# UJSE-2020-0093 Response to Reviewers

I would like to thank the associate editor and the reviewers for their constructive comments and suggestions. Below I outline how I addressed each comment in the revised manuscript.

## Editor

*As you may know, I am not a fan of the phrase "statistically significant;" see* [*https://www.tandfonline.com/doi/full/10.1080/10691898.2019.1702415*](https://www.tandfonline.com/doi/full/10.1080/10691898.2019.1702415)*, or see the attachment. On page 3 you write that you want students to "better understand the nuances of statistical significance." Here "significance" means "small p-value." Then on page 5 you refer to an apparent difference in groups and ask "is it an important (i.e. significant) difference?" Here "significant" seems to mean "important."   
  
A couple of lines later on page 3 you write "To understand whether the observed (data) plot provides evidence of a significant difference, we must understand the behavior of our test statistic under the null hypothesis." It is not clear -- to me at least -- what "significant" means here. I think the question is "Is there a difference -- not necessarily an important or large difference, but a difference?" Thus, you might want to write about "significant evidence of a difference" but you should not write about "evidence of a significant difference." A hypothesis test does not test whether a difference is large/important, only whether a difference exists.   
  
This confusion over significance and importance, and the fact that a hypothesis test only leads to a conclusion about evidence of a (possibly very small and unimportant) difference, is part of what prompted me to write my editorial last winter. I invite you to re-write page 5. You might even consider the phrase "statistically discernible" if you think that would help readers, and their students, keep straight what a hypothesis test does, and doesn’t do.*

**Response:** Thank you for pointing out this confusing language. You are correct about the common confusion over significance and importance, and I should have taken more care in my original text. I have revised the paper in an effort to clarify the language. I rewrote passages on page 5 using “statistically discernible difference” to clarify my meaning. I also rewrote my careless error using important and significant synonymously.

## Associate Editor

*I noticed you used the words “prove” and “validated.” I recommend not using words as strong as those. For example, in the first paragraph of your introduction, you said, “simulation-based pedagogies for the first course have been proposed and validated.” They have been tested, there is evidence that suggests the methods benefit students, but I would not say it has been “validated.”*

**Response:** Thank you for pointingthis out. I have edited the language. *I understand this is not a research paper, but can you provide any evidence of usefulness of the lineup protocol? Even if it is informal observations such as how students responded, were they surprised, how did they react?*

**Response:** I have included informal observations from my classes in the implementation section, which is now called “Implementation and student reception.”  *The lineup protocol is used throughout your paper. However, I had no idea what the lineup protocol was. You started giving an example in section 2, but I was struggling to visualize what it actually was. At first, I was imagining a list of things that students should follow, and then I was wondering if it was something for teachers to help them guide students. It was not until after reading section 2.1 that I started to understand what was going on. Looking back at your abstract, there was some description there, but not enough for me to understand what the purpose was. I suggest having a short description, definition, and/or purpose in your introduction of what the lineup protocol is. That could help readers grasp what your paper is about.*

**Response:** I have added a short description in the introduction in an effort to clarify what the lineup protocol is for the reader. *Your writing is very strong. I have some minor suggestions:*

* *Abstract, fourth sentence starting with “Specifically, the lineup protocol…”: This sentence should be re-worded.*

**Response:** This sentence has been revised, along with the sentence that preceded it. I hope that the revision improves the clarity of the abstract.

* *Introduction, second paragraph, first sentence: I suggest saying “In addition to changes in how we…”***Response**: I have made this change.
* *Section 2, first sentence starting with “As outlined by Cobb…”: This sentence should be re-worded.***Response**: I have reworded this for clarity, and upon further reflection removed the reference to Cobb, since it is more generally known that this is the structure.
* *Section 3.1.1, second sentence starting with “We have found…”: This is a long sentence. Would it be appropriate to delete the part after the parenthesis?***Response:** I agree and have deleted the part after the parenthesis.
* *Figure 5, can you provide the answer?***Response:** I added the answer in the sentence referencing this plot in the paper.

## Reviewer 1

*This “lineup protocol” is an interesting and intriguing idea that I had never heard about. I think the author does a good job of providing examples of how the lineup protocol work. However, as an instructor who might consider using this method (similar to much of the JSE audience), I am left wondering:*

*#1: What happens when the plot lineup results in an interpretation that can be kind of subjective? For example, what happens when one plot stands out to some students, but not others?  
#2: Related to the above question, what are some examples of how you scaffold students along this process?*

*Regarding point #1, I would have liked to see more discussion about how instructors can deal with this subjectivity, or what to do when students are “supposed” to be able to identify a plot that stands out, but they can’t. (It was sometimes a bit hard for me to pick out the observed data plot from the lineup – imagine how much harder it must be for students.)*

**Response:** I added subsection 5.3 in the implementation section to provide the reader with some ideas about how to deal with this subjectivity.

*Regarding point #2, I appreciate that supplemental materials are being provided with a tutorial on creating the lineups and Shiny apps. I am also wondering if a few sample activities (even just one or two) could be provided with an example of questions to ask students to scaffold them throughout the process, and foster discussion/reflection. Section 5 left me wanting to know more about how the scaffolding is done. A few questions were provided in p. 18 lines 50-51 but I would like to know a bit more about how the questions fit into a larger activity. If there is space to expound section 5 just a bit to talk more about the “how” of these activities, and then some sample activities in the supplementary materials, that would be fantastic.*

**Response:** Section 5.2 of the revision provides some additional detail to provide some of the “how” of these activities and two full-length activities have been added to the supplements along with an instructor guide. These activities are full-length versions of the first two examples discussed in the paper.

*Otherwise, I think mostly minor revisions are needed.*

*Detailed, smaller edits:*

* *Abstract: p. 1, line 16 says “Lineups have ‘proven’ valuable...” which to me implies that there is some sort of research evidence of the effectiveness of lineups, when that’s not what you are discussing here. Maybe it’s just the statistics instructor in me not liking the word “prove” :) but I wonder if a less strong claim can be made in that sentence.***Response:** I have revised the abstract and no longer use the word “proven”. Instead, I emphasize that lineups can be used in those situations.
* *P. 2, Line 29: capitalize Shiny***Response:** Shiny is now capitalized.
* *P. 3, line 17: the word “of” does not have a modifier (of what?)***Response:**This was an error and has been deleted.
* *P. 3, lines 24-33: What happened to section 5 in the description? All other sections are mentioned.***Response:** This paragraph was revised to include mention of Section 5.
* *P. 4, lines 26-28 say: “If the observer is able to identify the observed plot in a lineup, then this provides evidence against the null hypothesis.” Given the recent p-value controversy about making “evidence against the null” a black and white issue, I wonder if instead, something along the lines of this could make sense: “The more distinguishable the observed plot is from the lineup, the more evidence this provides against the null hypothesis...”***Response**: I have replaced the phrase following your suggestion. It now reads: “The more discernible the observed plot is from the lineup, the more evidence this provides against the null hypothesis.” I used “discernible” rather than distinguishable to consistent with language changed elsewhere in the paper, and to follow the Editor’s suggestion.
* *P. 6, line 35: The Figure 2 description claims the observed data is in panel #4. But this plot does not look like the observed data in Figure 1, so I am confused....is Figure 2 using a different set of observed data? (If it’s not panel 4, then p. 6 line 38 and p. 8 line 26 will have to refer to a different panel as well.)***Response:** This error has been corrected. The data are displayed in panel #10, and the paper has been updated to reflect this.
* *P. 7, line 38 refers to “assigned roles” which leaves me wondering what types of roles could be used. I don’t expect much elaboration on this, but maybe a very brief list in parentheses (e.g., reader, recorder, time keeper) would clarify for readers who are less familiar with cooperative learning.***Response:** I added a brief list of roles I use in my classes, along with a reference.
* *P. 18, line 40: I liked the Sesame Street analogy and I wish this had instead been introduced sooner to give a clearer picture of the activity, such as in section 2 where the “police lineup” idea is introduced.***Response:** I’m glad that this analogy resonated with you. I have added this analogy in the introduction based on a suggestion from the AE.

## Reviewer 2

*Typo, I believe, around lines 15-17 on page 3 of the actual manuscript: "a formal distribution of is not necessary to introduce the core ideas behind hypothesis testing." I believe that something is missing after of, or that the of should be removed.*

**Response:** The “of” was a typo and has been deleted in the revision. *Summary at the end of Section 1 does not mention Section 5 Implementation, which seems odd given the structure of the paragraph.*

**Response:** This paragraph was revised to include mention of Section 5. *In Figure 2, the boxplot in #4 does not seem to match the data boxplot from Figure 1. Plots 10 and 15 look more like Figure 1.*

**Response:** This error has been corrected. The data are displayed in panel #10, and the paper has been updated to reflect this. *In construction of the field the author(s) reference the null setting multiple times. However, I suspect that there is actually a missing detail. The null only comments on means, while the randomization also assumes that under the null the distributions have equal shapes. This is an important distinction that is often missed in randomization techniques but affects student intuition. It seems the author(s) ought to at least acknowledge that there are additional details which are being ignored for better focus on the main point of the hypotheses. This is the most consequential point of revision in my view, but could be addressed quickly with a sentence or two of acknowledgement when introducing the creation of the null plots for the lineup and a quick comment that additional conditions are ignored in other settings.*

**Response:** You are absolutely correct that additional conditions are needed and glossed over when initially honing student intuition. I added a technical note under point #3 in Section 2 when I introduce the null distribution to clarify this point. *Generally, the manuscript is written in the first person plural. The parenthetical comment on page 8 at the end of Section 3.1.1 is one exception.***Response**: This inconsistency has been fixed.