Supplementary Document for
"What Can We Learn from a Semiparametric Factor Analysis of Item Responses and Response Time? An Illustration with the PISA 2015 Data"

Contents

A Parametric Fittings

A.1 Baseline Model

The baseline parametric model features simple linear normal factor analysis models for log-transformed response time (RT) variables, standard item response theory (IRT) models for item responses, and a bivariate normal density for the latent slowness and ability. Given the latent slowness X_{i1}[1], the log-RT variable $Y_{ij} \in \mathbb{R}$, $j = 1, \dots, m_1$, is assumed to be normally distributed. The conditional density of $Y_{ij} = y$ given $X_{i1} = x$ is denoted

$$f_j(y|x) = \phi(y; \mu_j(x), \sigma_j^2), \tag{S1}$$

in which $\phi(\cdot; \mu, \sigma)$ is a generic notation for the density of $\mathcal{N}(\mu, \sigma^2)$, and the mean function

$$\mu_j(x) = \zeta_j + \gamma_j x \tag{S2}$$

is assumed to be linear. There are three free parameters for each continuous MV: the intercept ζ_j, the common factor loading γ_j, and the unique variance σ_j^2. Given the latent ability X_{i2}, the item response function (IRF) for a discrete item response $Y_{ij} \in \{0, \dots, C_j - 1\}$, $j = m_1 + 1, \dots, m$, can be expressed as

$$\mathbb{P}\{Y_{ij} = y | X_{i2} = x\} = \begin{cases} \dfrac{1}{1 + \sum_{c=1}^{C_j - 1} \exp(\zeta_{jc} + \gamma_{jc} x)}, & y = 0, \\ \dfrac{\exp(\zeta_{jy} + \gamma_{jy} x)}{1 + \sum_{c=1}^{C_j - 1} \exp(\zeta_{jc} + \gamma_{jc} x)}, & y = 1, \dots, C_j - 1, \end{cases} \tag{S3}$$

in which ζ_{jc} and γ_{jc}, $c = 1, \dots, C_j - 1$, are referred to as the intercept and slope parameters. When $K = 2$, Equation S3 reduces to the IRF of the two-parameter logistic (2PL) model, which is specified for dichotomous responses in the baseline model. Finally, the joint density for the two latent variables (LVs) $(X_{i1}, X_{i2})^\top$ is assumed to be bivariate normal with density

$$h(x_1, x_2) = \phi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \\ \sigma_{21} & 1 \end{bmatrix} \right), \tag{S4}$$

in which $\phi(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for a bivariate normal density with a mean vector $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$. Both X_{i1} and X_{i2} are marginally distributed as $\mathcal{N}(0, 1)$; therefore, the conditional MV densities in the baseline model are directly comparable with those in the semiparametric model.

[1] We adopt the same notational convention as we used in the main document.

A.2 Updated Model

The semiparametric fitting reported in the main document indicates that a more flexible model is needed to capture the complex dependencies among the observed item responses and RT. Therefore, we proceed to modify the baseline model as follows. For the log-RT variables, we specify a nonlinear normal factor model: While the conditional distribution $Y_{ij}|X_{i1}$ is still characterized by Equation S1, the mean function is now a quintic polynomial of form

$$\mu_j(x) = \zeta_j + \gamma_{j1}x + \gamma_{j2}x^2 + \gamma_{j3}x^3 + \gamma_{j4}x^4 + \gamma_{j5}x^5. \tag{S5}$$

To match the semiparametric setup, we require the mean log-RT function of item CM442Q02 to be monotonically increasing. This is achieved by imposing positivity on the derivative of Equation S5 via the parameterization

$$\mu_j'(x) = \omega_j \left[1 - 2\upsilon_{j1}x + (\upsilon_{j1}^2 + \tau_{j1})x^2\right] \left[1 - 2\upsilon_{j2}x + (\upsilon_{j2}^2 + \tau_{j2})x^2\right], \tag{S6}$$

in which $\omega_j, \tau_{j1}, \tau_{j2} > 0$ (Elphinstone, 1985; Falk & Cai, 2016). The corresponding quintic polynomial coefficients can be obtained by straightforward algebra. While testlet responses are still modeled by Equation S3 in the updated model, dichotomous item responses are now modeled by the four-parameter logistic (4PL) model with IRF

$$\mathbb{P}\{Y_{ij} = 1|X_{i2} = x\} = \varpi_j + \frac{(\varrho_j - \varpi_j)\exp(\zeta_j + \gamma_j x)}{1 + \exp(\zeta_j + \gamma_j x)}, \tag{S7}$$

in which $\varpi_j, \varrho_j \in [0, 1]$, $\varpi_j < \varrho_j$, are the lower- and upper-asymptote parameters. Finally, the latent slowness and ability are assumed to jointly follow a mixture of two independent bivariate normal distributions:

$$h(x_1, x_2) = \pi\phi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}, \begin{bmatrix} 0.5^2 & \\ \sigma_{21}^{(1)} & 0.75^2 \end{bmatrix}\right) + (1-\pi)\phi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \begin{bmatrix} \mu_1^{(2)} \\ \mu_2^{(2)} \end{bmatrix}, \begin{bmatrix} \sigma_{11}^{(2)} & \\ \sigma_{21}^{(2)} & \sigma_{22}^{(2)} \end{bmatrix}\right). \tag{S8}$$

In Equation S8, the free parameters are the mean and (co)variance parameters $\mu_1^{(2)}$, $\mu_2^{(2)}$, $\sigma_{11}^{(2)}$, $\sigma_{21}^{(2)}$, and $\sigma_{22}^{(2)}$ for the second latent class, the covariance parameter $\sigma_{21}^{(1)}$ for the first latent class, and the class membership probability π. The means and variances for the first latent class are arbitrarily set in order to identify the LV scale. Note that the joint density characterized by Equation S8 does not have $\mathcal{N}(0, 1)$ marginals. To facilitate comparison with the semiparametric fitting, let

$$X_{id}^* = \Phi^{-1}(H_d(X_{id})),\ d = 1, 2, \tag{S9}$$

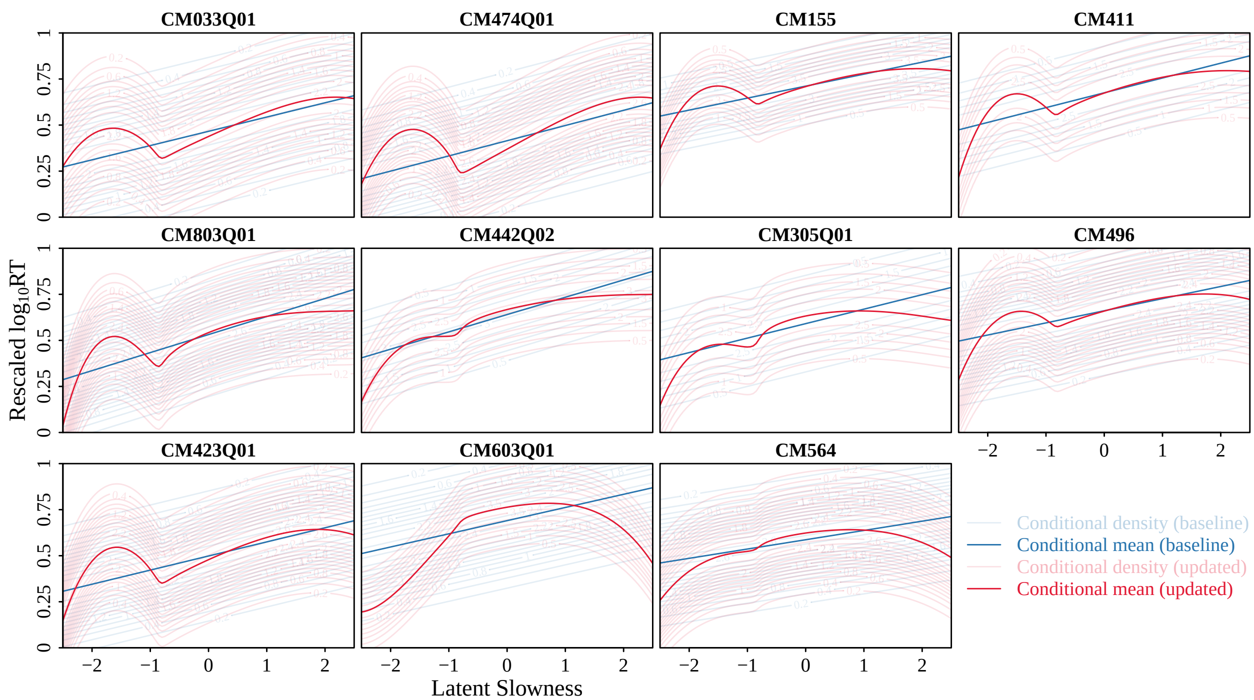

Figure S1: Estimated conditional densities and means for log-10 response time (RT) variables (rescaled to [0, 1]) from the two additional parametric models. Each panel corresponds to a single item/testlet. Conditional densities of manifest variables given the slowness factor are visualized as contours in lighter colors. Estimated conditional means for both the are superimposed as solid curves in darker colors.

in which H_d stands for the dth marginal distribution function under the joint density (Equation S8). By the probability integral transform, both X_{i1}^* and X_{i2}^* are $\mathcal{N}(0, 1)$ variates. In the sequel, all the density estimates based on the updated model are plotted for the standard normal X_{i1}^* and X_{i2}^* unless otherwise specified.

A.3 Results

Parameters of the two parametric models were estimated using an R implementation of the expectation-maximization (EM) algorithm. The intractable marginal likelihood functions were approximated by an outer-product rectangular quadrature, in which there are 21 equally spaced points ranging from -5 to 5 per dimension. The EM algorithm was terminated when the log-likelihood change between consecutive iterations is less than 0.001.

We performed five-fold cross-validation and computed the empirical risk defined in a fashion similar to Equation 23. The empirical risk for the baseline model is 5.17 with a standard error (SE) of 0.16; the empirical risk for the updated model is 4.77 with an SE of 0.15. Compared to the empirical risk of the semiparametric model, i.e., 4.06 with an SE of 0.03 as shown in the last panel of Figure 1, the empirical risks for the two parametric models are substantially higher and thus imply poorer fit.

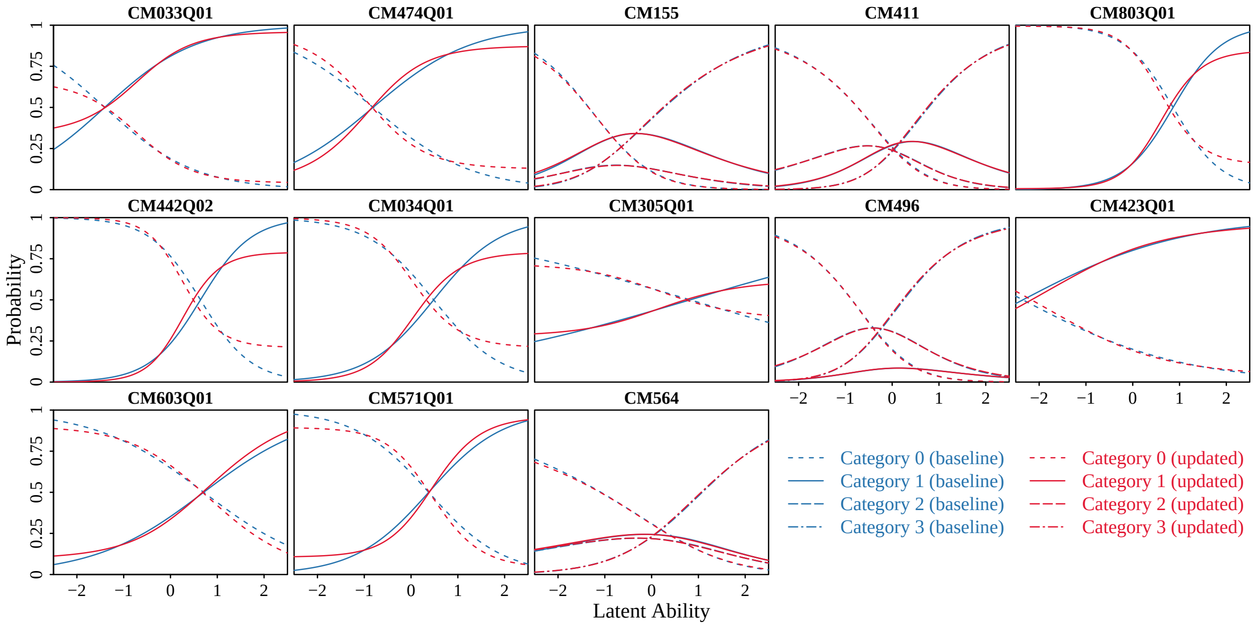

Figure S2: Estimated conditional densities for discrete response variables, also known as item response functions (IRFs), from the two additional parametric models. Each panel corresponds to a single item/testlet. Curves for different categories are shown in different line types. The two models are distinguished by different colors.

The estimated conditional densities and conditional mean functions for the log-RT variables are depicted in Figure S1. On the one hand, the linear estimates (blue) of the conditional mean functions are quite different from the (transformed) quintic estimates (red) and the semiparametric fitting (Figure 3 in the main document), which contributes to the poor model-data fit of the baseline model. On the other hand, the (transformed) quintic and the semiparametric fittings are more or less aligned in their basic shape. The funky shapes of the quintic fittings around $X_{i1}^* = -1$ are consequences of the LV transformation (Equation S9).

The estimated IRFs for item/testlet responses are displayed in Figure S2. For all the testlets and a majority of dichotomous items, there is little discrepancy between the two parametric fittings, which also closely resemble the semiparametric fitting (Figure 4 in the main document). Exceptions include the non-trivial upper asymptotes obtained for items CM803Q01, CM442Q02, and CM031Q01, and the non-trivial lower asymptote for item CM033Q01.

In Figure S3, the estimated bivariate LV densities from the two parametric models are displayed and contrasted. In the baseline model, the latent slowness and ability are forced to follow a bivariate normal distribution. We therefore obtain a nearly zero inter-factor correlation (about 0.06) and thus a very weak η^2 (less than 0.01). The

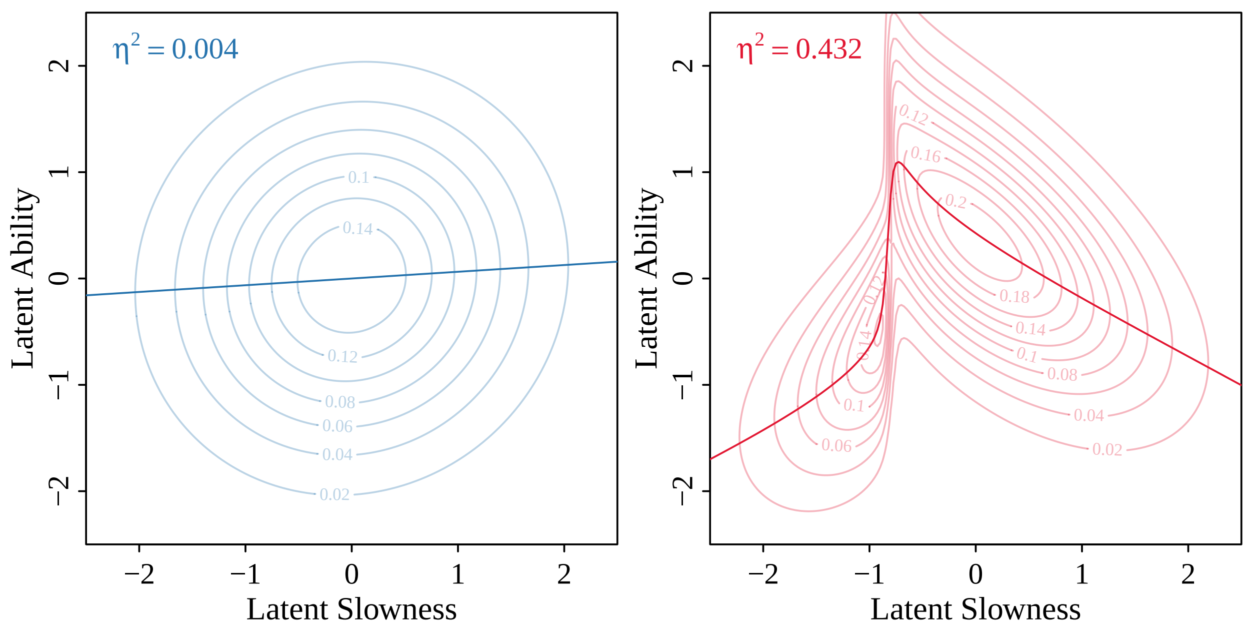

Figure S3: Estimated joint density for the slowness and ability factors (contours in lighter colors) and the conditional mean of ability given slowness (solid curves in darker colors). The population η^2s for predicting latent ability by latent slowness are printed at the upper-left corner of each panel. Left: Baseline model. Right: Updated model.

two-component mixture density in the updated model allows us to capture the nonlinear relationship between the latent slowness and ability. The resulting η^2 value for the updated model (0.43) is also close to what we have obtained in the semiparametric fitting (0.45), indicating that the latent ability can be effectively predicted by the latent slowness. Nevertheless, it is noticed that the estimated bivariate LV density in the updated model, after being transformed to have standard normal marginals, still differs from the estimate in the semiparametric fitting (Figure 5). This discrepancy is likely resulted from the limited flexibility of the two-component mixture density.

B Relationship Between Log-RT and Ability Based on the Semiparametric Model

In Figure S4, the item-level log-RT variables are plotted against the expected *a posteriori* (EAP) scores of the ability factor computed from the semiparametric simple-structure model. Inverted U-shape relationships are observed for a number of items, and monotonic relationships are observed for the rest. However, the association between the log-RT and the ability scores is in general weak.

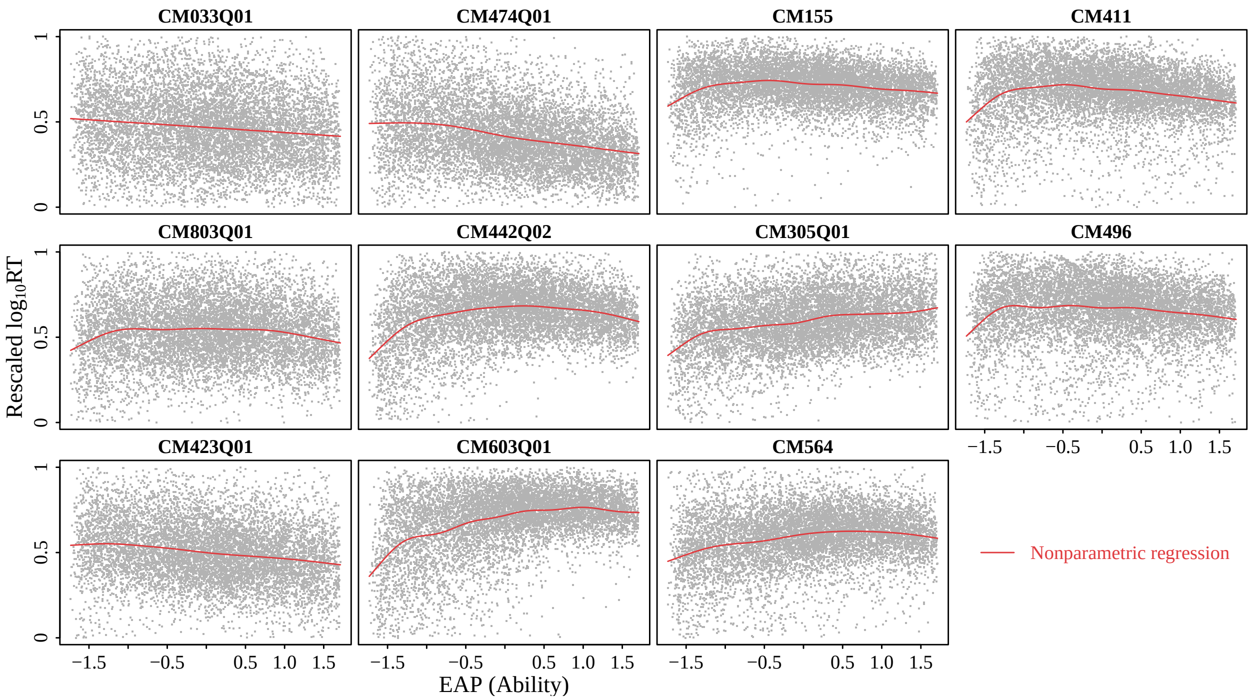

Figure S4: Rescaled log-transformed response time (RT) plotted as functions of expected *a posteriori* (EAP) scores for ability. Each panel corresponds to a single item/testlet. Smoothing spline regression lines are superimposed to visualize the trend.

It is also possible to describe the relationship between an item-level log-RT variable and the latent ability by a predictive distribution and the corresponding mean function. In the simple-structure model, the log-RT variables Y_{ij}, $j = 1, \dots, m_1$, are not directly related to the latent ability X^*_{i2}. But because we can predict the latent slowness X^*_{i1} by X^*_{i2} and X^*_{i1} is indicated by the log-RT variables, we can examine the predictive distribution of $Y_{ij} = y$ given $X^*_{i2} = x^*_2$ characterized by the density

$$p_j(y|x^*_2) = \int_{\mathbb{R}} f_j(y|x^*_1)h(x^*_1|x^*_2)dx^*_1, \tag{S10}$$

in which $h(x^*_1|x^*_2) = h(x^*_1, x^*_2)/h(x^*_2)$ is the conditional density of X^*_{i2} given X^*_{i1}. The predictive mean of Y_{ij} at $X_{i2}^* = x_2^*$ is then

$$\mathbb{E}(Y_{ij}|X_{i2}^* = x_2^*) = \int y p_j(y|x_2^*) dy. \tag{S11}$$

We plot Equations S10 and S11 in Figure S5. Both Figures S4 and S5 suggest that item-level log-RT variables cannot be reliably predicted by the latent ability.

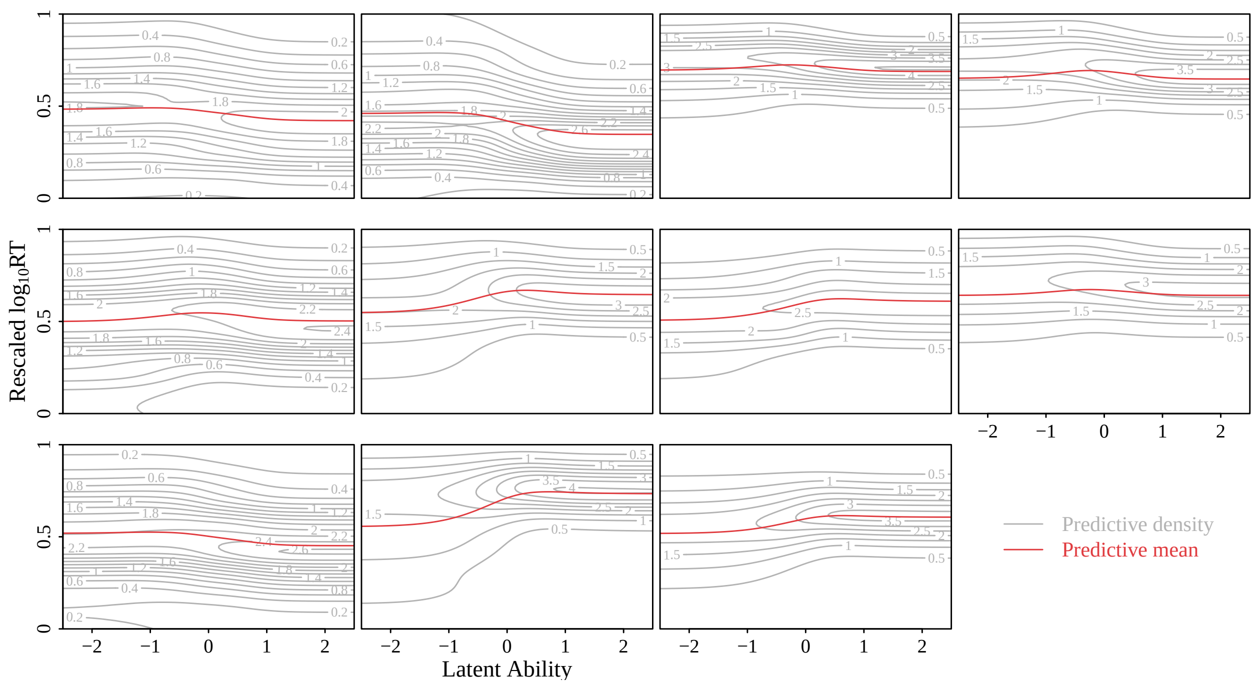

Figure S5: Predictive densities (gray contours) and mean functions (red solid lines) for log-transformed response time (RT) by latent ability. Each panel corresponds to a single item/testlet.

References

Elphinstone, C. D. (1985). *A method of distribution and density estimation* (Unpublished doctoral dissertation). University of South Africa.

Falk, C. F., & Cai, L. (2016). Maximum marginal likelihood estimation of a monotonic polynomial generalized partial credit model with applications to multiple group analysis. *Psychometrika*, *81*(2), 434–460.