Response to Referees

General Response

I would like to thank both Referees for their helpful and insightful comments, which will help contribute to an improved manuscript.

Both Referees, but especially Referee 2, highlighted that more observation and evaluation data would enhance this study. Much of the information used is anecdotal, gathered from over 5 years of exhibited SeriousGeoGames Virtual Reality activities, but this was not formally collected and recorded (I consider this information to still be useful and of value). I agree with the two Referees and would like to take the opportunities provided by two upcoming events (in September and October 2019) to conduct more formal evaluations, before submitting a revised manuscript.

In line with Referee 2's suggestion to expand the literature survey, particularly in the area of evaluation methodologies, I intend to discuss different levels of evaluation. For example, the example references provided by Referee 2 refer to evaluations of events (a science festival itself, albeit as part of a larger cultural festival), whilst this study concerns the evaluation of a single exhibit within the event (the science festival), and sometimes even a single activity within an exhibit within a event. I would like to briefly look at the relationship between evaluating activities, exhibits, and events. Whilst it is possible to evaluate at individual activity level, this will run the risk of leaving the public with a "survey fatigue" if many exhibitors are doing so (and in too much detail) and compromise the achievement of objectives.

Another point raised by both Referees was regarding the SeriousGeoGame (SGG) 'model'. I think by using the term model I have overplayed its significance – the idea of combining science festivals, video games and virtual reality, with real research data and/or models, was not formulated as evidence suggested that this would work - although individually they have each been shown to be effective for public engagement with science. Instead, it is a preferred design concept behind the development of the activities. Alongside this, a point Referee 2 raised – *Flash Flood!* isn't a game. It has no objectives, no narrative, nor any way to win. Instead, as they pointed out, it is a visualisation. The SGG concept does not require the applications to be full blown games, but to feature elements someone would identify as being related to video games. In *Flash Flood!'s* case this is the explorable 3D environment and the use of gaming controller.

In response to the comments of both Referees therefore, I propose a revised manuscript including the following additions –

- Explanation of the design concepts behind SeriousGeoGames and *Flash Flood!*
- Review of literature in activity, exhibit, and event level evaluations
- Additional evaluation of at least one further event using methods identified in the literature review
- Increased explanation of gaming terminology
- Better explanation of the level of evidence behind findings

I have made direct responses to the Referees below.

Referee 1 – Laura Hobbs

General comments I really enjoyed reading this; it's an important contribution to a field that it can be difficult to get work published in due to the constraints of data collection, and it also makes some very

important points about that. It is also important to consider that when discussing results and explain how conclusions have been drawn in the light of sometimes limited data. For example, more presentation and discussion of observational and quantified data from event feedback are needed to support statements made - this would be interesting to see. At the moment I'm wondering what the evidence is behind some of the findings, and about the finer details of how evaluation data were collected and analysed. This is research in its own right; some work is probably needed to pull that information together (or, perhaps, rephrase some of the text so that some elements are more suggestions than assertions). Some of the gaming terminology in particular needs to be made more accessible to those without the relevant background technical knowledge or interest, as it's entirely possible that readers may be interested in using games for communication, but not themselves be gamers.

However, the description of the game and it's development use are valuable in their own right and are rightly a strong focus of the paper. It's great to see work in this field and I'll look forward to reading (and citing, I'm sure) the final paper. Specific comments and technical corrections - please see attached pdf.

Thank you, Laura. I enjoyed reading your comments and I think the manuscript will benefit from your experience in this field. I will make response to your line by line comments.

Line 12 - What do you mean by best components? For what purpose, and in whose view?

"best components" is too descriptive a term. It is a design decision rather than an evidenced-one, so I will change this to "features selected from each".

Line 28 - This is really interesting, and a great way to make your point. There could be more exploration here here about why people are more engaged with risks from zombies than flooding.

I agree, I found this connection, originally made by John Curtin from UK Environment Agency, fascinating. I'm unsure whether the introduction to this paper is the correct place to explore this further but it certainly would be an interesting piece of work.

Line 31 - Why? What takes precedence?

This is my opinion so I will remove this statement – it links to the points made later in the paragraph that geomorphology can influence flood risk but the processes are not included within the flood forecasting.

Line 42 - This clause doesn't make sense - should it be WITH the science being well reported?

I've written this poorly – The point I was trying to make (from Clarke et al, 2017) is that the News will often report of geomorphology issues, e.g., landslides or river erosion, but won't use the terminology of geomorphology. This results in the public having little perception of geomorphology as a discipline. I will rewrite this section.

Line 47 - This feels like it could have a citation.

Again, this is my opinion and I shouldn't have included this. I will remove the statement.

Line 52 - There needs to be a brief explanation of VR here (and that you're referring to immersive VR), and perhaps a note that it will be described in more detail later.

Agreed – and also that I mean what other papers would specifically call immersive virtual reality... I will add a brief explainer here and a link to a wider description later.

Line 64 - Not everyone will understand what this means; it merits an explanation, even if just briefly in parentheses.

Will add "(software package used to design and develop video games)".

Line 68 - Not everyone will know what an Oculus Rift is.

Will change to "The scene was viewed using an immersive VR headset, aka Head Mounted Display (HMD) – the model used was the Oculus Rift Developer Kit 2."

Line 69 - Why did it become obsolete?

Long story... basically, we no longer have the development project files to update it so will only work with the much older Oculus Developer Kit 2 headset. It can only work with legacy drivers for the headset (0.6.0 at the latest), and an older legacy version of Windows (the PC we used updated itself and the application no longer worked).

I will change this comment to reflect it relies on legacy software rather than it being totally obsolete.

Line 75 - Is this essential? If so, why? What is the evidence that you've drawn on to come to this conclusion? Were there other models with different features for comparison?

It is not essential – it is a design decision to combine these elements. Each element has shown success in their own right, as discussed later in the manuscript, so I wanted to combine them.

Line 78 - What do you mean here? Detailed datasets, or broader (but not basic) links to research? It's possible to engage people and spark interest in research using broader links, so if its the former, why, and how do you know?

Again, this was a design decision. I wanted people to directly interact with elements from our research, either by manipulating a numerical model (as in *Humber in a Box*), or exploring a landscape built from survey data (as in *Flash Flood!*). Along with the comment above, I will rephrase this section to make it clear that the SeriousGeoGame model is a design choice

I do think this increases the publics engagement with them but I do not have the evidence for that – this would be another potential area for further.

Line 86 - Absolutely. This paragraph makes a very important point that often gets lost when considering analysis of these projects, which can present difficulties in publishing. This recognition of the value of planting that seed is hugely valuable too.

Thank you, this is the core point I am trying to make with this manuscript.

Line 142 - In the case of Minecraft at least, there are analogies to real world settings and processes too - not to try and engineer a link to our work, but there may be some useful references for background information in these: https://jcom.sissa.it/archive/18/02/JCOM_1802_2019_N01, https://www.bgci.org/files/Worldwide/Education/Roots_PDFs/Roots15.1(med).pdf (pp. 20-23), https://eos.org/wp-content/uploads/2018/10/Nov-18_magazine.pdf?x64125 (pp. 25-29).

Thank you for these, they are useful. I think there might be future work into the use of gamifying reality too!

Line 168 - Great description - it'd be useful to have mention here of the difficulties that some people have with VR experiences.

Indeed, I will add a paragraph here on this.

Line 231 (Figure 3) - This needs a brief outline of the changes visible, for those who aren't used to identifying them.

Agreed – "The flood stripped away the vegetation adjacent to the river and eroded through meanders creating a straighter river. The lighter colours after the flood are deposited stones, pebbles, and gravels, transported from upstream. Post-flood the channel, with lots of loose sediment and no vegetation to hold it together, the river is shallower, wider, and more dynamic going from a single channel to multiple channels."

Line 262 - It'd be good to explain why here.

This links to the previous sentence. Will add "For example," to beginning of this sentence.

Line 277 – is

Will change.

Line 297 - Not everyone is going to understand the AAA terminology - it either needs explaining or a different, more generic term (higher-end, maybe).

The use of the term is important. Will add "(big budget game by a major games design studio)"

Line 307 (Figure 5) - I can't see Figure 5

Apologies... will put this in.

Line 315 - Why?

This was a design decision. Partly I was fed up with the sound of my own voice, partly because Jess is awesome, but mainly as I feel video gaming is a traditionally male-dominated area that I wanted the initial gatekeeper to the virtual experience to be a woman.

From watching behaviours, a lot of people simply click through the opening menus without thought so will hear the default voice over from Jess. When they think and select, generally, males go from Chris and females go for Jess. I don't have formal evidence for this, however.

Line 331 - So it wasn't necessary to use the simulation in order to be able to use the handout?

The handout included information relating to *Flash Flood!* and the science behind it so had enough on it that it could read separately. The activity included needed the simulation but would work with the Desktop and YouTube versions (it was made before we had the YouTube version, but we were still giving them out once we had it).

Line 344 - How? To be able to draw the conclusions made, there needs to be more information given about the observational data collected.

This is anecdotal evidence. As part of conversation, myself and other staff will ask how they found it and overwhelmingly the response is positive, although we have not formally collected and recorded these. I will revise this to make it clearer that this is anecdotal.

I plan to do perform more detailed evaluations at future events and include in the revised manuscript.

Line 352 - How many people voted, and how many votes did the stand get?

NERC did not provide a breakdown of the votes but I will check with them if they hold this information.

Line 356 - Do you have any quantification of positive vs. negative comments?

I have no quantification of this, but the ratio would be highly skewed towards positive (we rarely get more than 1 or 2 negative comments each time we exhibit).

Line 371 - So is this counted as an SGG too? If so, how does that fit with the criterion of using VR? (Or if that's not an essential criterion, that needs to be made clearer in the text).

This is not an SGG but an activity within the Earth Arcade project. Earth Arcade allows us to run SGGs alongside non-SGGs, and we've found that the combination of different types of activities, and levels of engagement, compliment each other well. Again, no formal evaluation of this, but definitely something I'm looking into.

Line 375 - Extra (

Need to include a link here.

Line 386 - How many attendees were there?

From the University's evaluation of the whole SciFest - visitors to the Discovery Zone of 45 exhibits, of which Earth Arcade was one, were 3,039. We did not count how many of those came to the Earth Arcade itself.

Line 386 - Were these whole comments, or individual elements which may have occurred within one piece of feedback left by a single attendee?

This was a mix of both. I will provide a table of all comments as an appendix.

Line 389 - They're being asked to give a positive repsonse, so you would expect positive answers here. It'd be interesting to see some more quantified content analysis here, although the data may be too limited for this to be analysed in depth.

This is true and I had hoped for more answers across the other boards. The sample size, and the general nature of the comments, means that there isn't much more to be extracted from this data.

Line 413 - Yes, absolutely - can't stress the content of this paragraph enough.

This is crucial – especially as we had several activities just within our Earth Arcade space which we could have individually evaluated too!

Line 455 - There needs to be more elaboration on what was collected here and what it contained. What observational information was collected? How many events were these data collected at - were there more than those described here, which the text seems to imply? Is there any quantification to support that it was overwhelmingly positive?

I will edit to make this clearer that this is our perception and anecdotal but supported by data gathered from 2018 Hull SciFest.

I will perform more detailed observation and evaluation data at future events.

Line 457 - How much privacy/anonymity did people have when giving this feedback - could this have affected the results?

The boards were not monitored by staff but were in the space, so possibly the public may have felt pressured not to provide negative feedback.

Line 458 - This should be drawn together above, in the evaluation section - and also discussed with the caveat that participants were asked questions that framed responses in a positive way - they weren't asked what they didn't like, or given a completely open choice on what to give feedback on.

I will move this to the Evaluation and include a caveat and discussion.

Line 481 - How many?

We have a conversation count from a previous event where we engaged 199 people in one day using two sets of *Flash Flood!*. I will include this into the Evaluation or headcounts from future events.

Line 487 - Comments to the event organisers, or at the stands?

This was from verbal feedback from the organisers from NERC to the exhibitors, but also conversations we had with parents where they told us they'd never been able to visit previously. I have no formal record of this. I will change comments received to conversations.

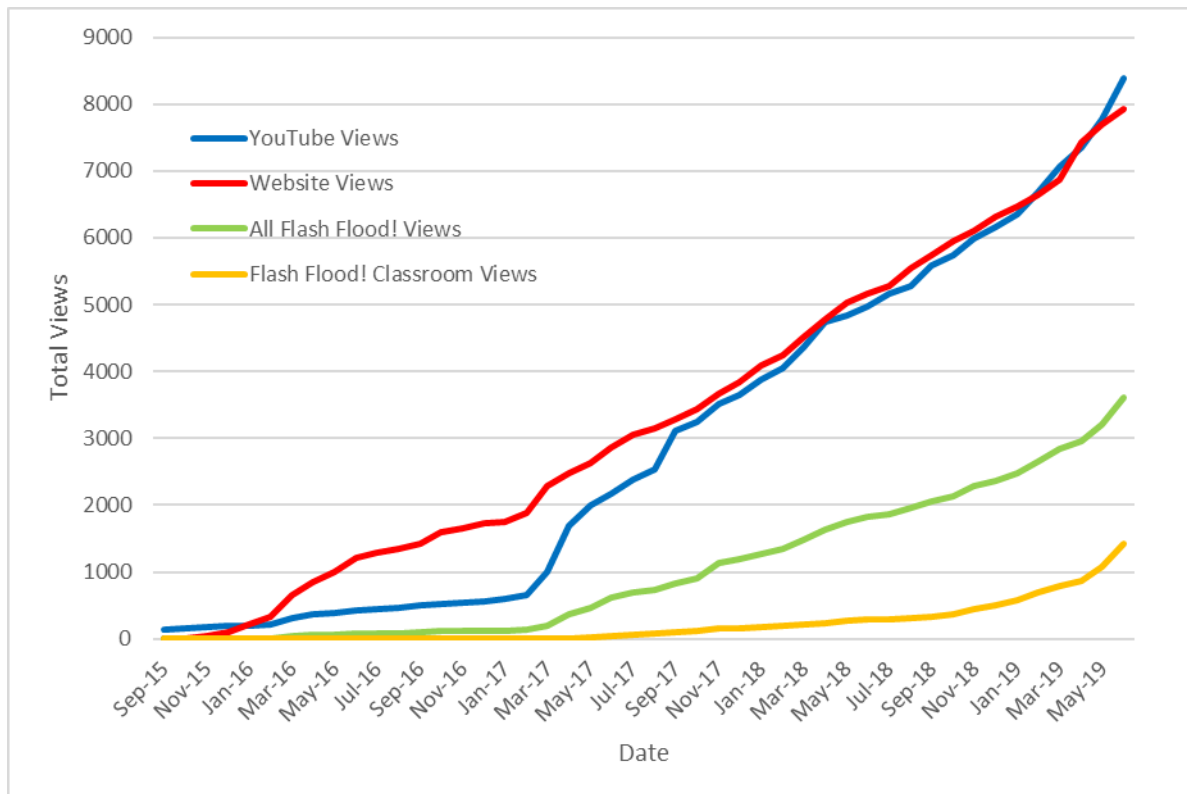
Line 491 - Is the video game element likely to be the most familiar of the three to attendees - and how does this intersect with how able they feel to make comments?

It is most familiar, but also the least developed, so this feedback is legitimate. It is a visualisation and not a game, so some expect an element of competition or a way to win. As much as I would like to include these elements, I can justify adding them in on a cost-benefit basis for the objectives we have.

Line 497 - This is important - the efficacy of gaming in drawing people in (the same happens if they get a glimpse of Minecraft!).

Absolutely – another thing a lot of kids used to ask is "is it Minecraft?", which has changed to "is it Fortnite?" over the last year.

Line 522 - Is this an actual existing criticism, or a potential one?

This is my criticism as I can see the potential of using the application to achieve more ambitious learning objective (just not as a festival exhibit in five minutes). In fact, the Classroom version of the Flash Flood! video has seen a rapid increase in views since I did this analysis and will include this in the Evaluation of the revised manuscript. An expanded version of Figure 7 to cover this period -

I'd like to think this is because it is getting use for teaching but I have no way of evaluating this unless people self-report this…

Line 527 - Definitely.

Line 530 - Absolutely - it's a paper in its own right (I'd like to read it though if you ever get the data).

Hoping to one day do this.

Line 533 - How many? Thousands is a bit too vague! It would also help to know for comparison of your sample size for feedback collection.

I wouldn't know as we have not kept records. From the conversation count from the Natural History Museum, we spoke to 199 people in one day using two sets of kits.

Referee 2 – Anonymous

This paper presents an intriguing idea around creating a visualization of a flood using VR to represent an often overlooked aspect of flood risk, to engage with participants of a science festival through encouraging their curiosity and sense of fun. It is a well written paper that describes the process of designing the game and testing it in a science festival scenario, however despite the interesting concept there are a couple of flaws I would like to see addressed before publication.

Firstly I think there is scope to improve the literature section, both in terms of quantity and source. Additionally I would like to see more description of how this simulation operates as a game, as it appeared from the descriptions to be more of a visualization? It would be great to get more information here. The figures could do with a bit of refining (particularly the graphs) as although they

presented interesting ideas I found them to be a bit confusing. In particular the presentation of the 'model' which I thought needs to be far clearer. With a little work, the visualization of a model would be really useful in this paper to help the reader understand the main premise.

My greatest concern, however, was with the evaluation. Although the author stated that science festivals are not conducive to evaluation, that is not actually the case if the appropriate evaluation method is chosen beforehand and designed carefully into the activity. For a really good example please see the works of Sardo and Grand (Science in Culture: Audiences Perspectives Engaging with Science at a Summer Festival, 2016 and What Works in the Field? Evaluating Informal Science Events, 2017).

Even if you wold prefer not to evaluate in a science festival, it is still possible to evaluate this game in other environments specifically designed for evaluation, which would provide a much more empirical as opposed to anecdotal data-set.

Having said that, I do think the idea and the effort that have gone into the game design are worthy of publication, but I think some more robust evaluation of the game needs to be done before that can happen. When that happens I look forward to reading the results!

I'd like to thank the anonymous Referee for their comments and suggestions. They have raised several valid points which I have addressed in the General Comments above.

I have made further responses to the line by line comments below -

Line 46 - This needs a reference.

This comment was in reflection to points raised in the first paragraph about low sign up to warnings, will make this clearer.

Line 48 - I am not sure I am comfortable with using a blog post as a reference here.

This isn't a blog post, it's a commentary article published in the journal Earth Surface Processes and Landforms - doi:10.1002/esp.4129

Line 55 - What about section one?

I will include a note about "In the rest of this section,"

Line 58 - 'conclusions -are presented- in Section 6'

This is correct, I will amend.

Line 70 - Why did it become obsolete?

Long story… basically, we no longer have the development project files to update it so will only work with the much older Oculus Developer Kit 2 headset. It can only work with legacy drivers for the headset (0.6.0 at the latest), and an older legacy version of Windows (the PC we used updated itself and the application no longer worked).

Line 70 - How were these data collected?

Anecdotal information from many events exhibiting the activity.

Line 70 - I don't know what model you are referring to here - clarification would help me understand this section.

The model the rest of this paragraph describes. I think calling it a model however is creating confusion and I will change such references to "design concept".

Line 72 - Why is this the case? Where is the data or literature to support this assumption?

I am outlining our preferred design concept. I think the term model is causing the confusion.

Line 75 - How does the game design interact with the learning objectives and the evaluation design?

This will depend on the individual application produced. For future applications I hope that these will be better integrated and embedded but the *Flash Flood!* application is designed to be fun and inspire curiosity, i.e., there are no learning objectives.

Line 75 (Figure 2) - This is a very confusing diagram. Firstly it is not how a Venn diagram works as the circle with the research data/models covers several of the intersections, which also have no additional detail in them. What is the cross over between video games and science festivals? What about video games and virtual reality? It feels like this would be a great place to put a reflection on the literature and theory of this study, what are the learning objectives, or methodology for example, but the image certainty needs work. Additionally what do you mean by research data/models? Those are very different things.

I will edit the description to remove reference to Venn diagrams. This is intended as a simple visualisation of the elements together, with research data and/or models embedded at the core of the applications produced. It is a design concept, not really a model, so will change all this terminology.

Line 84 - This is a challenge, however in this sentence I would refer to studies which demonstrate the difficulties of doing longitudinal studies on attitudes and behaviour change. If you struggle to find examples from informal science communication and education environments there are plenty of examples from formal environments.

As behaviour change was not an objective for the project, I think a review of the literature here would be superfluous.

Line 87 - I am still confused by the model.

I have erroneously used the term model here in a way it can be read to refer to the design concept in Figure 2. Instead, it is referring to the activity acting to facilitate a conversation with the public and will rephrase this.

Line 89 - So are you measuring the objectives of the VR as a tool itself or as a facilitation for engagement with scientists? These are different things.

They are different things and to some extent *Flash Flood!* does both. I am not evaluating them separately as there are no learning objectives at the science festivals, but this would be an interesting thing it disaggregate in future work when I do try and achieve learning objectives.

Line 97 - This needs clarification - I think what you mean is about formally supported face-to-face interactions with the public to discuss their research, as science communication happens in many environments often informally, and you are also missing the huge interactions researchers have through social media and other platforms.

I take the point and will remove the second part of this sentence.

Line 98 - Are there any more up to date figures? A lot has changed in 5 years.

I will update with figures from http://sciencefestivals.uk/rounding-up-the-network-in-2018/

Line 113 - 'that'

Will change.

Line 116 - I would expect an Ipsos MORI ref here for their own data?

The report the reference refers to is no longer available publicly. This remains at 3 % in the most up to date report (2014).

I will edit to "In a 2011 MORI poll showed that only 3 % of the UK population attended a Science Festival in the previous year (Jensen and Buckley, 2014), and this remained at 3 % for the latest MORI poll in 2014 (MORI, 2014)." https://www.ipsos.com/sites/default/files/migrations/en-uk/files/Assets/Docs/Polls/pas-2014-main-report.pdf

Line 123 - 'including'

Will edit.

Line 134 - Is there no UK data for this? Cultural differences can be significant.

Level of ownership in the UK looks much lower at 8.15 million (https://www.barb.co.uk/tv-landscape-reports/tracker-games-consoles/) out of 27.2 million households (~30%). However, the US report includes smartphones and tablets used to play video games on – 36% of households had a console. I will change this section to "Video games are popular, with 30 % of UK households owning a gaming console (BARB, 2019), and 36 % for US households (Entertainment Software Association, 2018). These figures do not cover the numbers of people who play games on PCs, smartphones, and tablets."

Line 146 - Explain this term?

The term is previously described in Line 68.

Line 185 - I feel like the way this is written is down-selling the importance of your work, because despite the fact that these phenomena are rare, sharing understanding about the devastating consequences is actually even more important!

I agree and will highlight this more.

Line 195 - Is there no literature for this other than a talk?

The reference is to the paper published from the talk – I included the mention of Kristie's talk as I felt it highlighted the importance of the concept that she chose the platform to talk about it. I will include "and associated paper" before the reference to clarify this.

Line 197 - This needs a reference

Fryirs (2016) – will make this clearer.

Line 202 - This section needs better referencing

I will make this clearer that it is summarising Fryirs (2016).

Line 229 - How fast does recovery normally take, is there a standard justification for this timeline?

This is very much dependent on the landscape itself and whether there have been subsequent events. I don't think there has been a study to establish the time frames, probably because the recovery times would likely be in the order of decades.

Line 257 - I'm confused by the definition of user vs operator, it would help for this to be clarified.

User refers to the member of the public, and operator the staff member running the demonstration. I will change this to "public participant" and "staff operator".

Line 273 - This term should be explained for those unfamiliar with video game terminology

Agreed, will change to – "…but it still required a sudden, unrealistic, change in the 3D environment, with the public participant's virtual presence temporarily removed and replaced back into the scene (referred to as 'respawning' in gaming),…"

Line 284 - Explain this term?

Will add "(big budget game by a major games design studio)"

Line 331 - This is only anecdotal - needs data

This is anecdotal and will make this clearer - I consider this useful information gained from several years of running these exhibits. I plan to perform more detailed evaluations at future events to provide more formal data.

Line 335 - This is very tenuous for a large event, where that comment may have been referencing a different stand, even if there were no other VR practitioners there.

I consider that this did refer to the *Flash Flood!* exhibit but can understand the point and will remove "the gogls".

Line 344 - This section solidifies a concern I am having that this is less of game than a visualisation or simulation. There don't appear to be any objectives that can be achieved or activities that the player can do. Perhaps reframing this activity as such may help with evaluation?

It is a visualisation. It uses elements from video games to give it a game-like feel without being a full game in its own right. I will make this clearer earlier in the manuscript and changing any mentions of it as "game" to "application".

Line 361 - Lonely parenthesis.

Will edit.

Line 365 - Is there any demographic data available here for context? Number of attendees, location, source of attendance etc?

There are for the Hull SciFest event as a whole, but not for those who came to the Earth Arcade exhibit. This is the issue we have with full evaluation – we have an exhibit of 5 activities, in a Discovery Zone of 45 exhibits, within the wider Hull SciFest event including shows and workshops.

Line 371 - Can you share this data?

I will include a spreadsheet of response as an appendix to the revised manuscript.

Line 379 - This section should be edited for clarity - there are too many examples. Some kind of analysis would help condense this?

I will edit this section and also refer readers to the data I will put within an appendix.

Line 400 - I disagree, evaluation can be done in a science festival setting, when designed appropriately - please see suggested references in the review. If you would prefer not to do that however, it would be useful to do some more structured evaluation in a different environment either before hand or after?

I don't disagree that evaluation can be done in a Science Festival setting, and indeed for both Into the blue and Hull SciFest the organisers did perform evaluations of the events (of a similar nature to those described in the references suggested). However, the issue is performing evaluations of individual activities and exhibits within the events.

It would be possible to perform a structured evaluation in a different environment, but this manuscript is describing the festival setting in particularly which is often busy, noisy, time restricted, and in competition with other exhibits.

I will perform more structured evaluations at future events and include these into the revised manuscript.

Line 408 - However you won't be able to tell from the user data if they used a VR interface here, correct? So the interaction with the visualisation without the immersive experience of VR would change the way the player interacts with it.

This is true, this isn't a level of detail provided by YouTube. However, the viewing figures are used to give an indication of people seeking further information about *Flash Flood!* so whether they viewed in VR or not doesn't matter.

Line 413 - The monthly views for what specifically?

The monthly views of the SGG website and YouTube channels – will include this information here.

Line 429 - I'm struggling to connect this to curiosity. There are some really great measures of science curiosity and it's importance in science learning/engagement (see Dan Kahan's work) but I'm not sure what measure you are demonstrating here and how you can quantify if it's actually curiosity you are looking at.

I am using the increase in views to the SGG website YouTube channel as an indicator that people are seeking further information about the activities. That seeking further information is curiosity. There was an increase in views which can be attributed to the exhibit.

Line 438 - Again I would disagree - there are tools and approaches that can help you do this.

But are these appropriate for activity level evaluation within a larger event? This level of surveying of the public will become intrusive.

Line 449 - I would question how you can be certain of this, as with the data you have presented I wouldn't know for sure, perhaps you could add these additional analytics.

I am unsure what additional analytics are being referred to? This line is summarising the data explained in the paragraph beginning Line 424.

Line 454 - You can check this data using analytics and if you have already done so I think it would be valuable to share it.

I will include a brief analysis of these analytics.

Line 459 - How does this match with the single-user focus of a VR simulation?

We use multiple sets of equipment and can perform five demos at once. Each is 6 minutes long, so up to 50 demos an hour. As demos are automated, the researchers exhibiting can converse with those watching, explaining what is happening.

Line 499 - So this suggests that the true value of the simulation is as a fcailitation tool? This is an interesting perspective and one that would be easy to evaluate.

Linked to above – *Flash Flood!* does both. The person in the headset is being engaged by the application, and as it is automated this releases the exhibitor to engage the audience watching. I will expand on this in the design and discussion sections of the revised manuscript.

Line 503 - Is that in real life or in the game? That may add a more 'game-like' feel to it if it were more interactive?

This would be in real-life. It is something which has to be appropriate to the event so would be expensive to build several versions of the application.

Line 507 - But that would be a very effective way to get some good evaluative data.

But would that data be transferable to the busy, hustle and bustle, of a science festival? They are too different settings with different types of engagement.

Line 765 (Figure 6) - I'm really not sure what this pie chart is trying to show, I wold actually assume this is more of an assessment of the questions themselves than of the activity. 'What will you do?' is quite a difficult question to answer, which might be why is got the least number, as opposed to 'what do you like' which gets the most. I would hesitate to use this as demonstration of the effectiveness of the game.

The point about effectiveness of the questions is a good one, and also raised by Referee 1. The evaluation was conducted for all of the activities within the Earth Arcade exhibit. I will remove this chart in place of the full spreadsheet of responses.

Line 766 (Figure 7) - I would suggest that both figure 7 and figure 8 show the same thing, and I'm not sure you need them both. Figure 7 in particular is very good at showing the increase in users over time. If you want the additional data you could aggregate the two using a timeline which identifies the festivals and promotional events which stimulated the increase in views.

Figure 7 shows the growth of views over the full lifetime of the activity. Figure 8 shows a snapshot of 2017 for use in analysing the November 2017 event, Into the blue. The full year was provided for context, showing the spikes for the release of the *Flash Flood!* YouTube video and the NERC Planet earth article. I will edit Figure 7 for the revised manuscript to include dates of key events – putting all the events on would make it too busy.