Measurement in Educational Research
Volume 5, Issue 1, 2025, 7-15
Available online: https://ejournal.ressi.id/index.php/meter

IRT parameter estimation with Bayesian MCMC methods for small samples in Islamic schools

Muhammad Ali Gunawan [1, a] *, Nor Syamimi Mohamed Adnan [2, a], Ari Setiawan [3, b]

[1] Sekolah Tinggi Agama Islam Ki Ageng Pekalongan. Jl. Simpang Tiga, Pekalongan, 51181, Indonesia
[2] Universiti Malaysia Perlis. Kuala Perlis, 02600 Arau, Perlis, Malaysia
[3] Universitas Sarjanawiyata Tamansiswa. Jl. Batikan, UH-III Yogyakarta 55167, Indonesia
[a] m.aligunawan@staikap.ac.id; [b] syamimiadna@unimap.edu.my; [c] ari.setiawan@ustjogja.ac.id
* Corresponding Author.

Received: 8 January 2025; Revised: 9 February 2025; Accepted: 24 April 2025; Published: 26 April 2025

Abstract: This study aims to estimate item parameters in Item Response Theory (IRT) using the Bayesian Markov Chain Monte Carlo (MCMC) method in the context of Islamic schools in Pekalongan Regency/City, where small sample sizes pose a challenge. Unlike conventional methods such as maximum likelihood estimation, which tend to yield biased results with limited data, Bayesian MCMC incorporates prior knowledge and contextual information to improve estimation accuracy. Simulated datasets with varying sample sizes (30, 100, 300, 1000) and item numbers (10, 25, 30, 40) were used to compare the performance of Bayesian MCMC with traditional IRT methods. The results show that Bayesian MCMC produces more stable and accurate estimates, particularly in small-sample conditions. These findings suggest that Bayesian approaches are effective for psychometric analysis in Islamic education settings. The study concludes that Bayesian MCMC is a valuable method for improving the robustness of item parameter estimation in limited-data contexts.
Keywords: Bayesian Methods; Educational Assessment; Islamic Schools; Item Response Theory (IRT); Markov Chain Monte Carlo (MCMC).

How to Cite: Gunawan, M. A., Adnan, N. S. M., & Setiawan, A. (2025). IRT parameter estimation with Bayesian MCMC methods for small samples in Islamic schools. *Measurement In Educational Research*, *5*(1), 7-15. https://doi.org/10.33292/meter.v5i1.378

INTRODUCTION

Item Response Theory (IRT) has played a crucial role in educational assessment, particularly in modeling the relationship between students' latent traits (such as abilities or knowledge) and their responses to test items. The application of IRT offers numerous advantages, such as providing more detailed diagnostic information about individual performance and improving the quality of test design (Embretson & Reise, 2013). However, the use of IRT models typically requires large sample sizes to produce reliable parameter estimates. In educational contexts where data collection is limited, such as in rural or small-scale schools like many Islamic schools in Indonesia, obtaining sufficient sample sizes is often a challenge. This issue becomes particularly important in evaluating student performance in under-resourced or specialized educational environments (J. P. Fox, 2010). Addressing this limitation is critical for improving the quality of educational measurement in these settings, ensuring that all students are assessed equitably regardless of institutional size.

Despite the well-documented benefits of IRT, there is a significant gap in the research regarding its application in contexts with small sample sizes. Traditional IRT models are often sensitive to sample size, and small samples can lead to unstable parameter estimates, making it difficult to draw meaningful inferences about student ability (Linden & Hambleton, 1997). Current research tends to focus on large-scale educational settings, leaving small-scale or specialized schools, such as Islamic schools, underrepresented in the literature (Cao et al., 2022; Zhang et al., 2022). In particular, studies that explore advanced statistical methods capable of compensating for small sample sizes, such as Bayesian estimation techniques, are limited. Addressing this gap is crucial for ensuring that smaller educational institutions can benefit from the sophisticated measurement capabilities of IRT without sacrificing accuracy due to sample size constraints (Hambleton et al., 1991).

Recent advances in Bayesian methods, particularly the use of Markov Chain Monte Carlo (MCMC) for parameter estimation, have shown promise in improving IRT models' performance with small samples (Fabreti & Höhna, 2022; Hanada & Matsuura, 2022; South et al., 2024). Bayesian methods differ from traditional maximum likelihood approaches by incorporating prior information into the estimation process, allowing for more robust parameter estimates when data is limited (Chater et al., 2020; J. P. Fox, 2010; Pane et al., 2018; van de Schoot et al., 2021; Wilson et al., 2022). Some studies have demonstrated that the use of robust priors, such as the Cauchy prior, can mitigate the biases typically associated with small sample sizes, offering more reliable results(Assaf et al., 2021; Kaikkonen et al., 2021; Kaplan, 2021; Khosravi-Farmad & Ghaemi-Bafghi, 2020; Lotfi et al., 2022; Lyle et al., 2020; Nishio et al., 2023; Srivastava & Xu, 2020; Stone & Zhu, 2015; Taka et al., 2020). Additionally, the use of MCMC sampling in Bayesian frameworks has been increasingly applied in educational research, particularly in contexts where data is scarce or expensive to collect (Bürkner, 2019; Mohammadi & Rezaei, 2020; Natesan, 2011; Plummer, 2024; Vasishth et al., 2023). However, while promising, these methods are not yet widely adopted in educational settings like Islamic schools, where the combination of small samples and unique educational frameworks presents additional challenges(Houts et al., 2018; Lüdtke et al., 2021; Rainey & McCaskey, 2021; Vaheoja, 2019).

Given the gaps in the application of IRT in small sample contexts and the potential of Bayesian methods to address these issues, this study seeks to answer the following research question: How can innovative Bayesian MCMC methods improve IRT parameter estimation in small sample settings, specifically within Islamic schools in Indonesia? This question is grounded in the need to explore statistical methods that maintain the robustness of parameter estimates even when sample sizes are below traditional thresholds. By focusing on Islamic schools, this study also aims to provide insights into the unique challenges faced by smaller educational institutions that may lack the resources or access to large-scale testing programs (Avvisati, 2020; Deonovic et al., 2020).

This study is significant for several reasons. First, it seeks to address a major gap in the current literature by investigating how advanced Bayesian methods can be adapted to educational contexts with small samples. Second, by focusing on Islamic schools in Indonesia, the study provides insights into a unique and underrepresented educational sector, where traditional assessment methods may not be as effective. The findings have the potential to inform policy decisions and improve assessment practices in similar contexts globally, particularly in rural or specialized educational settings. Moreover, the study contributes to the broader field of educational measurement by exploring the practical applications of innovative statistical methods in real-world educational environments (Linden & Hambleton, 1997).

The primary hypothesis of this study is that Bayesian MCMC methods with robust priors, such as Cauchy priors, will produce more accurate IRT parameter estimates compared to traditional maximum likelihood estimation in small sample contexts. Specifically, it is expected that the Bayesian approach will result in more stable estimates of item difficulty, discrimination, and guessing parameters, even when sample sizes are significantly smaller than typically required for IRT models. This hypothesis is based on previous research demonstrating the effectiveness of Bayesian methods in compensating for small sample sizes through the incorporation of prior information (Assaf et al., 2021; Chater et al., 2020; J.-P. Fox, 2010; Kaikkonen et al., 2021; Lotfi et al., 2022; Pane et al., 2018; van de Schoot et al., 2021; Wilson et al., 2022). Furthermore, this study hypothesizes that the application of these methods in Islamic schools will demonstrate their practical utility in real-world educational assessment, ultimately leading to improved decision-making processes in student evaluation (Chang-Tik, 2022; Shepard et al., 2018).

METHOD

This study employs a Bayesian Markov Chain Monte Carlo (MCMC) approach to estimate item parameters within the framework of Item Response Theory (IRT), focusing on the unique context of Islamic education. Simulated datasets with varying sample sizes—30, 100, 300, and 1000—and different test lengths of 10, 25, 30, and 40 items allow us to systematically assess the impact of these variables on the effectiveness of Bayesian MCMC compared to traditional Maximum Likelihood Estimation (MLE) techniques. The simulations use a two-parameter logistic model (2PLM) to capture essential item characteristics such as discrimination and difficulty, omitting the guessing parameter present in the 3PLM (Hambleton et al., 1991). These combinations of sample sizes and test lengths, including (30, 10), (100, 25), (300, 30) and (1000, 40), reflect the specific educational settings and learning outcomes of Islamic schools, while latent trait distributions simulate the diversity of student populations. Prior knowledge parameters are informed by recent 2022-2024 student learning outcome reports, ensuring that the priors are rooted in empirical data and relevant to the assessment practices in Islamic education (Islam et al., 2021; Sukenti et al., 2021).

Data analysis is conducted using the rjags package in R, implementing Bayesian MCMC algorithms to estimate item parameters. MCMC chains are run for a sufficient number of iterations to ensure convergence, with diagnostic tools used to assess stability and mixing. The posterior distributions generated from the MCMC process provide robust estimates of the uncertainty surrounding each parameter (Hanada & Matsuura, 2022; Neklyudov et al., 2020; Plummer, 2024). We evaluate the performance of Bayesian MCMC estimates by comparing them to MLE across various sample sizes using metrics like bias, root mean square error (RMSE), and coverage probabilities (Lüdtke et al., 2021). Sensitivity analyses explore how different priors affect parameter estimates in small samples, emphasizing the robustness of the Bayesian approach, particularly in the context of Islamic schools (Sukenti et al., 2021).

RESULTS AND DISCUSSION

The results of this study indicate that MLE consistently has higher Bias and RMSE values compared to MCMC, especially for the parameter a in P1, where MLE Bias reaches -3,353, while MCMC Bias is significantly lower at -0.846. This suggests that MLE struggles with accurately estimating certain parameters, leading to larger deviations from the true value. In contrast, MCMC demonstrates greater precision, with Bias values consistently closer to zero and lower RMSE, indicating more accurate and reliable parameter estimations.

Table 1. Comparison of Bias and RMSE for MLE and MCMC Parameter Estimations

Parameter	MLE		MCMC	
	Bias	RMSE	Bias	RMSE
(a)P1	-3,353	10,609	-0,846	0,944
(b)P1	0,273	1,591	0,538	1,152
(theta)P1	-1,540	6,100	-0,359	1,071
(a)P2	-0,865	2,991	-0,980	0,982
(b)P2	0,004	0,955	-0,352	1,174
(theta)P2	-0,431	1,973	-0,024	0,991
(a)P3	-1,459	4,163	-1,004	1,004
(b)P3	0,273	1,170	-0,159	0,849
(theta)P3	-0,593	2,666	0,027	0,995
(a)P4	-0,909	1,499	-1,008	1,008
(b)P4	-0,258	1,082	0,198	1,054
(theta)P4	-0,584	1,290	0,020	0,980

For instance, the MLE-RMSE for *a* in P1 is 10,609, whereas MCMC-RMSE for the same parameter is only 0.944, highlighting MCMC's ability to reduce error. Across other parameters and conditions, such as *b* in P2 and θ in P3, MCMC continues to show better performance with lower Bias and RMSE. These results suggest that MCMC is the preferred method for complex, multidimensional estimations in this research, as shown in Table 1.

Figure 1, represents the differences between the MLE and MCMC estimations across all treatments.

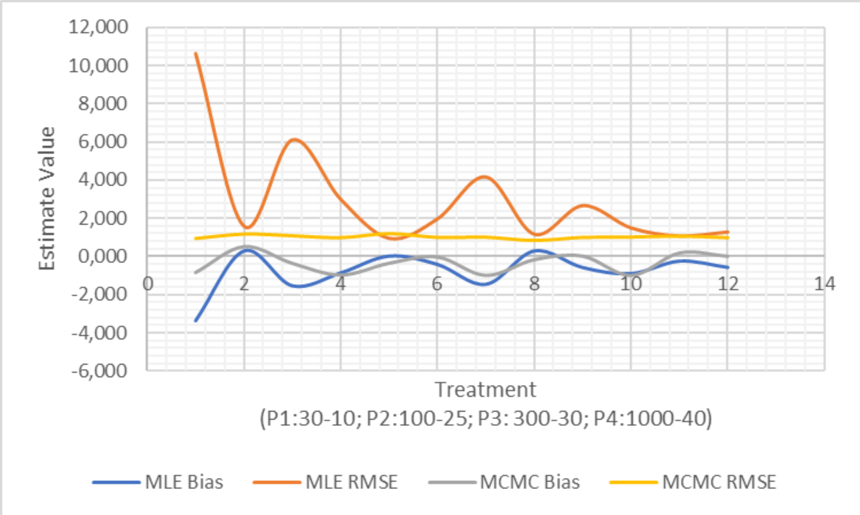

Figure 1. Comparison of Bias and RMSE Between MLE and MCMC Parameter Estimation Methods

Figure 1, show that for Treatment 1, MLE exhibits substantial bias and error. The MLE Bias (blue line) starts at a value below -4,000, indicating a significant underestimation of the treatment effect. In contrast, the MLE RMSE (orange line) reaches a peak of nearly 10,000 for Treatment 1, signifying considerable error in estimation. This suggests that MLE struggles to estimate the true effect size accurately under certain treatment conditions, especially in the early stages. The combination of high bias and high RMSE in Treatment 1 reflects the poor performance of MLE for that specific treatment. As the treatments progress (from Treatment 2 onward), the MLE bias reduces slightly, but it remains variable, with slight overestimations and underestimations in other treatments.

From Treatment 2 to Treatment 4, the MLE method's bias improves, fluctuating closer to zero, but it still shows some inconsistency. The MLE RMSE reduces from its peak at Treatment 1 but still exhibits high values between 2,000 and 3,000, indicating lingering estimation errors. In Treatment 3, for instance, the MLE bias approaches 0, but RMSE is still around 2,500, meaning the method is more accurate in estimating the treatment effect but still prone to error. These fluctuations imply that while MLE can deliver more accurate estimates for some treatments, it is not uniformly reliable across all treatments. The high RMSE values observed in multiple treatments reinforce the instability of the method, particularly for treatment effects with high variability or complex interactions.

On the other hand, the Bayesian MCMC method provides far more stable results across all treatments. The MCMC Bias (gray line) consistently stays close to zero, fluctuating between -500 and 500 across all treatments, indicating minimal bias in estimating the true effects. In Treatment 1, where MLE demonstrated significant bias and error, MCMC Bias is only slightly negative, around -200, demonstrating that MCMC estimates the treatment effect far more accurately. Similarly, the MCMC-RMSE (yellow line) remains low and steady, consistently hovering around 500 across all treatments. This stability shows that MCMC offers more reliable treatment effect estimates, minimizing both bias and error compared to MLE.

The findings of this study on item response theory (IRT) parameter estimation using Bayesian Markov Chain Monte Carlo (MCMC) methods have profound implications for Islamic education, particularly given the unique curricular demands faced by Islamic schools. These institutions operate under a dense curriculum that includes general subjects such as mathematics, science, and language arts while also requiring deep engagement with religious studies, including the Qur'an, Hadith, Fiqih, and Aqidah Ahlaq. This dual focus creates a distinct educational environment where assessment practices must effectively measure both secular and religious knowledge, making the need for reliable measurement methods even more critical.

The integration of religious education alongside general subjects results in a complex and multifaceted learning environment. Students must achieve proficiency in diverse disciplines while developing a solid foundation in their faith, which poses significant challenges for educational assessment. The superior performance of MCMC in providing accurate estimates of IRT parameters is particularly relevant in this context, especially given that Islamic schools often cater to small class sizes and diverse student populations. Traditional estimation methods like Maximum Likelihood Estimation (MLE) can yield biased results, as demonstrated in this study. MCMC's ability to produce reliable parameter estimates can help educators develop more equitable assessments that accurately reflect students' capabilities, ensuring that all students are evaluated fairly across both secular and religious subjects.

In comparison to the findings of this study, several previous research efforts have demonstrated similar advantages of using Bayesian MCMC methods over traditional approaches like Maximum Likelihood Estimation (MLE) in educational measurement. For instance, (Asosega et al., 2022; Candès & Sur, 2020; Urban & Bauer, 2021) highlighted that MLE tends to perform poorly in scenarios involving small sample sizes or high-dimensional data, resulting in high root mean square error (RMSE) values, as also seen in this study. In contrast, MCMC consistently yields lower RMSE and bias values (Tolba, 2022), indicating its superior capacity for parameter estimation under such challenging conditions. The present research echoes this result, particularly for parameter a in Treatment P1, where MLE's RMSE exceeded 10,000, while MCMC remained under 1,000, reaffirming MCMC's robustness.

Additionally, the findings align with (Alsefri et al., 2020; Buyl & De Bie, 2020; Herrera et al., 2022), who similarly observed that Bayesian methods minimize bias in parameter estimation, especially when handling sparse data. This study corroborates that MCMC delivers more stable and less biased estimates, crucial in Islamic schools, where diverse student populations and small class sizes often complicate assessment accuracy. (Hoofs et al., 2018; Tian et al., 2024; van de Schoot et al., 2021) also support these findings, suggesting that MCMC's ability to incorporate prior distributions allows for more stable and nuanced estimation of parameters, even in sparse or multidimensional data contexts.

On the technical side, (Nijkamp et al., 2020) also found that MCMC provides superior stability in parameter estimates compared to MLE, particularly across multiple treatments and diverse educational models. The current study's results, showing consistently lower RMSE values for parameters b and θ across different treatments, further bolster these findings. However, the computational challenges associated with Bayesian MCMC, highlighted by (Fabreti & Höhna, 2022; Hanada & Matsuura, 2022; Neklyudov et al., 2020), are consistent with this research. They caution that while MCMC improves accuracy, its implementation can be resource-intensive, a limitation that this study also acknowledges. The resource-intensive nature of MCMC may pose challenges for large-scale assessments, as noted in previous studies (Papamarkou et al., 2024). However, the potential benefits of enhanced accuracy and reliability in educational measurement could outweigh these challenges, particularly in high-stakes environments where precision is critical.

CONCLUCION

The results of this study indicate that Bayesian MCMC methods can significantly enhance measurement and assessment practices within Islamic education. By effectively addressing the complexities of a curriculum that integrates both secular and religious subjects, MCMC offers a reliable framework for evaluating student performance. This approach not only fosters more accurate assessments but also aligns with the goals of Islamic education cultivating knowledgeable, ethical individuals prepared to navigate both academic and moral challenges. As Islamic schools evolve, adopting innovative assessment methods will be crucial for meeting diverse student needs and contemporary educational demands.

However, this study has several limitations that must be acknowledged. Conducted within a specific sample of Islamic schools in Pekalongan, the findings may not be generalizable to other regions with different curricular frameworks. Additionally, while the sample sizes were adequate, they may not fully capture the variability among student populations. The study primarily focused on mathematical literacy, leaving the integration of other subjects underexplored. Future research should investigate the applicability of MCMC methods across a wider range of educational outcomes, including non-cognitive skills and character development, and consider longitudinal studies to track performance over time. Finally, cross-cultural studies could yield valuable insights into the effectiveness of Bayesian methods in various educational contexts, ultimately informing best practices for educational measurement in Islamic education and beyond.

Contribution of the Article to the Related Field of Study

This article contributes to the field of psychometrics by developing a Bayesian Markov Chain Monte Carlo (MCMC) approach for parameter estimation in Item Response Theory (IRT), particularly in the context of Islamic schools in Pekalongan Regency/City, which face challenges related to small sample sizes. This approach enhances estimation accuracy by leveraging prior distributions based on contextual knowledge and addresses the biases often encountered in traditional methods such as maximum likelihood estimation. The study is highly relevant to Islamic education, offering practical guidance for data-driven educational evaluation in resource-constrained settings. Furthermore, it enriches the literature on Bayesian methods in psychometric analysis, promoting the adoption of more robust and flexible approaches for IRT model analysis while providing a foundation for future research in educational assessment.

REFERENCES

Alsefri, M., Sudell, M., García-Fiñana, M., & Kolamunnage-Dona, R. (2020). Bayesian joint modelling of longitudinal and time to event data: A methodological review. *BMC Medical Research Methodology*, *20*(1). https://doi.org/10.1186/s12874-020-00976-2

Asosega, K. A., Iddrisu, W. A., Tawiah, K., Opoku, A. A., & Okyere, E. (2022). Comparing Bayesian and Maximum Likelihood Methods in structural equation modelling of university student satisfaction: an empirical analysis. *Education Research International*, *2022*. https://doi.org/10.1155/2022/3665669

Assaf, A. G., Tsionas, M., Kock, F., & Josiassen, A. (2021). A Bayesian non-parametric stochastic frontier model. *Annals of Tourism Research*, *87*. https://doi.org/10.1016/j.annals.2020.103116

Avvisati, F. (2020). The measure of socio-economic status in PISA: a review and some suggested improvements. In *Large-Scale Assessments in Education* (Vol. 8, Issue 1). Springer. https://doi.org/10.1186/s40536-020-00086-x

Bürkner, P.-C. (2019). *Bayesian item response modeling in R with brms and Stan*. http://arxiv.org/abs/1905.09501

Buyl, M., & De Bie, T. (2020). *DeBayes: a Bayesian method for debiasing network embeddings*.

Candès, E. J., & Sur, P. (2020). The phase transition for the existence of the maximum likelihood estimate in high-dimensional logistic regression. *Annals of Statistics*, *48*(1), 27–42. https://doi.org/10.1214/18-AOS1789

Cao, H., Lee, C.-J., Iqbal, S., Czerwinski, M., Wong, P., Rintel, S., Hecht, B., Teevan, J., & Yang, L. (2022). Large scale analysis of multitasking behavior during remote meetings. *IEEE*

International Conference on Program Comprehension, *2022-March*, 36–47. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Chang-Tik, C. (2022). Introduction: collaborative active learning—strategies, assessment and feedback. In C. Chang-Tik, G. Kidman, & M. Tee (Eds.), *Collaborative Active Learning: Practical Activity-Based Approaches to Learning, Assessment and Feedback* (pp. 3–32). Springer Nature. https://doi.org/10.1007/978-981-19-4383-6

Chater, N., Zhu, J. Q., Spicer, J., Sundh, J., León-Villagrá, P., & Sanborn, A. (2020). Probabilistic biases meet the bayesian brain. *Current Directions in Psychological Science*, *29*(5), 506–512. https://doi.org/10.1177/0963721420954801

Deonovic, B., Bolsinova, M., Bechger, T., & Maris, G. (2020). A Rasch model and rating system for continuous responses collected in large-scale learning Systems. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.500039

Embretson, S. E. ., & Reise, S. P. . (2013). *Item response theory*. Taylor and Francis.

Fabreti, L. G., & Höhna, S. (2022). Convergence assessment for Bayesian phylogenetic analysis using MCMC simulation. *Methods in Ecology and Evolution*, *13*(1), 77–90. https://doi.org/10.1111/2041-210X.13727

Fox, J. P. (2010). *Bayesian item response modeling theory and applications*. Springer Science+Business Media. LLC. https://doi.org/10.1007/978-1-4419-0742-4

Fox, J.-P. (2010). *Bayesian item response modeling: Theory and Applications*. Springer Science+Business Media, LLC. https://doi.org/DOI 10.1007/978-1-4419-0742-4

Hambleton, R. K., Swaminathan, H., & Rogers, J. H. (1991). *Fundamentals of item response theory (measurement methods for the social science)*. Sage Publication, Inc.

Hanada, M., & Matsuura, S. (2022). MCMC from scratch: A practical introduction to Markov Chain Monte Carlo. In *MCMC from Scratch: A Practical Introduction to Markov Chain Monte Carlo*. Springer Nature. https://doi.org/10.1007/978-981-19-2715-7

Herrera, P. A., Marazuela, M. A., & Hofmann, T. (2022). Parameter estimation and uncertainty analysis in hydrological modeling. In *Wiley Interdisciplinary Reviews: Water* (Vol. 9, Issue 1). John Wiley and Sons Inc. https://doi.org/10.1002/wat2.1569

Hoofs, H., van de Schoot, R., Jansen, N. W. H., & Kant, Ij. (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. https://doi.org/10.1177/0013164417709314

Houts, C. R., Morlock, R., Blum, S. I., Edwards, M. C., & Wirth, R. J. (2018). Scale development with small samples: a new application of longitudinal item response theory. *Quality of Life Research*, *27*(7), 1721–1734. https://doi.org/10.1007/s11136-018-1801-z

Islam, M. S., Hasan, M. K., Sultana, S., Karim, A., & Rahman, M. M. (2021). English language assessment in Bangladesh today: principles, practices, and problems. In *Language Testing in Asia* (Vol. 11, Issue 1). Springer. https://doi.org/10.1186/s40468-020-00116-z

Kaikkonen, L., Parviainen, T., Rahikainen, M., Uusitalo, L., & Lehikoinen, A. (2021). Bayesian networks in environmental risk assessment: A Review. *Integrated Environmental Assessment and Management*, *17*(1), 62–78. https://doi.org/10.1002/ieam.4332

Kaplan, D. (2021). On the quantification of model uncertainty: A Bayesian perspective. *Psychometrika*, *86*(1), 215–238. https://doi.org/10.1007/s11336-021-09754-5

Khosravi-Farmad, M., & Ghaemi-Bafghi, A. (2020). Bayesian decision network-based security risk management framework. *Journal of Network and Systems Management*, *28*(4), 1794–1819. https://doi.org/10.1007/s10922-020-09558-5

Linden, W. J. van der., & Hambleton, R. K. . (1997). *Handbook of modern item response theory*. Springer.

Lotfi, S., Izmailov, P., Benton, G., Goldblum, M., & Wilson, A. G. (2022). Bayesian Model selection, the marginal Likelihood, and Generalization. *39th International Conference on Machine Learning*, 1–25.

Lüdtke, O., Ulitzsch, E., & Robitzsch, A. (2021). A comparison of penalized maximum Likelihood estimation and Markov Chain Monte Carlo techniques for estimating confirmatory factor analysis models with small sample sizes. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.615162

Lyle, C., Schut, L., Ru, B., Gal, Y., & Van Der Wilk, M. (2020). A Bayesian perspective on training speed and model selection. *34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, 1–13.

Mohammadi, M., & Rezaei, J. (2020). Bayesian best-worst method: A probabilistic group decision making model. *Omega (United Kingdom)*, *96*. https://doi.org/10.1016/j.omega.2019.06.001

Natesan, P. (2011). A review of bayesian item response modeling: Theory and applications. In *Journal of Educational and Behavioral Statistics* (Vol. 36, Issue 4). https://doi.org/10.3102/1076998611411919

Neklyudov, K., Welling, M., Egorov, E., & Vetrov, D. (2020). *Involutive MCMC: a Unifying Framework*.

Nijkamp, E., Hill, M., Han, T., Zhu, S.-C., & Wu, Y. N. (2020). On the anatomy of MCMC-Based Maximum Likelihood Learning of Energy-Based Models. *The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)*. www.aaai.org

Nishio, M., Kobayashi, D., Matsuo, H., Urase, Y., Nishioka, E., & Murakami, T. (2023). Bayesian multidimensional nominal response model for observer study of radiologists. *Japanese Journal of Radiology*, *41*(4), 449–455. https://doi.org/10.1007/s11604-022-01366-y

Pane, D. N., Fikri, M. EL, & Ritonga, H. M. (2018). Bayesian statistics the fun way. In *Journal of Chemical Information and Modeling* (Vol. 53, Issue 9).

Papamarkou, T., Skoularidou, M., Palla, K., Aitchison, L., Arbel, J., Dunson, D., Filippone, M., Fortuin, V., Hennig, P., Miguel Hernández-Lobato, J., Hubin, A., Immer, A., Karaletsos, T., Emtiyaz Khan, M., Kristiadi, A., Li, Y., Mandt, S., Nemeth, C., Osborne, M. A., & J Rudner, T. G. (2024). *Position: Bayesian Deep Learning is Needed in the Age of Large-Scale AI*.

Plummer, M. (2024). *Annual Review of statistics and its application simulation-based Bayesian Analysis*. *12*, 49. https://doi.org/10.1146/annurev-statistics-122121

Rainey, C., & McCaskey, K. (2021). Estimating logit models with small samples. *Political Science Research and Methods*, *9*(3), 549–564. https://doi.org/10.1017/psrm.2021.9

Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Classroom assessment principles to support learning and avoid the harms of testing. *Educational Measurement: Issues and Practice*, *37*(1), 52–57. https://doi.org/10.1111/emip.12195

South, L. F., Riabiz, M., Teymur, O., & Oates, C. J. (2024). *Annual review of statistics and its application postprocessing of MCMC*. https://doi.org/10.1146/annurev-statistics-040220

Srivastava, S., & Xu, Y. (2020). *Distributed Bayesian Inference in Linear Mixed-Effects Models*.

Stone, C. A. ., & Zhu, Xiaowen. (2015). *Bayesian analysis of item response theory models using SAS®*. SAS.

Sukenti, D., Tambak, S., & Siregar, E. (2021). Learning assessment for madrasah teacher: strengthening islamic psychosocial and emotional intelligence. *Al-Ishlah: Jurnal Pendidikan*. https://doi.org/DOI:10.35445/alishlah.v13i1.552

Taka, E., Stein, S., & Williamson, J. H. (2020). Increasing interpretability of bayesian probabilistic programming models through interactive representations. *Frontiers in Computer Science*, *2*. https://doi.org/10.3389/fcomp.2020.567344

Tian, Q., Lewis-Beck, C., Niemi, J. B., & Meeker, W. Q. (2024). Specifying prior distributions in reliability applications. *Applied Stochastic Models in Business and Industry*, *40*(1), 5–62. https://doi.org/10.1002/asmb.2752

Tolba, Ahlam. (2022). Bayesian and non-bayesian estimation methods for simulating the parameter of the akshaya distribution. *Computational Journal of Mathematical and Statistical Sciences*, *1*(1), 13–25. https://doi.org/10.21608/cjmss.2022.270897

Urban, C. J., & Bauer, D. J. (2021). A deep learning algorithm for high-dimensional exploratory item factor analysis. *Psychometrika*. https://doi.org/doi:%2010.1007/s11336-021-09748-3

Vaheoja, M. (2019). *Finding equivalent standards in small samples*. Springer International Publishing. https://doi.org/10.1007/978-3-030-18480-3_9

van de Schoot, R., Depaoli, S., Gelman, A., King, R., Kramer, B., Kaspar, M., Tadesse, M. G., Vannucci, M., Veen, D., Willemsen, J., & Yau, C. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, *1*(1). https://doi.org/10.1038/s43586-020-00003-0

Vasishth, S., Yadav, H., Schad, D. J., & Nicenboim, B. (2023). Sample size determination for bayesian hierarchical models commonly used in psycholinguistics. *Computational Brain and Behavior*, *6*(1), 102–126. https://doi.org/10.1007/s42113-021-00125-y

Wilson, K. J., Williamson, S. F., Allen, A. J., Williams, C. J., Hellyer, T. P., & Lendrem, B. C. (2022). Bayesian sample size determination for diagnostic accuracy studies. *Statistics in Medicine*, *41*(15), 2908–2922. https://doi.org/10.1002/sim.9393

Zhang, Z., Tian, J., Zhao, Z., Zhou, W., Sun, F., Que, Y., & He, X. (2022). Factors influencing vocational education and training teachers' professional competence based on a large-scale diagnostic method: a decade of data from China. *Sustainability (Switzerland)*, *14*(23). https://doi.org/10.3390/su142315871

Conflict of Interest Statement: The Author(s) declares that the research was conducted in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.

Measurement in Educational Research is an open access and peer-reviewed journal published bt Research and Social Study Institute, Indonesia

Open Access