Please see our response to each reviewer comment below in **BOLD**, below each reviewer comment.

I read this article with interest. The authors should be appreciated for attempting to shed light into an area that we academics often consider a secondary responsibility, namely creating an effective classroom learning experience.

The article is written in clear language that makes it easy to read and understandable by an international user of English language.

We thank the reviewer for their kind words regarding the manuscript.

I think educators are almost unanimous these days that it is of critical importance that clear definition of Intended Learning Outcomes (ILOs) or Learning Objectives is critical for ensuring good learning outcomes. (Whether we all practice it all the time is another matter!). Another almost common-sense guidelines is that the assessments (and learning activities) should be aligned to those ILOs (as proposed by constructive alignment [1].

Reading this article, I failed to find a list of well-defined ILOs. Indeed authors list in Fig. 2 (Also in Table 2, which they do not refer to in the text – the 'table 2 they refer in bottom of page 6337 should be table 3.) they list what they refer to as 'nine overall learning outcomes', but these are not specific enough for me to know what were the specific, testable, verifiable goals behind the section in question.

This article would definitely benefit by stating a well defined set of learning objectives (see TeachOnline site of ASU [2] among many others for good practice).

This will shed light also to the appropriateness of the assessment instrument used. More on that later.

We agree that more clear learning objectives would potentially aid in the understanding of the manuscript, and give a better sense of the specific content instructed here. We will thus not only adjust the incorrect table numbering and referencing, but will also better articulate the learning outcomes already included in the manuscript. We will also better direct the reader to Table 2 which gives concrete examples of the learning objectives.

Recent literature has shown a large number of uses of the term 'T-shape'. While at the conceptual level these uses agree, the precise meaning varies greatly among the different uses (especially on the 'breadth' aspect). The definition I found in the article is in the abstract, which requires 'professional breadth combined with technical depth'. Upon reading the article, I wondered whether the important findings of this article are related to the T-shape idea.

While an interactive tool (DMDGC) will definitely provide a more absorbing learning experience, I fail to find how it provides 'T-shaped' learning. Overall it is my view that this article will be more effective if it does not discuss the notion of the 'T-shape' but focus on the learning quality differences of the two approaches – a worthwhile objective in itself.

We do agree that the notion of the 'T-shape' is used somewhat inconsistently in different contexts. However, we have tried to be as specific as possible regarding how our mapping of the different 'legs' of the 'T' are implemented here. It is our contention that the professional breadth is really an analogue for the understanding of the roles of a professional hydrologist, whereas the technical depth comes from understanding the concepts themselves. While certainly somewhat abstract, pragmatically speaking these legs seem to clearly represent different types of knowledge. Further, we would resist the urge to further specify the legs of the 'T' much more, as this was an introductory Earth Science course, and not a class specific to hydrology alone, and thus any such specification might be an overreach on our part related to the domain. We again agree that the relations of these legs of the 'T' to the learning outcomes may have not been as clear as possible (related to the Reviewer's earlier point), so we will attempt to make this connection more definite by rearticulating the learning outcomes. We do also agree that the learning quality differences should indeed be the primary focus here and will attempt to constrain our discussion of the results to focus more on the learning itself, as the reviewer suggests.

The authors do not provide the learning material used in the two cases. The article should provide supplements with or links for the learning material in order for the reader to understand the link between the learning experience and the outcomes discussed in the. I was able to find online [3,4] which I suspect are the material used for DMDGC case, but I failed to find the material used for PP case.

While we would like to provide the PP materials, these materials are copyrighted and published by Kendall-Hunt, and thus we cannot freely release them without violating said copyrights. However, we have already included the correct reference to the materials in the Reference section should a reader seek to purchase the materials themselves. That said, we are attempting to find a way to potentially release these materials in a way that maintains compliance with Kendall-Hunt's copyrights, although we cannot guarantee anything at this point unfortunately.

I have to admit that I did not read though the material in [3,4], but upon looking at them, I could not see how they will enable the students to better answer questions like Q3 and Q4 (table 1). They authors should attempt to explain what aspects in the interactive material that resulted in students answering such questions better.

The only information regarding PP material is in page 6335 (around line 25). This is a calculation to determine whether a channel will flood before and after urbanization occurs in a watershed. How does completion of such an exercise prepare students to answer questions like Q3 or Q4? If that does not prepare the students in anyway what so ever, then is it logical to test students for that and arrive at the conclusions listed?

The page 6336 (lines 9-10) lists essentially what was different between the two treatments. Then I fail to see how one can explain how that can explain the differences of marks for questions like Q3 and Q4 (or goals 7, 8 and 9).

The factual material about the role of NOAA, USGS, and other agencies, is contained in the standard lecture material that accompanied the laboratory instruction. As such, it is important to note that this factual content was taught equally to both groups, and was not explicitly instructed in the laboratory exercises themselves. It is our contention, however, that the DMDGC exercise provides these learners a more tacit and explicit

understanding of these job duties, such that when asked to answer such questions they were more easily able to recall and connect said job duties to specific agencies. In other words, their understanding of what these agencies do became less abstract and indirect, as they themselves became more familiar with the job duties themselves. While certainly speculative, the large group differences (and the fact that both groups were equally exposed to the factual information) seem to support this suggestion. We will make sure to emphasize the speculative nature of this interpretation in a revision, and would certainly be open to other interpretations.

I was a bit intrigued by the way analysis was presented. It would be nice to see the pre and after treatment scores for each question rather than presenting the analysis for each 'learning outcome'. This would provide a more straight forward way for the reader to evaluate the findings. Further the authors do not provide any indication about the pre-treatment results (other than the fact that it was used as covarient in the ANCOVA analysis).

We would be happy to include the pre-treatment results (by learning outcome), and will do so in the revision.   However, we must emphasize that comparisons between groups at the pre-treatment time point are not exactly warranted, and any initial group differences are indeed already accounted for by the current method of analysis. For example, given the quasi-experimental nature of the current study, such a comparison at pre-treatment would only demonstrate that the groups are potentially different based on enrolled lab section. If this were the only time point of measurement (it is not), or if these differences were not accounted for relative to later measurements (they are), this might be cause for concern. However, we would argue that these initial knowledge levels are only important relative to their final standing in the course (i.e., how much did they learn, controlling for their initial levels of knowledge). The current analysis does directly examine the amount learned, while also simultaneously controlling for initial knowledge levels. We feel that this is the most appropriate way to frame the current results as it directly evaluates the effectiveness of the instructional manipulation in the lab sections, and not pre-existing group differences. We would also add that this method of analysis is typical in most educational research.

Further, we do feel it more straightforward to maintain the discussion of results relative to the learning outcomes. For example, if we were to discuss the results strictly relevant to each question, as each question taps multiple types of content knowledge (see Figure 2 for an example), we believe that this would make the pattern of results more confusing for readers.

Some sort of graphical representation of that results (e.g. box-plots) could have been useful.

While technically redundant with the information presented in Table 3 (which already includes measures of central tendency and variance), we would be happy to include such a graphic to facilitate the demonstration of effect.

Information about how the students were selected for the two types of treatments is also missing (randomly?).

We would kindly direct the reviewer to section 2.1 where this information is already available; there were multiple laboratory sections of a single course, and each section was randomly assigned to either the control or treatment group.

As indicated in the beginning I find this a useful and intersecting study. However, it needs considerable shaping up in order for it to become genuinely useful for the wide readership. I hope the authors would take up the challenge of revising it.

We again thank the reviewer for their kind words and input, and hope that our revisions will effectively address the reviewer's concerns.