Dear Editor,

We were happy to learn that our manuscript was conditionally accepted for publication in the AJAE. Here are details on how we responded to the remaining comments and suggestions:

- To reduce the length of the main text by 1 page, we decided to put the section on heterogeneous treatment effects in the online appendix. We considered this the best way to shorten the manuscript without affecting the flow of the text. We decided to keep the part on cost-effectiveness as is, as all reviewers seemed to be very happy about that part.
- 3 out of 4 reviewers wondered why we choose to run regressions on separate sub-samples to test heterogeneity of treatment effects (reviewer 2.3, reviewer 3.8, reviewer 4.1). They argue that it would be much more informative to use an interaction term of the treatment indicator with an indicator of the group for which a different effect is expected. This allows one to see if differences are statistically significant. We completely agree with this and re-ran the regressions using interactions (Tables 15 and 16 in the Online Appendix). The analysis confirms previous findings that learning effects from the video intervention are higher among less educated farmers, and that inputs promoted in the video are adopted more by farmers that live closer (i.e. less than median distance) to agro-input dealers.
- We did not follow the suggestion of reviewer 1 to report Lee bounds (instead of Manski bounds as we now do). The reviewer notes that Lee bounds are widely accepted these days and that they are less conservative than Manski bounds. We feel that, for the case of testing robustness, it is better to be conservative; as our results broadly hold under this conservative approach, we see little reason to change to Lee bounds.
- Reviewers also asked us to acknowledge a few things. For instance, reviewer 2 asks to acknowledge the fact that the Manski bounds did sometimes switch signs between lower and upper bound. In addition, reviewer 2 notes that despite the fact that we find only 1 percent of control group households that called into the IVR, there are still many calls on the call log that we could not identify; some of these could be control farmers, so there may be more contamination. We have acknowledged these caveats as suggested.
- Reviewer 3.2 agrees that discussion of potential mechanisms is helpful, but wants us to be more cautious when considering the second knowledge question (on combining different inputs) as evidence of a specific learning mechanism. We have therefore removed reference to inspirational channels here, as we agree other channels may be at play as well.
- Reviewer 3.3 notes that for the knowledge index, the last question should not have been included in the index as it was not expected to have an effect. This was already the case in the original manuscript, but we have now indicated this in the note of Table 2.
- Reviewer 3.4 cautions that the effect of IVR on hybrid seed adoption may be picking up noise. We have added this observation, further cautioning the reader in interpreting results from the IVR intervention.
- Reviewer 3.5 notes that the production index is composed of variables that do not seem to fit into a family as nicely as other variables that were combined into indices. We agree, but the main reason why we insist on using also an index here was to account for multiple hypothesis testing.
- Reviewer 3.6 suggests removing LATE estimates. Initially, we did not want to include the LATE analysis, but in previous rounds of the review process, all reviewers were of the opinion that it is important to include these results, despite the issues we mentioned. So we think it is better to leave them in. Most of the discussion on LATE is in the appendix anyway, and the part that discusses low compliance is still useful for the interpretation of the intent-to-treat effects of the IVR intervention.

- Reviewer 3.7 takes issue with the example we give when arguing that some questions are more prone to experimenter demand effects than others and hence, consistency in results over different questions is reassuring. The reviewer does not seem to disagree with this argument, but seems to suggest the example is not very good. We thus removed this example.
- Reviewer 3.8 suggests to use interactions to test heterogeneity, which we have now done. We have rewritten the part that seems to suggest that fertilizer adoption is more susceptible than seed adoption (this was not what we wanted to say). We also indicate that we looked for heterogeneity in other characteristics, but no significant differences were found.
- Reviewer 3.9: the reviewer finds the cost-benefit analysis very helpful, but notes that we only look at short-run effects; if farmers also adopt in the future, effects may be much larger. The reviewer notes he/she would have loved to see a discussion on this. We feel that saying something about the long run would be too speculative. Also, we find that even in the short run, the video intervention is profitable. We therefore decided to leave this section as is.
- Reviewer 4.2. notes that there is now more detail about the design of the study, data collection and how variables were constructed. However, he/she thinks more can be done, especially with respect to how household level variables are constructed from answers of both spouses. As advised, we now refer to our discussion paper that focuses on the gender dimension of the study, which has information on variance of responses between the spouses. In addition, the final version of the paper now also refers to additional online material (which previous versions could not do because of the double-blind review process). This included a link to a public repository on GitHub that tracked through time all code and data needed to replicate the analysis. From this, it should be fairly easy to find out these technical details.
- To get an idea of the likelihood of spillovers, reviewer 4.3 wants to have an idea of the number of households in the average village. Villages consist of on average 600 households.
- Reviewer 4.4: we actually did collect baseline information in a dedicated baseline. This has been removed.
- Reviewer 4.5 wonders if significance stars are correct in table 4. They are: these stars are for p-value thresholds that also consider multiple hypothesis testing.
- Reviewer 4.6 flags and incomplete sentence. This sentence was removes in response to reviewer 3.2.