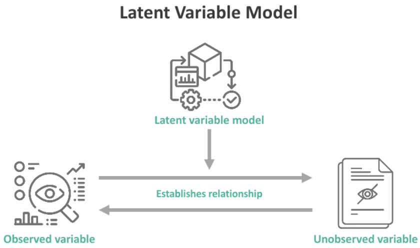
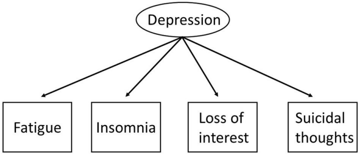
Bayesian Factor Mixture Modeling with Response Time for Detecting Careless Respondents

Lijin Zhang[1], Esther Ulitzsch[2,3], Ben Domingue[1]

[1] Graduate School of Education, Stanford University
[2] Centre for Educational Measurement, University of Oslo
[3] Centre of Research on Equality in Education, University of Oslo

- In psychology and education, many important traits—like anxiety or motivation—can't be directly observed.
- Scales have been extensively used to investigate latent variables in social science research.
- Latent variable models, such as confirmatory factor analysis (CFA), are commonly used to study latent traits.





Careless Respondents

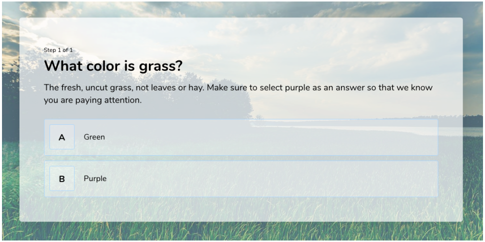
The effectiveness of survey data depends on the assumption that responses accurately represent the latent constructs.

- 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).

Influences of Careless Respondents

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

Attention Check Items



1. After I attend a party, I need to spend a lot of time alone to restore my energy levels

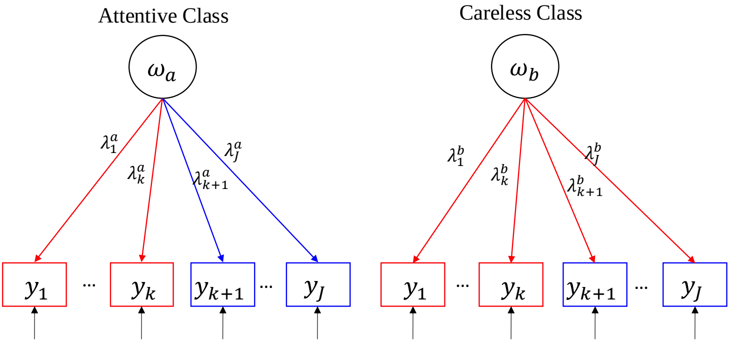
Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
5	4	3	2	1

2. I feel energized when I attend large social gatherings

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
1	2	3	4	5

Item Wording

Factor Mixture Modeling

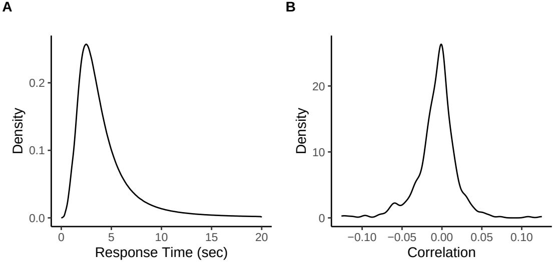


Factor Mixture Model based on Item Wording

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

Response Time

- 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., 2024; https://datapages.github.io/irw/)

Goals

- Despite the prevalent usage of Confirmatory Factor Analysis (CFA) for scale data analysis, the integration of response time in CFA remains underexplored.
- We seeks to fill this gap by incorporating response time metrics within the CFA framework to detect careless respondents.

Assmuptions

- Attentive respondents' answers are influenced by both item and person characteristics. In contrast, for careless respondents, item responses and response times are independent of these characteristics.
- 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.

The Proposed Model

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

Attentive Group:

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

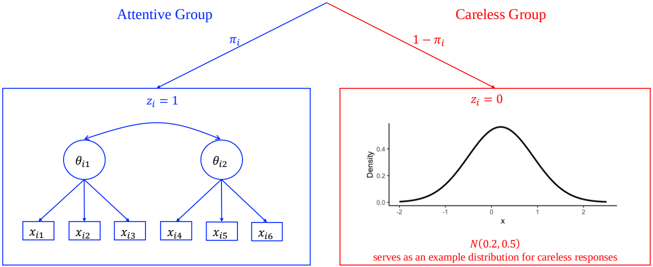
Careless Group:

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

π_i: probability that i-th respondent belongs to attentive group:

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

Idea



- Identifying specific types of careless behavior is not the focus of this study.
- It serves as a minimally structured residual class, capturing any systematic deviation from attentive responding.

Simulation Study: Goals

Assess the effectiveness of the proposed model

- Examine the convergence, classification precision, and estimation accuracy of the proposed model

Simulation Design

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

Model Estimation

$$
\begin{aligned}
(\pi_i, 1-\pi_i) &\sim Dir(1,1) \\
\mu_j, \mu_y^b, \beta_j, \mu_t^b &\sim N(0,100) \\
\psi_{\epsilon,j}, \sigma_y^{b2}, \phi_\tau, \psi_{\gamma,j}, \sigma_t^{b2} &\sim Inv-Gamma(0.01, 0.01) \\
\delta_j, \lambda_j &\sim TN(0,100,0,)
\end{aligned} \tag{4}
$$

- JAGS (Plummer, 2004).
- Burn-in iterations: 10,000 - 100,000.
- Model convergence is assessed using the estimated potential scale reduction (EPSR) index (Gelman, 1996): EPSR < 1.1.
- Upon convergence, we generate additional 20,000 MCMC samples for model estimation.

Evaluation Criteria

- Model Convergence
- Estimation Accuracy
  - Relative Bias; Root Mean Square Error (RMSE)
- Classification Accuracy

		Estimated		
		Attentive	Careless	
True	Attentive	TP	FN	$\text{FNR} = \frac{FN}{TP+FN}$
	Careless	FP	TN	$\text{FPR} = \frac{FP}{FP+TN}$
		$\text{Precision} = \frac{TP}{TP+FP}$		$\text{Accuracy} = \frac{TP+TN}{N}$

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		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.

Estimation Accuracy

Table 4: Estimation Results of the Proposed Model.

π	0.05				0.1				0.15				0.2			
β_j	U(0.95-1.2)		U(1.2-1.45)		U(0.95-1.2)		U(1.2-1.45)		U(0.95-1.2)		U(1.2-1.45)		U(0.95-1.2)		U(1.2-1.45)	
	RB(%)	RMSE	RB	RMSE	RB	RMSE	RB	RMSE	RB	RMSE	RB	RMSE	RB	RMSE	RB	RMSE
β_j	3.59	0.07	3.89	0.08	2.54	0.06	2.55	0.07	1.29	0.07	1.51	0.07	1.18	0.06	1.37	0.06
λ_j	5.78	0.06	5.17	0.06	2.21	0.05	1.98	0.05	1.92	0.05	1.83	0.05	2.05	0.05	1.99	0.05
μ_{j1}	**-10.51**	0.08	-9.06	0.07	-5.20	0.06	-4.55	0.06	-3.57	0.05	-3.20	0.05	-1.23	0.05	-0.91	0.05
ϕ_τ	**10.72**	0.07	5.78	0.05	3.64	0.05	1.15	0.05	3.97	0.05	2.11	0.05	1.62	0.05	0.04	0.04
$\psi_{\epsilon,j}$	-4.54	0.03	-3.95	0.03	-2.26	0.03	-1.78	0.03	-1.71	0.03	-1.43	0.03	-1.19	0.03	-0.94	0.03
$\psi_{\gamma,j}$	1.84	0.08	1.72	0.08	2.14	0.08	1.95	0.08	1.72	0.08	1.59	0.08	1.51	0.08	1.40	0.08
μ_y^b	-9.39	0.14	-8.87	0.14	-1.18	0.06	-0.82	0.06	-1.22	0.06	-1.12	0.06	-0.77	0.05	-0.78	0.05
σ_y^{b2}	**-47.43**	0.96	**-44.42**	0.91	**-18.21**	0.40	**-16.14**	0.36	-9.65	0.23	-8.72	0.21	-6.05	0.16	-5.46	0.15
μ_t^b	8.96	0.10	**14.73**	0.14	0.97	0.04	1.84	0.04	0.38	0.03	0.83	0.03	0.57	0.03	0.67	0.03
σ_t^{b2}	**61.49**	0.32	**57.79**	0.30	**14.02**	0.08	**12.11**	0.08	6.98	0.05	6.52	0.05	4.04	0.03	3.74	0.03

Other Simulation Studies

- Overly consistent careless behaviors
- Benefits of incorporating response time
- Benefits of accounting for careless responses in mediation analysis

Empirical Study

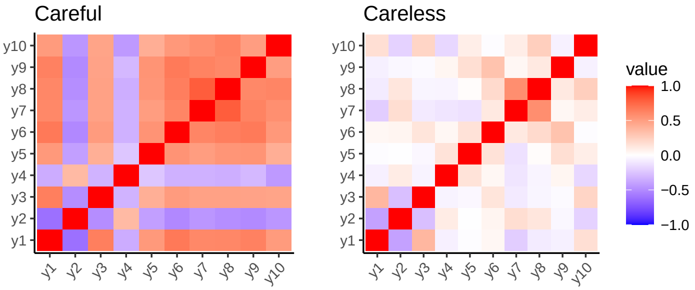
- Emotional Stability Scale
- $N = 1000$
- Number of Items: 10 (8 negatively worded items)
- Five-point Likert Scale

The single-factor model did not adequately fit the data (BCFI = 0.921, BTLI = 0.882, BNFI = 0.918).

Results

The proposed model identified 261 respondents as careless.

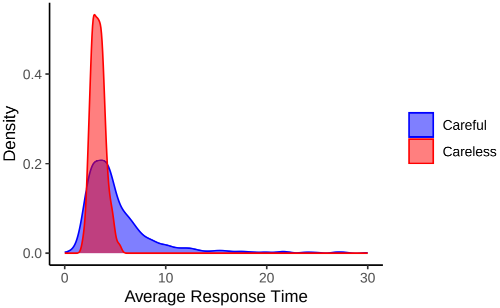
Upon removal of these careless respondents, the fit of the CFA model to the data improved significantly (BCFI $= 0.955$, BTLI $= 0.930$, BNFI $= 0.952$).



Results

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 attentive group.



Summary

- We integrate response time into 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.
- Extension: Careless Respondents Detection for Visual Analogue Scale Data (Zhang, Domingue, Vogelsmeier, Ulitzsch, 2024).

Thank you!

Slides:
lijinzhang.com/share/250416_fmm.pdf

Preprint: