

We thank Anna Hicks for taking the time to review the manuscript ([gc-2020-43: *Using paired teaching for earthquake education in schools*](#)). We have updated the manuscript as suggested. The detailed response is shown below in **blue**.

1. Teaching experience: The authors talk about incorporating teaching strategies into the videos but provide no information about the level of teaching knowledge and experience of those developing and presenting the videos. Are any of the authors teachers or researchers in education? They are all based in geoscience departments, but it is not clear about their teaching experience. I believe the lead author may have considerable experience. That being said, were teachers consulted before making the videos? Did the authors understand the curriculum in both countries and seek to co-develop these videos with the teachers and/or curriculum developers with whom they engaged? How did the authors know that earthquake education in schools needed supplementing? These are important reflections that should be in the paper. Beyond seeing making videos as a 'good idea' to them, can the authors demonstrate evidence for need? The study follows MIT Blossoms protocols and 5E Instructional Model but the authors do not reflect on this in the lessons learned. It therefore comes across as a retrospective 'shoehorning' into a published methodological approach. I was also confused about whether 5E lessons plans or plans developed by Mohadjer (2010) were used. Please clarify in section 2.1.

Response: Thank you for these excellent points. We have revised the manuscript text to address these concerns. Below, we explain each concern separately.

Authors' experiences - All video lessons were developed under the guidance of the lead author (with over two decades of experience related to K-12 science education and outreach), with input from the MIT BLOSSOMS team including faculty and teachers who have published >100 STEM videos following the paired teaching technique. Other authors (Matthew Kemp, Sophie Gill, and Sebastian Mutz) are also experienced with developing and teaching STEM related educational activities in both formal and informal settings including K-12 schools, museums, and public outreach activities conducted by universities in the UK and Germany. Finally, the video scripts and all the classroom activities that are done as part of these videos, were adapted from published lesson plans (Mohadjer et al., 2010). These lesson plans have been tested, evaluated and improved by hundreds of teacher volunteers (members of Teachers Without Borders) across China, Tajikistan, Afghanistan, India, and Haiti. The video modules described in the manuscript are modified from these existing (and already evaluated) lesson plans, but differ in format and pedagogical approach (i.e., they're published as video lessons using the scientist/teacher paired teaching technique).

On supplementing the curriculum – Our video lessons are not intended to replace an existing curriculum but rather to support the teaching of concepts related to earthquake science, hazards, and safety through interactive, hands-on activities and discussions that teach students to think critically. These discussions/activities are ignited and guided by a guest scientist and the in-class teacher. It is important to note that our video lessons are developed for the global school community, and do not target or adhere to the educational standards of a specific region or country. However, we encourage (and provide examples) to interested teachers on how to contextualize the content according to their local environment. We do this in the video segment titled "Teacher's Guide" at the end of each video. To clarify this, we have modified the manuscript (section 2.1.1).

In addition, we have observed that in both countries, topics related to natural hazards are textbook-driven and often discussed briefly, with no/little active learning exercises to encourage critical thinking. Particularly in Tajikistan, rote learning and classroom management predominate. Furthermore, inviting local/international scientists to co-teach with classroom teachers is rare in Tajikistan and some parts of the UK. Our video lessons,

therefore, are not only an opportunity for igniting active learning in schools, but also allow for "virtually" bringing scientists into school classrooms without additional costs/resources. Concerning the 5E lesson plans (Mohadjer et al., 2010), we have modified the manuscript to clarify the confusion raised by the reviewer. The content of the 5E lesson plans were used only as a guide by the video teachers when developing the video lessons. Therefore, the 5E method was not used in the production of the video lessons (hence no discussion of it in the manuscript), but the content of the lesson plans was adapted, modified, and improved by the video teachers (under the guidance of the lead author) to make them suitable for paired teaching. Therefore, the main pedagogical approach used in this manuscript is the paired teaching method which is discussed extensively in the method section (section 2.1) as well as in Larson and Murray (2017) which is referenced in the manuscript. The reason the 5E instructional model was mentioned in this manuscript (section 2.1) is to differentiate the video lessons (focus of this manuscript) from the lesson plans of Mohadjer et al.

2. Ethics and consent: Where are the ethics assessments and the consent forms for this study? Working with minors and collecting data from them requires consent and ethical assessments. How will the data be used and stored? Could the authors urgently provide some information about this in accordance with GDPR.

Response: Ethical approval for this study was sought and received from the participating schools and institutional partners that coordinated this effort in 2018-2019. The ethical procedures were designed to adhere to current standards of assent and consent regarding in-school research and to provide participants with anonymity. As a low-risk, school-based study focused on learning outcomes from regular teaching activities, the consent of the school-based stakeholders, i.e. the principals and teachers, were considered sufficient to proceed with the pre- and post-tests. All participation was voluntary, and students were given the opportunity to assent or refuse participation at both the pre- and post-test points in time. The pre- and post-tests were anonymous with students creating their own codenames, which were known only to them, that were used to match the pre- and post-tests for analysis purposes. No sensitive or identifying information was collected, and the anonymous data are stored in a secure location within the European Union that is password protected, in accordance with GDPR regulations. The data will be destroyed upon completion of this research project.

3. Cultural aspects: Please could the authors provide some reflections on the cultural dimensions of this study? Empirical evidence is fine. For me, observations would be as useful, if not more so, than the numbers! I am intrigued for example about the student responses to the causes of earthquakes between the countries. Why haven't the authors explored this? More reflections on why the Tajik teachers did not engage would also be useful for readers.

Response: Thank you for these important points. We discuss each separately below. The manuscript text is also modified to include this reflection:

Cultural dimensions - This study was not designed to assess the cultural differences in students' understanding of earthquakes. However, our earlier work (Mohadjer et al., 2010) gave additional insights into the cultural dimensions of earthquake knowledge among school children in Dushanbe. Unlike our 2010 study, we documented no mention of myths, legends, or religious explanations in Tajik students' responses to our interview questions probing them on earthquake causes and preparedness measures. This could be due to (1) different data collection methods (anonymous survey questionnaires vs individual/group interviews), (2) differences in students age/grade level (the 2010 participants were 2-3 years older than the 2018 participants), and/or (3) may reflect changing earthquake perceptions over the last decade. As for differences between students' responses from UK and Tajikistan, it is clear that the Tajik students (unlike those from UK) lacked the correct scientific terminologies

(e.g., plate tectonics), but could associated processes such as mountain building and volcanism to causes/locations of earthquakes. Living in an earthquake-prone country that is mountainous could be one factor explaining their answers. This was discussed in section 4 under "causes of earthquakes".

Low teacher engagement - Thank you for this point which was also raised by another reviewer. As for why the Tajik teachers did not engage, we believe this is due to their unfamiliarity with an interactive/collaborative classroom culture, and their level of comfort operating within it. Below, we explain what we learned from this observation. The manuscript text is also modified to include this reflection:

The paired teaching video lessons are designed to be a complete resource. They include video segments, teacher's guides, downloadable handouts and lists of other resources relevant to the topic. Therefore, no teacher training should be needed in order for teachers to use these videos. However, teachers are encouraged to view the videos and familiarize themselves with the content before using them in their classrooms. This study, however, reveals that these videos may not be seen as a complete resource by some teachers. While the UK teachers tested the videos with minimal input from video creators, teachers in Tajikistan asked to observe classroom testing of the videos. This request was made despite the fact that the teachers were offered (i) training to deliver the videos, and/or (ii) the option to co-teach the video lesson with experienced instructors. The textbook-based classroom culture (typical of schools in Tajikistan) may explain why Tajik teachers did not want to actively engage in video testing. The study, therefore, shows that the paired teaching pedagogy is not a "one size fits all" teaching approach, and depends on the classroom culture and teacher's comfort operating within it.

4. Statistical information: The statistics are distracting and not really that helpful for the reader. Whilst I appreciate the authors are trying to demonstrate a rigorous evaluation of the videos, the reader is yearning for an interpretation and discussion. Much of the statistical information could be moved to an appendix, or tabulated simply.

Response: Thank you for the comment. The statistical test results are included in a supplementary table (Table 2b). However, since we used statistical tests to compare results across and in between each group (UK vs Tajikistan) and to determine if differences observed were significant, we think including this information will help some readers. To increase manuscript readability, statistical information appears consistently at the end of each subsection (e.g., 3.2.1-3.2.4) in the result section, and are not mixed with the description/discussion of results. We hope this is a satisfactory answer.

Major points for reflection - An opportunity has been missed to develop science communication theory or incorporate social scientific methods into the study. - The authors throw around the word significant in the paper, but care should be taken as to whether it really is or not. - I like that the videos were broken up into segments, but it would have been beneficial to receive feedback from the teachers and students as to the desired length of the videos – they seem pretty long as a whole (12-24 minutes per video – 10-20 hours teaching time in total) - The study seems to be more focused on evaluating knowledge gain or change and has failed to uncover what it actually was that caused that knowledge change – was it the words, the presentation, the models, the presenter? If the authors have this information, it would be useful to include, or even provide some observations and reflections that might lead to further study. – Regarding the students feeling that the questionnaire was an exam, this is absolutely my experience too, particularly in areas where cultural differences between the students and the visitor are quite obvious. It would be really useful for the authors to reflect a bit on this, and ideally put it in the paper – might it be related to perceived or actual power, positionality, cultural norms etc?

Response: Thank you for these reflections. We exercised care when using the word significant, always clarifying what we meant (e.g., 95% level). This is also why we included the results from our statistical tests in the manuscript and referred the readers to the supplemental information for more details. As for collecting feedback on desired video length, we have done this (please see section 4.2: Teacher feedback). In general, we observe that both the UK and Tajik teachers used the videos flexibly, skipping sections irrelevant to students' lives or shifting to lesson plans when the technology failed. As mentioned earlier in this documents, the video lessons are designed for a global audience. Therefore, teachers are encouraged to use/modify the videos according to the needs of their students and their local environment.

As for evaluating the causes of change in knowledge, this is discussed in section 4.1. We observed significant changes in students' knowledge of three concepts (earthquake location forecasting, Earth's interior and non-structural hazards). Each of these concepts and related changes are discussed in more detail in section 4.1.

Concerning the questionnaire being treated like an exam by Tajik students, we have added more reflection on this, as suggested by the reviewer, to section 4.3 (curriculum evaluation). This study shows that while using questionnaires with students may be a time/cost-efficient way of collecting information anonymously, if not designed carefully, the students may interpret the questions differently, as it was the case with some students in Tajikistan. This is especially important if the students are not experienced with participating in anonymous questionnaires where giving honest (as opposed to correct) information is crucial. Our recommendation, therefore, would be to use questionnaires as one of several methods for collecting and evaluating data. Amongst these methods, we recommend conducting face-to-face interviews with students when possible and/or arrange for group interviews. In this way, if a question is misinterpreted, the interviewer can rephrase it. In addition, recording the sessions may provide important insights into student-student and student-teacher interactions that enhance or hinder learning. We have added these recommendations to section 4.3.

Minor points to address:

Line 10 The opening line of the abstract is conjecture. Suggest a rewrite.

Response: Thank you for the comment. We agree, and have taken out this sentence.

Line 13 "coupled with activities carried out by local classroom teachers". Suggest a rewrite – the teachers aren't carrying out the activities, and what do you mean by local classroom? Isn't teachers enough?

Response: We have revised this sentence for more clarification: "[...] coupled with activities carried out under the guidance of classroom teachers."

Line 21 I'd suggest replacing views with knowledge – you weren't evaluating their opinions. Although that would have been interesting. We have made this change.

Line 25 What makes a classroom culture suitable? Who says so? Suggest a rewrite.

Response: We have revised the text to clarify this: "[...] (e.g., levels of teachers' participation and classroom culture) [...]"

Line 46 empowerment of communities

Response: Thanks for catching this typo.

Line 66 Be consistent with language – choose either teaching-duet or paired teaching

Response: We have replaced "teaching-duet" with "paired teaching" throughout the manuscript except in section 2.1 (line 80) where we make a reference to the MIT

BLOSSOMS (the original developer of this method which they sometimes refer to as teaching-duet).

Line 75 Typo on through

Response: We fixed this typo.

Line 87 How did you develop an iterative process? There did not seem to be any iteration in the UK example – did the video teacher actually talk to the teacher in-situ and tweak approaches? In the Tajik example the video teacher was the in-situ teacher, so again, no iteration.

Response: We agree with the reviewer, and have taken out this term. The final sentence reads, "The passing of teaching between the in-class and video-teachers is a type of blended learning referred to as the paired teaching."

Line 128 How do you know that the teacher continues to keep the students interested?

Response: Good point. We don't know. We have revised the sentence, "The teacher continues by stating that despite not being able to drill deep into the Earth's interior, [...]".

Line 160 You collected data on the first and last days of video implementation but the testing periods were very different between the UK and the Tajik studies (50 and 5 days respectively). Can you reflect a bit on this, maybe in a later section?

Response: This is correct, and we have revised the manuscript to add our reflection to section 2.2.4 (school settings). The difference between the testing periods was due to different teaching schedules teachers had to follow. While the UK teachers were more flexible with their program and could spread out the video lessons in a 50-day period, the lead author of this manuscript had to follow a restricted schedule of five days for video testing in Tajikistan. It is possible that the differences in the testing periods between the two groups (UK vs Tajikistan) influenced the study results. For example, due to time constraints, some in-class activities had to be shortened or skipped in Tajik classrooms. Furthermore, technological issues in Tajik classrooms (e.g., system malfunction, missing cables, etc.) posed additional challenges to video testing by shortening video testing period.

Line 214 Section 2, not 3.

Response: We corrected this.

Line 293 How can you have a 20% increase in student understanding of the Earth's interior. Please clarify or omit this.

Response: Thanks for this comment. We have modified the manuscript to clarify this. The sentence now reads, "[...] indicate an increase of at least one score point in their understanding of [...]".

Line 303 Might be a personal thing, but the word naïve feels a belittling – can the authors just say a basic conceptual framework?

Response: We agree and have changed the word naïve to a basic conceptual framework throughout the manuscript.

Lines 328-29 I like where the authors have put numbers of students in brackets, in really helps the reader. Consistency with this throughout the paper would be welcome.

Response: We have revised the text to be consistent throughout the manuscript.

Lines 397 I'd argue that the authors have not produced a series of DRR educational materials, only a few related to DRR and most were about hazard education.

Response: We changed this to "a series of earthquake education video lessons".

Line 409 constraints (typo)

Response: Thanks for catching this. We fixed it.

Line 411 Fix non-structural hazards? Could a more appropriate word be found?

Response: We are not aware of any other terminology that can be used to describe non-structural hazards. The term "falling hazards" is sometimes used to refer to non-structural hazards. However, this term is not entirely accurate as not all non-structural hazards are falling objects (e.g., tables not fixed to the floor can shift during earthquake shaking and cause injuries if they collide with people). In section 3.2.2., we have defined non-structural earthquake hazards as those "caused by the furnishings and non-structural elements of a building (e.g., suspended ceilings and windows)". We hope that this explanation is helpful to the readers.

Line 411 non-seismically active regions

Response: We made this change.

Line 420 "Nearly all Earth scientists agree that public outreach is important." How do the authors know this? Where's the evidence/reference? Otherwise just conjecture.

Response: We agree and have change this sentence to "The importance of public engagement activities is increasingly recognized by scientists, funding institutions, and policymakers", and provided additional references to support it: NSF, 2015; Rauws, 2015; European Union, 2002.

Line 460 I'd suggest changing the word ensure to support.

Response: We have made this change.

Line 460-466 These examples are ok, but are they really related to K-12 and higher education? How about researching some serious games such as Earth Girl and Hazagora?

Response: The reviewer is correct. These examples (i.e., Hazard Ready, Central Asia Geohazards Database, and GEM products) are not related to K-12 and higher education, but they are tools developed by scientists to enhance access to geohazards information. If incorporated into DRR educational materials, they have the potential to effectively serve as educational tools for increasing public knowledge on geohazards. The paired teaching approach, as discussed in this manuscript, is one way to turn such scientific tools into classroom video lessons. As for including examples of serious games such as Hazagora, we have added this information to section 1 (introduction). Thanks for bringing our attention to these and similar resources.