Outcomes assessment pitfalls: challenges to quantifying knowledge gain in a sex education game [version 1; peer review: 1 not approved]

Elena Bertozzi1, Amelia Bertozzi-Villa2, Swathi Padankatti3, Aparna Sridhar4

1Game Design & Development, Quinnipiac University, Hamden, Connecticut, 06518, USA
2Institute for Disease Modeling, Bellevue, Washington, 98005, USA
3Department of Pediatrics and Neonatology, Sundaram Medical Foundation, Chennai, India
4Department of Obstetrics and Gynecology, David Geffen School of Medicine, Los Angeles, CA, 90095, USA

Abstract

Background: We describe challenges associated with incorporating knowledge assessment into an educational game on a sensitive topic and discuss possible motivations for, and solutions to, these challenges.

Methods: The My Future Family Game (MFF) is a tool for collecting data about family planning intentions. The game was expanded to include information about human anatomy and sexual reproduction. To assess the efficacy of the game as a tool for teaching sexual education, we designed a pre-post study with assessments before and after the game which was deployed in three schools in and around Chennai, India in summer of 2018.

Results: The pre-post process did not effectively assess knowledge gain and made the game less enjoyable. Although all participants completed the pre-test because it was required to access the main game, many did not complete the post test. As a result, the post-test scores are of limited use in assessing the efficacy of the intervention as an educational tool. This deployment demonstrated that pre-post testing has to be integrated in a way that motivates players to improve their scores in the post-test. The pre-test results did provide useful information about players’ knowledge of human anatomy and mechanisms of human reproduction prior to gameplay and validated the tool as a means of data collection.

Conclusion: Adding outcomes assessment required asking players questions about sexual anatomy and function with little or no introduction. This process undermined elements of the initial game design and made the process less enjoyable for participants. Understanding these failures has been a vital step in the process of iterative game design. Modifications were made to the pre-post test process for future deployments so that the process of assessment does not diminish enthusiasm for game play or enjoyment and
motivates completion of the post-test as part of gameplay.

**Keywords**
games for health, serious games, sexual education, outcomes assessment, family planning, India

**Corresponding author:** Elena Bertozzi (elena.bertozzi@quinnipiac.edu)

**Author roles:** Bertozzi E: Conceptualization, Funding Acquisition, Investigation, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; Bertozzi-Villa A: Data Curation, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; Padankatti S: Investigation; Sridhar A: Funding Acquisition, Investigation, Methodology, Project Administration, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work is supported by the Bill and Melinda Gates Foundation through its Grand Challenges in Global Health program which funded the development of the My Future Family Game [OPP1161938]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Copyright:** © 2020 Bertozzi E et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Bertozzi E, Bertozzi-Villa A, Padankatti S and Sridhar A. Outcomes assessment pitfalls: challenges to quantifying knowledge gain in a sex education game [version 1; peer review: 1 not approved] Gates Open Research 2020, 4:73 https://doi.org/10.12688/gatesopenres.13129.1

**First published:** 30 Jun 2020, 4:73 https://doi.org/10.12688/gatesopenres.13129.1
Introduction
The acceptance of games as useful and effective tools for collecting data, educating players, and achieving positive behavior change is growing due to an increase in rigor in the deployment and assessment of applied games (Coover et al., 2017; Zammitto, 2009). Embedding outcomes assessment within the game itself is often described as an important design principle in building games, largely due to the fact that most games incorporate some form of player feedback and metrics as part of gameplay (Ifenthaler et al., 2012; Van Staaldruinen & de Freitas, 2011). There are situations, however, in which such assessment is quite difficult.

The My Future Family Game (MFF) game was initially developed as a tool for collecting information about family planning intentions among adolescents in Mysore, India in 2017. The original idea was to gather information about desired family size and spacing, influencers of the decision-making process, and other data points. Focus group participant feedback during early stage planning was crucial to the success of the project. Researchers determined that although sex education is included in the standard curriculum for adolescents, many young people do not have basic knowledge about human reproduction (Bertozzi et al., 2018). Including this information in the game would strongly motivate adolescents to play, and was supported by parents and educators as a way of communicating sensitive information.

The first beta of the game was successfully tested on 480 adolescents in summer of 2017 and proved to be a very effective tool for gathering information from a population about which little accurate information is available from other sources (Bertozzi et al., 2018).

Discussion of human anatomy and behavior regarding sex and reproduction is problematic in India (Ismail et al., 2015). Many adolescents receive very little information from their parents or teachers due to cultural taboos (Khubchandani et al., 2014). In designing the MFF game, we were very careful to introduce explicit material slowly and through a process in which it was revealed in context. The game was constructed so that at each point where explicit material is available for the player, the player was asked whether or not they wanted to see it, and then if they agreed, the material was presented in a context that made sense based on the information being gathered.

For example, when players were asked information about when they planned to start dating a possible partner, they were provided with information about the anatomy of the opposite sex. When they were asked about the age they planned to marry, after consenting (Figure 1), they were given information about how intercourse works via the animation in Figure 2.

Post-game questionnaires and interviews demonstrated that the game was well-accepted by student players. Analysis of the pilot deployment suggested that the game could function not only as a method of data collection about family planning intentions, but also as a means of assessing preexisting knowledge of and educating players about sex and reproduction (Bertozzi et al., 2018). This created a challenge for the development team because in order to see if the game was effective at teaching adolescents about sex, we needed to know how much they knew about it before they played the game. Given cultural taboos around discussions of sexuality, it was difficult to do so without undermining some of the care that had gone into introducing the topic in the game. The process of designing and deploying a knowledge assessment with the MFF game encountered several pitfalls.

Methods
School selection and gameplay protocol
The second deployment of the MFF game was in Chennai, India as part of research conducted by Dr. Swathi Padankatti and her team from the International Alliance for the prevention of AIDS in collaboration with the U.S. based game development

---

The project was funded by a Grand Challenges in Global Health grant (https://gcgh.grandchallenges.org/grant/childbearing-intentions-and-family-planning-game).

Due to cultural taboos in India which would have made it impossible to deploy the game at all, same-sex marriage was not an option in the game.
In addition to collecting data about family planning intentions, the research and development teams had two additional goals: to overcome challenges identified during the first deployment, and to assess knowledge gain via a pre-post testing framework.

**MFF (original and modified versions of the game are available here: [https://osf.io/gtfu5/wiki/home](https://osf.io/gtfu5/wiki/home) (Bertozzi-Villa et al., 2020). The apks can be installed on any Android tablet or phone.)**

**Challenges from first deployment**

Following the initial deployment of the game, issues with the original study protocol were identified and addressed for this second deployment.

During the early stages of pilot deployment, teachers stayed in the room during gameplay. They often gave stern instructions on how to behave and ordered students to follow the instructions of the researchers. We realized that this made it impossible for students to experience playing the game as play. Due to the presence of their teachers it felt more like a test that they were preparing for it when they reached the game.

A usability issue encountered during the first deployment was that players lacked familiarity with the drag and drop interface commonly used on smartphones and tablets. The design team determined that the pre-test was the perfect opportunity to teach players how to use drag and drop so that they would be prepared for it when they reached the game.

The third main issue identified in the pilot involved the post-game questionnaires. These were paper forms filled out by students after playing the game, asking students to qualitatively self-assess knowledge gain and provide feedback on the process of gameplay. While these questionnaires provided valuable feedback and indicated high rates of self-assessed knowledge gain, they were not efficient data-collection strategies. Because forms were filled out on paper, response rates were low and it was not possible to link student feedback to specific test-takers. Positive self-assessment of knowledge gain was encouraging, but not a rigorous method for determining game efficacy. The absence of an evaluative framework for the game was the primary motivation for development of the pre-post testing process.

**Development of pre-post assessment**

The key educational content of each milestone of the game is outlined in Table 1.

To assess knowledge of these questions while training students on a drag-and-drop interface, the pre- and post-games were designed to show male and female figures in outline, with internal organs visible. A series of 14 anatomy questions covering the full scope of in-game content was presented in a sequence of views. Players answered questions in the pre-test by dragging a word representing a concept (usually with an animation to help explain it) to the correct location on an image (Figure 3). The structure and content of pre-and post-game tests is identical.

The assessment was designed to correlate with the way information is delivered in the different milestones in the game. In the first deployment we noted that players had a difficult time understanding where different organs were located in the body

<table>
<thead>
<tr>
<th>Table 1.</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Milestone</strong></td>
<td><strong>Content</strong></td>
</tr>
<tr>
<td>1: Puberty and bodily functions (same sex as player)</td>
<td>Hair growth, menstruation or ejaculation</td>
</tr>
<tr>
<td>2: Reproductive anatomy</td>
<td>Identification of the internal reproductive organs of male and female bodies and their functions</td>
</tr>
<tr>
<td>3: Puberty and bodily functions (opposite sex as player)</td>
<td>Hair growth, menstruation or ejaculation</td>
</tr>
<tr>
<td>4. Anatomy of intercourse</td>
<td>Act of heterosexual coitus via union of penis and vagina</td>
</tr>
<tr>
<td>5. Fertilization</td>
<td>Movements of eggs and sperm, fertilization of egg by sperm.</td>
</tr>
</tbody>
</table>
and what their functions were. In the assessments, we were careful to depict both male and female bodies as a whole at the start of the assessment. The view then zooms in to just the abdomens of the male and female bodies. We added the whole person views in the top right and left of the screen so that players could understand which view of the body was presented to them. It is very difficult to understand how organs are laid out relative to other organs. For example, in the female body, it can be difficult to show the positions of the three apertures of the urethra, vagina, and rectum relative to one another. The additional views were added to minimize this confusion.

Our hope was that the layout of the assessment prior to play would prepare players to approach the anatomy section of the game where they have to drag and drop each body part to its correct location (Figure 4). During the first deployment of the game, it was clear that some players did not understand the difference between the front and side views of the anatomical drawings. In addition to adding the side views in the upper right and left corners of the assessment, we also incorporated them into the minigame. These views update as each organ is dragged into the correct location in the front view.

Analysis
During gameplay, tablets kept timestamped records of every user input. Data on pre- and post-test responses were saved in .csv format for statistical analysis. These datasets include information on the tablet used, the school in which the game was deployed, the self-reported gender of the player, and a unique user id for each run-through of the game. The pre/post data contains no other personalized student data.

All analyses were run in R version 3.6.0 (Bertozi-Villa, 2020). Overall pre- and post-test scores, as well as the percent of students who responded correctly to each question, were calculated from individual responses. On the post-test, players who responded “not sure” to every question were logged as having a “null” post test. Score differences between groups were assessed via two-sample t-tests, and pre- to post-test score changes were assessed via one-sample t-tests.

Ethics and consent
The study design was approved by the Institutional Review Board of the Sundaram Medical Foundation, Dr. Rangarajan Memorial Hospital, Chennai, India (IRB # IEC-09/1/2018).

Informed verbal consent was obtained from the principals of the three participating schools following consultation and a gameplay demonstration with each one. Consent was not obtained from student participants. The board deemed oral consent would suffice for the principals, and as the game covers topics which are part of the curriculum, participants’ consent was not needed.

Results
The goal of this analysis was to test if embedding the game within a pre/post assessment would accurately assess how much players had learned over the course of the game. Unfortunately things did not turn out the way we planned. Reasons for this are outlined in detail below. We did determine that the pre-test is a useful way to gather information about student knowledge of male and female anatomy and some sexual functions. The results demonstrate that the assessment tool is very helpful in demonstrating which schools are doing better and specifically which topics are better understood.

Pitfall one - Pre-test made the game feel like a test
As described in the methods, considerable care was taken in the game design to encourage a sense of play and remove the
To counter this, the researchers repeatedly stated that they should just answer what they knew and then go on to the game, but this clearly affected the experience. We learned that in future deployments, we need to add more context and less pressure to the pre-test to ensure players understand that they will not be criticized or penalized for not knowing the answers.

Pitfall two – Language issues with terms for sexual parts and functions
To be as accessible as possible to players at any reading comprehension level, the game includes as little text as possible and communicates most information through graphics, audio and animation. This is especially important when discussing information about sexuality because these terms may not be familiar to students. However, the inclusion of the pre-post assessment introduced a great deal of technical vocabulary in English before the gameplay began. All of the schools included in the study had instruction in English, but it was unclear if terms like testicles, ovaries, urine and feces were well understood by players (Figure 5). Although education about sexual functions is technically part of the educational curriculum for all students in India, the content is not actually taught in many schools due to cultural reluctance to discuss sexuality. In the results that follow, it is possible that some of the variation in scores on the pre-test is related to differences in knowledge of terms rather than differences in knowledge of sexual/anatomical functions.

Pitfall three – Low participation in the post-test
While the transferal of the post-test questionnaire into a digital framework did allow for personalized tracking of results, there were challenges to collecting post-game information. Qualitative feedback from the deployment team indicated that, when students came to the end of the game, and saw the same screen they had seen earlier for the pre-test, many simply dragged the tiles to “Not Sure” because it was the fastest way to get to the final screen. Others simply put down the tablet which meant that researchers had to exit the player from that game session (with no responses to the post-test questions) to reset the tablet for the next group of students. Due to the fact that we do not know exactly what happened in all the cases where there appear to be random answers to the post-test, we cannot determine how many students actually answered the questions.
We failed to provide players with a compelling reason to want to engage in the final assessment, which will be corrected in the next deployment.

**Assessment of pre-test scores**

A total of 419 students in three schools completed the pre-test and main game. The schools were selected based on scheduling availability and willingness to participate. The researchers from the IAPA had previously worked with these schools on AIDS education initiatives. Across all schools, the pre-test score was 33.5% on average (SE 1.15%), with substantial variation between schools. In particular, students at School 2 (who had sexual education as a formal part of their curriculum) performed significantly better than students at Schools 1 or 3 (two-tailed t-test *p*<0.001). Pre-test scores were not significantly different between male and female students at any school (Figure 6).

Across all schools, students scored slightly better on pre-test questions relating to the anatomy of their own sex compared to the opposite sex, but this effect was not statistically significant (Figure 7). For female anatomy questions, 34.4% (SE 1.86%) of female respondents answered correctly, compared to 30.6% (SE 1.6%) of male respondents (two-tailed t-test *p*=0.13). For male anatomy questions, 33.3% (SE 1.89%) of female respondents answered correctly, compared to 36.3% (SE 1.8%) of male respondents (two-tailed t-test *p*=0.25).

As shown in Figure 8, the only questions for which a majority of responses were correct were “Where is urine excreted from a male?” (53.0% correct) and “Where does a lining build up to prepare for pregnancy?” (50.6% correct). For eight of the remaining 12 questions, the correct answer received a plurality of responses, but not a majority. The four questions for which the most frequent response was not the correct answer were “Where sperm exit the body?” (plurality answer “Not Sure”, 29.8%), “where menstrual blood is excreted?” (plurality answer “Not Sure”, 27.7%), “The organ that becomes erect before intercourse?” (plurality answer “Vagina”, 33.7%), and “Where urine is excreted from a female? (plurality answer “Vagina”, 35.6%).

**Pre-post test assessment**

As described above, assessment of knowledge gain was complicated by the large number of students who did not complete the post-test or who rushed through it, answering “not sure” to all questions (173 students, 41.3% of total). We refer to this group as having a “null” post-test. While it is not possible to assess knowledge gain among those with a null post-test, among the 246 (58.7%) students who did attempt the post-test we find an average 6.27-point score gain between pre- and post-tests (95% CI 3.8-8.75, one-sample t test, Figure 9).

A question-by-question breakdown of pre- vs post-test result among those who attempted the post-test shows the largest knowledge gain around topics of intercourse, egg storage, and sperm movement (Figure 10).

**Discussion**

Our initial response to the results was dismay. It appeared that the game was not a useful tool for teaching players because overall there was very little change between pre- and post-test...
Figure 6. **Pre-test scores by school and sex of respondent.** Bars represent the standard error of the mean. N=419.

Figure 7. **Pre-test scores by type of question and sex of respondent.** Bars represent the standard error of the mean. N=419.
Figure 8. Pre-test responses by question. The highlighted bar shows the correct answer in every instance. N=419.

Figure 9. Violin plot of score distributions in the pre- and post-test, for those students who completed both tests (N=246). Points represent means, and bars represent two times the standard error.
results. Discussions with the deployment teams and more detailed analysis of the results produced a more nuanced understanding of what happened. When the data for players who did complete both the pre- and post-tests was analyzed separately, there was a modest but notable increase in knowledge. Additionally, we determined that the pre-test was a useful tool for assessing prior knowledge.

We learned a great deal about the difficulty of creating effective pre-post assessments for a game that includes sensitive topics. Adolescents offered a game of this type are already nervous and excited about it. The process of setting up a context in which their current knowledge is assessed needs to be approached carefully. We encountered several pitfalls that complicated the assessment process and which affected the validity of the assessment data. We are able to conclude that using a game to assess current knowledge of reproductive anatomy and processes can be very effective. In order to assess knowledge gain after gameplay, students need to be motivated to fully engage in the post-test assessment. For future deployments of the game, we plan to change the deployment protocol to address the issues discussed in this report and better integrate the pre-post testing process in the overall experience.

It is standard practice in applied game development to seamlessly integrate assessment into the existing structure of the game (Klopfer et al., 2018; Serrano-Laguna et al., 2018). As we have shown, this is difficult in a game that deals with a sensitive topic. Our plan going forward is to address this challenge openly in the introduction to the game experience. After players open the tablet, we will have an animated character appear who discusses the fact that what will follow is a game about sexuality and that this is a difficult topic for many people to talk about. After normalizing the idea of embarrassment, the character

Figure 10. Question-by-question comparison of pre- and post-test responses, by proportion, excluding those with “null” post-tests (N=246).
will then introduce the idea that knowledge is power and that the game will help players learn about things that are important to their future. Then the pre-post will be presented as a challenge…” let’s see how much you know now and then see if after you play the game you know all the answers to things you didn’t know before.” Hopefully, with this context, we will avoid the pitfalls of our Chennai deployment.

Conclusion
This deployment demonstrated that a game-based tool can be an effective means of gathering information. We learned that many adolescents in these schools lack basic knowledge of human anatomy and sexuality, especially given that the students chosen had already received baseline training in HIV prevention and are likely better informed than other students. The deployment also provided us with important information for improving the tool.

Data availability

Underlying data
Open Science Framework: Outcomes Assessment Pitfalls: Challenges to Quantifying Knowledge Gain in a Sex Education Game. https://doi.org/10.17605/OSF.IO/WMHCD (Bertozzi-Villa et al., 2020)

This project contains the following underlying data
- prepost.csv (Questions and responses to all pre- and post-tests administered, along with timestamps and other metadata)

Extended data
Pre- and post-test data were analyzed and visualized using R version 3.6.0. All code is available from GitHub (https://github.com/bertozzivill/india-family-planning) and archived with Zenodo (http://doi.org/10.5281/zenodo.3822455 (Bertozzi-Villa, 2020))

Data are available under the terms of the Creative Commons Zero “No rights reserved” data waiver (CC0 1.0 Public domain dedication).

Software availability
An installable and playable version of the game and all data used for analysis is publicly available at Open Science Framework, as described below.

Archived source code at time of publication: https://doi.org/10.17605/OSF.IO/WMHCD (Bertozzi-Villa et al., 2020)

License: MIT

Acknowledgements
We are grateful to the International Alliance for the Prevention of AIDS deployment team:

Nivetha Jagadeesh, Sheema Gopi, Arulraj Louis, N. Gayathri.

This project would not have been possible without the work of the game development team: Zachary Kohlberg, Christopher Blake, and Jacob Kohlberg.

References

Coovad MI, Winner J, Bennett W, et al.: Serious Games are a Serious Tool for Team Research. International Journal of Serious Games. 2017; 4(1). Publisher Full Text
Thank you very much for the opportunity to review this paper. In it, the authors consider how to integrate knowledge assessment into a sexual health game intervention entitled 'My Future Family Game' delivered to young people in Chennai, India.

This game-based intervention aims to collect information about family planning intentions. Here, the authors consider the efficacy of a pre-post assessment and the post-test assessment in particular. They describe the development process and some of the pitfalls of the approach they took.

Overall, the paper is interesting and I enjoyed reading about the logistics and decision making that the team underwent. Working internationally on game development can be difficult and it was interesting to see how the team tackled field research. That said, this seems to be a report about processes and decisions that were made rather than a generalizable paper couched in the larger literature.

Specific points:

Introduction:
- If this paper is about game assessment, then the introduction should provide a literature review on that topic.
- The introduction turns into a description of pilot game design and development. Instead, this information perhaps belongs in the methods section.

Methods:
- The methods section should be the planned study or evaluation. Instead it seems almost like a results section in that you state how many people were involved etc.
The section on challenges from first deployment again seems like the result section. Regardless, it is a bit confusing as to how this relates to an assessment of knowledge.

Results:
- It seems that the many pitfalls are avoidable and others have successfully done game-based assessment. It is not clear whether these are particular to the ways the researchers set up the assessment or inherent to assessment, I would argue for the former.

Discussion:
- Statements such as “our response to the results was dismay” is quite casual and instead perhaps the discussion could be used to review the literature on game assessment.
- It is not clear that pre post assessment issue is particular to sensitive topics and can arise for other reasons

Is the work clearly and accurately presented and does it cite the current literature?
No

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: ASRH, game-based learning

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.