

Editor comments following peer review.

Overall assessment

The topic of CRM and the design of the engagement tool, GreenDealz, are great contributions to the GC community and science engagement practitioners in general.

The final GreenDealz kit seems very well made. I congratulate the authors on their work.

However, all three reviewers (RC1, RC2, RC3) raised concerns regarding the clarity and completeness of the methodology description and the validity of statistical interpretations. This needs to be addressed comprehensively. There is also a need to improve the readability of the manuscript, as mentioned by all reviewers. I concur with all these conclusions.

My overall recommendation to the authors is to make the methodology section more robust, particularly describing the decision-making process between iterations for transparency and to help conceptualize the study's aim to design a toolkit and embed an evaluation method.

I also suggest considering including the feedback from the audience as suggested by Reviewer 2 (RC2) and Reviewer 3 (RC3). I further recommend presenting the results in a more qualitative nature, given the limitations and representativeness of the sample pools. This is in line with RC3 comment #2.

Recommendations

1. Readability: The text is at times hard to follow. Please remove redundancies to enhance the clarity of the text.
2. Structure: The reason to present methods and results together is clear and well founded. However, it is difficult to follow the iteration process the way it is currently written, particularly as the flow is interrupted by lengthy paragraphs of statistical analysis.

I propose some specific suggestions below.

3. Methods and results

There could be more transparency and clarity in the description of the iteration phases, including the decision-making process.

- a) For example, it is not clear how the prompts were classified from easy to difficult, nor how the prompts were chosen. It would be relevant to make explicit

how the scientific knowledge presented in GreenDealz was selected from the list of literature sources.

- b) As another example, it is not clear why there was a change in questionnaire wording from the Dingle to the Cork festivals. Namely from,
- “A3: *The EU is self-sufficient when it comes ...*” (Dingle)
- to
- “A3: *I am surprised that the EU is not self- sufficient when it comes...*” (Cork),

since this changed from a neutral statement to an opinion (*I am surprised*). I would assume this could affect the response in a survey? This is not discussed in the text.

- c) Figure 1 introduces the workflow from scoping to designing phase, as well as characteristics of the design (first under Overview, Line: 110).

The information of the sample size (n) could be added to Figure 1 to make the methods more transparent.

I would further recommend the figure be split into two: (1) a Fig. 1 visualising the phases of the iterative workflow, the corresponding festivals (with festival details as suggested by RC1 and RC3) and participant numbers, and (2) a Table X with the results/conclusions from the iteration process, including the main changes that were made in each step from scoping phase to refinement phase. This table could help summarise the results from the iteration process, which is not yet clear, as commented by the referees. This would conceptualize study’s aim to design a toolkit and embed an evaluation method.

- d) Reviewer 3, similarly, asks for “qualitative insights into audience engagement”, to which the authors respond that it falls outside the scope of the paper. However, from the statement:

- Line 521: “*From the scoping to the refinement phase, development of the PE toolkit GreenDealz has been driven by audience feedback, responses and field observations.*”

and in alignment with the comments by RC2 and RC3, I also strongly suggest considering providing an overview if not an analysis of the audience feedback.

This will also serve to make the method section more transparent, and additionally, it will offer the readers a learning process on engagement tool design (not as a final conclusion of effectiveness, but rather as a best practises method).

4. Statistics and representativeness

- Line 128-129: *“For the data collection carried out at live events, sampling and recruitment was representative of and realistic to the free-flowing nature of the festival environment.”*

- a) That recruitment was representative is not evident as we are presented limited information on the different festivals. Do the festivals provide demographic information to which your samples could be compared? For example, most of the Dingle festival participants (88%) had a university degree (compared to ~50% as mentioned in the Irish Central Statistics Office report you refer to). How representative is this sample to (1) assess the change in engagement of an educational toolkit? and (b) overall representative of the “general publics” in festivals?

In alignment with RC1 and RC3, I agree the study would benefit to expand the table section to include the festival characteristics, the festival setup and the summary of participants in each setup (as I mention in comment 3c). Indeed, as also RC2 notes, it would help clarify and improve the methods section to include the decision-making process at each iteration and a discussion on the limitations.

- b) I agree with RC2 comment #4. I do not see the validity of the university student validation (n= 9, 89% female). The response given to RC2 explains the student and festival groups represent the “control” and “in-festival” settings, respectively; however, these two groups represent different segments. I question whether this small student sample is relevant, particularly in the argument in Lines 363-364, which observes an increase in subjective knowledge in the student sample, which is not seen in the festival group (National Ploughing Championships).

On the other hand, the control “match up” game in National Ploughing Championships was very clear. Thus, I would suggest reconsidering if the student study is relevant in this context, as it may add unnecessary complexity to the text.

- c) I agree with Reviewer 2 comment #5. I also question the extent to which the statistics can be used to make the conclusions in the text. I thank the authors for their reply to R2 indicates a statistician may be called in during the revision.

Technical comment.

Order of in-text citations: if they are placed in order of relevance, please leave them as they are. Otherwise, please arrange them chronologically.