We thank the reviewers for the constructive comments and suggestions. Our responses are detailed in the following.

Reviewer 1: We will add the interpretations of the latent domains. The first domain is characterized by good mood, less guilt feeling, and better functional and physical status; the second domain is mainly depicted by less anxiety and better mood; the third domain is characterized by better appetite and less weight loss; higher value in the fourth domain indicates higher severity in insomnia and distress. We will add more discussions in broader impact, e.g., potential applications to observation studies, behavioral sciences and econometrics to learn effective personalized interventions.

Reviewer 2: In this work, we demonstrated the proposed method with randomized trials; however, our method can be easily modified and applied to observational studies under the assumption of no unmeasured confounding. This assumption, also known as ignorability, is commonly assumed in learning heterogeneous treatment effects for observational studies, and it ensures the identifiability of causal effects conditional on observed confounding variables $\mathbf{X}$ and $\mathbf{Y}_0$. Since $\mathbf{X}$ and $\mathbf{Y}_0$ are included and adjusted in our model, their confounding effects can be removed. Additionally as suggested, matching or weighting using lower-dimensional learned latent $\mathbf{Z}_0$ can better remove confounding when $\mathbf{Y}_0$ is high-dimensional. Thus, our algorithm allows more efficient matching due to the reduced dimensionality in feature variables by using the lower-dimensional latent variables.

To our best knowledge, our method is the first to incorporate multi-domain outcomes as opposed to a single scalar outcome in learning individualized treatment rules (ITRs), and we combine measurement models and ITR learning under an unified loss where the latent states and ITRs are simultaneously learned with an iterative procedure. The proposed model is very general; it can be applied to other areas where reward outcome is latent and needs to be inferred from many measurements, e.g., behavioral sciences and consumer preference in marketing for effective interventions. Additionally as suggested, we will add discussions in related work on latent variable models, e.g., hidden Markov models. Since they only estimate latent states but do not consider learning treatment rules, one straightforward approach is to conduct a two-step procedure. But the two-step procedure will not be as efficient as our approach which borrows strength from data before and after treatment with shared representation parameters while preserving the consistency of the latent constructs.

Reviewer 3: The choice of the predefined $g()$ function of the latent states may depend on application but a common choice is the total sum similar to aggregating sub-scales of instruments commonly used in psychiatry [3]. We described in Section 4.2 (starting from line 283 and Supplementary material Section C) about how to ensure $g(\mathbf{Z}_1)$ is meaningful to optimize. In summary, based on the loading parameters, the latent constructs are interpretable and we scored each latent domain so that their directions are aligned, which ensures that their sum score is meaningful to optimize. To achieve identifiability of the latent variables, we have controlled their directions and exchangeability in model fitting (Section 4.1 starting from line 225 and Supplementary Section A). Essentially, for each latent construct we set initial values of the loading parameters for one observed item, which is analogous to fixing one loading path per latent variable in factor analysis. Identifiability of the latent constructs was achieved under this value initialization: loading parameters in $P(\mathbf{Y}_0 \mid \mathbf{Z}_0)$ were recovered and we obtained a prediction accuracy of 100% for the latent $\mathbf{Z}_0$ under training sample size of 1000 and 2000 (described in Section 4.1 Simulation results, in the last paragraph).

The validity of the latent constructs is important but not a serious concern in our application because the instrument (HAMD) we used to infer latent states have already been validated, shown to exhibit adequate psychometric properties [2] and widely used in clinical settings. The latent constructs are the lower-dimensional representations of these instruments. We have empirically shown that our latent constructs lead to improved value function evaluated by other external outcomes not used in training (Table 2). In addition, we consulted psychiatrists on the interpretations of the latent constructs and will discuss validation on external datasets in the paper. For the broader impact, it is possible that the latent constructs from poor instruments can have certain bias in terms of their correlations with racial or social-economic factors. However, the risk of such bias is not unique to our method, and our application uses carefully chosen instruments to minimize potential bias. As suggested, we will discuss this point in the Broader impact.

For related work, we will add relevant methods from psychometrics as suggested. In fact, although our model for $P(\mathbf{Y}_0|\mathbf{Z}_0)$ and $P(\mathbf{Y}_1|\mathbf{Z}_1)$ is similar to measurement models with latent constructs, our method is fundamentally different from the traditional two-step procedure where latent constructs are estimated first and then ITRs based on them are identified. In contrast, we adopt a unified approach where the latent constructs and ITRs are simultaneously learned. This is more efficient and borrows strength from data before and after treatment with shared representation parameters which preserves the validity of the latent constructs. About clarity, since our method is a simultaneous approach, we introduced the key concepts for ITR first and then break down to each component of the model. We will add a paragraph to guide reading. For reproducibility, we have submitted the codes for reproducing all experiments as Supplementary and all necessary details for reproducing the results are presented in the paper and Supplementary.

Reviewer 4: 808 patients from STAR*D were included (Section 4.2 under Data and model). We will provide more information about data distribution, e.g., mean age is 43, with 59% females. As suggested, we will discuss validating on different psychiatric disorders as future work. Since measurement models have been successfully applied to other mental disorders [1], we expect our method to perform well on them. Abbreviations and short descriptions for the performance scores are provided in Section 4.2 under Data and model. We will include them in Figure 3 for the ease of reading.

[1] E. S. Barratt. Factor analysis of some psychometric measures of impulsiveness and anxiety. *Psychological reports*, 16(2):547–554, 1965.

[2] G. A. Fava, R. Kellner, F. Munari, and L. Pavan. The hamilton depression rating scale in normals and depressives. *Acta Psychiatrica Scandinavica*, 66(1):26–32, 1982.

[3] N. Rose, W. Wagner, A. Mayer, and B. Nagengast. Model-based manifest and latent composite scores in structural equation models. *Collabra: Psychology*, 5(1), 2019.