Referee report:
A framework for mentored data science research

The manuscript described their framework for mentoring 2 undergraduate senior students doing data science research. The authors laid about the framework components, connections to the 3 goals in Nolan and Temple Lang (2010) and added one more goal on reproducible research. Student, mentor, and scholarly outcomes were presented. In discussion, the authors described benefits of the current framework, and potential improvement areas for future iterations.

While the manuscript offers an interesting and useful framework for undergraduate data science research, there are many areas of this work which need improvement to become a meaningful contribution to the Special Issue.

Major comments:

1. The structure of the manuscript is not very clear, and the non-labeling of each sections and subsections has made it even harder for readers to follow. I would first recommend giving clear section and subsection labels so the readers do not have to go back and forth to understand what each section is about. For example, in Methods, it is very confusing to go from the framework implementation (with several subsections) and then all of a sudden four sections of the 4 goals. The organization is too loosely done.

2. Many sections and subsections have repetitive material, which indicates either a restructuring and deleting unnecessary repetition, or reduce some subsections into bullet points. For example, the one paragraph under "Broaden statistical computing to include emerging areas" is almost identical to the material in "Examples", which is then unclear to the readers why a separate subsection is needed, except for the purpose of listing and discussing each of the 4 goals. While I think it is good to discuss 4 goals one by one in the context, the authors should not repeat the same material multiple times.

3. My recommendation on getting more concise on this whole Methods section is to first have a subsection to describe the two examples so the readers have a context of what types of research questions the students are working on. Then, a subsection on the framework implementation with the nested several subsubsections in detail. After that, maybe one subsection on project 1 and discuss all 4 goals about project 1, and then another subsection on project 2 with goal discussions. I think this structure could greatly reduce the current repetition of material, and improve the readability of this section.

4. A few times, the authors did not give reference when a term first appears, but rather, provided references later. The most troubling one is about backward design, especially in the Discussion section, the authors provided the 2005 reference and a flowchart. The authors could consider giving the backward design reference in the Introduction section, and moving the flowchart into the Appendix and reference it in the Introduction. Another case is the latent Dirichlet allocation on page 6 line 129, a reference is needed here, or wherever LDA is mentioned the first time if the authors decide to restructure the manuscript.

5. To me, the tables are not space efficient. For example the Framework Overview table (there is no label either), the authors could collapse the three rows of details column of Question formulation into one. There is no clear benefit of listing every details item in its own row. Same for the table of Prioritizing Key Terms of Figure 1 of @nolan2010computing (there is no table label here, and the @nolan2010computing seems to be a bib problem).

6. Somewhere early one, it would be good to add the students' background so the readers can know what kind of statistics and computing experience these two students had had before this mentored research experience. I think this can help readers a lot when evaluating and adopting this framework.

Minor comments:

1. Page 1 line 20: not sure why the word "trainees" is used, unless this is a training program.
2. Page 2 line 34: "can have" only use one.
3. Page 2 line 48: as mentioned before, add reference to backward design, and possibly for active learning too.
4. Page 2 line 51: "we've" should be "we have".
5. Page 3: table title has an extra row with "t".
6. Page 4, line 76: "For the most appealing scientific hypotheses", how do you define "most appealing", and what about the "not most appealing" hypotheses? Are they dropped from the project?
7. Page 4, line 78: "Skill in translating..." should be "The skill in translating...".
8. Page 5, line 106, "we've" should be "we have".
9. Page 5, line 115: "doesn't" should be "does not".
10. Page 6 line 129: as mentioned before, add reference to LDA. Also, describe what LDA does (you can move the material on page 7 line 174 here).
11. Page 7 line 156: "latent dirichlet allocation" should be "latent Dirichlet allocation".
12. Page 7, line 174: "a bayesian" should be "a Bayesian".
13. Page 8, line 199: "didn't" should be "did not".
14. Page 10, line 230: "the modern data scientist" should be "a modern data scientist".
15. Page 10, line 243: "We've" should be "we have".
16. Page 10: table "Prioritizing..." has an extra row in the title of "xxx".