Outcomes assessment pitfalls: challenges to quantifying knowledge gain in a sex education game [version 1; peer review: 2 not approved]

Elena Bertozzi, Amelia Bertozzi-Villa, Swathi Padankatti, Aparna Sridhar

1 Game Design & Development, Quinnipiac University, Hamden, Connecticut, 06518, USA
2 Institute for Disease Modeling, Bellevue, Washington, 98005, USA
3 Department of Pediatrics and Neonatology, Sundaram Medical Foundation, Chennai, India
4 Department 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
Introducción

La aceptación de los juegos como herramientas útiles y efectivas para la recopilación de datos, la educación al jugador, y la promoción de cambios de comportamiento positivos está creciendo debido a un aumento en la rigurosidad en la implementación y evaluación de aplicaciones de juegos (Covert et al., 2017; Zammitto, 2009). La evaluación de resultados incorporada en el juego es a menudo descrita como un principio de diseño importante en el desarrollo de juegos, ya que la mayoría de los juegos incorporan algún tipo de retroalimentación del jugador y métricas como parte de su juego (Ifenthaler et al., 2012; Van Staalduinen & de Freitas, 2011). Existen situaciones, sin embargo, en las que esta evaluación es bastante difícil.

La aplicación My Future Family Game (MFF) fue inicialmente desarrollada como una herramienta para recopilar información sobre las intenciones de planificación familiar entre adolescentes en Mysore, India en 2017. La idea original era recopilar información sobre el tamaño deseado y la separación de familias, influenciadores del proceso de toma de decisiones, y otros datos relevantes. La retroalimentación de los participantes durante las fases iniciales de planificación fue esencial para el éxito del proyecto. Los investigadores determinaron que a pesar de que la educación sexual está incluida en el currículum estándar para adolescentes, muchos jóvenes no tienen conocimientos básicos sobre la reproducción humana (Bertozzi et al., 2018). Incluir esta información en el juego, motivaría a los adolescentes a jugar, y fue apoyada por los padres y los educadores como una forma de comunicar información sensible.

La primera beta del juego se probó con éxito a 480 adolescentes en verano de 2017 y demostró ser un muy efectivo medio para recopilar información de una población para la que se poca información precisa está disponible (Bertozzi et al., 2018).

La discusión sobre la anatomía y el comportamiento relacionado con la sexualidad y la reproducción es problemática en India (Ismail et al., 2015). Muchos adolescentes reciben muy poca información de sus padres o maestros debido a tabus culturales (Khubchandani et al., 2014). En el diseño del juego MFF, nos comimos muy cuidadosos para introducir el material explícito de manera gradual a través de un proceso en el que se revelaba en contexto. El juego estaba construido de manera que en cada punto donde el material explícito estaba disponible para el jugador, el jugador era consultado sobre si o no se quería ver el material, y a continuación, si se aceptaba, el material se presentaba en un contexto que hacía sentido con la información que se estaba recopilando.

Por ejemplo, cuando los jugadores fueron consultados sobre la edad en la que planeaban casarse, después de la edad de consentimiento (Figura 1), se les proporcionó información sobre cómo el coito funciona a través del arte animado en la Figura 2.

Discusión de la anatomía humana y el comportamiento relacionado con la sexualidad y la reproducción es problemática en India (Ismail et al., 2015). Muchos adolescentes reciben muy poca información de sus padres o maestros debido a tabus culturales (Khubchandani et al., 2014). En el diseño del juego MFF, nos comimos muy cuidadosos para introducir el material explícito de manera gradual a través de un proceso en el que se revelaba en contexto. El juego estaba construido de manera que en cada punto donde el material explícito estaba disponible para el jugador, el jugador era consultado sobre si o no se quería ver el material, y a continuación, si se aceptaba, el material se presentaba en un contexto que hacía sentido con la información que se estaba recopilando.

Para un ejemplo, cuando los jugadores fueron consultados sobre la edad en la que planeaban casarse, después de la edad de consentimiento (Figura 1), se les proporcionó información sobre cómo el coito funciona a través del arte animado en la Figura 2.

Metodología

Selección de escuela y protocolo de juego

El segundo despliegue del juego MFF fue en Chennai, India como parte de la investigación realizada por Dr. Swathi Padankatti y su equipo de la Alianza Internacional para la prevención del VIH en colaboración con el equipo de desarrollo de videojuegos del U.S. que recibió el apoyo de la Alianza Internacional para la prevención del VIH.

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 company.
team (Dr. Bertozzi’s group at Quinnipiac University) and Dr. Aparna Sridhar at U.C.L.A.’s School of Medicine. Dr. Padankatti and her team identified three schools willing to participate in the study who could provide a total of 419 student players. Schools were selected based on the research team’s pre-existing relationships with administrators, with whom they had previously worked on AIDS education initiatives.

For game deployment, a set of 30 android tablets and headsets were set up in a school classroom, and groups of students in the target age group successively cycled through to play the game and discuss their experience. Groups were not segregated by sex. To ensure comfort and privacy, students were able to move freely around the room to find their preferred space to sit and play. Upon beginning the game, players are first asked to indicate their sex and age, after which the pre-game quiz is triggered prior to initiating the main game. The post-game quiz, with exactly the same structure and questions as the pre-game, appears after completing the game. Given that this was the first field deployment of the revised game, the deployment team reported issues to the U.S. development team after each play session. The issues were collected and organized into topics to be addressed in future revisions of 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/ (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 required to engage in. To encourage a sense of play, the research protocol was modified early in the first deployment to ask instructors to leave the room during gameplay. Additionally, language was added to the introductory scripts, encouraging students to play the MFF game as a game – they should only do the parts of it that they wanted to, and could stop playing at any time. This protocol was extended into the second deployment.

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>Milestone</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<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 (Bertozzi-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
pressure associated with an examination. By introducing a pre-test, however, we recreated the circumstances under which the experience of play was potentially undermined. Students were invited into the room to play a game. However, after they are welcomed to the game, they are presented with an assessment. The deployment team reported that some students were concerned that they did not have the “right” answer and wanted to be able to go back and correct their previous answers during the pretest. Given that the Indian system of education heavily relies on test scores and impactfully rewards those who test well, these students appeared very motivated to “do well” as soon as they realized it was an assessment.

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.

Figure 4. Screenshot of anatomy minigame.
intentionally. 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 on average a 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
Coovert MD, 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
Reference Source
Reference Source
Claudia Regina Furquim de Andrade
Department of Physiotherapy, Speech Therapy and Occupational Therapy, Faculty of Medicine FMUSP, University of Sao Paulo, Sao Paulo, Brazil

I would like to thank you for the opportunity to review the manuscript entitled “Outcomes assessment pitfalls: challenges to quantifying knowledge gain in a sex education game” that has been submitted to Gates Open Research.

The study investigated and discussed the challenges associated with incorporating educational assessment before and after the use of an educational game entitled 'My Future Family Game', delivered to students of three schools in and around Chennai, India. The authors investigated the efficacy of a pre- and post-test assessment. I consider this topic relevant and interesting, and in need of further investigation, especially with the application of experimental designs. Therefore, the authors are commended for the novel and trendy approach. However, I note several methodological concerns. Among these, relevant literature regarding the use of educational games, particularly concerning outcomes assessment, is not reviewed (and was not used to guide the research design); inclusion and exclusion criteria, for both participants and schools, are not specified and may not have been controlled; and the study methods are not described in sufficient detail. For these reasons, I am concerned that it is not possible to fairly review the results and conclusions reached. Furthermore, I would suggest publishing this manuscript as a research note, since it mostly describes the unexpected observations and lab protocols.

Comments on the Introduction:
- In general, the background/rationale for this study is not sufficiently developed as written. The study is not hypothesis-driven and predicted outcomes are not provided.
- The introduction, albeit interesting, is essentially a description of the game used for this study, detailing its design and development.

Comments on the Methods:
- I believe most of the information concerning the game development and pilot studies could
be presented, in short, in the methods section, with citations to the previous published studies. In addition, the methods section should clearly and succinctly explain the study methods, so that it can be duplicated.

- Inclusion and exclusion criteria, for both participants and schools, are not provided (for example, were the schools chosen only by convenience? What about the student participants in each school, how were they selected?). Demographic information about included participants is not very substantial either. The design of the pre- and post-test assessment is also not clearly described.

Comments on the Results:
- The manuscript depicts a lack of interest from many participants in completing the educational assessments. I suggest that the authors elaborate on what they believe was causing this lack of interest. Could it be influencing the results?

Comments on the Discussion:
- The manuscript would benefit if the authors discussed their findings on relevant literature. The authors did not cite and engage with the pertinent literature regarding educational assessment.

**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?**
No

**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:** I am a Full Professor at the Faculty of Medicine at the University of São Paulo, and published 155 articles in indexed journals and 376 papers in the annals of events. I have additional 96 publications, between book chapters and books, and 361 items of technical production. Between 1989 and 2020 I participated in 24 research projects, 22 of which I coordinated. I am also the coordinator of three specialization courses.

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.

Author Response 28 Oct 2020

Elena Bertozzi, Quinnipiac University, Hamden, USA

Thank you very much for the detailed and thorough feedback in this review. We have implemented your suggestions as detailed below.

1. Added literature review.
2. Added inclusion and exclusion criteria.
3. The study is not hypothesis driven. The paper seeks to add to the scarce literature in the field relating to assessment of the efficacy of serious games to educate players about sexual anatomy and reproduction.
4. The introduction and other sections of the paper have been revised to clarify the paper's purpose and arguments.
5. The results section was edited to clarify the reasons for the lack of participation in the post test and add suggestions to resolve the issues we encountered.
6. The discussion section was edited to clarify connections to the pertinent literature.

The paper was reorganized to clarify the purpose of the study and how publication of the issues we met and resolved could be useful to other researchers.

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

Reviewer Report 17 August 2020

https://doi.org/10.21956/gatesopenres.14316.r29297

© 2020 Gilliam M. This is an open access peer review report 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.

Melissa Gilliam
Department of Obstetrics and Gynecology, University of Chicago, Chicago, IL, USA

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.

---

Author Response 28 Oct 2020

**Elena Bertozzi,** Quinnipiac University, Hamden, USA

Thank you for this thoughtful review and for the suggestions that guided our revisions to the paper.

**Specific Points:**

Intro: We added a literature review and analysis of results which we used to focus the introduction of the paper. The introduction now clarifies that the topic of the paper is refining the process of outcomes assessment for a serious game, specifically one that addresses sexual anatomy and reproduction.

Methods: We reorganized the paper so that the methods section explains the process that we undertook based on the first and second deployments of the game and problems associated with the addition of the pre/post assessment.

Results: There is a growing body of research on game-based assessment, however there is little published work in the area of games related to education about sex and reproduction. The assessment of knowledge gain without in person interviews is complicated by numerous factors and our paper seeks to provide guidance in this evolving area.

Discussion: We have revised this section to remove unprofessional language and clarify the connections between the assessments and issues relating to serious game deployments.

The paper has been reorganized for clarity.

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