engagement data, SGSEEM requires the framework to support the following measures based upon system log data:

- participation rate
- number of players per day
- play time of a player per day
- submissions of all player per day

- social interaction of all player per day
- website errors per day

The participation rate measures the percentage of users who used the game based on the total eligible players. In the serious game context, it indicates the level of involvement or awareness of the serious matters. The number of players and play time per day measure how frequently the players interact with the game. The submissions per day measures the rate of serious game specific activities (online or real world) that players completed, while the social interaction per day measures the rate of social interactions happened in the game between the players. At last, the website errors per day measures the rate of errors encountered by the players while using the game website. In general, with the opposite of website error measurement, the higher value of these measurements are, the higher engagement level the system has.

2. System Admin Efficiency

System administrators simply install the framework and dependent libraries, do backups, and so forth. SGSEEM assesses the efficiency of the system admin stakeholder through interviews involving the following questions:

- How much time did you require to install the system, including the dependencies?
- How much time did you require to maintain the system?
- What problems did you encounter?
- Did you find it difficult to admin the system? What was difficult?

After the interview data is acquired, the evaluator will perform qualitative data analysis, which involves transcribing (if the interview data is in audio format), categorizing and coding the description of reported problems or difficulties. Once the categories of problems are coded, the evaluator will correlate the reported time data and the coded categories to identify the area of strength (less time spent) and weakness (problems and difficulties).

3. Game Designer Efficiency

A game designer uses the serious game framework to design and create a serious game. A serious game framework always provides certain tools or interfaces to game designers with the hope that these will simplify the design of a game. Such tools might involve configuring global settings for the game, such as how long will the game run, who are the players, and how to design individual game elements.

SGSEEM measures the efficiency of a serious game framework with respect to the Game Designer stakeholder by addressing the following two questions: (a) How much time is required to design an instance of a serious game using the framework? and (b) How many, and how problematic are the errors that designers encounter during the design process?

SGSEEM requires the evaluator to first identify the sequence of tasks to be carried out by the game designer, then acquire two sets of assessment data. The first data set is the system log data for the interaction between the game designer and the serious game framework. This data is used to determine how much time it takes a designer to complete a certain design task using the game framework, and any errors encountered.

A second set of data is obtained by interviewing the designer and asking the following questions:

- How much time did you spend to complete each design task?
- What problems did you encounter?
- Did you find it difficult to configure? What was difficult?
- Did you find it difficult to design a specific game? Which one, and what was difficult?

The interview data will be transcribed (if audio recording), categorized and coded. The time data from the interviews will be validated by the system log data, while the descriptive problem categories will be correlated with the system log time and error data to assess the areas of strength and weakness of each category.

4. Game Manager Efficiency

A game manager uses the serious game framework to manage the serious game that the game designers created. It is possible that a game manager is also the game designer. Serious game frameworks normally provide certain interfaces for the managers to manage the game. This may involve managing player submissions, monitoring the game state, entering manual resource data, notifying winners of the game, etc.

SGSEEM measures the efficiency of a serious game framework with respect to the game manager stakeholder with the following questions: (a) How much time is required to manage an instance of a serious game using the framework? and (b) How many, and how problematic are the errors that managers encounter during the design process?

Similar to the evaluation of game designer experience, SGSEEM requires the evaluator to identify the tasks for managing the serious game, then analyze two sets of data to assess game manager efficiency. The first set of data is the system log data for the interaction between the game manager and the serious game framework. This log data is used to determine the time it takes a manager to complete management tasks using the interface, and any system error he or she encountered.

A second set of data is obtained by interviewing the managers to answer the following questions:

- How much time did you spend to complete each managing task?
- What problems did you encounter?
- Did you find it difficult to manage? What was difficult?

The analysis is similar to the game designer evaluation analysis. In the case of the same person(s) sharing the roles of game designer and game manager, when performing interviews, the evaluator should clarify what stakeholder role an interviewing question applies to.

5. Developer Efficiency

The developer stakeholder is different from the game designer stakeholder, in that the game designer stakeholder tailors the framework without requiring any software development, while the Developer stakeholder enhances, corrects,

and extends the system by manipulating code.

To investigate how efficient it is to understand, extend, and debug a serious game framework, SGSEEM assesses how much time it takes to develop an enhancement to the game framework, and how many errors are encountered during the process. This is accomplished by interviewing the developer(s) to answer the following questions:

- How much time did you spend developing and debugging an enhancement to the game framework?
- What problem(s) did you encounter?
- Did you find it difficult to understand, extend and debug the system? What was difficult?

Similarly, the descriptive data will be categorized and coded. The evaluator will analyze the time and problem the developers reported. The areas of strength (less time spent) and weakness (problems and difficulties) are identified from developers' perspective.

6. Researcher Efficiency

Finally, the researcher stakeholder is the one who uses the serious game framework to investigate questions about gaming in general, human computer interaction, etc.

To investigate how efficient it is to do research with the system, SGSEEM assesses how much time it takes to use the system for specific research queries, and how many errors are encountered during the process. The evaluator interviews the researcher(s) to answer the following questions:

- How much time did you spend to collect the research data for a specific topic?
- What problems did you encounter when collecting the data?
- Did you find it difficult to collect data from the system? What was difficult?

CASE STUDY OF MAKAHIKI

Now that we have described SGSEEM, this section will discuss a case study of how we applied SGSEEM to evaluate Makahiki, our serious game framework for sustainability.

Makahiki in Brief

Makahiki is an innovative serious game framework for sustainability whose development began in 2010. It combines both software and hardware infrastructure for the development of sustainability challenges. Makahiki explores one section of the design space where virtual world game mechanics are employed to affect real world sustainability behaviors.

Makahiki consists of a configurable game framework that can be customized to the needs of different organizations. It includes a library of pre-built game "widgets" that implement a variety of game mechanics. Using the widgets, an organization can create a custom sustainability challenge in which players can compete individually or in teams to earn points and reduce consumption of resources such as water or energy. Figure 4 illustrates the architecture of Makahiki. Figure 5 shows a few examples of the games implemented in the Makahiki game library. More detailed description of Makahiki can be found in [8].

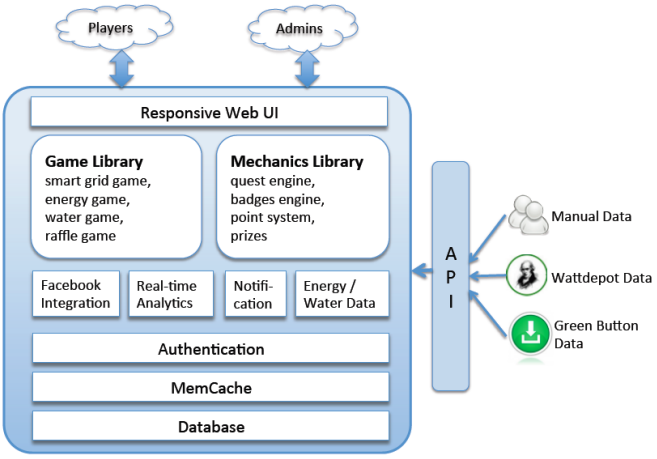

Figure 4. Architecture of Makahiki

(a) Smart Grid Game (b) Energy Goal Game

Figure 5. Makahiki Game Library

Experiences with Makahiki

We have used Makahiki to create four different Kukui Cup Energy Challenges. Kukui Cup Energy challenges were held at the University of Hawaii (UH) in 2011 and 2012 for over 1,000 first year students living in the residence halls. Hawaii Pacific University (HPU) held a Kukui Cup Energy challenge in Fall 2012 for about 200 students. An international organization called the East-West Center (EWC) held a Kukui Cup Energy and Water challenge for approximately 600 international residents living in their residence halls. Since the halls did not have internet-enabled meters, resource consumption data had to be entered by the game managers manually.

The successful creation of serious game challenges by three different organizations provides evidence that Makahiki can be successfully tailored to the needs of different organizations. First, UH and HPU used different metering infrastructure, and EWC collected their resource data manually. Second, while UH and HPU challenges involved only energy consumption data, the EWC challenge involved both energy and water consumption data. Third, the IT infrastructure at UH and HPU provided authentication services using CAS (Central Authentication Service) and LDAP, while EWC used the built-in Django authentication. Fourth, the user interface was customized to "brand" each challenge with the logo, thematic elements, and the education contents of the sponsoring organizations.

Besides the real world usage of Makahiki in the series of Kukui Cup challenges, we performed in-lab evaluation ex-

periments in 2013. Makahiki was used in a serious game development course in Spring semester of 2013 at the Information and Computer Sciences Department of the University of Hawaii at Manoa. There were a total of 8 students who participated in the experiments. The participants were either senior undergraduates or graduate students majoring in Computer Science. During the course, the students installed Makahiki, configured and designed a serious game instance with Makahiki, and finally developed an enhancement to the Makahiki framework. We asked the students taking the course to voluntarily participate in the evaluation experiments of Makahiki, using SGSEEM.

Evaluation of Makahiki

Player Effectiveness Evaluation

We evaluated player effectiveness regarding the 2011 Kukui Cup Challenge at the University of Hawaii at Manoa. There were over 1000 eligible players for this challenge, who were mostly first year college students living in the resident halls. The challenge lasted for 3 weeks. Makahiki recorded detailed logging data from every interaction between the players and the website.

To assess the effectiveness of the game in improving player literacy in sustainability, we conducted two energy literacy surveys, one before the challenge (pre-game) and one after the challenge (post-game). 24 players completed both surveys. Out of the total 19 energy literacy questions, the average number of questions answered correctly is 7.54 before the challenge, and 8.96 after the challenge. This result indicates an 18% improvement on the energy literacy. We also surveyed non-players as a control condition, and found that their literacy did not change, indicating that the improvement in player literacy was indeed due to the game.

To assess the effectiveness of the game in producing positive change in sustainability behaviors, we recorded and analyzed energy consumption data before, during and after the challenge. Before the challenge, an energy usage baseline was established. During the challenge, compared to the baseline, 12 out of the total 20 teams reduced their energy consumption, with the highest reduction of 16.1%. However, 3 teams actually increased their energy consumption, with the highest increase of 11.7%. Overall, the average reduction of the 20 teams was very low—approximately 2%.

To assess player engagement of the game, we calculated a variety of engagement metrics. The results are shown in Figure 6:

Measurement	Value
Participation rate	29
Number of players per day	40 (avg), 130 (max), 1 (min)
Play time per day	15 mins (avg), 8.8 hrs (max), 1 min (min)
submissions per day	111
social interaction per day	101

Figure 6. Makahiki Engagement Metrics

The participation rate of this challenge is 29%, which is very good compared to other sustainability challenges. Over the course of the challenge, an average player spent about 15 minutes per day on the website. One player spent 8.8 hours on one day. There were an average of 111 activity submissions and 101 social interactions between players per day.

In summary, SGSEEM indicates that Makahiki can be successful in achieving player engagement and literacy improvement. SGSEEM could not provide evidence of positive change in behavior.

System Admin Efficiency Evaluation

System admin efficiency evaluation was determined using an in-lab experiment. Students in the serious game class were tasked with installing the Makahiki system into their local computers. In order to understand how much time it takes to install Makahiki and what problems might be encountered, we designed a Google form explaining the steps required to install Makahiki. We asked the students to record the time they spent completing each step and the problems they encountered. We also asked the students to provide feedback about their installation experiences in the form of blog posts.

The Google form data shows that the average total time to successfully install Makahiki was 1.4 hours, with a maximum time of 2 hours and the minimum time of 0.9 hour. The main problem the students encountered during the installation was the difficulty in configuring the Postgres database which is one of the dependent technologies of the Makahiki system.

We calculated the time for each task from the data we collected from the Google Form, and coded the descriptive feedback reported by the students in both the Google Form and their blog posts. If the student reported problems in one of the tasks, we assigned the task a rating value of -1 (negative); if the student gave positive comments for the task, we assigned the task value of 1 (positive). We generalized the local installation tasks into 6 categories which are common to a serious game framework. We aggregated all the ratings from the feedback of the 8 students that participated in the experiment. Figure 7 shows the result of the analysis:

Category	Average Time (minutes)	Feedback rating
Setup runtime environment	10.0	1.4
Install dependencies	29.4	1.3
Install and configure database	37.5	0.1
Download the software	7.5	0.4
Install the software	26.8	1.0
Start the server	2.5	1.0

Figure 7. Makahiki Installation Analysis

The "Install and configure database" category has the longest average time of any category and the lowest average rating. This reflects the issues encountered by students during the configuration process.

In summary, SGSEEM identified database installation as a weak point in installation. Otherwise, SGSEEM indicates generally positive results regarding Makahiki with respect to installation.

Game Designer Efficiency Evaluation

We also used the in-lab experiment to evaluate the game designer efficiency of Makahiki. One of the class assignments for the students in the experiment was to design a serious game using the Makahiki framework. We asked the students to follow specific design steps and record the time required and any problems encountered during their design process, using a Google Form similar to the one used for the system admin evaluation. In addition, students were asked to provide feedback about their design experiences in the form of blog posts.

The game designer efficiency evaluation was generalized into 8 categories corresponding to distinct types of administrative tasks and game design planning. We aggregated all the ratings from the feedback of the 7 students that participated in the experiment. Figure 8 shows the result of the analysis:

Category	Average Time (minutes)	Feedback rating
Update Cloud Instance	7.9	-0.1
Configure Challenge Settings	46.5	-1.0
Resource Goal Game Design	5	0.3
Smart Grid Game Design	107.9	-1.1
Top Score Game Design	10.7	0.1
Raffle Game Design	7.9	0.4
Badge Game Design	13.6	-0.1
Action Submission Management	6.4	0.1

Figure 8. Makahiki Configuration and Management Analysis

In summary, SGSEEM revealed two shortcomings with Makahiki configuration: "Smart Grid Game Design" and "Configure Challenge Settings". Issues encountered in "Configure Challenge Settings" included 1) difficulty in accessing certain configuration options via the Makahiki menu system, 2) a bug in the processing of Ajax queries caused by consecutive clicks on the same interface button, and 3) a bug that prevented users with passwords containing capital letters from logging in. Issues encountered in "Smart Grid Game Design" included 1) difficulty in implementing dependencies between game activities, 2) a lack of documentation on the predicate system used to define dependencies between game activities, and 3) difficulty in generating event attendance codes for game activities.

Game Manager Efficiency Evaluation

We used the 2012 Kukui Cup Challenge at the Hawaii Pacific University (HPU) to evaluate the game manager efficiency of Makahiki. We interviewed the game manager of the HPU Kukui Cup challenge, who is also the game designer of the challenge. We asked him about his game management experiences using the Makahiki admin interface, as outlined in the game manager section of the SGSEEM.

The interview took place after the challenge and was audio-recorded. We transcribed the audio recording. The data shows that the game interface was easy for him to use to manage the challenge. He made sure that player submissions were either approved or rejected within 12 hours. He also discovered a useful feature in the approval interface without help from the Makahiki support team. The only problem he reported was that after the competition ended, he discovered that some of the analytics disappeared. This was identified by the Makahiki development team as a software bug and has since been fixed.

In summary, SGSEEM uncovered few problems with Makahiki game management.

Developer Efficiency Evaluation

We evaluated developer efficiency using an in-lab experiment. One of the class assignments for the students in the experiment was to develop an enhancement to Makahiki. This involved setting up a development environment, following the tutorial to create a "Hello world" widget using Makahiki, and finally, developing an enhancement to extend the functionality of Makahiki.

The students were asked to submit their development source code to the public source code repository (GitHub) and write a blog post to discuss their efforts to complete the development activity.

All 8 students reported that the first task of creating the simple "Hello world" widget was easy, while the enhancement development was hard. Only one student successfully completed all 5 required features, while the rest successfully completed 1 or 2 features. The main problem students reported was the lack of documentation for the development libraries. One student stated in his blog that he decided to choose Makahiki framework to develop his own serious game because of Makahiki's features and possibility of reducing development effort by using the framework.

In summary, SGSEEM reveals significant problems with developer efficiency. Analysis is still ongoing regarding the specific causes of problems and how best to address them.

Researcher Efficiency Evaluation

Several researchers are currently using Makahiki and the Kukui Cup challenge to perform research. As of this time, we have not formally evaluated their experiences regarding the role of researcher stakeholder. We do plan to interview them and analyze the data as outlined in the the SGSEEM researcher stakeholder experience section.

CONCLUSIONS

We have developed a serious game framework evaluation method called Serious Game Stakeholder Experience Evaluation Method (SGSEEM). SGSEEM evaluates serious game frameworks from the perspective of different stakeholders' experiences: player effectiveness, system administrator efficiency, game designer efficiency, game manager efficiency, developer efficiency, and researcher efficiency. These experiences are evaluated qualitatively and quantitatively to determine the effectiveness and efficiency of a serious game framework.

We also applied SGSEEM to Makahiki. The results of the evaluation show both strengths and weaknesses in this framework. Most importantly, the evaluation has provided actionable insight into how to improve the framework for system

administrators, developers, and game designers. We now understand Makahiki far better than we did before the application of SGSEEM.

Our use of SGSEEM also reveals concerns with the evaluation method itself. For certain stakeholders, we took advantage of a course on serious game design to obtain fairly detailed quantitative data about, for example, game design efficiency. While we feel confident of these results, the effort required to collect the data was substantial. On the other hand, for other stakeholders such as game managers, we only had access to a single person who could provide insight from that perspective. While easier to collect, the small sample size limits our confident in the data. We are considering ways to augment the method with a "confidence" value that helps others better interpret the findings. We also hope to apply SGSEEM to a different serious game framework in order to gain additional insights into the strengths and limitations of this approach.

ACKNOWLEDGMENTS

Omitted from review version.

REFERENCES

1. Brewer, R. S. *Fostering Sustained Energy Behavior Change And Increasing Energy Literacy In A Student Housing Energy Challenge*. PhD thesis, University of Hawaii, Department of Information and Computer Sciences, March 2013.
2. Creswell, J. W. *Research design: qualitative, quantitative, and mixed methods approaches*, 2nd ed. ed. Sage Publications, Thousand Oaks, California, 2003.
3. De Freitas, S., and Oliver, M. How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Computers & Education 46*, 3 (2006), 249–264.
4. Deterding, S., Dixon, D., Khaled, R., and Nacke, L. From game design elements to gamefulness: Defining "gamification". In *Proceedings of MindTrek* (2011).
5. Harteveld, C. Triadic game evaluation: A framework for assessing games with a serious purpose. In *Workshop of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems* (2010).
6. Hodge, C. Dorm energy competitions: Passing fad or powerful behavior modification tool? Presentation at the 2010 Behavior Energy and Climate Change conference, November 2010.
7. Johnson, P. M., Xu, Y., Brewer, R. S., Lee, G. E., Katchuck, M., and Moore, C. A. Beyond kWh: Myths and fixes for energy competition game design. In *Proceedings of Meaningful Play 2012* (October 2012), 1–10.
8. Johnson, P. M., Xu, Y., Brewer, R. S., Moore, C. A., Lee, G. E., and Connell, A. Makahiki+WattDepot: An open source software stack for next generation energy research and education. In *Proceedings of the 2013 Conference on Information and Communication Technologies for Sustainability (ICT4S)* (February 2013).
9. Mayer, I. Towards a comprehensive methodology for the research and evaluation of serious games. *Procedia Computer Science 15*, 0 (2012), 233 – 247.
10. McGonigal, J. *Reality is broken: Why games make us better and how they can change the world*. Penguin Press, 2011.
11. Petersen, J. E., Shunturov, V., Janda, K., Platt, G., and Weinberger, K. Dormitory residents reduce electricity consumption when exposed to real-time visual feedback and incentives. *International Journal of Sustainability in Higher Education 8*, 1 (2007), 16–33.
12. Reeves, B., and Read, J. *Total engagement: using games and virtual worlds to change the way people work and businesses compete*. Harvard Business School Press, 2009.
13. Sherrod, A. *Ultimate 3D Game Engine Design & Architecture*. Charles River Media, Inc., 2006.
14. Song, S., Lee, J., and Hwang, I. A new framework of usability evaluation for massively multi-player online game: Case study of "World of Warcraft" game. In *Human-Computer Interaction. HCI Applications and Services*. Springer, 2007, 341–350.
15. Zyda, M. From visual simulation to virtual reality to games. *IEEE Computer 38*, 9 (Sep 2005), 25 – 32.