Civic Engagement Evaluation Using Bayesian GRM and CFA

Wei Huang and Jujia Li

Abstract: This study aimed to explore the impact of multicultural exchange experiences on college students' global civic engagement by examining a civic engagement measurement of a youth (age between 19 to 32) exchange program. The program brought together twenty participants from the United States and India for a five-week exchange program to promote their civic attitudes, ethics, and skills. In response to the small sample size, the researchers utilized a Bayesian IRT method incorporating means and standard deviation from previous studies to examine the item parameters in the program evaluation survey. The study employed a Graded Response Model (GRM) and a Bayesian confirmatory factor analysis (CFA) to evaluate participants' latent traits. The research findings indicated that there is difference between the two groups. Bayesian IRT model and GRM with hierarchical priors should be recommended for future studies with a small sample size, given their accuracy, cost, and time effectiveness. This study contributes to the literature by highlighting the need for global civic engagement and promoting multicultural exchange experiences to prepare students to be global civic leaders.

Introduction

As early as in 2011, the US Department of Education, in collaboration with the Association of American Colleges and Universities (AAC&U), emphasized the need for a more informed, engaged, and socially active citizenry in response to the demands of our dynamic and turbulent century. To this end, they drafted "Core Competencies" for civic engagement in academic programs.

Over a decade later, in the midst of the pervasive pandemic, racial and political unrest, educators and researchers have recognized that one of the top priorities in higher education is to promote a just and equitable world. But the question is how to prepare the college graduates to become global civic leaders who will collaborate globally on civic activities to change this world to be a better place? How can they develop the strategic awareness required for effective civic engagement, with attention to their own civic knowledge, attitudes, skills, and behavior?

One potential solution to this question is explored through our new exchange program, which brings together twenty ambitious young adults from the United States and India. The five-week exchange program offers necessary insight and training to develop global civic leaders for the future. Among its objectives is the advancement of civic engagement competencies, while thematically preparing and training participants (undergraduates, graduate students, and new K-12 educators) who are dedicated to making a difference in the lives of others in local, national, and global communities.

However, before addressing how to train these young civic leaders, we must first consider how to evaluate their civic engagement competencies. This article will focus on the evaluation of the latent variable using Bayesian estimation with a very small sample size. By doing so, we aim to contribute to the ongoing discourse on the best practices for evaluating civic engagement competencies and provide insights into the assessment of such competencies within the context of our exchange program.

Theoretical Framework

Literature Review

In light of the need for civic learning and the gap in civic education, educators and researchers have been taking initiatives to promote civic learning and civic engagement measurement in the quality learning outcome (Baumann et al., 2014). Civic learning and civic engagement benefit students in their belonging to the community, increasing social networks, building social capital, and employment opportunities (Evans & Kilinc, 2013; White, 2020). 68% of the chief academic officers surveyed from the 433 member institutions recognized the importance of including civic learning as an essential learning outcome (AAC&U, 2011).

However, research and assessment data did that there were “civic empowerment gap” (Baumann et al., 2014) for students from underprivileged or marginalized communities (Kawashima-Ginsberg, 2013). There has been even less research to address the global civic engagement when it should be not examined on a local scale, but from a global perspective.

In many countries, educators aim to promote civic engagements among students at the local and national levels with related civic knowledge, skills, and attitudes for success as a local leader. However, educators are struggling to merge civic engagement and global learning (Ajaps & Obiagu, 2020; Lorenzini, 2013). Therefore, there is a need to examine how multicultural exchange experience and civic engagement learning can facilitate young leaders’ global civic engagement.

Research Questions

Due to the small sample size, we can utilize the Bayesian IRT method to determine the civic learning and civic engagement difference between American and Indian participants, prepare a credible measurement for future research on the improvement of individuals’ civic learning and engagement in this International cultural exchange activity. Our research questions are:

- Considering the small sample size, can we measure the participant civic engagement attitude, behaviors and skills through a self-reported survey?
- Is there any difference between Indian participants and American participants in their civic engagement competencies?

- Do the young people improve global civic engagement in terms of civic attitude, behaviors and skills based on their multicultural exposure, and international exchange experience?

Method

Study Design

The purpose of this study is to investigate and compare Graded Response Model (GRM) parameter estimation and Confirmatory Factor Analysis (CFA) with MCMC methods:

A multilevel Bayesian item response theory (IRT) model is superior to conventional models in terms of its fit to the data and its ability to use information (May, 2006). Natesan et al. (2016) researched the impact of three prior distributions: matched, standard vague, and hierarchical in Bayesian estimation parameter recovery in two and one parameter model.

Bayesian Graded Response Model

We evaluate participants' responses to item j according to its degree of intensity of favorableness to the statement in the measurement of attitude. All the possible responses to item j can be classified into 5-point scale categories arranged in the order of intensity. The Graded response model (GRM) for Item response theory (IRT) uses a two-step process to obtain the probability that a participant responds to a particular category. The first step is to model the probability that an participants' response falls at or above a particular ordered category given θ.

$$P_{jk}^{*}(\theta) = \frac{e^{Da_j(\theta - b_{jk})}}{1 + e^{Da_j(\theta - b_{jk})}} \quad (1)$$

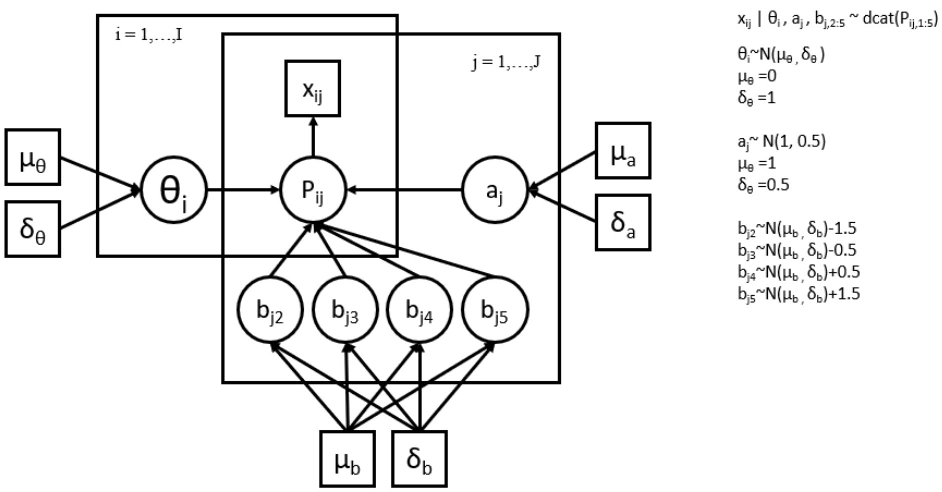
The second step of the GRM is to find the category response functions, which indicate the probability of responding to a particular category given θ based on equation:

$$P_{jk}(\theta) = P_{jk}^{*}(\theta) - P_{j,k+1}^{*}(\theta) \quad (2)$$

This study uses the mirt package in R to build graded response model and checks validity of items.

Figure 1

The graphical equation of Bayesian IRT model



In this study, all the observable score of participants' latent traits about civic engagement are denoted θ_i, as well as five parameters of items are denoted as a_j , the discrimination, and $b_{j,2:5,}$ four thresholds of locations/difficulties. The prior distribution of each parameter are the most crucial in a Bayesian method used to estimate post parameters. However, in most research about civic engagement and related topics, only mean and variance are recorded and published. The participants' latent traits of interests, denoted θ_i, are commonly assumed random effects having normal distribution, $\theta_i \sim N\ (1,\ 0)$, in which i is the order number of each participant. As shown in the figure 1, all the collected means and variance of each items/questions were inputted as the middle point of four locations $b_{j,2:5}$. All $b_{j,middle}$ will follow the normal distribution $b_{j,middle} \sim N(\mu_b ,\ \delta_b)$, and the distribution was divided into five equal parts with four thresholds points, $b_{j2} \sim N(\mu_b , \delta_b)-1.5$, $b_{j3} \sim N(\mu_b , \delta_b)-0.5$, $b_{j4} \sim N(\mu_b , \delta_b)+0.5$, and $b_{j5} \sim N(\mu_b , \delta_b)+1.5$. Based on the previous research, discrimination, aj were assumed random effects having normal distribution, $a_j \sim N$ (1, 0.5). In both a and b parameters of item, j is the order number of items.

This research used Bayesian method to estimate graded response method, the observable scores are following $x_{ij} \mid \theta_i , a_j , b_{j,2:5} \sim dcat(P_{ij,1:5})$, in which dcat( ) is a categorical distribution.

Bayesian Confirmatory Factor Analysis

Although graded response model (GRM) model and offer a reliable analysis on participants' latent traits, confirmatory factor analysis (CFA) is another important analysis tool for many areas of the social and behavioral sciences. It belongs to the family of structural equation modeling techniques that allow for the investigation of causal relations among latent and observed variables in a priori specified, theory-derived models (Hancock & Mueller, 2001).

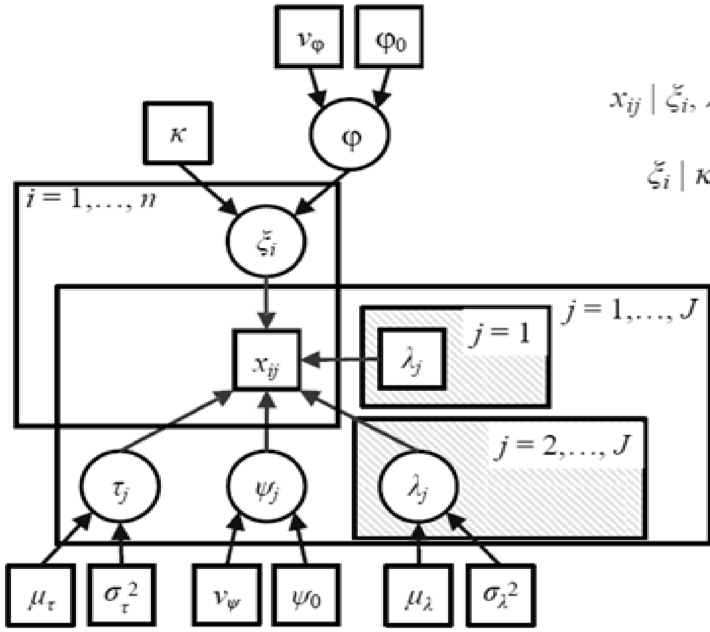
A unidimensional CFA can be estimated using a MCMC method, and the predicted scores can be calculated by:

$$x_{ij} = \tau_j + \lambda_j \xi_i + \psi_j \quad (3)$$

Based on the CFA framework, the latent trait ξ_i is explaining the observed responses as a function of the factor loading λ_j plus interception τ_j and some sort of measurement error ψ_j.

Figure 2

The graphical equation of Bayesian CFA model



$$x_{ij} \mid \xi_i, \lambda_j, \psi_j, \tau_j \sim N(\tau_j + \lambda_j \xi_i, \psi_j)$$

$$\xi_i \mid \kappa, \varphi \sim N(\kappa, \varphi)$$

$$\kappa = 0$$

$$\varphi \sim \text{Inv-Gamma}(5, 10)$$

$$\tau_j \sim N(\mu_j, \sigma_j^2)$$

$$\lambda_1 = 1$$

$$\lambda_j \sim N(1, 10)\, j > 1,$$

$$\psi_j \sim \text{Inv-Gamma}(5, 10)$$

A unidimensional CFA can be estimated using a MCMC method, and the predicted scores can be calculated by:

$$x_{ij} = \tau_j + \lambda_j \xi_i + \psi_j \tag{3}$$

Based on the CFA framework, the latent trait ξ_i is explaining the observed responses as a function of the factor loading λ_j plus interception τ_j and some sort of measurement error ψ_j.

In Bayesian framework, participants' latent value is denoted by ξi, which is randomly sampling as distribution with mean ($\kappa_i = 0$) and variance ($\varphi_i \sim$ inverse gamma (5, 10)), in which i is the number of participants. Meanwhile, we applied means (μ_j) and variances (σ_j) of each item collected from previous research and input the distribution of $\tau_j \sim N(\mu_j, \sigma_j^2)$ as interception of model, in which j is the number of items. In this research, we fixed loading to 1 at the first item, and use $\lambda_j \sim N(1, 10)$ to randomly collect samples for the slope of model from second item to the last one. Finally, we set the error of model as $\psi_j \sim$ inverse gamma (5, 10). In figure 2, the equation in bayesian method, can be converted to $x_{ij} \mid \xi_i, \lambda_j, \psi_j, \tau_j \sim N(\tau_j + \lambda_j \xi_i, \psi_j)$.

Estimation

This international exchange program is now at the first year of its total three-years-long period. The reasons why we choose Bayesian IRT method is: (a) This group has a small sample size with 10 American students and 10 Indian students annually; (b) The scale test of civic engagement is developed on several tests, which can offer priors for items in this test; (c) We can continuously use previous stage results as priors to next-stage research, adjusting the estimation of model.

The priors are collected from previous similar research's distribution. Three prior choices were considered for each of the two Bayesian techniques. The matched priors were: θ, b and a_k.

To compare the accuracy of parameter estimates with respect to four estimation methods (MCMC, VB, CML and MML) and two prior settings (matched prior and hierarchical prior), seven separate analyses of variance (ANOVA) were conducted for graded response model data.

Sample description

Participants were 20 young university undergraduate and graduate students aged from 19 to 32 who participated in a five-week exchange program between the U.S. and India. 10 participants were from Alabama, U.S. and 10 from different provinces in India. 18 of the 20 participants competed the survey.

Measures

While there is increased interest in the need to support the development and measurement of civic learning and engagement, information about instruments that measure civic engagement is not easily accessible. We adapted the survey framework from Tedeschi et al. (2021) technical review report of civic engagement surveys and took out 12 items with prior information from previous studies as is shown in Table 1 below.

Table 1

Item No.	Item	Literature	Mean	SD	Sample Size
1	*I enjoy working in groups or on projects with people with backgrounds and experiences that are different from mine.*	Wilson-Daily et al., 2018	3.698	0.672	High school (N=1592)
2	*I can make a positive difference in my community.*	Syversten et al., 2015	3.55	0.74	High School (N=1138)
3	*It makes me angry when I think about the conditions some people have to live in.*	Flanagan et al., 2007	3.76	0.85	High school N=598
4	*Being concerned about state and local issues is an important responsibility for everybody.*	Saban, 2018	3.16	0.627	College N=204
5	*Being actively involved in community issues is my responsibility.*	Chung, 2011	3.58	0.87	African American 129
6	*It is important to me to help those who are less fortunate.*	Syversten et al., 2015	3.36	0.7	High School (N=1138)
7	*I have stood up for a classmate who was being picked on.*		3.45	0.62	
8	*When I see or read a news story about an issue, I try to figure out if they're just telling one side of the story.*		2.92	1.01	
9	*I am good at leading others to reach a goal.*		3.29	0.74	
10	*I am a hard worker.*		3.89	0.56	
11	*I treat others the way I want to be treated.*		4.16	0.61	
12	*When I work with others, I think about what is best for my team.*		3.95	0.66	

All the items taken from different research surveys were assigned to measure civil engagement including attitude, behavior and skills (Wilson-Daily, 2018; Flanagan et al., 2019; Saban, 2018; Chung, 2011; Syversten et al., 2015).

The authors used a five-point Likert scale (strongly disagree 1, disagree 2, neither agree nor disagree 3, agree 4, strongly agree 5) to determine the level of agreement on the items. No scale items were reverse-scored. A higher overall score indicates a higher civic engagement.

The split-half consistency was tested and the average interitem r was .17. Internal consistency of the twelve items was investigated using Cronbach's alpha. Results indicated that the alpha for the total scale was equal to .63.

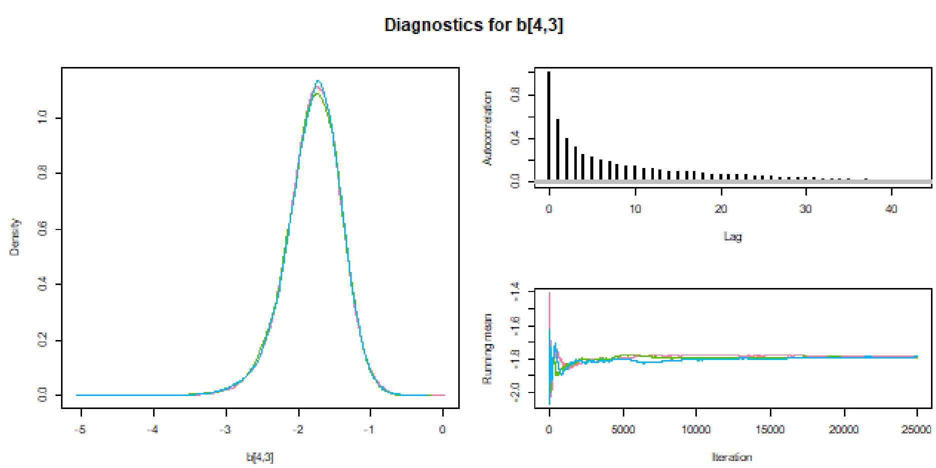
Estimation of GRM using MCMC

MCMC is a method of simulating random samples from any theoretical multivariate distribution. In recent years, Gibbs sampling, was effectively applied to IRT problems (Albert, 1992). In this research, we applied rjags package in R to estimate GRM model with MCMC method.

A three-chain MCMC was run for 25000 iterations. We found all parameters of items and persons are converged from 10000th iteration in Gelman-Rubin test, all the estimated points are equal to the upper confidence interval.

Figure 3

Convergence of Second Difficulty Threshold in Item 4



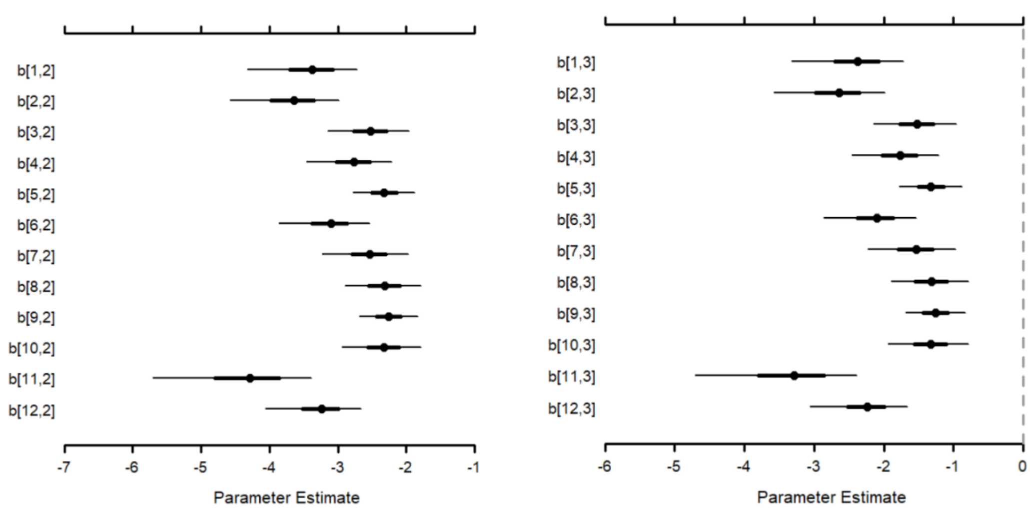
Take the most severe non-convergence situation of second threshold of Item 4 as an example, well convergence began from 10000th iteration.

Item parameter

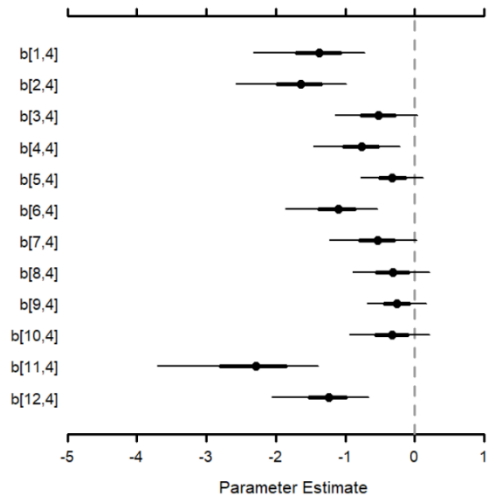
In figure 5-(a), difficulty or location of item 1, item2, item 11 and item 12 are located between -4 to -3, a comparatively low position, meaning the three items are easy for student get a higher score. In all four sub-figures, the phenomenon is similar. The last figure 5-(d) addresses that item 1, item 2, item 6, item 11, and item 12 have negative value for their difficulty parameter b_5 are too low, a potential problem to test participants' latent traits on these interests. Most participants, even whose civic engagement ability is lower than an average lever, can show a high scale answer in item 11.

Figure 5

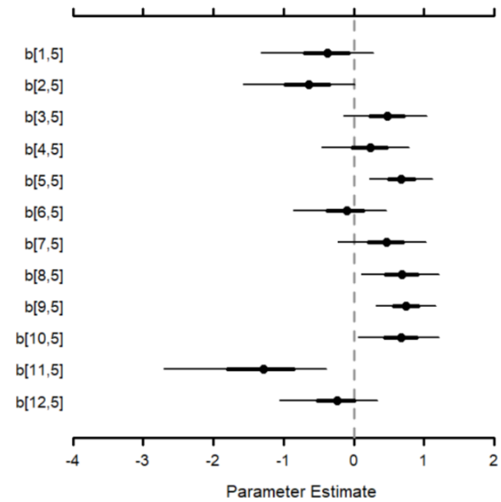
Difficulty thresholds, b_2, b_3, b_4, and b_5 of 12 items



(a) Mean of b_2 (b) Mean of b_3



(c) Mean of b_4



(d) Mean of b_5

In Table 2, all twelve discriminations of items fall between [1.74, 3.18]. Item 9 has the extreme discrimination, 3.18, greater than others, which means that we need take a further analysis on this item to check whether all the participants have a close performance on this question.

Table 2: Discrimination, a of 12 items

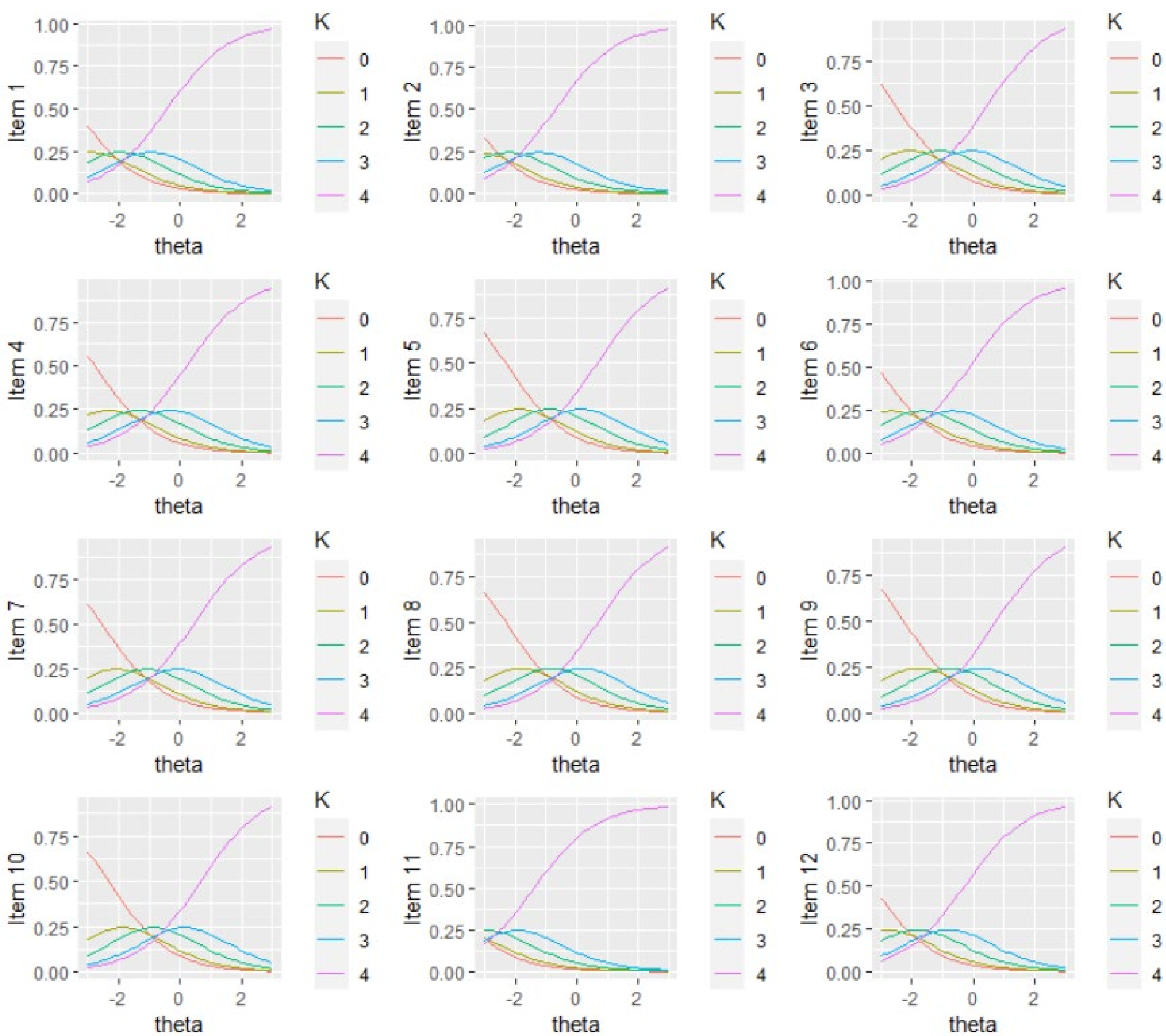
DISCRIMINATION	MEAN	SD	2.50%	50%	97.50%	RHAT	N.EFF
A_1	1.81	0.68	0.73	1.73	3.33	1	8402
A_2	2.67	0.93	1.12	2.58	4.69	1	10980
A_3	1.74	0.54	0.87	1.68	2.95	1	10597
A_4	2.01	0.67	0.92	1.93	3.53	1	9001
A_5	2.92	0.82	1.54	2.85	4.72	1	10398
A_6	2.21	0.78	0.96	2.11	3.96	1	8571
A_7	1.86	0.61	0.89	1.78	3.27	1	9094
A_8	1.75	0.5	0.92	1.7	2.85	1	11226
A_9	3.18	0.83	1.76	3.11	4.98	1	10946
A_{10}	1.88	0.56	0.96	1.83	3.16	1	9911
A_{11}	2.56	0.93	1.04	2.47	4.62	1	13738
A_{12}	2.34	0.83	0.98	2.25	4.18	1	8995

Item Characteristic Curve (ICC)

Using all parameters of item estimated by MCMC, twelve item characteristic curves are displayed in figure 6. Generally, most items are skewed towards right side of zero, meaning they are easy for participants, with normally distributed latent ability, to achieve higher scores. The extreme item is the 11th, with all thresholds are cumulated at from -3 to -2. This item obviously needs to be checked and adjusted in next-step research.

Figure 6

ICC curves for 12 items

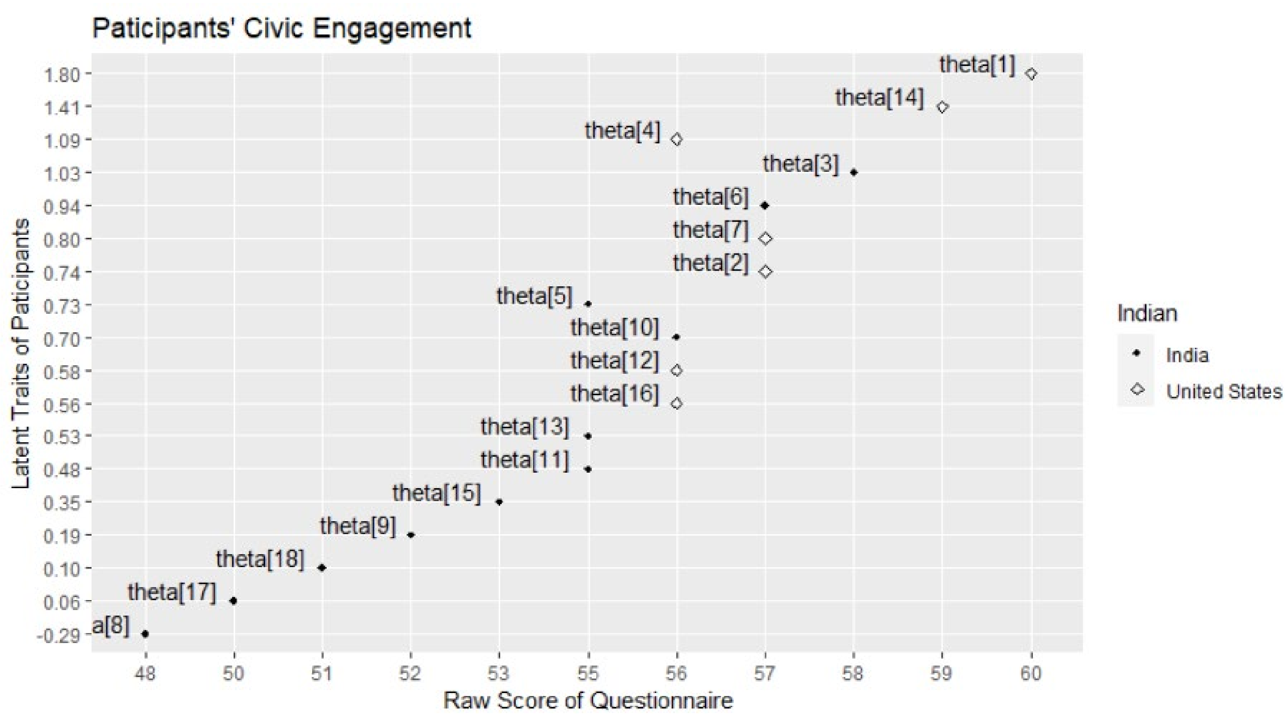


Persons' parameter

In the result of MCMC estimation, participants' raw scores of a survey were converted to the latent traits of interest, which gave a more reliable method to analyze civic engagement of participants. In figure 7, black dot indicates the value of Indian participants and white rhombus shows the value of participants from the United States. Generally, all the participants show a linear regression relationship between the raw scores and predicted latent trait value. The participants from the United States, the rhombus symbols, have a comparative greater value in both raw score and latent traits. Although the No.2, No.6, and No.7 participants have same raw score 57, they are detected a different latent trait value, in which the No.6 participant from India gets 0.94, higher than the two American participants (0.80 and 0.74). Likewise, on raw score 56, No. 10 participant from India has a higher latent trait value than No. 12 and No.16 participants from the United States. However, there is an outlier, No.4, who has an extremely high value 1.09, with better performance at some certain items, compared to other participants.

Figure 7

Participants' raw score of questionnaire (x axis) and latent traits (y axis) in BGRM model

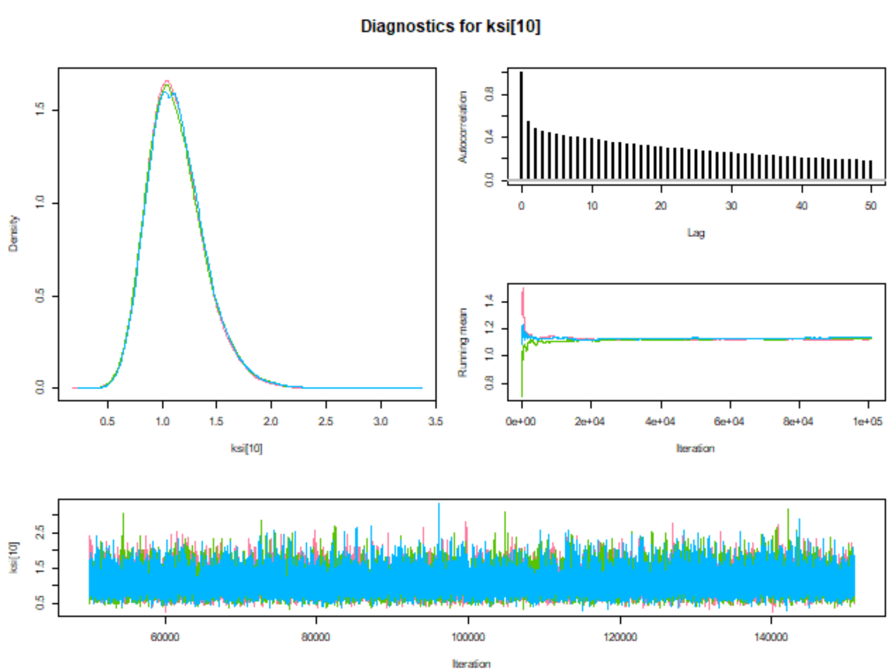


Estimation of CFA using MCMC

Like Bayesian GRM model, Bayesian confirmatory factor analysis (CFA) also offers an alternative to frequentist CFA, for example, maximum likelihood estimation for the assessment of reliability and validity used as educational and psychological measures (Hoofs, et al., 2018). In this study, Gibbs sampling method were applied to estimate the Bayesian CFA model. After several times test, 150000 iterations with 100 thin was an ideal setting for three-chain MCMC, which finally addressed stable convergences in all parameters (ξ_i, λ_j, τ_j, and ψ_j). This result passed a Gelman-Rubin test, in which all the estimated points were equal to the upper confidence interval.

Figure 8

Convergence of No. 10 student's latent civic engagement (ξ_{10})



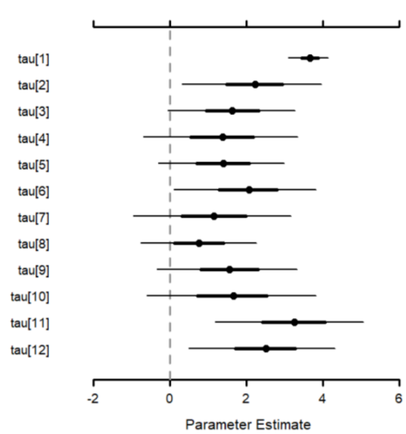
Take the No.10 students' ξ_{10} as an example. No. 10 was the most difficultly converged parameter in this model. It showed a well-converged graph began from 20000th iteration.

Item parameter

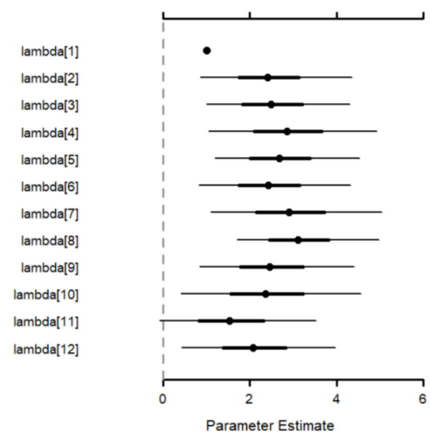
In the Gibbs sampling, the first item's factor loading / slope (λ_1) was fixed at 1. Hence, in Figure 6-(b), λ_1 is shown as a dot lying at 1. Except the first item, item 11 has the lowest factor loading value (λ_{11}) located at Mean 1.6 and the second highest intercept value only after item 1, in figure 6-(a). It means that item 11 is comparatively not a reliable item to discriminate participants' latent traits of civic engagement, a similar result of Bayesian GRM model.

Figure 9

Intercepts and slopes of 12 items in Bayesian CFA model



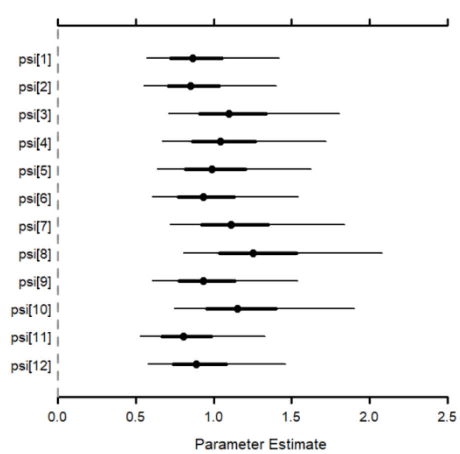
(a) Intercepts of 12 items (τ_j)



(b) Slopes of 12 items (λ_j)

Figure 10

Errors of 12 items in Bayesian CFA model



In figure 10, item 8 and item10 having high error (ψ_j) and variance, require us to pay more attention to the estimation of participants' latent traits in this model.

Persons' parameter

In figure 11, x-axis is participants' raw score of questionnaires and y-axis is predicted latent traits of interests. Black dot is the value of Indian participants and white rhombus is the value of participants from the United States. Similar in the BGRM model, all the participants in BCFA model also fit a linear regression relationship between row score of questionnaires and predicted latent trait value. Due to some participants are located at same position in figure 8, we need check the exact value of these variances in table 2. The No.2, No.6, and No.7 participants getting same raw score 57 have the same predicted latent traits value 1.18. No.4, No.10, No.12, and No.16 participants who are scored 56 have the same predicted latent traits value 1.15. In conclusion, BCFA gave a more linear regression model, but offered less information about discrimination of participants' latent trait.

Figure 11

Participants' raw score of questionnaire (x axis) and latent traits (y axis) in BCFA model

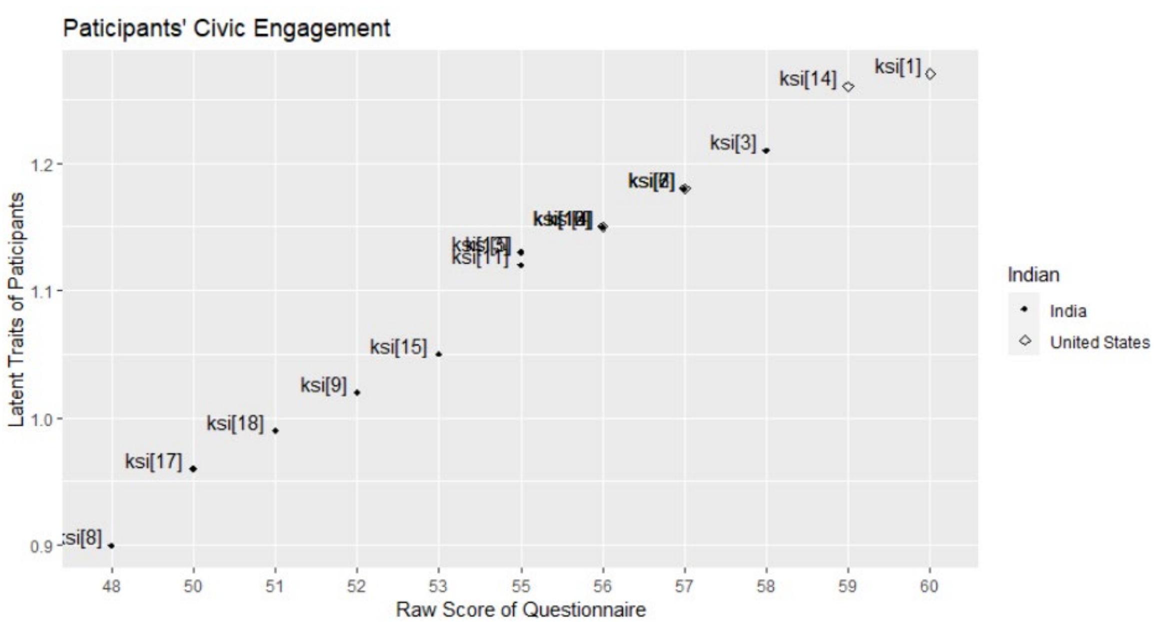


Table 3

Participants' latent traits (ξ_i) in BCFA model

Latent traits	Indian	Raw	Mean	SD	95% HPD lower	95% HPD Upper	Effective Size
ksi[8]	1	48	0.90	0.24	0.46	1.37	1186.02
ksi[17]	1	50	0.96	0.24	0.52	1.46	1192.04
ksi[18]	1	51	0.99	0.25	0.52	1.48	930.59
ksi[9]	1	52	1.02	0.25	0.57	1.54	975.75
ksi[15]	1	53	1.05	0.25	0.58	1.55	926.80
ksi[5]	1	55	1.13	0.26	0.65	1.66	815.25
ksi[11]	1	55	1.12	0.26	0.64	1.64	772.67
ksi[13]	1	55	1.13	0.26	0.65	1.65	810.34
ksi[4]	0	56	1.15	0.27	0.65	1.69	812.82
ksi[10]	1	56	1.15	0.27	0.66	1.68	753.97
ksi[12]	0	56	1.15	0.27	0.67	1.70	731.83
ksi[16]	0	56	1.15	0.27	0.65	1.67	706.59
ksi[2]	0	57	1.18	0.27	0.70	1.74	769.46
ksi[6]	1	57	1.18	0.27	0.68	1.71	719.08
ksi[7]	0	57	1.18	0.27	0.69	1.72	727.58
ksi[3]	1	58	1.21	0.27	0.73	1.78	738.53
ksi[14]	0	59	1.26	0.28	0.74	1.81	665.82
ksi[1]	0	60	1.27	0.28	0.74	1.81	670.52

Model Fit Comparison

There is more than one definition of DIC and WAIC. Celeux et al. (2006) provide 8 variants of DIC; Gelman and Vehtari (2013) provide 2 definitions of WAIC. The definition of DIC (Celeux et al, 2006)) used in JAGS differs from other DIC variants. Martyn Plummer (the author of JAGS) uses the following definition:

$$DIC_{JAGS} = \overline{D(\theta)} + 0.5var\{D(\theta)\} \tag{4}$$

where $\overline{D(\theta)}$ is the posterior mean deviance,

$$\overline{D(\theta)} = \frac{-2}{M}\sum_{m=1}^{M}\sum_{i=1}^{N} logp(y_i|\theta_i^{(m)}) \tag{5}$$

where N is the [data] sample size and M is the posterior sample size, and the expectation is taken over the unknown parameter(s), θ, and D_n is the deviance. This is also the definition that is used in the R package R2jags.

In BGRM model, DIC is 295.5, with penalty (51.8, 347.3) while in BCFA model, DIC is 485.4, with penalty (45.66, 531), which means the BGRM is better than BCFA in this research.

Conclusion

In this study, we used Bayesian IRT estimations to compare GRM and CFA models. Both models generated similar results in item parameters and person parameters.

Incorporating informative priors from previous studies in this analysis, MCMC addressed the small sample problem by intervals taken from percentiles of the posterior distribution for each parameter. Take item 11 as an example, in traditional frequency test, item 11 had to be deleted because there was no variance and item-total correlation cannot be calculated. However, with informative priors from larger sample size, MCMC can mitigate the small size sample issue by getting more information than the data provide.

Compared to CFA, GRM model gives more information on the difficulty thresholds of each item. For instance, Figure 6 the ICC curves show item 11 as non-discriminative for it is extremely skewed to the right, so participants, even with low civil engagement traits, will be highly likely to choose "strongly agree" in their responses to this item.

GRM also shows a clearer map for patterning the person parameters. Figure 7 using GRM model fits in a linear regression line with most Indian participants located on the lower end of the line and all U.S. participants on the upper end of the line. When several of them have the same raw score, their latent traits value can still be differentiated from each other.

Limitations of this study lie in the item selection, dealing with missing data and incomplete demographic data collection. The original survey had 34 items from a civic readiness survey item inventory with four categories (Tedeschi, 2021). We only chose 12 of them for the convenience of finding prior means and SDs in the literature. Therefore, the internal consistency of the 12 items was not good enough (Cronbach's $\alpha = 0.63$). One participant with a few missing data were deleted without analysis.

Future studies are expected to develop the model from item-level to test-level. A multi-dimensional model will be investigated so as to explain the latent traits in a more detailed hierarchical construct.

References

Association of American Colleges and Universities. (2011). *The LEAP vision for learning.* Retrieved from http:// leap.aacu.org/toolkit/wp-content/ uploads/2010/12/LEAP-Vision_ Summary.pdf.

Ajaps, S. O., & Obiagu, A. N. (2021). Increasing civic engagement through civic education: A Critical Consciousness Theory Perspective. *Journal of Culture and Values in Education*, *4*(1), 64-87.

Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of educational statistics*, *17*(3), 251-269.

Baumann, P., Millard, M., & Hamdorf, L. (2014). *State civic education policy framework*. Education Commission of the States.

Chi, B., Jastrzab, J., & Melchior, A. (2006). *Developing indicators and measures of civic outcomes for elementary school students* (CIRCLE Working Paper No .47). Center for Information and Research on Civic Learning and Engagement.

Chung, H. L., & Probert, S. (2011). Civic engagement in relation to outcome expectations among African American young adults. *Journal of Applied Developmental Psychology, 32*(4), 227-234.

Evans, R. T., & Kilinc, E. (2013). History of pace-based education in the social studies field. *Journal of Social Sciences*, 6(14).

Flanagan, C. A., Syvertsen, A.K., & Stout, M. D. (2007). *Civic measurement models: Tapping adolescents' civic engagement (CIRCLE Working Paper No. 55)*. Center for Information and Research on Civic Learning and Engagement.

Hancock, G. R., & Mueller, R. O. (2001). Factor analysis and latent structure: Confirmatory factor analysis. *International encyclopedia of the social and behavioral sciences*, 5239-5244.

Hoofs, H., van de Schoot, R., Jansen, N. W., & Kant, I. (2018). Evaluating model fit in Bayesian confirmatory factor analysis with large samples: Simulation study introducing the BRMSEA. *Educational and psychological measurement*, *78*(4), 537-568.

Kahne, J. E., & Sporte, S. E. (2008). Developing citizens: The impact of civic learning opportunities on students' commitment to civic participation. *American Educational Research Journal, 45*(3), 738–766.

Kahne, J., Crow, D., & Lee, N.-J. (2013). Different pedagogy, different politics: High school learning opportunities and youth political engagement. *Political Psychology, 34*(3), 419–441.

Kieftenbeld, V., & Natesan, P. (2012). Recovery of graded response model parameters: A comparison of marginal maximum likelihood and Markov chain Monte Carlo estimation. *Applied Psychological Measurement*, *36*(5), 399-419.

Lorenzini, M. (2013). From global knowledge to global civic engagement. *Journal of Political Science Education, 9*(4), 417-435.

May, H. (2006). A multilevel Bayesian item response theory method for scaling socioeconomic status in international studies of education. *Journal of Educational and Behavioral Statistics*, *31*(1), 63-79.

Natesan, P., Nandakumar, R., Minka, T., & Rubright, J. D. (2016). Bayesian prior choice in IRT estimation using MCMC and variational Bayes. *Frontiers in psychology*, *7*, 1422.

Saban, M. V. E., & Saban, G. A. S. (2020). Sense of Civic Responsibility of National Service Training Program Completers and Non-Completers in a University Setting in the Philippines. *International Forum Journal, 3(*2), 163-177.

Syvertsen, A. K., Wray-Lake, L., & Metzger, A. (2015). Youth civic and character measures toolkit. Search Institute. Retrieved from https://www.search-institute .org/downloadable/Youth-Civic-Character-Measures-Toolkit .pdf

Tedeschi, S. E., Brodersen, R. M., Schramm, K., Haines, M., Liu, J., McCullough, D., Eide, M., & Cherasaro, T. (2021). *Measuring Civic Readiness: A Review of Survey Scales*. REL 2021. Regional Educational Laboratory Central. Retrieved from http://ies .ed .gov/ncee/edlabs .

White, C. (2020). Wielding social media in the cyber-arena: Globalism, nationalism, and civic education. *Research in Social Sciences and Technology, 5*(1), 1–21.

Wicks, R. H., LeBlanc Wicks, J., Morimoto, S. A., Maxwell, A., & Ricker Schulte, S. (2014). Correlates of political and civic engagement among youth during the 2012 presidential campaign. *American Behavioral Scientist, 58*(5), 622–644.

Wilson-Daily, A. E., Kemmelmeier, M., & Prats, J. (2018). Intergroup contact versus conflict in Catalan high schools: A multilevel analysis of adolescent attitudes toward immigration and diversity. *International Journal of Intercultural Relations, 64*, 12-28.