A Beta Mixture Model for Careless Respondent Detection in Visual Analogue Scale Data

Lijin Zhang[1], Ben Domingue[1],

Leonie V.D.E. Vogelsmeier[2], Esther Ulitzsch[3,4]

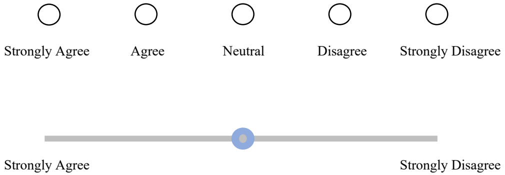
[1] Graduate School of Education, Stanford University
[2] Department of Methodology, Tilburg University
[3] Centre for Educational Measurement, University of Oslo
[4] Centre of Research on Equality in Education, University of Oslo

Introduction

Self-report scales for measuring latent constructs.

- Visual analogue scale, or slider scale, allowing responses along a continuous line.
- High sensitivity to variations in the factor, allowing for the detection of subtle shifts in perceptions or emotional states

Example: Are you inclined to keep in the background on social occasions?



Likert Scale and Visual Analogue Scale

Careless Respondents

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

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).

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 latent variable modeling (Voss, 2023; Woods, 2006).

Aim

Despite increasing use, psychometric research on VASs has lagged behind.

- Aim: Provide an approach that is tailored to detect careless responding in VAS data

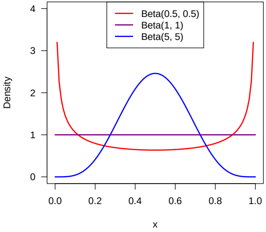
We integrate existing measurement models for VASs (Noel & Dauvier, 2007) with mixture IRT or factor models for identifying and modeling careless responding.

Beta Item Response Model

VAS Data

- Converted to values between 0 and 1 (proportion of distance on the scale)
- Continuous but bounded $\rightarrow$ traditional CFA unsuitable

Beta IRM (Noel & Dauvier, 2007): $X_{ij} \sim \text{Beta}\left(\alpha_{ij}, \beta_{ij}\right)$



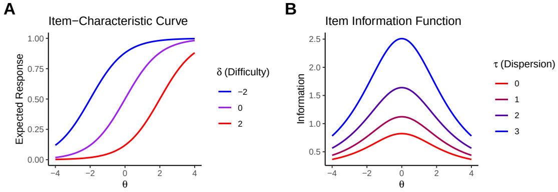
Beta Item Response Model

α_{ij} tendency toward high (near 1) β_{ij} tendency toward low

$$
\begin{aligned}
\alpha_{ij} &= \exp\left(\frac{w_j\theta_{id[j]} - \delta_j + \tau_j}{2}\right) \\
\beta_{ij} &= \exp\left(\frac{-(w_j\theta_{id[j]} - \delta_j) + \tau_j}{2}\right)
\end{aligned}
\tag{1}
$$

- latent trait $\theta_{id[j]}$
- item wording w_j
- difficulty δ_j
- dispersion τ_j, governs the response variability

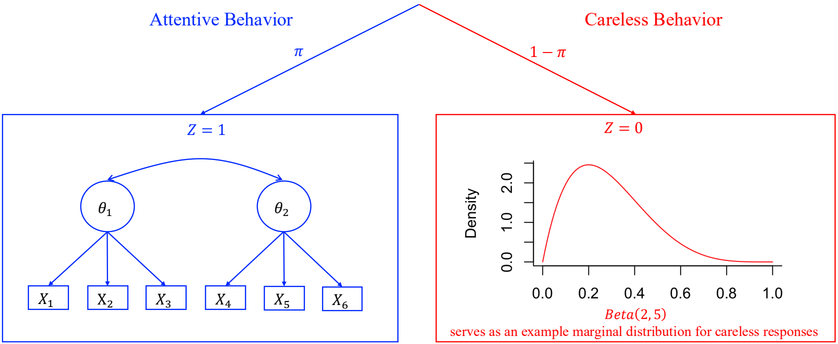
Beta IRM



A higher τ yields a more peaked information function, meaning the item provides more information at specific trait levels.

- Beta IRM models data on (0, 1) scale directly
- Offers a straightforward interpretation of the parameters
- Traditional factor models assume normality, which may be violated (e.g., skewed or boundary-clustered responses)

The Proposed Model



The Proposed Model

$$
\begin{aligned}
f(x_{ij}) = \pi_i \cdot f(x_{ij} \mid z_i = 1) + (1 - \pi_i) \cdot f(x_{ij} \mid z_i = 0) \\
X_{ij} \mid Z_i = 1 \sim \text{Beta}\left(\alpha_{ij}, \beta_{ij}\right) \\
X_{ij} \mid Z_i = 0 \sim \text{Beta}(m, n)
\end{aligned} \tag{2}
$$

Careless Component

- Careless behaviors can vary across respondents
- Any careless component model will most likely be misspecified
- The goal is not to distinguish subtypes of careless behavior

The careless component of the mixture model is conceptualized as a residual class with minimal structure

- Designed to capture any response pattern that systematically deviates from attentive behavior
- The Beta distribution of the careless component is assumed to "absorb" different careless response patterns
- Reflects the marginal distribution of all careless patterns

Simulation Study

The aim of the simulation study was to evaluate the performance of the proposed model in two aspects.

- its accuracy in detecting careless respondents exhibiting different behavioral patterns
- its estimation accuracy of model parameters of the attentive response model that are adjusted for the occurrence of careless responding.

Simulation Design

1) Random response pattern at extremes: Beta$(0.5, 0.5)$.
2) Overly consistent response pattern: Careless respondents consistently selected values near one end of the slider scale—either the left (unif$(0, 0.2)$) or the right (unif$(0.8, 1)$) —regardless of item wording.
3) Random response pattern at midpoint: Responses clustered around the midpoint of 0.5.

The proportion of careless respondents: 0.05 - 0.15

Simulation

Prior setting for attentive probability

$$
\begin{aligned}
(\pi_i, 1-\pi_i) &\sim \text{Dirichlet}(\nu(\pi_{\mathcal{P}}, 1-\pi_{\mathcal{P}})), \\
(\pi_{\mathcal{P}}, 1-\pi_{\mathcal{P}}) &\sim \text{Dirichlet}(1,1) \\
\nu &\sim \text{Cauchy}(0,5)
\end{aligned}
\tag{3}
$$

Classification:

- a ranking approach using the population-level proportion of attentive respondents ($\pi_{\mathcal{P}}$).
- we ranked π_i values, and identified the least likely $(1-\pi_{\mathcal{P}}) \times N$ respondents as careless based on their individual probabilities.
- This approach allows the classification to reflect the estimate of overall attentiveness.

Results

The model exhibited high convergence rates for all conditions ($\geq$ 99%).

Table 2: Classification Results Based on $\pi_{\mathcal{P}}$

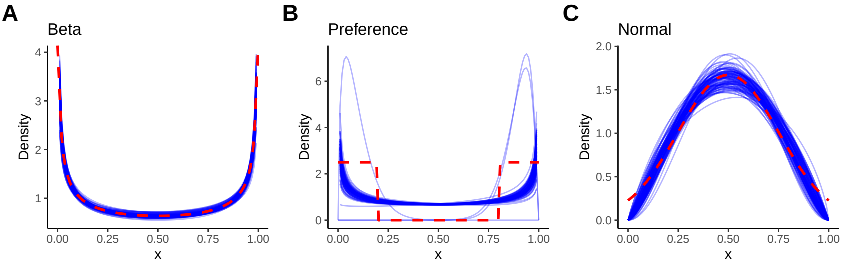
$1 - \pi_{\mathcal{P}}$	Pattern	Accuracy	Sensitivity	Precision	FPR	FNR
0.05	Beta	0.984	0.985	0.998	0.045	0.015
	Preference	0.969	0.970	0.997	0.053	0.030
	Normal	0.963	0.966	0.995	0.095	0.034
0.10	Beta	0.984	0.985	0.997	0.031	0.015
	Preference	0.961	0.961	0.996	0.037	0.039
	Normal	0.950	0.953	0.990	0.086	0.047
0.15	Beta	0.980	0.985	0.991	0.048	0.015
	Preference	0.955	0.952	0.995	0.031	0.048
	Normal	0.952	0.950	0.993	0.040	0.050

Results

Table 3: Estimation Accuracy of Item Parameters and $\pi_{\mathcal{P}}$

		δ_j		τ_j		$\pi_{\mathcal{P}}$		σ_θ	
$1-\pi_{\mathcal{P}}$	Condition	$r(\delta_{true}, \delta_{est})$	RMSE	$r(\tau_{true}, \tau_{est})$	RMSE	RB	RMSE	RB	RMSE
	Beta	0.996	0.088	0.981	0.190	-0.011	0.012	0.026	0.053
0.05	Preference	0.996	0.092	0.974	0.224	-0.026	0.026	0.040	0.063
	Normal	0.997	0.090	0.980	0.200	-0.027	0.030	0.029	0.053
	Beta	0.996	0.089	0.980	0.203	-0.009	0.011	0.038	0.065
0.1	Preference	0.995	0.102	0.974	0.230	-0.034	0.033	0.041	0.070
	Normal	0.996	0.097	0.977	0.216	-0.035	0.037	0.042	0.067
	Beta	0.996	0.086	0.976	0.217	-0.004	0.012	0.031	0.059
0.15	Preference	0.996	0.101	0.972	0.245	-0.042	0.039	0.046	0.067
	Normal	0.996	0.103	0.977	0.216	-0.041	0.037	0.049	0.067

Results



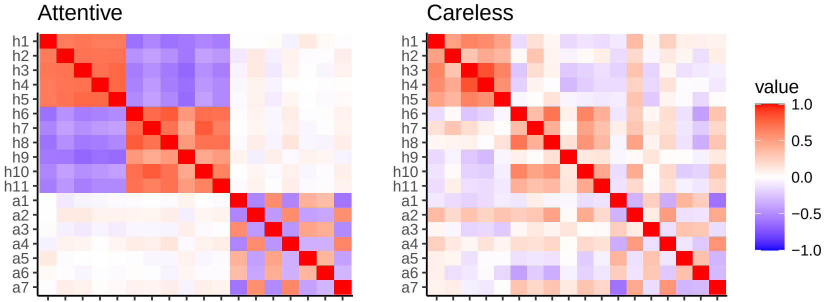
A
Beta
Density
4
3
2
1
0.00
0.25
0.50
0.75
1.00
x
B
Preference
Density
6
4
2
0
0.00
0.25
0.50
0.75
1.00
x
C
Normal
Density
2.0
1.5
1.0
0.5
0.0
0.00
0.25
0.50
0.75
1.00
x

Empirical Study

$N = 851$, Constructs Measured:

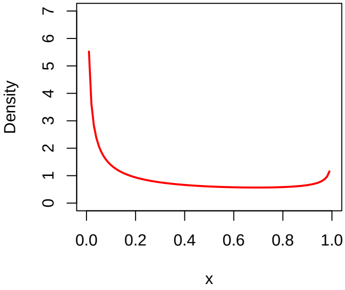
- Height: Perceptions/experiences of physical stature; 11 items (h1–h11), 6 negatively worded
- Autonomy: Sense of decision-making freedom, 7 items (a1–a7), 3 negatively worded

The population-level proportion of careless respondents $(1 - \pi_{\mathcal{P}})$ for the VAS was .08 ($N = 69$), with a 95% credibility interval of [.069, .092].

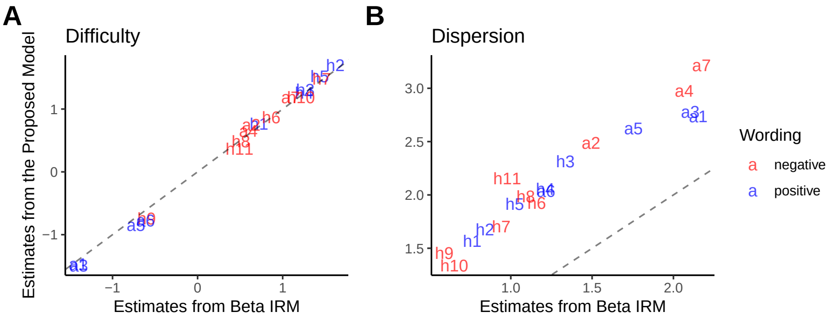


Empirical Study

The parameters for the Beta distribution of careless responses, with $m = 0.388$ (95% CI: [0.357, 0.420]) and $n = 0.728$ (95% CI: [0.640, 0.823]), indicated that, marginally, careless respondents tended to select the left side of the slider scale



Empirical Study



Increased dispersion implies enhanced item informativeness, indicating an improvement in item psychometric properties upon accounting for careless responses.

Takeaways

- We proposed a Beta Mixture Item Response Model designed to identify careless respondents in VAS data.
- We demonstrated the effectiveness of the proposed model.

Preprint: `https://doi.org/10.31219/osf.io/tp6df_v1`

- We also conducted a simulation study to explain why mixture CFA is unsuitable for VAS data.
- Additional analysis in the empirical study also revealed a higher proportion of careless respondents in VAS compared to Likert scale data.
- Forthcoming Study: Comparing Careless Responding in Likert and Visual Analogue Scales in Ecological Momentary Assessment (Ulitzsch et al., 2025)

Thank you!

Contact: lijinzhang.com

Preprint:
https://doi.org/10.31219/osf.io/tp6df_v1