Plan Overview

A Data Management Plan created using DMPonline

Title: Does a workshop to create visual novels elicit bias regulation?

Creator:Dídac Jiménez Torras

Principal Investigator: Dídac Jiménez Torras

Data Manager: Dídac Jiménez Torras

Project Administrator: Dídac Jiménez Torras

Affiliation: Other

Template: DCC Template

ORCID iD: 0009-0007-3565-437X

Project abstract:

The aim of this study is to examine whether creating a visual novel (VN) video game contributes to the ability to self-regulate and detect cognitive biases. A “visual novel” video game is like a digital comic, but interactive.

To conduct the study, I have already contacted a civic centre where I intend to conduct a 10-session workshop on creating visual novels (individual, not with groups). My intention is to cover several steps to making a visual novel, with an emphasis on storytelling, identity, and iteration, encouraging students to give mutual feedback while learning about art, music, and interaction.

I plan to disseminate the results in a scientific article. This study is being conducted in Barcelona, in Catalan and Spanish.

ID: 183515

Start date: 07-10-2025

End date: 09-12-2025

Last modified: 11-09-2025

Does a workshop to create visual novels elicit bias regulation?

Data Collection

What data will you collect or create?

The main data results will be the answers to a 27-question survey (21 questions belong to the instruments). The questionnaire will be released in .xlsx, .csv, .ods, and perhaps .pdf files. The chosen formats allow versatility and interoperability and can be visualised in many repositories without needing to download external software.

No data exist at the moment of making this DMP.

Expected size/volume should be around 1 Mb.

Additional data will come from observation rubrics and audio interviews.

The chosen formats are easy to store, and although audio may be somewhat prone to corruption, I will seek generate a transcript with either Hugging Face, Audacity's whisper AI model, or YouTube's subtitles feature. If it comes down to YouTube, I will make sure to remove the anonymised audio file once the subtitles are generated.

How will the data be collected or created?

The data will be collected using a Microsoft Forms survey (or Qualtrics, if possible). Naming will not be compressed into codes, files will instead have self-explanatory descriptive titles. Similarly, the resulting data from this project will follow an exploratory folder structure.

Versions will be tracked through the repositories (ResearchBox, and OSF), although no revisions are expected. If published, then we will issue a note to update the article data.

Statistical analysis integrity will be ensured by following a dummy coded file, ready to be used by just introducing new data. The file was stress tested to have thresholds that don't give a random hit after 50 trials with 50 data random points. You can check it here:
https://claude.ai/public/artifacts/9475c6b8-e4ab-4e9e-a61e-eab5b6f2a53b

Documentation and Metadata

What documentation and metadata will accompany the data?

Only the titles, questions, and whether a question is reversely coded should be needed to understand the data. This information will be included in the dummy coding statistical analysis sheet, or in a separate coding book. Either way, both documents will be together, so the information to repeat the analysis will be there (also, to facilitate the process, the provided data will be normalised on a scale from 0 to 1).

The documentation will be created manually, and may include additional comments to guide through the analysis process. Metadata will include a title, format, size, file language, scale, keywords, and a brief definition with links to the original instruments following OSF's standard model. Note that a complimentary .txt may be included too.

Ethics and Legal Compliance

How will you manage any ethical issues?

Regarding the main data source, the survey has a question asking for particular consent for data sharing. Participants who do not adhere to it will be used in the analysis but will be removed from the final open data.

Identity will be protected through a pseudo-anonymisation process. Although I will know the names of the participants, each one will receive a Roman alphabet letter, which I will not know — the blinding process will be as follows: A research colleague to be determined will generate a random list of letters using Random.org, and will give me a document or the cutouts of the various letters arranged in order. The papers will be placed in an envelope or other non-transparent paper, with a number (the order in which they will be delivered). Only I will be able to see this number, and I will deliver, in order of rank, the papers, according to the numbers.

The students who participate will put their letter (but not the number) of the papers, in the survey, completing the blinding process, so that neither I nor other researchers can link the data to a specific person (except for my research colleague to be determined, who in any case will not know or have seen the students).

No "sensitive" data will be handled per se, as the digitised version will already have the filtered "anonymised" results. Nevertheless, I will download it from Microsoft Forms and store it in my Windows 11 computer, which is not isolated from the Internet. I will then share the files through Telegram, which is not encrypted. Upon publishing, the files will be deleted and uploaded on repositories.

When it comes to the diary observation rubrics and the presumed interviews to elaborate exploratively, although I will personally know the participants, I won't identify them, neither on the record, nor on the uploaded files, and I won't give demographic clues therein to extrapolate who they were. I will keep the files on my personal devices. I will aim to anonymise the sound of the interviews as soon as I make them, to avoid personification if I were to get hacked or stolen.

When it comes to ethical issues and formal consent, I am pursuing approval by UOC's committee, and will share my contact details every one of the 10 sessions, in case there are doubts about the research process.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

I assume as the main researcher I am the "owner" of the data whereby it came from my particular investigative efforts. Nevertheless, the data will be licensed with the most open licence (CC0) I am allowed to provide once it is published in a journal.

I don't put any restrictions in the re-use of data by third parties for research purposes, but data scrapping, AI, and corporate entities should avoid using the data for secondary analysis.

Data won't be postponed or restricted to seek patents as this will be a mainly "confirmatory" study in terms of whether to do a follow-up (turning into explorative to inform the rest of my thesis), without "new" discoveries, rather I'll test if creating a VN can help with arrange one's biases.

Storage and Backup

How will the data be stored and backed up during the research?

I have sufficient internal and cloud storage, the survey responses shouldn't be more than 1Mb, the rubrics around 5Mb, and the audio files around 75Mb.

The data will be uploaded two times, once the survey results are drawn, and once the analysis is over. It won't be periodically backed up, as it won't be updated. The documents are not meant to be "alive" or "revised", they are analysed, and then it's over.

I will check periodically (every 6 months) during the 5 years after publication whether the cloud services (OSF, ResearchBox, and Playbook) are still online, ensuring the links work, and finding apt replacement (Zenodo, Figshare) if they don't.

Recovery and/or backup in case of incident will be resolved by the technical and IT teams of UOC and/or the services providers, in a case-by-case basis, contacting them to ensure problems get resolved.

How will you manage access and security?

Once both surveys are answered, the data will be downloaded to the computer of the principal investigator (me, Dídac), and will be uploaded to the OSF repository (https://osf.io/5hqem/?view_only=5b48b88363594319a7a176b3e2c171eb) and ResearchBox (https://researchbox.org/4546&PEER_REVIEW_passcode=SPGYSG) of the study. With this double authentication, once the study is accepted for publication, I will delete both the original Forms questionnaire to avoid data usurpation in Microsoft's private cloud, as well as my local files — which I will upload to my personal cloud on the Playbook platform.

As stated, I plan to keep the survey data, in .xlsx and .csv format (less than 1Mb expected). As per the observation rubrics, they will be stored in .pdf or .png format depending on whether I physically print them (less than 5Mb expected). The presumed optional interviews would be stored after audio distortion in .txt and .wav format (75Mb expected per file). They will be maintained on the aforementioned platforms as long as they exist, and I commit to periodically reviewing (once every 6 months, with an alarm) that they remain online for the 5 years following the completion of the study.

There exist data risks (mainly getting infected with malware, and the cloud storage providers becoming malicious and manipulating or stealing the data). These risks will be managed on their due time by finding alternative storage places.

Collaborators will have access to the data via Telegram messaging or through email, at djimeneztorr@uoc.edu, or didacjt@hotmail.com.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

The data will be openly accessible for as long as the platforms to store it in exist, unless they get updated to a decentralised system, or otherwise prohibited for some reason.

The data of participants that haven't given permission to have their results published will be destroyed as soon as the article study with the results gets published.

The data will be used to adapt, promote or discard the creation of visual novels as an educational process/workshop. Knowing the specific percentages that have changed of the various constructs can

be used to fit the activity into specific educational modules or processes.

What is the long-term preservation plan for the dataset?

Once again, data will be stored in OSF and Research Box (public), and Playbook (private). No monetary costs come associated with uploading the data and sharing it on these platforms. Nevertheless, the process of generating, tagging, coding, and explaining the process to ensure it follows FAIR Open Access principles is estimated to take around 75 hours. 25 hours for preparation, 25 for analysis, and 25 for archival, revision, contacting, transfer, blinding, and other procedures.

Data Sharing

How will you share the data?

Data will be shared through online repositories (OSF and ResearchBox) linked in the article with a persistent identifier along the research results. The data will also be available in the appendices. Potential users will be able to find the data in said purported published article.

During the process, two research collaborators that are to be determined as well as my two supervisors, Daniel Aranda and Joan Pons, will be able to see the data.

Are any restrictions on data sharing required?

No restrictions or difficulties are expected beyond the possible hesitation of participants to openly share their information. If any participant sends me a message to remove its contributions, I will do that as soon as I see it.

No special system, algorithm, or biological sample/project is being undertaken here, so I deem non-disclosure and exclusivity agreements to be overstated.

I expect the analysis to take about 1 week, but since I will be busy with other studies, I project the manuscript won't be ready until past mid-2026, with a projection to have the paper published in 2027.

Responsibilities and Resources

Who will be responsible for data management?

I, Dídac Jiménez Torras, am the sole main researcher in this study. This is not meant as a collaboration or a big test, but rather as a small sample trial. I am responsible for the implementation of this simple DMP, which consists in the stewardship of a single pseudo-anonymised survey results file.

What resources will you require to deliver your plan?

I assume this point particularly relates to studies dealing with clinical or otherwise interview dense data, which can be heavy and difficult to maintain. Nevertheless, as I anticipated, this study won't be backed up in institutional university storage facilities.

No special hardware is required, beyond the computer room in which the workshop will take place, and some printed paper or phone to store the observations. No special software is required either, beyond the engine and tools to make the VN.

I don't foresee needing experts or training for this particular project.