Review:

I reviewed the second revision of your submission today and found that your responses to most of my points satisfactory. I am pleased to conclude that your submission can be accepted if you can clarify in your next revision the following issues in a timely manner.

1. In my last round of review, I asked "If Model 1 involves all background variables in the imputation model but the DI cases only have the 6 DI variables, how was the imputation carried out? What values are specified into the imputation model for the other background variables to do the imputation? This should have been made clear". In your response and in the paper, you merely confirmed that the DI cases are missing most of the background variables and that this is the reason why it does not work well, but you did not answer my question. If the imputation model takes the form of a predictive distribution conditioned on given background variables' values but the DI cases do not have all the values on the background variables, then how are the proficiency values of the DI cases imputed? To be more concrete, if the imputation model has y = x1+x2+x3 + error, but the DI cases only have x1 values but not x2 and x3 values, how do you impute y? what x2 and x3 values are plugged in? In addition, why does Model 1 tend to underestimate?

Response: Note that for doorstep interview cases, data for background variables other than the six doorstep-specific variables are missing by design. In the latent regression model, all non-doorstep interview background variables were coded with a "missing" category (e.g., gender includes three response categories: male, female, and missing), while the six available doorstep interview variables retained their actual values. This means the model conditioned estimates on both the known doorstep interview variables and the missing responses of other background variables.

For your concrete example of y = x₁ + x₂ + x₃ + ε, DI cases would have x₁ observed but x₂ and x₃ coded as missing, leading the model to estimate y based primarily on the missingness pattern rather than substantive background information.

1. On the top of page 6, you explained how the item scores were imputed, and wrote "from the path 1 posterior theta distribution (averaged over all of the existing path 1 respondents' posteriors) ". Does this mean the six doorstep variables of the DI cases are not used in this imputation and all DI cases were just drawn from the same mixture distribution? Why is it done this way?

Response: In imputing these scores for locator items for doorstep interview cases, data from background variables were not used. These background information in addition to the imputed cognitive items were both used in estimating the proficiencies for those doorstep interview cases.

1. You confirmed in your response that the first four bullet points under Results (bottom of page 6) are based on theoretical considerations (model specifications). In light of this, your writing should make it clear that these bullet points are predicted behaviors of these models based on theory, rather than conclusions drawn from Figure 2. I would suggest that you move them to before Section 4 and make them a separate "Section 3.4 Predicted performance". Then at the beginning of Section 4 you briefly discuss how Figure 2 is consistent with such predicted behavior. (e.g., Model 2 has similar results with and without imputed cognitive data, which means double dipping is unnecessary; Model 3 without cognitive data tends to overestimate as you explained in your response to my last round of review, etc).

Response: We appreciate this suggestion about clarifying the relationship between theoretical predictions and empirical results. The current organization tightly links predictions to their empirical validation. Major revisions could disrupt the logical progression from methodology to results.

In addition, please note the following message below regarding ORCID and affiliation formatting. Please revise accordingly in your next revision. Please submit your revision no later than 10am (NYC time) in the morning of Jun 20.  Please submit the revision you sent me today onto the OpenReview website. And also submit your next revision to the website.

Response: The revised documents were uploaded to the OpenReview website.

Although ORCID iDs are not required for DOI registration, they are encouraged. Authors can create one at [https://orcid.org/register](https://nam10.safelinks.protection.outlook.com/?url=https%3A%2F%2Forcid.org%2Fregister&data=05%7C02%7Cuali%40ets.org%7C2c130ef6fcd5422d0c9608ddaee9cfa8%7C0ba6e9b760b34fae92f37e6ddd9e9b65%7C0%7C0%7C638859041473481817%7CUnknown%7CTWFpbGZsb3d8eyJFbXB0eU1hcGkiOnRydWUsIlYiOiIwLjAuMDAwMCIsIlAiOiJXaW4zMiIsIkFOIjoiTWFpbCIsIldUIjoyfQ%3D%3D%7C0%7C%7C%7C&sdata=UgkDo3HUaRqLM0DuC6jN76imeF0PC305dyC9s%2FHKnOw%3D&reserved=0) and include it in their final manuscripts.

For LaTeX submissions:  
Add the following lines before the \title command in the template:  
\usepackage{orcidlink}  
\newcommand{\orcidauthor}[2]{#1~\orcidlink{#2}}

Then replace:  
\author{F. Author}  
with  
\author{\orcidauthor{F. Author}{0000-0001-1234-5678}}

For Word submissions:  
F. Author† (ORCID: [https://orcid.org/0000-0001-2345-6789](https://nam10.safelinks.protection.outlook.com/?url=https%3A%2F%2Forcid.org%2F0000-0001-2345-6789&data=05%7C02%7Cuali%40ets.org%7C2c130ef6fcd5422d0c9608ddaee9cfa8%7C0ba6e9b760b34fae92f37e6ddd9e9b65%7C0%7C0%7C638859041473503699%7CUnknown%7CTWFpbGZsb3d8eyJFbXB0eU1hcGkiOnRydWUsIlYiOiIwLjAuMDAwMCIsIlAiOiJXaW4zMiIsIkFOIjoiTWFpbCIsIldUIjoyfQ%3D%3D%7C0%7C%7C%7C&sdata=Biq%2FCXERzhPlpqUAuue6V%2Bj3vdaYk%2FTa71fqZ8FAeiQ%3D&reserved=0))

Response: Latex file was updated to include the ORCID.