Appendix of "A Style-Aware Polytomous Diagnostic Model for Individual Traits"

Yixuan Wang[a], Jiale Feng[a], Yue Huang[a], Xuruo Pan[a], Zhongjing Huang[b,*], Zhi Liu[a,c] and Hong Qian[a,d,*]

[a]Shanghai Institute of AI for Education, East China Normal University, Shanghai, China
[b]Faculty of Education, East China Normal University, Shanghai, China
[c]Shanghai Innovation Institute, Shanghai, China
[d]Key Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE, East China Normal University, Shanghai, China

The appendix consists of the following parts:

- Appendix 1 introduces the population-level distribution of traits inferred by SAPD in other datasets.
- Appendix 2 illustrates the visualization of the response style estimated by SAPD in other datasets.
- Appendix 3 demonstrates the validity and effectiveness of SAPD in diagnosing participants' trait states.
- Appendix 4 illustrates the detailed effect of hyperparameter α on other real-world datasets.
- Appendix 5 shows the case study about disacquiescence response style on Houston dataset.

1 Population-Level Trait Distribution (To Q3)

To further assess the interpretability and psychometric validity of SAPD on different datasets, we visualize the population-level distribution of the inferred trait levels. As shown in Figure 1, we select the commonly shared trait Responsibility from the Suzhou, Houston, and Moscow datasets, and additionally include the trait Empathizing Quotient from the EQSQ dataset. Each subfigure presents the predicted trait levels using SAPD as probability histograms overlaid with Gaussian density curves. The results show that most distributions follow a bell-shaped pattern, with the majority of participants concentrated around moderate levels and fewer individuals at the extremes. This aligns with the common assumption of normal distribution in psychometrics, supporting the validity of SAPD's trait inference at the population level. Figure 1(d), the distribution of the Empathizing Quotient trait skews slightly lower, which may be attributed to the sample size of the EQSQ dataset or the nature of this particular trait. In general, these visualizations demonstrate that SAPD yields psychologically plausible and interpretable trait estimates across all five datasets.

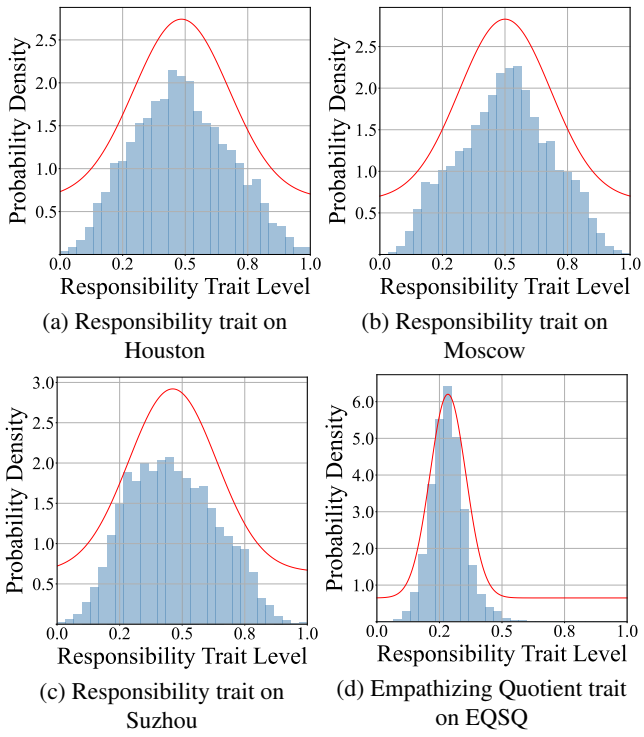

Figure 1. The population-level trait distributions inferred by SAPD on four datasets.

2 Visualization of Response Style (To Q3)

To further validate the generalizability and robustness of SAPD in modeling participants' response styles, we visualize the learned response style embeddings across four datasets: Houston, Moscow, Suzhou, and EQSQ. Leveraging *t*-SNE for dimensionality reduction, we project each participant's response style embedding onto a two-dimensional plane. As shown in Figure 2, the resulting distributions exhibit three distinct clusters that correspond to disacquiescence, midpoint, and acquiescence response styles. In particular, these clusters appear in a left-to-right progression, reflecting the semantic order of the Likert scale options from disagreement to agreement. This consistent spatial pattern across datasets indicates that SAPD effectively captures the ordinal nature of response options and differentiates between individual-specific stylistic tendencies. These visualizations provide strong empirical evidence that SAPD learns meaningful and interpretable style representations that generalize well across varied populations and assessment settings.

* Corresponding Authors.
Email: zjhuang@dedu.ecnu.edu.cn, hqian@cs.ecnu.edu.cn.

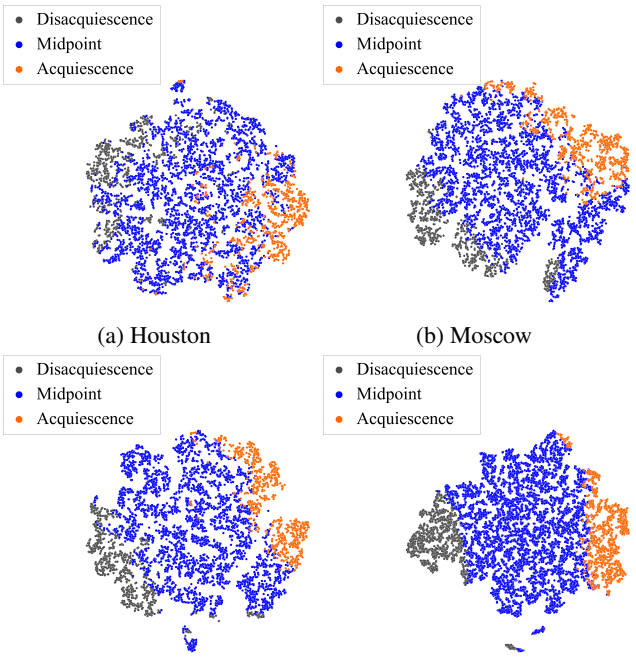

Figure 2. *t*-SNE visualizations of response style learned by SAPD on four dataset.

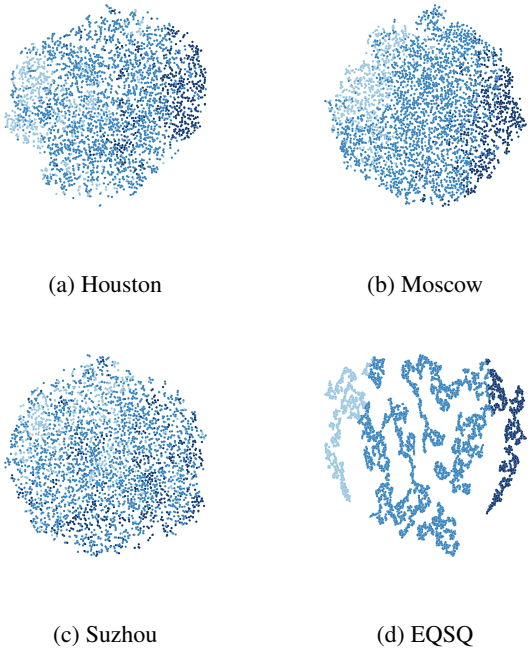

Figure 3. *t*-SNE visualizations of trait inferred by SAPD on four datasets.

3 Visualization of Trait (To Q3)

To further examine whether the trait estimates produced by SAPD are truly disentangled from response styles, we visualize the learned trait embeddings under the Midpoint response style across four datasets: Houston, Moscow, Suzhou, and EQSQ. As shown in Figure 3, although all participants share the same response style category, their representations of the trait in the two-dimensional *t*-SNE space form a smooth and continuous gradient. This indicates that participants exhibit diverse levels of latent traits even within the same style group, confirming that midpoint style does not imply midpoint ability. This pattern strongly supports the effectiveness of SAPD in separating individual response styles from underlying trait levels. The model not only mitigates stylistic bias during inference, but also preserves meaningful variation in participant abilities, demonstrating both accuracy and interpretability in trait diagnosis in diverse assessment contexts.

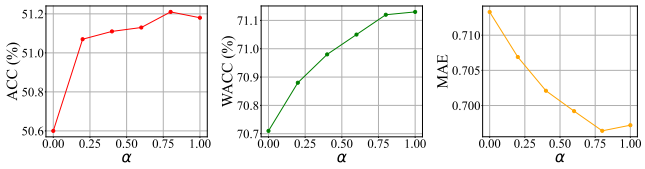

Figure 4. ACC, WACC, and MAE performance under different values of α on Suzhou dataset.

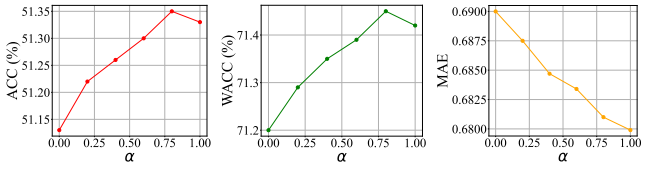

Figure 5. ACC, WACC, and MAE performance under different values of α on Houston dataset.

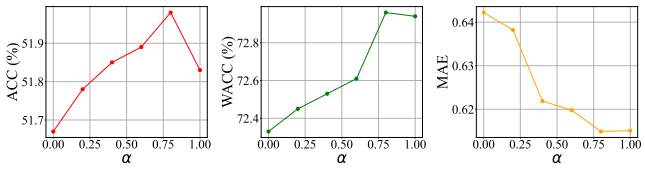

Figure 6. ACC, WACC, and MAE performance under different values of α on Moscow dataset.

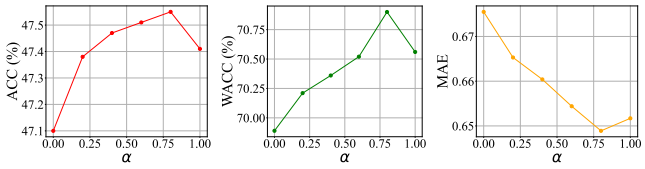

Figure 7. ACC, WACC, and MAE performance under different values of α on EQSQ dataset.

4 Supplement of Hyperparameter Analysis (To Q4)

We further supplement our hyperparameter study by analyzing the impact of α on four additional datasets. The coefficient α balances the contribution of $\mathcal{L}_{EMD}$ to the joint objective. We vary α over the range $\{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$, and the results in Suzhou, Houston, Moscow, and EQSQ are presented in Figure 4, Figure 5, Figure 6, and Figure 7. In all four datasets, performance across all three metrics, ACC, WACC, and MAE, consistently peaks at $\alpha = 0.8$. This indicates that placing appropriate emphasis on ordinal consistency significantly improves the accuracy of the prediction. Together with

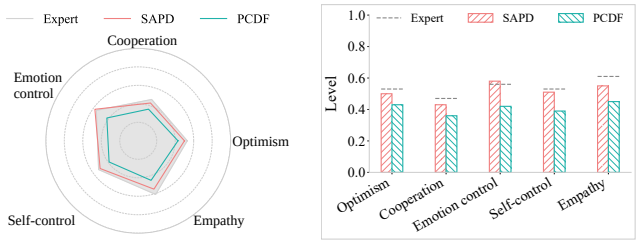

Figure 8. Case Study of the typical participant with dis-acquiescence on Houston dataset.

results of BIG5, all five datasets reach optimal performance when $\alpha = 0.8$, validating our choice of this value as the default setting in SAPD.

5 Case Study Analysis (To Q5)

We conduct a case study on the Houston dataset to further examine the effectiveness of SAPD in diagnosing participants with low response style bias (specifically, disacquiescence). As illustrated in Figure 8, the radar chart on the left compares the estimated scores in five social-emotional traits. The gray region denotes the ground-truth provided by expert, while the red and blue lines represent the predictions from SAPD and PCDF adapted to NCDM (NCDM-PCDF), respectively. SAPD demonstrates consistently closer alignment with expert ratings across all traits, outperforming PCDF, which relies solely on data recoding. The bar plot on the right quantifies the deviation from expert scores for each model, clearly showing that SAPD yields the smallest prediction error. These results confirm SAPD's superior accuracy and its robustness in correcting response-style bias for more trustworthy trait estimation.