Thank you for addressing my previous comments. I think the adjustments have improved the manuscript. I still have the following two remaining comments, however:

1. I appreciate the authors’ adjustments to the effect size calculation and see their point. Nevertheless, if dz and dav are mixed together in a single meta-analysis, then I think sensitivity analyses should be added that test if the results change if dz effect sizes are excluded. The authors note that differences between dz and dav are usually quite small, but this depends strongly on the correlation between the paired measures. When this correlation is high, the difference between dz and dav can be quite substantial. I realize that discarding effect sizes is far from an ideal solution, but at least it would give some (even if flawed) indication of the extent to which differences in effect size computation contribute to the results of the meta-analysis. Again, I think this is especially important for H3, which I think will almost certainly show that within-subject designs produce larger effect sizes, simply because of differences between dz and ds.
2. I might be missing something here, but it is not clear to me how multivariate three-level models can account for confounding between moderators (and I don’t think they automatically do that). Unless multiple moderators are included together in a single model, how could these models account for confounding between said moderators?