Proceedings of the 90th Annual International Meeting of the Psychometric Society, Minneapolis, United States (2025).

A Joint Factor–Topic Model for Mixed-Format Response Data

Yuxiao Zhang [*†] David Arthur [‡] Yukiko Maeda [†] Nielsen Pereira [†] and Hua-Hua Chang [†]

†Department of Educational Studies, Purdue University, West Lafayette, 47907, Indiana, USA
‡Department of Science and Mathematics, University of Washington Tacoma, Tacoma, 98402, Washington, USA
*Corresponding author. Email: zhan3971@purdue.edu

Abstract

Mixed-format response data that combine rating-scale items and open-ended text are increasingly common, yet these modalities are often analyzed separately. We propose a joint factor–topic model that integrates confirmatory factor analysis for rating-scale responses with an LDA-style topic model for text. The two components are coupled through a multinomial logistic regression in which latent factor scores predict document-level topic proportions, enabling coherent joint inference across modalities. The model is implemented in Stan and estimated via Hamiltonian Monte Carlo. We demonstrate the approach using survey data from 577 middle school students, where rating-scale responses measure academic and social giftedness and open-ended responses describe career interests. The model recovered two interpretable career-interest topics (technology/design-oriented vs. people-centered/helping professions) and revealed clear associations between the giftedness factors and topic prevalence. This framework extends topic models by allowing latent covariates inferred under a measurement model and broadens psychometric latent-variable modeling by incorporating text evidence.

Keywords: joint modeling, mixed-format response data, topic modeling

1. Introduction

Mixed-format response data are increasingly common in social science research. Surveys often include both rating-scale items and open-ended questions, while educational assessments routinely combine selected-response items with constructed-response items. Traditional psychometric analyses primarily target the numeric component (e.g., rating scales or item scores; Chang et al., 2021). In contrast, textual responses are typically converted into numeric features or scores, either by human coding or automated scoring methods, and then entered into psychometric models as observed variables (McCaffrey et al., 2022). This two-stage workflow can disconnect the measurement model from the original text, restricts direct information sharing across modalities, and treats extracted text scores as fixed despite uncertainty. These limitations motivate a cohesive modeling approach that incorporates text more directly within a psychometric framework.

Conceptually, two challenges must be addressed: (1) how to statistically represent text responses, and (2) how to integrate a text model with latent variable models used in psychometrics. Probabilistic topic models provide a principled approach to representing text (Blei, 2012; Maeda et al., 2025; Zhang et al., 2026). A seminal example is Latent Dirichlet Allocation (LDA; Blei et al., 2003), which models each document as a mixture of latent topics and each topic as a distribution over words. Topics are inferred from systematic word co-occurrence patterns across documents, and documents can be summarized by their topic proportions, enabling interpretable low-dimensional representations

of large text corpora. Topic modeling has also evolved beyond purely unsupervised discovery. For example, supervised LDA relates topics to an observed outcome (Blei & McAuliffe, 2007), and the Structural Topic Model (STM; Roberts et al., 2014, 2016) incorporates document-level covariates to explain variation in topic prevalence (and potentially topic content). These extensions support common social-science goals of connecting qualitative text to quantitative measures.

The focus of the current study is the second challenge: integrating topic modeling with psychometric latent variable models. Existing covariate-informed topic models typically assume covariates are observed, rather than latent constructs measured with error. Meanwhile, psychometric research has begun to explore the model that combines latent trait modeling with representations of constructed responses, including approaches that connect item response models with topic-like structures for text (Hong et al., 2021). However, existing approaches do not provide a flexible and general framework in which topic prevalence can be driven by latent factors estimated from rating-scale data, while allowing the text component to feed back into inference about those factors through a unified likelihood, thereby motivating the development of new joint models.

In this study, we propose a joint factor–topic model that integrates a confirmatory factor analysis (CFA) measurement model for rating-scale responses with an LDA-style topic model for open-ended text. This joint formulation propagates uncertainty across modalities and can mitigate attenuation relative to two-stage pipelines that regress topics on plug-in factor scores (Keith, 2014). Substantively, it enables researchers to examine how latent traits relate to latent themes in text within a single coherent model. We demonstrate the proposed approach using an empirical example involving students' giftedness measured by rating scales and career interests expressed in open-ended responses.

2. Proposed Joint Factor-Topic Model

Figure 1 illustrates the proposed joint factor–topic model. The lower component is a CFA for rating-scale responses that estimates each participant's latent factor score(s). The upper component is a topic model for open-ended text that infers latent topics and participant-specific topic proportions. The two components are linked through a multinomial logistic regression in which factor scores predict topic proportions. This coupling enables information sharing across modalities: rating-scale data can inform topic discovery, while the text likelihood updates factor score estimates. Rather than treating factors and topics as separate latent representations, the model provides a unified framework that jointly characterizes participants using both response formats, which can yield a more complete profile when the modalities capture complementary aspects of the same or overlapping underlying attributes.

In Figure 1, shaded nodes denote observed data. Specifically, $\mathbf{Y}_d$ denotes the vector of rating-scale item responses from participant d, and w_{dn} denotes the nth word token in participant d's open-ended response. Unshaded nodes denote latent variables and model parameters. For the CFA component, F_d is the vector of latent factor scores for participant d, $\boldsymbol{\Lambda}$ denotes the factor loading matrix, $\boldsymbol{\mu}$ denotes the vector of item intercepts, $\boldsymbol{\sigma}_Y$ denotes the vector of item residual standard deviations, and $\boldsymbol{\Phi}$ denotes the factor correlation matrix. For the topic model component, θ_d is the K-dimensional vector of topic proportions for document d, z_{dn} is the latent topic assignment for token n in document d, $\boldsymbol{\kappa}_k$ is the vector of topic-specific word logits, and $\boldsymbol{\psi}_k$ is the topic–word distribution for topic k. The matrix Γ contains the factor-to-topic effects in the multinomial logit link that connects the CFA and topic model components. The filled dot represents a hyperparameter: τ_κ is the scale parameter for the topic-word logits, which governs how concentrated the topic–word distributions are. The inner dashed rectangle indicates word-level operations, indexed by token $n = 1, \ldots, N_d$, where N_d is the number of word tokens in document d. The outer dashed rectangle indicates document/participant-level quantities, indexed by $d = 1, \ldots, D$, where D is the total number of participants/documents. Nodes outside the outer rectangle are global parameters shared across participants.

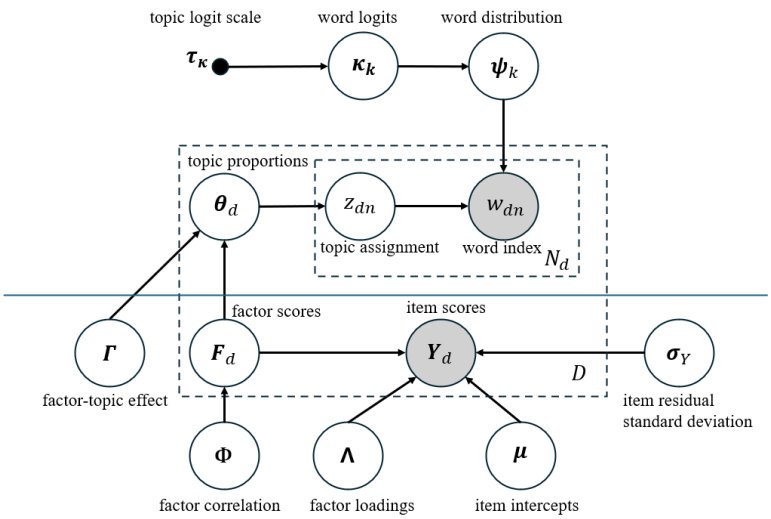


Figure 1. Directed Acyclic Graph of the Joint Factor-Topic Model

2.1 Measurement model for rating-scale responses (CFA)

The rating-scale responses are modeled using a standard CFA (Brown, 2015). For participant d and item j, the observed score Y_{dj} is assumed normally distributed around its model-implied mean,

$$\mu_{dj} = \mu_j + \sum_{l=1}^{L} \lambda_{jl} F_{dl}, \tag{1}$$

where μ_j is the intercept for item j, F_{dl} is participant d's score on factor l, and λ_{jl} is the loading of item j on factor l. Under a simple-structure CFA, each item loads on a single factor (i.e., $\lambda_{jl} = 0$ for all but one l for each j). The observed item score is then

$$Y_{dj} \sim N\left(\mu_{dj}, \sigma^2_{Yj}\right), \tag{2}$$

where σ^2_{Yj} is the residual variance for item j.

2.2 Text model for open-ended responses (topic model)

The text component follows the standard LDA-style generative story (Blei et al., 2003). For each token (i.e., word) n in document d, a topic indicator z_{dn} is drawn from the document's topic proportions θ_d, and the observed word w_{dn} is then drawn from the word distribution corresponding to that topic:

1. Topic assignment: $z_{dn} \sim \text{Categorical}(\theta_d)$;
2. Word generation: $w_{dn} \sim \text{Categorical}(\psi_{z_{dn}})$.

Although the token-level story matches LDA, our model differs in how θ_d and ψ_k are modeled. In basic LDA, document-topic proportions and topic-word distributions are assigned Dirichlet priors (with hyperparameters α and β), implying exchangeability across documents in the absence of covariates (Blei et al., 2003). In the proposed joint model, the topic proportions are linked to participants' latent factor scores via a multinomial logistic regression, so that rating-scale information can systematically explain variation in topic proportions across documents. In addition, we model

the topic-word distributions using a logistic–normal parameterization (i.e., softmax-transformed Gaussian logits), which provides a flexible alternative to Dirichlet priors and aligns with formulations used in more recent topic model extensions such as the STM (Roberts et al., 2016, 2019). Specifically, for each topic k, let $\boldsymbol{\kappa}_k = (\kappa_{k1}, \ldots, \kappa_{kV})$ denote the vector of word logits over the V-term vocabulary, with

$$\kappa_{kv} \sim N(0, \tau_\kappa^2), \qquad v = 1, \ldots, V. \tag{3}$$

Here, κ_{kv} is the logit for term v in topic k, and τ_κ controls the dispersion of these logits; larger values of τ_κ allow more extreme logits and hence more concentrated topic–word distributions. The topic-word distribution is then obtained by a softmax transformation,

$$\psi_{kv} = \frac{\exp(\kappa_{kv})}{\sum_{v'=1}^{V} \exp(\kappa_{kv'})}, \qquad v = 1, \ldots, V, \tag{4}$$

so that $\boldsymbol{\psi}_k = (\psi_{k1}, \ldots, \psi_{kV})$ is a valid probability vector over vocabulary terms.

2.3 Factor-to-topic regression

Topic proportions are linked to participants' latent factor scores through a multinomial logistic regression (i.e., a softmax link). Specifically, for document d and topic k, we define the linear predictor

$$\eta_{dk} = \gamma_{0k} + \sum_{l=1}^{L} F_{dl}\, \Gamma_{lk}, \tag{5}$$

and map the K linear predictors to a probability simplex via

$$\theta_{dk} = \frac{\exp(\eta_{dk})}{\sum_{k'=1}^{K} \exp(\eta_{dk'})}. \tag{6}$$

Here, γ_{0k} is a topic-specific intercept and Γ_{lk} quantifies the association between factor l and the log-odds of topic k relative to the other topics. This structural regression plays a role analogous to a structural path in structural equation modeling (Keith, 2014), with the key distinction that the outcome is a simplex-valued probability vector. The softmax transformation ensures that $\boldsymbol{\theta}_d = (\theta_{d1}, \ldots, \theta_{dK})$ is a valid topic proportion vector with positive elements summing to one, thereby defining a multinomial logistic regression of topic proportions on factor scores.

2.4 Estimation and implementation

The generative process in Figure 1 implies a joint likelihood for the rating-scale responses and the open-ended text. Let $\mathbf{Y} = \{Y_{dj}\}$ denote the rating-scale responses and $\mathbf{W} = \{w_{dn}\}$ denote the observed word tokens. Conditional on the model parameters (denoted collectively by $\boldsymbol{\vartheta}$), the two data sources are assumed independent, so the joint likelihood factorizes as

$$p(\mathbf{Y}, \mathbf{W} \mid \boldsymbol{\vartheta}) = \prod_{d=1}^{D} \left[\prod_{j=1}^{J} \mathcal{N}\left(Y_{dj} \mid \mu_{dj}, \sigma_{Y_j}^2\right) \prod_{n=1}^{N_d} \sum_{k=1}^{K} \theta_{dk}\, \psi_{k,w_{dn}} \right]. \tag{7}$$

Regarding the text component, for each word token w_{dn} in document d, the term $\sum_{k=1}^{K} \theta_{dk}\psi_{k,w_{dn}}$ marginalizes over the latent topic assignment and yields the token probability under the document-specific topic mixture. Multiplying this quantity over all N_d tokens gives the text likelihood contribution for document d. The joint likelihood for participant d is then obtained by multiplying the

text likelihood and the rating-scale likelihood, and the overall joint likelihood follows by taking the product over all D participants.

The joint model requires several identification constraints. In the CFA component, factor means were fixed to 0 and factor variances to 1, while all factor loadings were freely estimated with positivity constraints for sign identification. In the topic model component, the word logits $\boldsymbol{\kappa}_k$ were mean-centered within topic before the softmax transformation, which resolves the shift invariance for $\boldsymbol{\psi}_k$. In the factor-to-topic regression, the topic intercept vector $\boldsymbol{\gamma}_0$ and each row of Γ were mean-centered across topics to resolve the analogous softmax shift invariance in the multinomial logit for θ_d, thereby constraining each factor's effects across the K topics to sum to zero. Because closed-form estimation is not available for the joint model, we conducted Bayesian inference. The model was implemented in Stan (Carpenter et al., 2017) and estimated using Hamiltonian Monte Carlo (HMC) with the No-U-Turn Sampler (Gelman et al., 2013). Weakly informative priors were placed on model parameters. To address topic label switching, we applied topic anchoring using seed terms selected from an initial LDA fit. Posterior sampling was carried out via the `cmdstanr` package in R (Gabry et al., 2025). To improve computational efficiency, the text likelihood was parallelized within each chain using `reduce_sum`. Sampling quality and convergence were assessed using trace plots, $\widehat{R}$, and effective sample sizes (ESS; Gelman et al., 2013), whereas overall model fit was evaluated using posterior predictive checks.

3. Empirical Demonstration

3.1 Data and measures

We applied the proposed joint factor-topic model to survey data from 577 students attending seven public middle schools in the Midwestern United States. Students completed the HOPE Scale (Gentry et al., 2021), which assesses self-perceived giftedness in two domains: academic (6 items) and social (5 items). Items were rated on a six-point response scale. Students also answered an open-ended question: "What career(s) or career field(s) would you like to explore, and why?" The responses contained, on average, 21 words (median = 15; Q1 = 8; Q3 = 27).

Standard text preprocessing was applied via the `tm` package in R (Feinerer et al., 2008), including tokenization, removal of stop words and punctuation, and stemming (reducing words to their base forms). After preprocessing, the resulting document-term matrix comprised 577 documents (rows) and 1,038 terms (columns) for model fitting.

3.2 Model estimation and diagnostics

We fitted a two-factor, two-topic joint factor–topic model. The two-factor structure aligns with the academic and social domains of the HOPE Scale (Gentry et al., 2021). We used two topics to provide a parsimonious representation of career-interest themes (i.e., uni-dimensional scaling of text); with $K = 2$, the topic proportions are fully determined by the prevalence of one topic because proportions sum to one. In practice, the joint factor-topic model generalizes to alternative numbers of factors and topics.

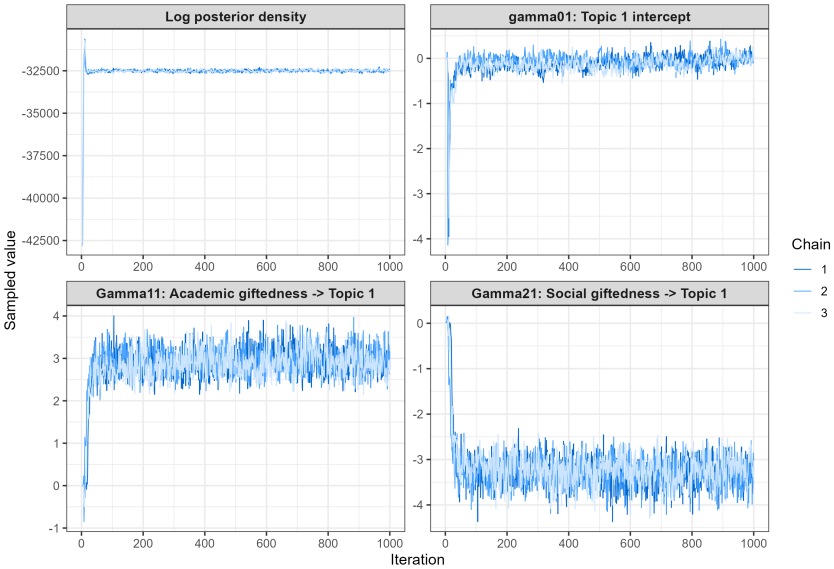
We ran three independent chains with random initialization, each for 1,000 iterations (500 warm-up), yielding 1,500 post-warm-up draws for inference. Posterior summaries for the key nonredundant global parameters are reported in Table 1, including posterior means, 95% credible intervals, and MCMC diagnostics. Convergence diagnostics indicated satisfactory mixing for the key parameters: trace plots showed stable behavior across chains after warm-up, $\widehat{R}$ values were close to 1.00 (all < 1.05), and effective sample sizes were at least several hundred for most parameters, supporting reliable posterior summaries. As an illustration, Figure 2 presents trace plots for the log posterior density and the nonredundant factor-to-topic regression parameters.

In addition to convergence diagnostics, we assessed model fit using posterior predictive checks. For the rating-scale component, all 11 observed item means, all 11 item variances, and 51 of 55

Table 1. Posterior Summaries for Key Global Parameters

Parameter	Mean	95% CrI	$\hat{R}$	Bulk ESS
Measurement model				
Academic giftedness factor loadings (standardized)				
$\lambda_{1,1}$	0.60	[0.53, 0.65]	1.01	891
$\lambda_{6,1}$	0.68	[0.63, 0.73]	1.01	557
$\lambda_{7,1}$	0.69	[0.64, 0.74]	1.00	756
$\lambda_{9,1}$	0.68	[0.63, 0.73]	1.00	588
$\lambda_{10,1}$	0.69	[0.64, 0.74]	1.00	622
$\lambda_{11,1}$	0.60	[0.54, 0.66]	1.00	834
Social giftedness factor loadings (standardized)				
$\lambda_{2,2}$	0.35	[0.27, 0.42]	1.00	1070
$\lambda_{3,2}$	0.52	[0.46, 0.58]	1.00	1187
$\lambda_{4,2}$	0.64	[0.58, 0.69]	1.01	859
$\lambda_{5,2}$	0.64	[0.58, 0.69]	1.01	866
$\lambda_{8,2}$	0.57	[0.51, 0.63]	1.00	1025
Factor correlation, ϕ_{12}	0.89	[0.85, 0.92]	1.01	551
Factor-to-topic regression (Topic 1: Technology/Design)				
Topic intercept, γ_{01}	0.06	[-0.20, 0.33]	1.03	81
Academic giftedness $\rightarrow$ Topic 1, Γ_{11}	2.95	[2.39, 3.60]	1.01	538
Social giftedness $\rightarrow$ Topic 1, Γ_{21}	-3.28	[-3.93, -2.71]	1.01	431
General giftedness (average) $\rightarrow$ Topic 1, β_{a1}	-0.34	[-0.58, -0.13]	1.02	443
Profile difference (academic − social) $\rightarrow$ Topic 1, β_{d1}	6.23	[5.10, 7.49]	1.01	491

Note. CrI = credible interval; ESS = effective sample size. Reported ESS values are bulk ESS. The CFA loadings reported here are fully standardized. Only nonredundant structural parameters are shown. Because the topic intercepts and each row of Γ were centered across topics, the corresponding Topic 2 coefficients are $\gamma_{02} = -\gamma_{01}$, $\Gamma_{12} = -\Gamma_{11}$, and $\Gamma_{22} = -\Gamma_{21}$. The reparameterized coefficients β_{a1} and β_{d1} represent, respectively, the effects of general giftedness (the average of the two latent factors) and profile difference (academic minus social giftedness) on Topic 1; by the same centering constraint, $\beta_{a2} = -\beta_{a1}$ and $\beta_{d2} = -\beta_{d1}$. Item intercepts and residual standard deviations are omitted for brevity.



Note. Only the nonredundant factor-to-topic regression parameters are shown. Because the topic intercepts and each row of Γ were centered across topics, $\gamma_{02} = -\gamma_{01}$, $\Gamma_{12} = -\Gamma_{11}$, and $\Gamma_{22} = -\Gamma_{21}$.

Figure 2. Trace Plots for the Log Posterior Density and Nonredundant Factor-to-Topic Regression Parameters.

pairwise item correlations fell within the 95% posterior predictive intervals, indicating generally good fit with only limited local misfit remaining in the item-correlation structure. For the text component, the replicated corpora captured the observed counts for all 16 checked high-probability topic words, indicating that the model recovered the main lexical patterns associated with the two latent career-interest topics.

3.3 Substantive findings

The CFA component supported the intended two-factor structure. Fully standardized loadings for the academic giftedness items ranged from .60 to .69, whereas those for the social giftedness items ranged from .35 to .64 (Table 1). The estimated factor correlation was .89, indicating that the two giftedness dimensions were strongly positively related, consistent with prior validation work on the HOPE scale (e.g., .87 in Gentry et al., 2021 and .90 in Pereira, 2021).

Figure 3 displays word clouds for the two topics, based on the 20 highest-probability terms within each topic (word size is proportional to within-topic term probability). Inspecting the top-20 terms and the 10 highest-prevalence responses for each topic, Topic 1 was interpreted as *Technology and Design-Oriented Careers*, characterized by terms such as *game*, *design*, *engineer*, *code*, and *program*. This topic reflects interest in creating and building technical products and is broadly aligned with STEM-related career interests. Topic 2 was interpreted as *Helping Professions and People-Centered Careers*, characterized by terms such as *people*, *help*, *medical*, *nurse*, *teacher*, and *psychology*, reflecting interests in caring and service-oriented careers (e.g., healthcare, counseling, and education). Table 2 provides representative responses for each topic.

The factor-topic regression also revealed a clear substantive pattern. Under the original parameterization, academic giftedness was positively associated with Topic 1 (Technology/Design), $\Gamma_{11} = 2.95$, 95% credible interval (CrI) [2.39, 3.60], whereas social giftedness was negatively associated with

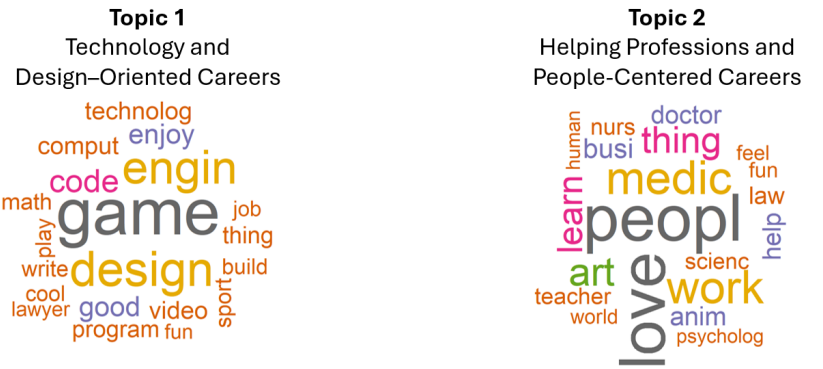


Figure 3. Word Clouds of the Top 20 Terms for Two Latent Career-Interest Topics.

Table 2. Representative Responses for Identified Topics

Topic ID	Topic Label	Representative Response 1	Representative Response 2
1	Technology and Design–Oriented Careers	Programming, coding, and game designing, because I like programming and creating things using technology.	Game design because it seems interesting and fun. Also, programming machinery, because it would be really cool to build something and program it to move.
2	Helping Professions and People-Centered Careers	I would definitely like to explore the psychology side of carers because, when I grow up, I would like to be a therapist who helps people with their problems and emotions… I want people to be comfortable and just feel like they have a shoulder to lean on, just as my best friend gives me. I want people to feel like if they have problems, everything will be okay. That is what I want to do.	I plan on going into the field of medicine, whether it be helping humans or animals. I have always had a passion for working in the medical field, and have always had more of a connection to humans than animals. There are so many animals in danger and are in need of help. Any kind of help to get them to a safe and loving environment, I feel like would require my help.

Note. Ellipses (…) indicate omitted text in the original response.

Topic 1, $\Gamma_{21} = -3.28$, 95% CrI [-3.93, -2.71]. By the topic-centering constraint, the corresponding coefficients for Topic 2 (People-Centered) are equal in magnitude and opposite in sign.

Because the two giftedness factors were strongly positively correlated in this sample ($r = .89$), we also re-expressed the regression for interpretive convenience in terms of two derived components: general giftedness, defined as the average of the two latent factors, and profile difference, defined as academic minus social giftedness. For Topic 1, the effect of general giftedness was small and negative, $\beta_{a1} = -0.34$, 95% CrI [-0.58, -0.13], whereas the effect of profile difference was large and positive, $\beta_{d1} = 6.23$, 95% CrI [5.10, 7.49]. Thus, students whose academic giftedness exceeded their social giftedness tended to emphasize technology/design-oriented career interests, whereas, by sign reversal, students whose social giftedness exceeded their academic giftedness tended to emphasize the helping/people-centered topic. The comparatively small average effect suggests that the dominant structural pattern was not a general giftedness effect, but rather a difference in students' relative academic versus social self-perceptions. These results are consistent with prior work linking self-perceived strengths to career interests (Marsh & O'Mara, 2008) and illustrate how the proposed joint model can simultaneously recover interpretable topics and quantify how latent traits are related to thematic patterns in open-ended responses.

4. Discussion

This study formulated a joint factor–topic model, implemented Bayesian estimation using HMC, and demonstrated the model on survey data combining rating-scale responses and open-ended text. The proposed framework jointly estimates latent factors and latent topics and links them, enabling inference on how individual differences in latent traits relate to thematic patterns in text. The joint inference allows information sharing across modalities and propagates uncertainty coherently to the trait–theme relationships.

Methodologically, the model extends covariate-driven topic modeling (Roberts et al., 2014, 2016) by allowing the covariates to be latent variables inferred from a measurement model, and it broadens psychometric latent variable modeling by incorporating text through a probabilistic topic model (Blei, 2012). The framework can be viewed in two complementary ways. From a structural modeling perspective, topics can be treated as latent constructs that are predicted by latent traits, yielding a generalized SEM that incorporates text-derived latent structure. From a measurement perspective, the text provides additional evidence about latent traits through latent themes: words are generated conditional on topics , which are in turn linked to factors. In our empirical demonstration, the model revealed interpretable career-interest topics and clear, opposing associations of academic versus social giftedness with topic prevalence, illustrating the value of integrating rating-scale and text data within a single probabilistic model.

Several limitations and future directions warrant attention. First, systematic simulation studies are needed to evaluate parameter recovery and robustness of the joint factor-topic model under varying conditions (e.g., number of topics, number of factors, document length, and factor–topic effect size). Second, although HMC is convenient and provides exact Bayesian inference, computational cost is nontrivial for larger corpora; exploring more specialized algorithms (e.g., variational inference or customized Gibbs sampling) would improve scalability. Third, extensions to common psychometric settings are promising, such as integrating other latent variable models (e.g., cognitive diagnosis models or latent class analysis) to link attribute profile or class membership to thematic patterns in text. Together, these directions can broaden the applicability of joint factor-topic modeling as a general framework for text-augmented psychometrics in social science research.

Acknowledgement

The empirical data used in this study were collected as part of the project *Closing Excellence and Opportunity Gaps for Students from Traditionally Underserved Populations in Gifted Education: A Multi-Tier Systems of Support Approach Program*, supported by the U.S. Department of Education, Javits Gifted and Talented Students Education Program (Award No. S206A190020).

We thank Dr. Vinayak Rao for helpful discussions and feedback during the early stages of model development.

Funding Statement This work received no specific grant/funding.

Competing Interests None.

References

Blei, D. M. (2012). Probabilistic topic models. *Commun. ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Blei, D. M., & McAuliffe, J. D. (2007). Supervised topic models. In J. C. Platt, D. Koller, Y. Singer, & S. T. Roweis (Eds.), *Advances in neural information processing systems* (pp. 121–128, Vol. 20). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2007/file/d56b9fc4b0f1be8871f5e1c40c0067e7-Paper.pdf

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993–1022.

Brown, T. A. (2015, January). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford publications. https://psycnet.apa.org/record/2015-10560-000

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). STAN: a probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Chang, H.-H., Wang, C., & Zhang, S. (2021). Statistical applications in educational measurement. *Annual Review of Statistics and Its Application*, *8*(Volume 8, 2021), 439–461. https://doi.org/https://doi.org/10.1146/annurev-statistics-042720-104044

Feinerer, I., Hornik, K., & Meyer, D. (2008). Text mining infrastructure in r. *Journal of Statistical Software*, *25*(5), 1–54. https://doi.org/10.18637/jss.v025.i05

Gabry, J., Češnovar, R., Johnson, A., & Bronder, S. (2025). *Cmdstanr: R interface to 'cmdstan'* [R package version 0.9.0, https://discourse.mc-stan.org]. https://mc-stan.org/cmdstanr/

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2013, November). *Bayesian data analysis* (3rd ed.). Chapman; Hall/CRC. https://doi.org/10.1201/b16018

Gentry, M., Pereira, N., Peters, S. J., McIntosh, J. S., & Fugate, C. M. (2021, September). *Hope Teacher Rating scale*. Routledge. https://doi.org/10.4324/9781003235569

Hong, M., Wilcox, K. T., & Cheng, Y. (2021). A constructed response model [PMID: 33263435]. *Multivariate Behavioral Research*, *56*(1), 150–151. https://doi.org/10.1080/00273171.2020.1852907

Keith, T. Z. (2014, December). *Multiple regression and beyond*. Routledge. https://doi.org/10.4324/9781315749099

Maeda, Y., Wang, X., Zhang, Y., Banks, J. B., & Kenney, R. H. (2025). Balancing human and machine coding: Evaluating the credibility and potential of topic modeling for open-ended survey responses. *Computers in Human Behavior*, *172*, 108703. https://doi.org/https://doi.org/10.1016/j.chb.2025.108703

Marsh, H. W., & O'Mara, A. (2008). Reciprocal effects between academic self-concept, self-esteem, achievement, and attainment over seven adolescent years: Unidimensional and multidimensional perspectives of self-concept [PMID: 18340036]. *Personality and Social Psychology Bulletin*, *34*(4), 542–552. https://doi.org/10.1177/0146167207312313

McCaffrey, D. F., Casabianca, J. M., Ricker-Pedley, K. L., Lawless, R. R., & Wendler, C. (2022). Best practices for constructed-response scoring. *ETS Research Report Series*, *2022*(1), 1–58. https://doi.org/https://doi.org/10.1002/ets2.12358

Pereira, N. (2021). Finding talent among elementary english learners: A validity study of the hope teacher rating scale. *Gifted Child Quarterly*, *65*(2), 153–166. https://doi.org/10.1177/0016986220985942

Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, *111*(515), 988–1003. https://doi.org/10.1080/01621459.2016.1141684

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R package for structural topic models. *Journal of Statistical Software*, *91*(2), 1–40. https://doi.org/10.18637/jss.v091.i02

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, *58*(4), 1064–1082. https://doi.org/https://doi.org/10.1111/ajps.12103

Zhang, Y., Pereira, N., Arthur, D., Castillo-Hermosilla, H., Ozen, Z., & Chang, H.-H. (2026). Using topic modeling in gifted education research: Drawing insights from open-ended survey responses. *Gifted Child Quarterly*, *70*(2), 257–269. https://doi.org/10.1177/00169862251378446