

Critique of Ooshima & Mitamura (2025) ‘Effects of perspective-taking training based on relational frame theory for cognitive empathy and emotional empathy: Differences in perspective-taking according to various theoretical approaches’

Ian Hussey

Ooshima & Mitamura (2025) reports a study on the impact of a perspective-taking (PT) intervention informed by Relational Frame Theory (RFT) on measures of cognitive perspective taking, specifically perspective taking. This critique highlights several issues with the design, data and results: several elements of the results are not computationally reproducible (sum scoring, outlier exclusion; possible swapping of conditions); no multiple testing corrections were applied, and when Holm corrections are applied all results are non-significant; results from analyses that appear to undermine the support for the hypothesis are not fully explicated; and the study lacked a control condition which undermines strong causal conclusions. On closer inspection, results do not support the hypothesis that RFT-inspired PT training increases cognitive empathy.

Omission of the non-significant result on the second outcome variable from the abstract

The article abstract states: “This study investigated the impact of perspective-taking training rooted in the relational frame theory on perspective-taking, as conceptualized by the organizational model of empathy. ... The post-training scores for the subscales measuring cognitive empathy increased significantly ($p = 0.031$)” (Ooshima and Mitamura, 2025, p. 1).

This statement about the results is incorrect. There were two cognitive empathy PT subscales, only one of which showed a significant effect: MSE PT ($p = .031$) and IRI PT ($p = .779$). The omission of the non-significant result on the much more widely used IRI scale from the abstract should be noted.

Absence of multiple testing corrections

Given that two measures of cognitive PT were used to test the same broad verbal hypothesis about the impact of RFT-informed PT training on cognitive empathy, multiple testing corrections are arguably warranted. Applying Holm corrections to the reported p values makes both non-significant: $ps = 0.062, 0.779$.

An alternative way of analyzing the multiple outcomes would be to combine both outcomes by calculating the precision-weighted-average Cohen’s d_{av} and its intervals. In the analyses below, I show that doing so produces a non-significant effect (precision-

weighted-average Cohen’s $d_{av} = 0.04$, 95% CI [-0.11, 0.19]).

Unclear statement about the alternative analyses and omission of their non-significant results

“...however, the increase in the score remained significant trend after adjusting for training performance.” (Ooshima and Mitamura, 2025, p. 1). This statement is somewhat unclear in terms of what is meant by a “significant trend” and what was adjusted for. My hopefully reasonable assumption is that the authors refer to the alternative analysis of both accuracy and reaction times on the cognitive empathy PT subscales using multilevel models. However, the results of those analyses showed that neither MSE or IRI subscales demonstrated significant change from pre to post for accuracy (MSE PT: $p = .144$; IRI PT: $p = .698$) or reaction time (MSE PT: $p = .062$; IRI PT: $p = .443$). Perhaps the marginal result for reaction times on the MSE are what are being referred to. Given that all four p values again tested the same broad verbal claim about the impact of RFT-informed PT training on cognitive empathy, multiple testing corrections are again warranted. Applying Holm corrections to the reported p values makes all non-significant (or indeed marginally significant): $ps = 0.432, 0.886, 0.248, 0.886$.

Questionable appropriateness of the reaction time analyses

Additionally, it's important to note that neither self-report scale, to my knowledge, were presented as speeded response tasks, nor is reaction times on a self-report empathy scale employed as a measure of empathy elsewhere in the empathy literature to the best of my knowledge, raising questions about the appropriateness of the reaction time analysis. Given that the article's stated aim is to investigate the impact of RFT-informed PT training on classical measures of cognitive empathy, analyzing the measure in a non-traditional way undermines the validity of the claim.

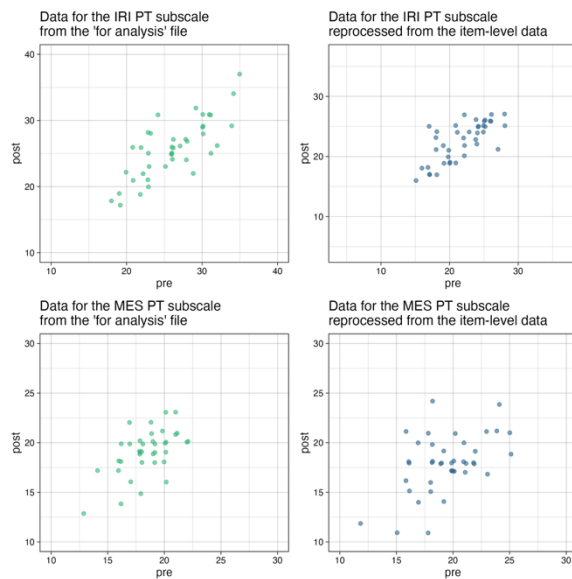
Summary of the hypothesis tests after adjusting for multiple comparisons

After adjusting for multiple comparisons, no statistically significant results were found on either cognitive empathy subscale in either method of analysis (all $ps > 0.062$).

Downplaying of evidence contrary to the hypothesis

The article's hypothesis is that RFT-informed PT training should increase cognitive empathy. This argument has implied specificity, i.e., that the PT training should not increase affective empathy; or perhaps that it should increase both, but not only cognitive empathy. The article does not acknowledge that results run contrary to this implied specificity: more significant results with larger effect sizes were observed for the MES Other-orientation ($p < .001$, Cohen's $d = 0.94$) and Affectedness subscales ($p = .001$, Cohen's $d = 0.58$; see Table 2 in the article).

Figure 1. Comparison of the PT subscales sum scores reported to have been used in the analysis vs. sum-scores recalculated from the item level data.



Sum scores are not computationally reproducible

In Figure 1, I demonstrate that the IRI PT and MES PT subscale sum score data reported to have been used in the analysis are not computationally reproducible from the item level data.

Separately, data for the MES PT subscale scores were provided in two different data files each on the OSF. However, verification shows that the pre and post conditions are swapped between the two files. It is unclear which represents the real data set (putting aside issues of reproducibility). If the wrong one was used in the analysis, this would switch the direction of the effect from increases in cognitive empathy to decreases in it (putting aside issues of multiple testing corrections).

Outlier exclusion strategy is not reproducible

The article states "For each subcategory of each index, we applied the interquartile range (IQR) method to identify outliers, which are data points that fall outside the IQR. We conducted the data analysis by excluding outliers upon observation. Outliers were excluded if they were observed. An overview of outlier counts revealed four ($n = 37$) for MES Perspective-taking, one ($n = 40$) each for Self-orientation and Other-orientation, and one ($n = 40$) for the conceptualization of the TSSQ" (Ooshima and Mitamura, 2025, p. 7). This is internally inconsistent: the interquartile range by definition includes 50% of participants, but only 1-4 participants were excluded from some subscales. No R code implementing the outlier exclusions is included in the authors' code on the OSF project. The exclusions therefore cannot be reproduced from the description in the article.

Absence of a control condition

Using pre-post intervention changes as evidence of the causal impact of the intervention on the outcome variable is widely critiqued as suffering from being confounded by time-related factors and cannot rule out alternative explanations such as demand, history effects, or regression to the mean. This is why randomized controlled trials must include a control condition, in order to provide an estimate of the causal impact of the intervention via the counterfactual of not receiving the intervention. In addition to concerns about the results after multiple testing corrections, the absence of a control condition serves to further undermine the weight of evidence the design is capable of producing.

Spin in the article vs. the poster

The authors' OSF project includes a poster presenting the results of the study at a conference. It is worth highlighting the difference in the presentation of the results in the article's abstract (which presents only the significant results) vs. their summary of results in the poster, where they state "There was no significant PT for IRI-J, but significant change in Perspective-taking in MES. Cannot conclude that there were impacts before or after training consistently." This is not technically at odds with the abstract's results, but arguably an important different in 'spin' of the results.

Transparency statement

All data and code are available at
<https://github.com/ianhussey/verification-of-ooshima-and-mitamura-2025>

Author notes

Ian Hussey, University of Bern, Switzerland
(<https://orcid.org/0000-0001-8906-7559>;
ian.hussey@unibe.ch).

References

Ooshima, Y., & Mitamura, T. (2025). Effects of perspective-taking training based on relational frame theory for cognitive empathy and emotional empathy: Differences in perspective-taking according to various theoretical approaches. *PLOS ONE*, 20(5), e0323120.
<https://doi.org/10.1371/journal.pone.0323120>