

# Bayesian Factor Mixture Modeling with Response Time for Detecting Careless Respondents

Lijin Zhang<sup>1</sup>, Esther Ulitzsch<sup>2</sup>, Ben Domingue<sup>1</sup>

<sup>1</sup> Graduate School of Education, Stanford University

<sup>2</sup> Centre for Educational Measurement, University of Oslo

April 25, 2024



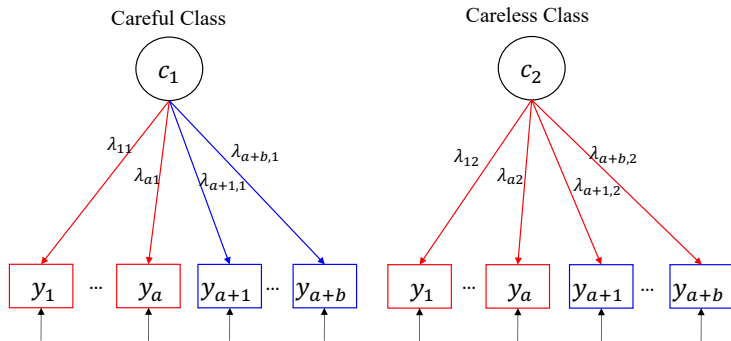
- ▶ Scales have been extensively used to investigate latent variables in social science research.
- ▶ The effectiveness of survey data depends on the assumption that responses accurately represent the latent constructs being studied.

## Careless Respondents

- ▶ Those whose answers are not the result of careful thought but rather lack of attention, misunderstanding, or lack of interest (Arias et al., 2020).
- ▶ Research into scales has revealed a widespread occurrence of careless behaviors, with reported rates varying from 3% to 50% (Meade & Craig, 2012).



- ▶ Careless responses are problematic because they are relatively unrelated to the relevant constructs.
- ▶ Biased findings. For instance, reduced correlation between variables (Kam and Meyer., 2015).
- ▶ Poor model fitting in CFA (Voss, 2023; Woods, 2006).

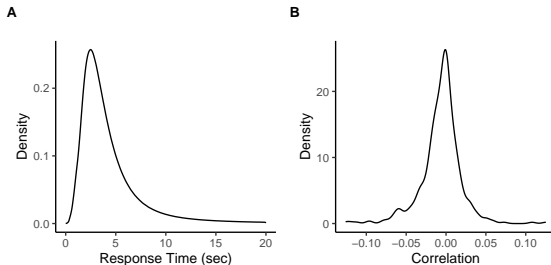


Factor Mixture Model based on Item Wording

Note: *Red/Blue blocks: Positively/Negatively worded items;*  
*Red/Blue lines: Positive/Negative loadings.*



- ▶ The duration required to respond to a question on a psychological scale tends to be short.
- ▶ Correlation between Time and Responses is around 0.



Time of 53,671,504 survey responses from IRW (Domingue et al., 2023)



- ▶ Despite the prevalent usage of CFA for scale data analysis, the integration of response time in CFA remains underexplored.
- ▶ Lack of comparative studies on models with and without response time



We seek to fill these gaps by incorporating response time metrics within the CFA framework to detect careless respondents.

- ▶ Simulation studies to demonstrate the effectiveness of the proposed model
- ▶ Simulation studies to illustrate the benefits of modeling response time
- ▶ Empirical study to show the application of the proposed model



For item  $j \in \{1, \dots, J\}$  and respondent  $i \in \{1, \dots, N\}$ , the model is formulated as follows:

Careful Group:

$$\begin{aligned}\log(t_{ij}^a) &= \beta_j - \tau_i + \gamma_{ij} \\ y_{ij}^a &= \mu_j + \lambda_j * \omega_i + \epsilon_{ij}\end{aligned}\tag{1}$$

Careless Group:

$$\begin{aligned}\log(t_{ij}^b) &\sim \mathcal{N}(\mu_t^b, \sigma_t^{b2}) \\ y_{ij}^b &\sim \mathcal{N}(\mu_y^b, \sigma_y^{b2})\end{aligned}\tag{2}$$

$p_i$ : probability that  $i$ -th respondent belongs to careful group:

$$\begin{aligned}\log(t_{ij}) &= p_i \cdot \log(t_{ij}^a) + (1 - p_i) \\ y_{ij} &= p_i \cdot y_{ij}^a + (1 - p_i) \cdot y_{ij}^b\end{aligned}\tag{3}$$





- ▶ For the careless group, both item responses and time are independent of the item/person characteristics, maintaining homogeneity of intercept ( $\mu_y^b$  and  $\mu_t^b$ ) and variances.
- ▶ Respondents who diligently read the item content and respond to the questions are likely to spend more time than those who exhibit less attention to details:  $\beta_j > \mu_t^b$ .



- ▶ Proportions of careless respondents ( $\pi$ ): 0.05, 0.1, 0.15, 0.2.
- ▶ Sample sizes: 300, 500, and 1000.
- ▶ Differences in log-RT between the careful and careless groups were controlled.

## Responses

- ▶ Careful: A single-factor model with ten items ( $\lambda_j = 0.8$ )
- ▶ Careless:  $y_{ij}^b \sim N(1, 2)$

## Response Time

- ▶ Careful: A log-normal model with intercept following  $U(0.95, 1.2)$  or  $U(1.2, 1.45)$ , and residual variance  $\psi_{\gamma,j}$  at 1 or 1.25.
- ▶ Careless:  $\log(t_{ij}^b) \sim N(0.7, 0.5)$ .



The model convergence rates are 100% for all conditions.

Table 3: Classification Accuracy of the Proposed Model.

	$N$	$\psi_{\gamma,j}$	$\pi = 0.05$	$\pi = 0.1$	$\pi = 0.15$	$\pi = 0.2$
Accuracy	500	1	0.92	0.97	0.98	0.98
Sensitivity			0.92	0.97	0.98	0.98
Precision			1.00	1.00	1.00	1.00
FPR			0.01	0.01	0.01	0.01
FNR			0.08	0.03	0.02	0.02
Accuracy	500	1.25	0.93	0.98	0.98	0.98
Sensitivity			0.92	0.98	0.98	0.98
Precision			1.00	1.00	1.00	1.00
FPR			0.01	0.01	0.01	0.01
FNR			0.08	0.02	0.02	0.02

*Note:*  $N$  denotes the sample size;  $\psi_{\gamma,j}$  indicates the residual variance of response time in the careful group; FPR = False Positive Rate; FNR = False Negative Rate.



We compared the proposed model with the model without response time:

$$\begin{aligned}y_{ij}^a &= \mu_j + \lambda_j * \omega_i + \epsilon_{ij} \\ y_{ij}^b &\sim N(\mu_y^b, \sigma_y^{b2}) \\ y_{ij} &= p_i \cdot y_{ij}^a + (1 - p_i) \cdot y_{ij}^b\end{aligned}\tag{4}$$

# Model without Response Time



A significant influence observed upon removing response time is the difficulty in achieving model convergence.

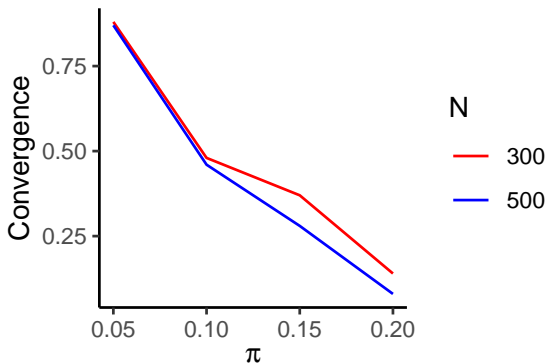




Table: Classification Accuracy of the Model without Response Time

$\pi$	0.05		0.1	
$\beta$	U(0.95-1.2)	U(1.2-1.45)	U(0.95-1.2)	U(1.2-1.45)
Accuracy	0.84	0.84	0.93	0.93
Sensitivity	0.84	0.84	0.92	0.92
Precision	1.00	1.00	1.00	1.00
FPR	0.02	0.02	0.02	0.02
FNR	0.16	0.16	0.08	0.08



Data came from the Item Response Warehouse (Domingue et al., 2023).

We selected 1000 samples from the emotional stability dimension of the Big-Five personality scale (Goldberg, 1992).

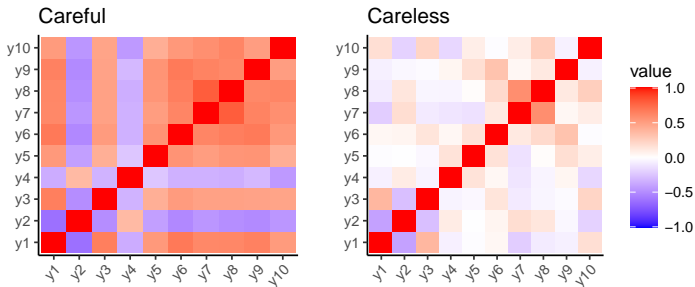
Five-point Likert scale with 2 positively worded items, 8 negatively worded items.

A CFA model with one factor was conducted using the lavaan package, revealing that the single-factor model did not adequately fit the data, as indicated by the fit indices (CFI = 0.922, TLI = 0.896, RMSEA = 0.098, SRMR = 0.046).



The proposed model identified 261 respondents as careless, characterized by a very weak correlation among all items.

Upon removal of these careless respondents, the fit of the CFA model to the data improved significantly (CFI = 0.957, TLI = 0.943, RMSEA = 0.084, SRMR = 0.035).

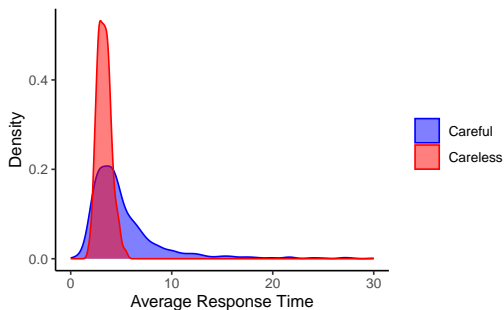






Respondents classified as careless took less time per item (3.31 seconds vs 6.46 seconds).

Time taken to respond by the careless group showed minimal fluctuation across ten items, with an average standard deviation (SD) of 1.51, compared to 7.96 in the careful group.





- ▶ We integrate response time in factor analysis to detect careless respondents by factor mixture modeling.
- ▶ It is independent of item wording, applicable to scales that do not include reverse-worded items.
- ▶ Simulation and empirical studies demonstrated its effectiveness and the benefits of modeling response time.



Thanks for Listening!

Slides:

[https://lijinzhang.com/share/240425\\_careless.pdf](https://lijinzhang.com/share/240425_careless.pdf)