

Responses to Comments of Reviewer #3

Summary: This is an interesting study that queries whether changing teaching modality impacts student-instructor interactions (really novel, very cool) and student learning. As an instructor, I really liked the emphasis of this study on interactions/questions – I think this is an interesting way to frame an analysis of this type. My only caveat to my comments is that I am not well versed in statistical analyses applied to educational assessments, so I am not able to comment on this part of the paper.

Major comments:

- 1. -I'd recommend revising and restructuring the introduction. The introduction has lots of good content, but felt a little disorganized, jumping around between general information and more specific hydrology information. There were also a few ambiguous statements in the introduction that I think could be sharpened (more in minor comments).**

The introduction will be revised. Specific changes are documented in the specific comment responses below (and in the responses to reviewers 1 and 2). We will add our research questions to the end of the introduction, as recommended by reviewer 1, and will significantly modify the introduction in accordance with other suggestions below.

- 2. -A minor point, but I consider it major, given the focus of the study – the term ‘student led learning’ is introduced in the introduction, but not defined or explained. I think it’s worth adding a few sentences to more clearly define this term and point to key references. This would broaden the introduction of this idea beyond only the ‘flipped classroom’ approach.**

Thank you for this comment. We agree that the term should be more clearly defined and delineated from other, sometimes similarly-used terms. We have added two references to help draw distinction between this term and others and revised the text as follows:

“There is broad evidence that student-led learning improves in-class experiences and retention of concepts in STEM education (de Jong, 2019). We define student-led learning to be a student-controlled learning process with continuous interaction and input from the instructor (Hoogenes et al., 2015), inclusive of both flipped classroom approaches and other project-based learning

paradigms. Student-led learning differs from other learning paradigms such as self-directed or self-regulated learning which tend to be unsupervised (Brydges et al., 2010).

- 3. -I'd strongly recommend moving Table S3 to the main text – otherwise it isn't clear from the methods what questions are being used for assessment.**

Thank you for this recommendation. We will move Table S3 to the methods for ease of referencing all questions in one place.

- 4. -While very interesting, my main concern is that this is only one year and one class of data. Thus, findings could be specific to those circumstances, and it is hard to say if this outcome would occur in another class and another year. However, I don't think that means this study should not be published. Instead, I'd encourage a thorough discussion of the limitations of this study in the discussion section.**

We will include a discussion on the choices we made for this experiment with regard to uncertainty and variability of results in the revised text, as detailed in our response to reviewer 2's comment #4.

Minor comments:

- 5. Lines 39 – 40: “A recent series of interviews with water resources professionals indicated that graduates lacked critical workforce skills” – Could you add a little more information here? As written, I think this statement could lead to some confusion.**

The referenced study found that industry professionals thought incoming employees lacked knowledge on use and interpretation of data and different modeling systems. These details will be added to the introduction.

- 6. Lines 49-50 aren't well integrated with the rest of the paragraph, which is about student-led learning – should these ideas come up later? Or could they be better connected to the rest of the paragraph? (Maybe move down to line 71?)**

Thank you for pointing out the lack of connectivity of this idea to the rest of the paragraph. We will move the reference to the following paragraph which similarly discusses data-driven analyses.

- 7. Line 59: What is meant by ‘more relevant material’? Could you be more specific here? Relevant to what and in what context?**

Thank you for this comment. Reviewer 1 similarly noted the ambiguousness of this phrasing. We will change the language to more clearly reflect the “career-relevance” of using teaching tools such as high-resolution datasets.

8. Line 83: Another challenge in what respect?

In this paragraph we intend to highlight a lack of research using individual-level data in contrast to the wide availability of studies that used aggregate-level perception data. Upon review, we agree that using the phrasing ‘another challenge’ alone, does not convey this current research gap. We will revise the phrasing to more clearly demonstrate this gap as follows:

“Prior studies place strong emphasis on group mean outcomes of hydrology courses (i.e., average perceptions, average assessment scores) rather than the outcomes of individuals (Gallagher et al., 2021; Knoben & Spieler, 2022; S. W. Lyon et al., 2013; Merck et al., 2021; Pérez-Sánchez et al., 2022), a current gap in hydrology education research.”

9. Line 100: 3000-level is institution specific – (my institution uses 300 level, for instance) – could you use another way to contextualize the course level that translates across institutions? Maybe just refer to this as ‘upper level’?

Reviewer 1 similarly pointed out this terminology challenge. We will adjust the language to represent, as recommended by reviewers 1 and 3 to be ‘upper-level’ and include expected courses previously taken.

10. Line 125: ‘on campus’ might be too colloquial – maybe ‘local’?

We will adjust the language as recommended.

11. Line 243: worth looking at interquartile ranges? Did you bring the lower grades up with the shift in teaching modality?

We used two-sample Kolmogorov-Smirnov (KS) tests to compare the distributions. This statistical approach tests for any changes in the distribution (including shifts in the extremes), not just the means or medians. None of the KS-tests showed a significant difference which suggested that there was no significant change in the lower grades.

12. Line 293: should this be ‘reported’?

Thank you for catching this error. Yes, we will correct it as recommended.

13. Line 294 – 296: I like this conclusion

Thank you for this positive feedback!

14. Line 298: have other studies done this? Is this an approach that is used in the educational literature?

In our experience (and review of the literature) it is common for instructors to track written assessments scores at the individual level, but only the average of student perceptions (e.g., mid- and end of-term student evaluations of teaching). To our knowledge, other studies have not presented an individual-level analysis of student perceptions in hydrology education, and this is a primary result of this study. The likert results (Fig. 3) and alluvial diagram (Fig. 4) together show that the mean outcome might remain constant, but the impact on individual students can be large (similar to written assessments, Fig. 7). We are recommending that this approach should be more widely adopted particularly because the changes in assessment scores did not match reported changes in perceptions (Fig. 7). We will clarify this point in the revised text.

15. Line 310: Yes – I think this is possible. I’ve seen students doing group work divide and conquer on assignments, meaning that they may miss out on learning because they have self-selected to do a portion of the assignment that doesn’t involve “x” activity.

We have similarly observed the “divide and conquer” approach. One group in particular was skilled in how they split up tasks. Although their final work was exceptional, each student seemed to only master one particular task. Given our IRB protocols, it is not possible to link this group to specific responses or grades as all results were anonymized. We were also not able to find any published studies that specifically looked into this effect. In contrast, most studies promote group learning as a panacea and do not consider the possibility for individual learning to lag that of the group. We are proposing to leave this wording as is so that we don’t overemphasize a point for which we don’t have supporting literature or direct observations from our study aside from a minor addition to point to the “divide and conquer” strategy : “...lowered student engagement *in individual assessment components*.”

16. Line 325: References got a little messed up here!

Yes, this was auto-generated by Zotero. I’ve checked the reference database and don’t see any obvious problems. We will go through the manuscript and manually correct issues with the references.

17. Figure 2: Could you update the legend to have spaces and be written text, not abbreviations?

Thank you for pointing this out. We will modify both figures 1 and 2 in accordance with this comment, and that of Reviewer 1. See modified figure with responses to Reviewer 1 comments.

18. Figure 3: Is it possible that interest in a career in hydrology merely increased through time, and not as a result of a particular approach to teaching? I don't think your study design allows you to separate temporal effects (if my assumption that the delivery timeline was lecture -> modeling -> design project), so it may be worth pointing this out (but my assumption may be incorrect).

Yes, this is possible; however, we do note that these teaching modalities were introduced in the second half of the course. Students already had 8 weeks of exposure to hydrology concepts and techniques to this point. We will include in our discussion of uncertainties the role of sequencing the teaching modalities as follows:

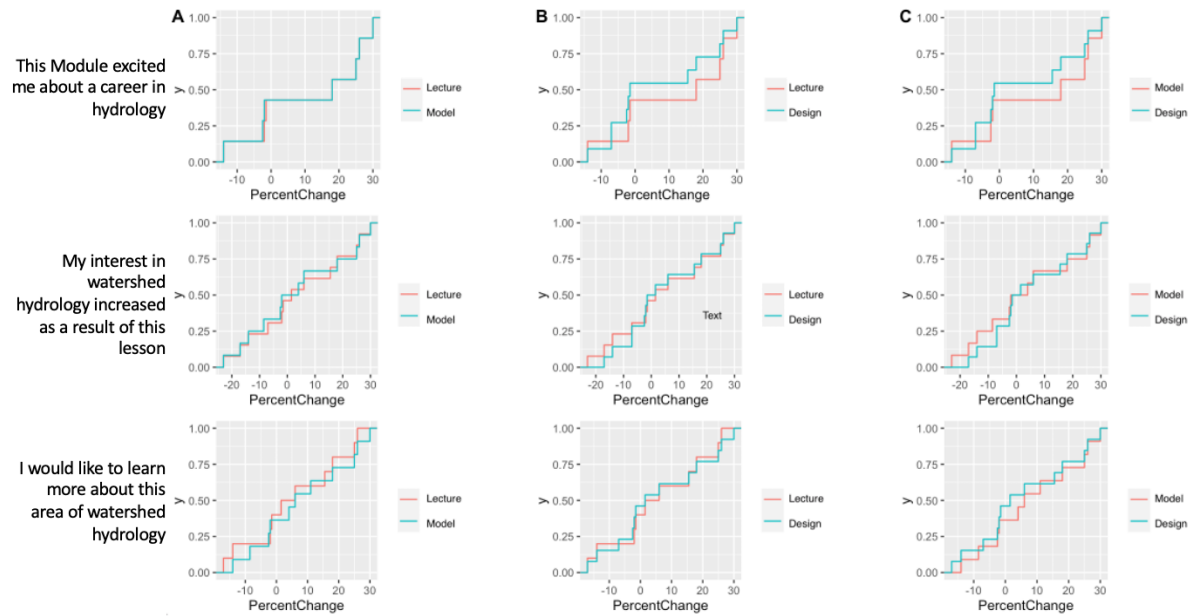
“These modality modifications were implemented in-sequence over the last half of the semester. This design was chosen such that students would already be comfortable with the hydrological concepts, instructor, and course environment and eliminate the ‘warm-up’ period to the new course environment. However, this study design does not allow for separation of the in-sequence influence of time from the results. We found that aggregate interest in careers in hydrology did not significantly increase with time, though individual students changed their perspectives with each modality, leading us to the conclusion that time was not a significant variable in these results.”

19. Figure 4: I love a good figure, but I struggled to see what the authors wanted me to see in this figure. Would there be some way to highlight a key message, or include a number or a few numbers with each graph, or even a summary of the key takeaway message in the caption?

As per recommendations by reviewer 1, we have modified Fig. 4 to include all modality comparisons (i.e. 3 columns instead of 2 columns of plots). The key conclusion here lies within the tracking of students who shifted their perceptions of the course dramatically between modalities. Despite average aggregate level conclusions that the three modalities were equally perceived (Fig. 3) There are certain students in each modality who dramatically shift their perceptions with modality. We will add detail to the results section and the figure caption to ensure that detail is not overlooked. We believe that together Fig. 3 and Fig. 4 demonstrate the necessity to have both aggregate-level and individual-level data analysis.

20. Figure 7: Could be moved to supporting information – I found this the least interesting! It was challenging to see anything in this figure. I also think this figure could be redesigned if you wanted to include it in the paper. For instance, add the question text above each section (so readers aren't flipping back and forth between different parts of the manuscript), and add significance level from the statistical test to each figure.

Thank you for this suggestion. We will move this figure to the supplemental material and modify the figure such that the questions are visible on the figure itself as displayed below:



Supplemental Figure: Cumulative distribution functions of student grades for students who responded to the career and interest questions with a '4' or '5' on the Likert scale of '1'-'5'. Questions refer to the Likert survey questions to which each response corresponds. Column A compares Lecture and Model modalities, column B compares Lecture and Design modalities, and column C compares Model and Design Modalities.