Thank you for the chance to respond to these comments.

R1:

* R1.1: Many of the sentences were excessively long and complex, making it hard to tease apart the major points being made.

In response to this feedback, we revised the introduction and literature review.

* R1.2: The motivation for the study did not seem clearly articulated.

We conjectured that work with data would be more engaging than other activities in light of past research on work with data and past research on engagement.

* The conceptualization of engagement as having challenge and competence components in addition to cognitive, affective, and behavioral dimensions is an interesting one. I would like more explanation on how challenge and competence aren't already encompassed by the cognitive dimension.

Thank you for this feedback. In response, we clarified how challenge and competence are distinct because of XXX. Also, emergent motivation theory (CZ) argues that...

* R1.3: At times, it seems like Emergent Motivation Theory is mixed with engagement, but motivation and engagement are related but different constructs.

Thank you for this helpful comment. We highlight that EMT is a theory that includes both motivation and engagement.

* R1.4: I suggest dropping RQ 4 for this paper, as it seems to be an extraneous analysis considering the population was so heavily URM. I'm a bit torn as I feel like this might be better as separate papers.

In response to this feedback, we have removed this research question.

* R1.5: There were several typos throughout that should be corrected.

We have carefully copy-edited the manuscript to address the typos contained therein.

R2:

* R2.1: It is very confusing about the definition of work with data and its connection with engagement.

In response to this feedback (as well as R1.2 and R1.3), we have extensively revised how we defined work with data and its connection to engagement. We also discuss this further in response to the next comment.

* The manuscript did not provide any review of literature on prior work related to engagement in work with data.

We have sought to address this in our review of past research related to how engaging work with data may be.

* The manuscript does not provide a sound rationale in setting up this study.

Thank you for this comment – in response, we have made extensive changes to the introduction and literature review (see also our response to R1.2). In particular, we emphasize that work with data may be an especially engaging practice in light of past research both on work with data and engagement.

* The theoretical and empirical meanings of cognitive, behavioral, and affective engagements and also challenge and competence was mentioned by briefly and should be discussed in details and depth.

In response to this comment, we have extensively revised this section of the manuscript. In particular, we describe theory and past research on engagement, emergent motivation theory in the context of challenge and competence beliefs, and motivation, as well as how they are distinct.

* Research questions 1 and 3 need to be reworded for clarity.

We have revised these research questions for clarity.

* Please remove the word "innovative" on page 15.

We have removed this word.

* The context needs to be described in main text instead of in an Appendix.

We response to this feedback, we have moved the text that was in the Appendix into the main text.

* The students were from 9 out-of-school STEM programs. Are they different programs? How can the study combine students from these programs considering the significant contribution of ESM is to situate research in specific contexts?

The nine programs are distinct: none of the students in one proram was enrolled in another. While program-level factors may matter and we note that while the students are from distinct programs, we believe that ESMstill allows the study of specific contexts, but less in terms of specific programs, and more in terms of instructional activities and youths’ subjective experiences in-the-moment.

* The demographics of 203 participants were atypical with majority being Hispanic and Black. Therefore, the authors should carefully interpret their findings since their study sample does not represent the typical STEM population.

We agree, and have emphasized this as a limitation to the study, and have made changes to our claims regarding how engagement and work with data are related in general.

* ESM data were collected 4 times a day for 2 days each week for week 2-4. Therefore, each student should receive 4x2x3=24 signals in total. First, 4 times a day signaling is quiet heavy considering they only spent around 3 hours per day. Students would have received a signal every 40 minutes. What's the response rate? How many total events were captured? How were these events distributed among students and across time?

Thank you for these comments. In response, we have made substantial additions to the data collection section. In particular, we describe how

* How were the video processed for analyses?

We have added details on how the video was saved, processed, and analyzed in the data analysis section. In particular,

* How were the units of analysis decided?

Thank you for this question. In response, we noted in the data analysis section that the units of analysis were decided on the basis of

* What was the inter-rater reliability for the video coding?

We have emphasized that the inter-rater reliability for the video coding was established through a process of XXX on p. XXX.

* On page 18, please provide details on the use of maximum value measurement approach, instead of asking readers to read "Authors and colleagues (2018b) for more details".

We have included these details in the section on the measures.

* There is a lack of details in term of data preparation. Any data screening process included in this study? Any outliers or missing data?

Thank you for this feedback. In response, we have added more information on the data screening as well as to how we checked the parametric assumptions of the data.

* On page 22, no justification was provided for the reason why the LPA solution was selected. Why did you choose the six-profile solution? What are the fit indices and quantitative measures used to justify this choice?

In response to this comment, we have added substantially more detail on the model selection carried out as a part of LPA.

* Although the authors mentioned that they used the different information criteria and a statistical test of the number of profiles in the data analysis section, no statistics was provided in the manuscript.

We have now included these in the manuscript.

* Also, please report the number of episodes in each profile.

We included this information in terms of the model-based predicted number of episodes in each profile.

* Please provide more details regarding your model specification and fit statistics for the mixed effects, multi-level regressions. Was it a logistic regression? If so, what was your reference group? What was the R-square of each model?

We included the detail that the model is a multi-level regression model. We also note that we ran the model as a multi-level logistic model (by hard-coding the posterior probabilities, instead of treating them as continuous), but the results were substantively the same (in terms of the statistical significance of all of the effects as well as their magnitude and direction). We also included fit statistics and the R-square values.