

Interactive comment on “Using video games for volcanic hazard education and communication” by L. Mani et al.

S. Anderson (Referee)

steven.anderson@unco.edu

Received and published: 9 February 2016

The topic at the heart of the research paper is an important one - how can we effectively educate the public with respect to geologic hazards. The approach presented in this paper - using video games as a mechanism for better understanding - is well-grounded and justified. However, the methods used in this study (to infer that understanding for students and adults was enhanced through the use of the video game) may not be valid and reliable and therefore it is impossible to know if the measurement of learning reported here is truly learning (and not just an increase because of a flawed assessment instrument) and repeatable.

The authors state correctly that the "main source of data collection was through the completion of knowledge quizzes before and after all of the sessions took place". How-

C1

ever, the actual assessment instrument was not contained in the manuscript so it was not possible to assess the main instrument of data collection. I infer from statements later in the manuscript that the instrument was multiple choice in nature, and there are well-known issues with using multiple-choice instruments for measuring learning. Julie Libarkin and her colleagues have published numerous studies over the past 15 years discussing the creation of multiple choice learning assessments for the geosciences that are both valid and reliable, and I would urge the authors to use these works as a guide for validating and modifying their test before they publish the data. In short, we can't be sure that the small gains shown on the exams are due to learning - the increases could be due to a number of issues related to validity and reliability discussed in the Libarkin papers.

In addition, it will probably be necessary to use the qualitative data that were collected, but not presented or discussed, to help ensure that that exams are valid and reliable. Pairing the qualitative with the quantitative data is an approach commonly used to create valid and reliable concept inventories in most scientific disciplines.

I found the discussion a bit underwhelming. Although the links to other studies were informative (for example, relating the greater gains for adult learners to other studies that found a similar result) there wasn't much discussion of how the students and adult learners were affected by the use of the game (other than just the increased post-test scores). I believe that the qualitative data that were collected but not presented should shed some light on some other changes that could occur through using the serious game (such as in increased or decreased desire to learn more about a volcano that has erupted or not erupted during their lifetime). In other words, is knowledge the only thing that changes as a result of the exercise, or are there some affective changes that also occur?

In summary, the topic of the paper, and the general approach (test an educational intervention with pre- and post- knowledge tests) is acceptable, but the assessment instrument has to go through a host of validity and reliability checks before we can

C2

have any confidence that any change evident in the data truly represents learning. Libarkin outlines a number of potential approaches that the authors can use to create a research-quality assessment exam. I also urge them to include the qualitative data that should help to not only validate their knowledge test, but also shed light on any learning that may occur by using the serious game. I realize that this will delay the publication of the work here substantially, but in the end it is far better to publish a study that truly tackles the issue in a way that is valid and repeatable than it is to simply have another seldom-read education paper in the world.

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-23, 2016.