6/28/2018

Dear Matthew Koehler, Jennifer Schmidt, Lisa Linnenbrink-Garcia, and Christina Schwarz,

I am writing with a memorandum of understanding regarding the required revisions to my dissertation. I outline my understanding of the specific revisions that are required. Thank you very much for this very valuable feedback and the opportunity to strengthen this work.

Sincerely,

Joshua Rosenberg

**Introduction**

1. **Improve the argument for the need for this particular study. Include this powerful rationale for the study in the abstract and literature review as well as throughout the manuscript (i.e., in the need for study section).**
   1. **Make the case that work with data empowers learners, turning learners from those consuming knowledge to those creating it.**
   2. **It also provides a capability that can be used across content areas, particularly in advanced coursework.**
   3. **Is involved in youths’ lives and can be a relevant context for learning, apart from contexts such as robotics and coding.**

**Literature Review**

1. Clarify from what sources the five aspects of work with data came from.
   1. I added this paragraph: “For instance, Wild and Pfannkuch (1999) consider the process in terms of identifying a problem, generating a measurement system and sampling plan, collecting and cleaning the data, exploring the data and carrying out planned analyses, and interpreting the findings from the analysis. Such a process is common in STEM content areas, particularly across statistics education research and is instantiated in standards for curricula: Franklin et al.’s guidelines for the American Statistical Association focus on the Framework for statistical problem solving: formulating questions, collecting data, analyzing data, and interpreting results (2007). The goals of this framework and its components are similar to Hancock et al.’s (1992) description of data modeling, the process of “using data to solve real problems and to answer authentic questions” (p. 337). Hancock et al. (1992) focus in on two goals, data creation and analysis, arguing that the former (data creation) is “the neglected counterpart of data analysis” (p. 339). Scholars have subsequently expanded Hancock et al.’s definition of data modeling to include six components: asking questions, generating measures, collecting data, structuring data, visualizing data, and making inferences in light of variability (see Lehrer & Schauble, 2004, for using this conceptualization of data modeling applied to the task of understanding how plants grow). The last of these components is crucial across all of the visions of data modeling reviewed here and distinguishes these processes from other aspects of data analysis: Accounting for variability (or uncertainty) is central to solving real-world problems with data and the process of data modeling.”
2. Re-order research questions #2 and #3 on p. 17.
   1. I re-ordered the research questions.
3. When the programs are described on p. 18, refer the reader to the appendix.
   1. I added the following sentence: “Add a coding frame for the STEM-PQA with the names, possible values (i.e., present or not present), description, and an example.”

**Method**

1. **Add a coding frame for the STEM-PQA with the names of the variables, possible values (i.e., present or not present), descriptions and examples.**
2. **Add a coding frame for the open-ended, qualitative coding with the themes, descriptions, and examples.**
3. *Clarify how the STEM-PQA aligns with the aspects of work with data.*
4. Report the reliability of the pre-interest measure.
   1. I added the following information: “The individual interest measure represented the mean of interest items across all relevant domains. Thus for some students, the mean was based on 3 items, while for others it was based on as many as 9 items representing all three domains (with Cronbach alpha values ranging from .77 - .86 for each domain specific interest scale)”

**Results**

1. **Provide more information about work with data and why it may be engaging to youth (in the abstract; on p. 38 with respect to the use of statistical and mathematical models versus the development of these models; and on p. 48 with respect to data modeling)**
2. *In the descriptive analysis, include the correlations between the aspects of work with data and the individual variables used to create the profiles.*
3. *Include Table 7.3 in the document instead of in the Appendix, but modify it to include only the AIC, BIC, SABIC & entropy, cell sizes, and BLRT.*
4. **Regarding *how* the six-profile solution as selected, move some of the discussion from the appendix. (See Lisa’s published work for example; mention I did analysis of six versus seven profiles in-text.)**
5. **Provide a richer description of the six profiles. Use a MANOVA to determine which variables differ across the profiles (and for which profiles). Use subscripts in a table with the mean values to indicate which differ.**
6. In the descriptions of the six profiles, report the percentage of responses in each profile. Related, state that entropies are high, so it is reasonable to extract the most likely profile membership.
   1. I added the percentage of responses associated with each profile to the descriptions of the profiles.
   2. I also added the following sentence: “Mention that any of the aspects of work with data versus none of the aspects of work with data and the interactive effects of youth characteristics and the aspects of work with data were examined but not found to be statistically significant (but do not include these in a table).”
7. *Improve Table 4.5, so that the betas and standard errors are labelled for each model; format the column (presently too wide) for the Engaged and Competent but not Challenged profile.*
8. Mention that any of the aspects of work with data versus none of the aspects of work with data and the interactive effects of youth characteristics and the aspects of work with data were examined but not found to be statistically significant (but do not include these in a table).
   1. I added the following two sentences: “Mention that any of the aspects of work with data versus none of the aspects of work with data and the interactive effects of youth characteristics and the aspects of work with data were examined but not found to be statistically significant (but do not include these in a table).”

**Discussion**

1. *Discuss more broadly what it means that this is happening in the context of a summer program, specifically in the limitations section.*
2. *Speculate about why some of the anticipated findings were not found, particularly by discussing reasons for why the activity does not matter very much: methodological, summer context, under-represented youth and equity issues, and work with data just is not very engaging.*

**Throughout the Manuscript or Overall**

1. Be careful about language use when discussing profiles; change any instances of profile membership to probability of profile membership at a particular moment.
   1. I made changes to the research question #2 results section to reflect this, as described in revision point 14 above. I also searched for other examples of this but did not find any instances.
2. Use the past verb tense throughout the methods, results, and discussion section. Use the first-person verb conjugation in these sections.
3. Carefully copy edit the manuscript or have the manuscript copy-edited.
4. Acknowledge that this is a secondary analysis of existing data, citing the STEM-IE NSF grant number (1421198).
5. Include NSF blurb relating to their independence from the findings of this research as an author’s note (“This material is based upon work supported by the National Science Foundation under Grant No. 1421198). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not reflect the views of the National Science Foundation.”