The lead author's first reaction to the reviewers' responses was one of encouragement. While the revisions required were indeed substantial, they signaled an intuitive way forward.

Reviewer #1's comments, which revolved around the critique that the paper was trying to accomplish two distinct goals which made it too ambitious, made it clear that we had to make the paper more coherent throughout. This resonated with Reviewer #2's first bin of critiques which about t*he way the study was framed conceptually and within the extant literature*.

This was the first and easiest thing to address collectively. We came up with five main points of modification that we believe addressed these first two concerns:

1. We revised the introduction to exclude anything that was not directly constructive to the argument.
2. Edited the "limitations of the study" following the critique from Reviewer #1
3. We re-framed the language in our paper away from "bad teachers" or "bad teaching" and were more explicit about the fact that we were talking about "the way students described teachers they rated poorly."
4. We sought to make a clearer connection between our study and educational policy
5. Perhaps most importantly, we set as an objective to retain a consistent argument throughout the paper. This meant that we sought to tie our results, discussion and conclusion, directly to the literature presented early in the paper.

Point 4 was also partly inspired by the positive comments made by Reviewer #3. It was, in fact, these positive comments that helped find a focus with which to accomplish Point 5. The importance of establishing this argumentative structure for addressing Reviewer #2's second bin of critiques cannot be understated.

Reviewer #2's critiques were brief but clearly communicated what the reviewer felt were the disconnections between the evidence and the argument. We focused the one sentence that encompassed Reviewer #3's thoughts:

*The authors are asking too much of the method. And we can see this in their results, which lack conceptual coherence.*

Addressing this began with a revision of the methodology. We changed our model selection criteria in two ways,

* one technical: from selecting the highest coherence to the moment where coherence score drops (a change made based on a technical but non-academic citation).
* one critical, but subjective: in selecting a model with fewer topics because key-words in models with more topics overlapped.

the reasoning and implications of these changes are discussed in the paper.

Additionally, we took advantage of the revision time and an increase in the computational power at our disposal to extend our analysis and experiment with more parameters and models. This resulted in what we believe to be a more robust body of statistical evidence.

However, perhaps the greatest revision was the process we set for our interpretation of the results. We used a more comprehensive selection of key-words and more extensive, though equally systematic, approach to for the sample reviews submitted by students.

Our Results, Discussion and Conclusions sections are completely different. Weaknesses in our approach are addressed early on and in the discussion, but we also believe that the evidence is considerably more coherent and conducive towards our policy argument. We are very proud of what we've accomplished with these revisions and thank the reviewers very much for their critique and insight.

We hope you enjoy the read.