TPM Vol. 32, No. S6, 2025
ISSN: 1972-6325
https://www.tpmap.org/

MIND THE GAP: STRATEGIES FOR MANAGING MISSING DATA IN PSYCHOMETRIC MODELING WITH SPARSE DATA FROM MAINSTREAM SCHOOL POPULATIONS

MUHAMMAD NAEEM SARWAR[1], ZAHIDA JAVED[1], SYEDA NAUREEN MUMTAZ[1], ZUBAIR YOUNAS[2], SYEDA RABIA BASRI[3], MUHAMMAD ASIF SHAHZAD[4], SUMERA BARI[5], SHAFAQ NAZ[6]

[1]DEPARTMENT OF STEM EDUCATION, UNIVERSITY OF EDUCATION, LAHORE, 54770, PAKISTAN
EDUCATION, LAHORE, 54770, PAKISTAN

[2]ASSISTANT EDUCATION OFFICER, SCHOOL EDUCATION DEPARTMENT, PUNJAB,
EMAIL: aeozubairyounas@gmail.com

[3]DEPARTMENT OF SPECIAL EDUCATION, UNIVERSITY OF EDUCATION, LAHORE, 54770, PAKISTAN

[4]DEPARTMENT OF EDUCATIONAL LEADERSHIP AND POLICY STUDIES, DIVISION OF EDUCATION, UNIVERSITY OF EDUCATION, LAHORE, 54770, PAKISTAN

[5]EDUCATOR DEPARTMENT OF SPECIAL EDUCATION, GOVT. INSTITUTE FOR SLOW LEARNERS, PUNJAB, PAKISTAN

[6]PRIMARY SCHOOL TEACHER, SCHOOL EDUCATION DEPARTMENT, PUNJAB, PAKISTAN

Abstract
Missing data can compromise the validity, reliability, and generalizability of psychometric assessment results, especially in mainstream school populations. This study examines the impact of missing data on educational assessments of students in mainstream education schools. A mathematics achievement test was administered to 275 students, and missing data patterns (MCAR, MAR, MNAR) were analyzed. Missing data were dealt with different approaches including classical techniques (mean imputations, listwise deletions), statistical techniques (Bayesian estimations, multiple imputations), and machine learning approaches (k-nearest neighbors, random forests). Using root mean square error, bias and recovery of psychometric parameters within IRT frameworks evaluated the techniques. Bayesian estimation and multiple imputations generally surpassed the other techniques, on most of the evaluated criteria. The results support the need of making educational assessments more reliable for underrepresented students and the need of data-informed decision making for more representative mainstream school populations.
Keywords: Missing Data, Educational Assessment, Multiple Imputation, Machine Learning, Data Integrity.

INTRODUCTION

Missing information is a prominent problem in educational evaluations, especially in mainstream school populations where learner differences, fatigue, and attention issues lead to non-responses. Not filling in the gaps will lead to spurious and untrustworthy interpretations of score results, which in turn will lead to unfair outcomes in education. The problem of missing data is common in psychometric modeling (Levy & Mislevy, 2017). Current assessments are based on the assumption that we have structured data on hand to measure not-so-latent traits, such as cognitive ability, personality characteristics, and a variety of psychological constructs (Epskamp, 2020). Missing responses can occur in large-scale educational and psychological testing for many reasons, including questions that may be bypassed, respondents may be fatigued, time limitations may exist, and respondents may not fully engage in the testing situation (Epskamp et al., 2018). If missing data are not appropriately handled, they can distort what a psychometric model thinks are the true values of its parameters, lead to false claims about the validity of the model's scores, and adversely affect decisions made by humans and/or algorithms based on these scores (Enders, 2022). Because many psychological and educational assessments are associated with high stakes, it is vitally important to ensure that missing data do not compromise the integrity and fairness of measurement outcomes. (Wu et al., 2015).

In order to preserve the functionality of psychometric models and avoid drawing erroneous conclusions, we must deal with and find solutions to missing information (McNeish, 2017a). The field of statistics is governed by certain

rules and principles and so we cannot avoid missing data. For example, in statistics, the technique of listwise deletion is widely used and accepted; however, we can end up cutting and trimming too much data, and so we run the risk of significantly decreasing the statistical power of the data. On the other hand, we might also consider some rudimentary techniques of imputation such as mean substitution. In such cases, we might see test scores result in a spike, but they don't really reflect the difference (Kline, 2023). In missing data analysis, there is also the expectation maximization technique, Bayesian estimating, and possibly the more widely reported technique of Multiple Imputation which are all centered on providing missing data with answers.

These methods do not just replace missing values with reasonable guesses; they make a series of assumptions about the data and the missingness mechanism and then use those assumptions to maintain the integrity of the data structure while imputing or replacing the missing values (Morin et al., 2016). The determination of the most suitable method for handling missing data depends on several factors. The principal factor is the missing data mechanism. Is it completely missing at random (MCAR), missing at random (MAR), or missing not at random (MNAR)? Other factors include the missingness itself and the model's complexity (Lei & Shiverdecker, 2020). Data that are not abundantly available and contain many missing responses pose special problems for modeling psychometrics (Dai, 2021). In adaptive testing, large-scale surveys, and longitudinal studies, not all respondents answered every item, which led to the creation of sparse datasets (Xiao & Bulut, 2020).

Data sparsity disallows the forming of meaningful distinctions between the test takers as the level of confidence in the estimation of parameters decreases (Huo et al., 2015). Difficulties surrounding model convergence and stability are exacerbated by sparse data too, particularly when advanced psychometric techniques are employed, such as item response theory (IRT) and structural equation modeling (SEM) that suffer from lack of statistical power with small sample sizes (Heine & Tarnai, 2015). Innovation is needed to tackle such challenges as we do smart data utilization, while accepting to work with some missing data.

We cannot reduce this bias entirely, but we can aim for strategies that optimize what we have obtained, which, in the context of this paper, are to find techniques that work well when applied to our sparse datasets. Three broad classes of strategies are available to us; imputation, prediction, and model-based correction (Epskamp et al., 2018). With these conundrums, the primary aim of this study was to systematically determine the effectiveness of a variety of missing data handling techniques when they are used in psychometric modeling. The methods being handled are divided into traditional and advanced varieties. The assessment criteria were ability estimation accuracy, bias reduction, and overall model performance. This work was conducted in two parts. First, part one uses empirical simulations to pit traditional and advanced techniques against one another. The next step is to do applied case studies to determine which techniques, in which situations, meet the best assessment criteria (de Chiusole et al., 2015).

The real value of this study goes far beyond the contribution to theory. It has value in the psychology and education assessment field. We live in times where the need for data-informed decision-making is imperative, especially in the context of talent identification and educational placement, or decisions made in the psychological assessment field. So the need to estimate, to predict, and also to do that fairly and accurately comes to the fore more than ever before. And it’s not only about those big, high-stakes decisions that get made and relied on test data. It’s also about the educational and professional opportunities and treatment of the test takers in the assessment process (Levy & Mislevy, 2017). When we do not have data for a particular test taker, we cannot estimate the latent trait that the test was designed to measure as well (or reliably and validly) as we could if we had data for that test taker. However, we live in a world "with missingness," and sparse data compound this issue. Few studies have compared the effectiveness of various robust estimation methods when data are missing. This study will contribute to that ongoing conversation (Guerra-Urzola et al., 2021).

2. Theoretical Background and Literature Review

A widespread problem in psychometric modeling is missing data, which undermines the correctness and legitimacy of our evaluations (Dai, 2021). When values in a dataset are not recorded and are instead lost because of "missingness." This can occur due to survey non-response, time and space limitations during an evaluation, or technical problems during the measures of data collection (Xiao & Bulut, 2020). If missing data are not properly addressed, they can distort the results and lead to biased estimates (Huo et al., 2015). The nature and pattern of missingness demand a careful analysis by researchers to figure out the best method for addressing it in the applications of psychometrics and, by extension, the best method for handling it in the construction and evaluation of measures of human behavior (Bulut & Kim, 2021).

In 1976, Rubin categorized missing data into three principal types. Missing Completely at Random (MCAR), missing at random (MAR), and missing not at random (MNAR) (Carpenter & Smuk, 2021). Complete missing at random occurs when the likelihood of missing data is completely independent of both observed and non-observed data (Enders, 2022). This kind of missingness does not cause any trouble at all because it does not introduce systematic bias and is therefore perfectly missing at random. Missing at Random (MAR) occurs when missingness is related to the observed data but not to the missing values (Mohan & Pearl, 2021). For example, respondents with lower cognitive ability scores were less likely to complete the test (Newgard & Lewis, 2015). The most

challenging type, MNAR, arises when the probability of missing data is affected by the missing values, which can lead to significant biases in the estimation (Pedersen et al., 2017). Understanding these differences is vital for the correct choice of strategies for dealing with missing data in psychometric modeling. Fatigue from taking tests is a prevalent reason for this behavior (Tang & Ishwaran, 2017). Respondents may not answer questions and instead leave them blank, especially when it comes to the portion of the assessment that is long and toward the end of that segment (Mohan & Pearl, 2021). Several sources may lead to missing psychometric data. Technical issues such as software malfunctions or poor data collection techniques can also cause missing data (Enders, 2022).

Item non-response is a leading cause of missing data, particularly in self-reported surveys. This occurred when participants failed to respond to certain survey items. These are not answered for one of two reasons: the items are too sensitive or the items are too difficult to respond to (McNeish, 2017b). Additionally, methods of adaptive testing, such as Computerized Adaptive Testing (CAT), may purposely exclude items based on the performance of an individual being tested. This produces a kind of missing data that must be managed with care (Wei et al., 2018). There are datasets with missing responses, and there are responses that are missing in these datasets (sparse data). This sparsity occurs in two well-known ways; when there are too many variables and insufficient data to fill them up or when we do not have enough variables and samples from our population to fill out the few variables we managed to put together. We try to do psychometrics with them, all the while knowing that precise parameter estimates are impossible because we have these two well-defined paths to imprecision (Lang & Little, 2018). Unlike datasets with missing values that occur randomly, sparse data usually results from intentional testing designs, such as matrix sampling, where different groups of respondents are given different sets of questions to reduce the length of the test. Large-scale surveys or longitudinal studies in which participants drop out over time can yield not only low response rates but also sparse data (Wang et al., 2019).

A structure that is frequently unfamiliar to the dataset builders. Sparse datasets contain a small number of adequately informative instances, which can also give rise to convergence issues when the model parameters are being estimated (Antholzer et al., 2019). Data that convey weak information can considerably affect how accurate and how sound the estimation of abilities is in psychometric mode (Patra et al., 2015). When a large amount of data is missing, the model can find it difficult to learn significant connections between things and respondent abilities, which can lead to standard errors that are too big and unreliable score interpretations (Jaritz et al., 2018). Also, the fairness of the test can be affected by having sparse data. If some of the questions on the test have no answers, this may distort the estimation of just how proficient the test taker is.

In other words, the answers from the missing test-taker would not have made them more proficient if they had taken the test on the test day (Graham & Van der Maaten, 2017). Researchers must meticulously assess how thin data affect the reliability of the tests, the validity of the constructs, and the generalizability of the results to ensure that the psychometric inferences drawn from the assessments carry any real meaning (Graham & Van der Maaten, 2017). According to Enders (2022), missing data in psychometric modeling must be dealt with using appropriate handling methods. The most traditional methods take a rather simple approach and use techniques like listwise deletion and mean imputation. More recently, and with the advent of better computational tools, we can use more sophisticated statistical methods that apply modeling to the psychometric data and employ machine learning techniques to the missing data problem (Lang & Little, 2018).

The effectiveness of any technique relies on the mechanism behind the missing data, the size of the dataset, and the psychometric model's indeterminacy. When missing data are present, what is the researcher to do? Should he or she simply use the data that were collected and "impute," or fill in the missing portions? Or should he or she conduct a more complex piece of statistical wizardry known as "full information maximum likelihood estimation"? Each approach has advantages and disadvantages (Kwak & Kim, 2017). As per Pepinsky (2018), list-wise deletion is one of the simplest techniques for dealing with missing data. When using this technique, any case with missing data for any variable is deleted. Although this method is straightforward and can easily be programmed into statistical packages, it has several drawbacks that often make it an unwise choice.

Instead of discarding all cases, pairwise deletion offers a middle ground by using all available data for each analysis. However, this allows us to sharpen our pencil without resorting to too much data imputation or carrying too much missingness for individual cases. (Wang & Aronow, 2023). Multiple Imputation (MI) is a widely used method for dealing with missing data. It is a more sophisticated approach that estimates missing values multiple times to produce several complete datasets (Austin et al., 2021). Different datasets allow us to dissect things more fully and obtain a clearer picture of what is going on. Using these together helps us better understand the problem space. We know from experience that doing things the way we are doing them now—using multiple imputations to handle the inevitable missingness when we are working with these datasets—helps us obtain much more trustworthy results. Still, many psychometricians doubt the psychometric soundness of their estimates because of the missing data issue, but that was a problem they had before (Blazek et al., 2021).

The Expectation-Maximization (EM) algorithm is a model-based method for iteratively estimating the missing values. The observed data were used to estimate the missing parameters in the EM algorithm. These constitute the E (expectation) step. The parameters are then refined in the M (maximization) step, and the EM iterates between these two steps until it converges. For applications in psychometrics involving IRT models, EM is not only effective but also remarkably more accurate than the interval methods frequently used for latent trait estimation. What is very important is this without discarding valuable data (Malan et al., 2020). Bayesian methods provide a way of dealing with missing data that is fundamentally different from the traditional approaches. This strategy is particularly useful when the achieved sample size is small and the analyst wishes to use any additional information

available to bolster the estimation process (Enders, 2022). The former refers to imputation methods based on poor information about the parameters of the model used for generating the imputed values and about the mechanism causing the missing data. As such, they can be safely used under nearly any model of the missing data mechanism. The latter, the high-in formativeness approaches, are methods that utilize good, strong, and specific information about both the model parameters and the missing data mechanism, leading to far less computationally intensive estimation (Wang et al., 2021).

Machine-learning techniques lead other alternatives in missing data handling in psychometric modeling (Emmanuel et al., 2021). They compute missing values using predictive algorithms. Advanced algorithms such as k-nearest neighbors (k-NN), deep learning, and random forest imputation do this. They are more adaptable and more proficient in managing the various types and distributions of variables in data as opposed to more traditional techniques. They are also very effective in large datasets (Ma et al., 2020). K-NN imputation identifies the k most similar instances in the data and uses their values to predict the missing values. It identifies the patterns in the data and makes no assumptions about the population of the data (Syahrizal et al., 2024). This is ideal when values are missing in certain ways and the reasons for data absence are similar to the other variables having values. Psychometric research has also employed deep learning to deal with missing data using auto encoders and generative adversarial networks (GANs). These models predict missing data with impressive precision because they are very good at identifying complex patterns in the data.

However, due to the size of the datasets and the amount of computational power needed, the applicability of deep learning methods for low-N psychometric research is limited at best (Jena & Dehuri, 2022). Imputation using the random forest method makes use of an ensemble of decision trees that are "grown" from a set of observed data (Hong & Lynn, 2020). Missing values are predicted from these trees and are assigned to the observations with missing values. This method assumes that the relationships we see in the observed data hold true for the unobserved data as well. This assumption may or may not be justified. We certainly hope it is, because if it is not, our imputed values are just "guesses" that will lead us to erroneous conclusions about our data and its structure (Feng et al., 2021).

RESEARCH OBJECTIVES

1. To identify the patterns and types of missing data in assessments of mainstream school populations.
2. To empirically evaluate the performance of missing data handling techniques.
3. To determine the most effective technique for recovering data with minimal bias and error.

RESEARCH METHODOLOGY

The sample comprised 275 students from mainstream schools in south Punjab, Pakistan. Participants were in Grades 5 to 7. Informed consent was obtained from schools and parents. A standardized mathematics achievement test based on the national curriculum was developed with 30 multiple-choice items. The test was pilot-tested for reliability and accessibility. Tests were administered in small groups under controlled conditions with teacher and psychologist support. Item-level responses were recorded digitally. Missing responses were coded as blank or skipped items. Observation logs were used to document any behavioral or fatigue-related issues. The nature of missingness was examined using Little's MCAR test, logistic regression for MAR identification, and pattern visualization. Missing data mechanisms were labeled accordingly.

Following data imputation techniques were applied

- Traditional: Listwise Deletion, Mean Imputation
- Statistical: Multiple Imputation (MICE), Bayesian Estimation
- Machine Learning: k-Nearest Neighbors (k-NN), Random Forest Imputation

Psychometric Modeling IRT (2PL model) was applied pre- and post-imputation using the ltm package in R. Parameter estimates (item difficulty, discrimination) and RMSE values were compared.

RESULTS AND ANALYSIS

Figure 1 presents a comparison of Root Mean Square Error (RMSE) across various imputation techniques. None of the approaches yielded RMSE higher than the others Multiple Imputation and Bayesian Estimation. The lowest RMSE (0.042 and 0.045) values were also indicators of faster computational speed and higher accuracy in recovering the true test scores. Conversely, Listwise Deletion and Mean Imputation yielded RMSE values greater than 0.07, indicating that poor performance were lost due to the leave out of the data. The MODERN BUSINESS INTELLIGENCE analytics techniques (such as data mining) yields moderate accuracy, where RMSE values fall higher than the traditional and below the contemporary statistical techniques. Techniques include k-NN, Random Forest.

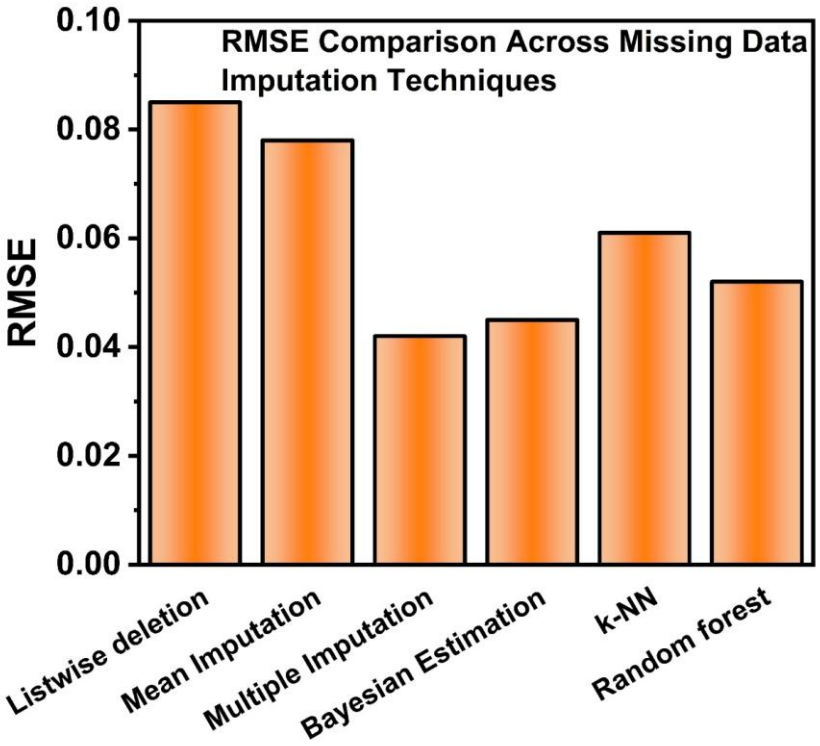

Figure 1: RMSE Comparison across Missing Data Imputation Techniques

The bias contributed by each technique in imputing missing data in shown in figure 2. Lower RMSE are aligned with the bias from Multiple Imputation and Bayesian Estimation, which are less than 2%, and are more reliable in maintaining the structure of the data. Conversely, Listwise Deletion and Mean Imputation are older methods which are more biased, being in the region of 7 - 8%, which can be detrimental to inferences. While more advanced methods from machine learning are expected to perform better, they still are more biased than the statistical methods.

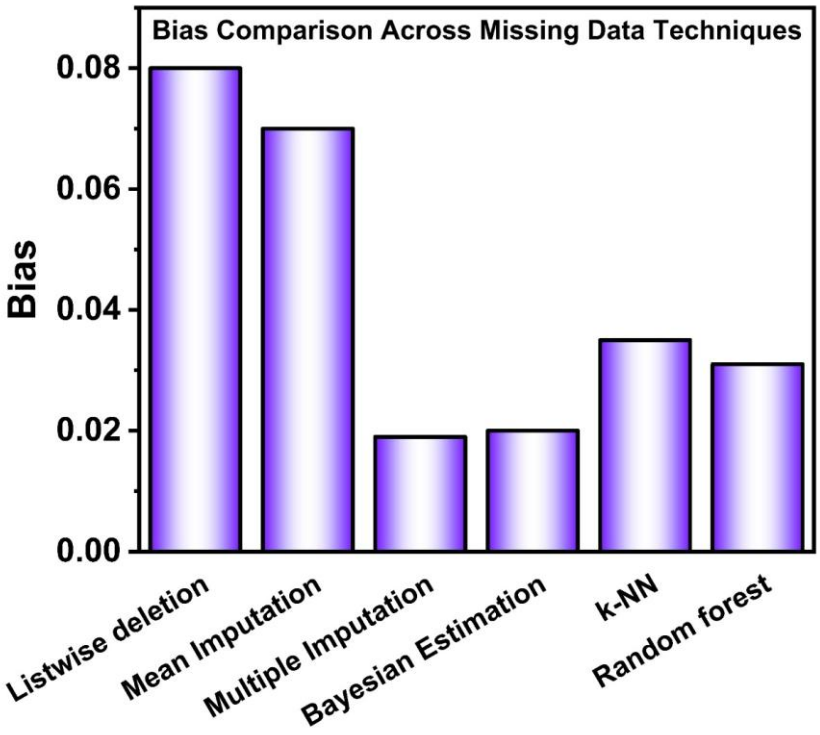

Figure 2: Bias Comparison across Missing Data Techniques

In the case of the data set distribution, Figure 3 indicates the absence of data types where the largest portion of the data is determined to be within the absence of data that is classified as missing at random (45%), and is followed by absents of data that is classified as missing completely at random (30%), and absents of data that is

classified as missing not at random (25%). The largest portion of absences of data indicates a sufficient number of missing data points was attributable to the unobservable variables, thus supporting the multivariate advanced techniques of slight absence data, and highly supports the absence data points that have been classified as missing at random types.

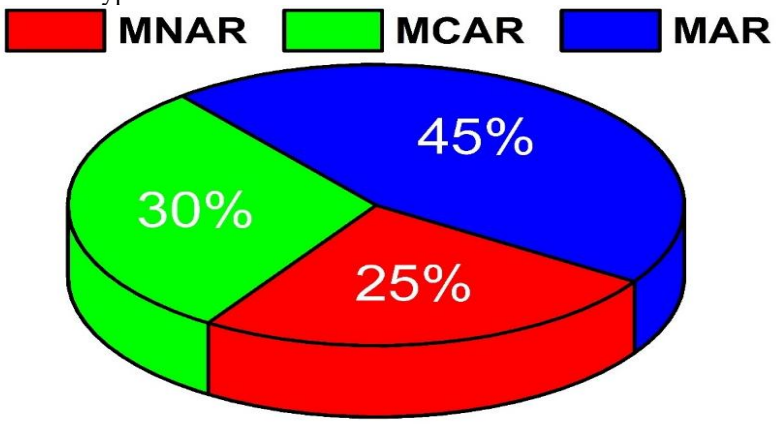

Figure 3: Distribution of Missing Data Types in the dataset

In Figure 4, we demonstrate imputation methods affecting item difficulty estimates in the IRT model. Again, Multiple Imputation and Bayesian Estimation showed minimal changes (only 0.03 and 0.04 average change in item difficulty estimates, respectively, and almost similar to 0). Meanwhile, Listwise Deletion and Mean Imputation increased item difficulty estimates (0.10 and 0.12, respectively) and may misestimate the student's competence. Machine learning impacts on difficulty parameters than better than traditional methods were, yet still were worse than statistical methods.

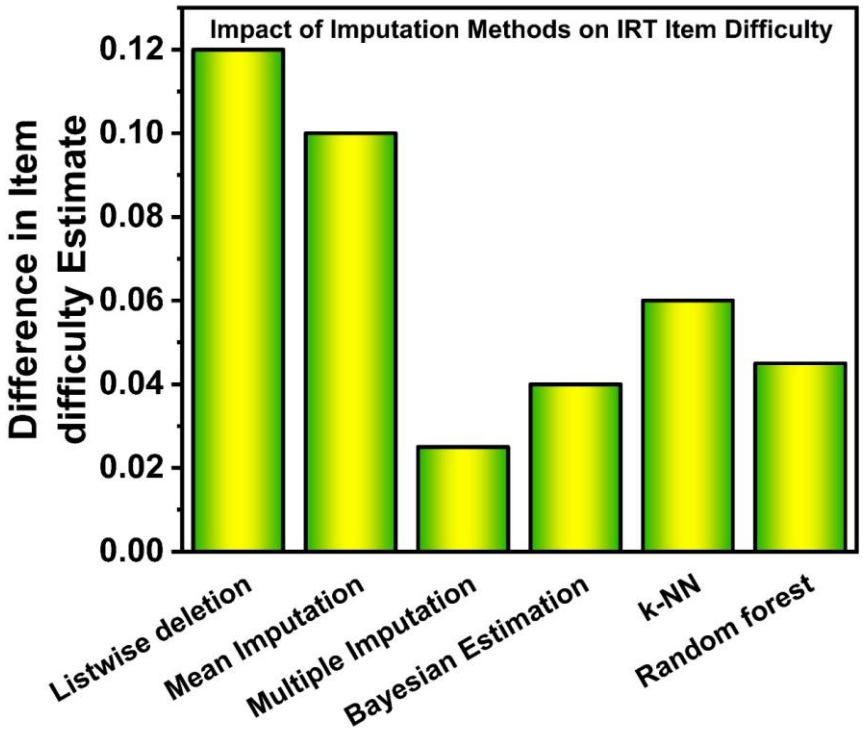

Figure 4: Impact of Imputation Methods on IRT Item Difficulty

The estimated reliability (Cronbach's Alpha) of the test shown in Figure 5. Based on the results we received, we can conclude that the best Internal consistency restoration of the dataset is from Multiple Imputation (0.87) and Bayesian Estimation (0.85) as they are the highest in reliability. The lowest estimates of reliability (less than 0.75) are obtained from Listwise Deletion and Mean Imputation as they eliminate the data variability. Machine learning methods are not that bad in reliability as they are around 0.81-0.83 and can be used as an alternative if the resources are available.

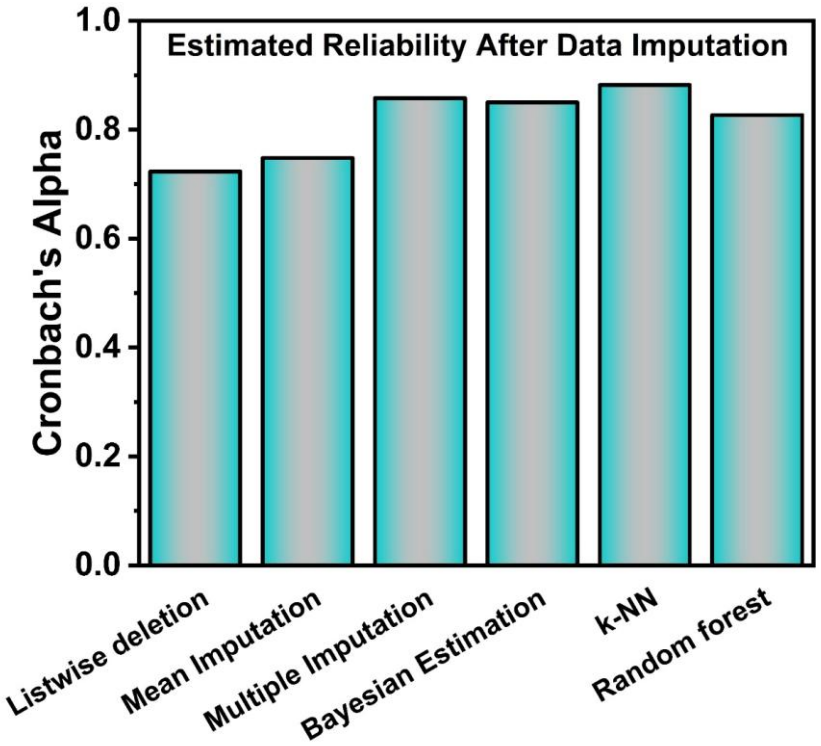

Figure 5: Estimated Reliability after Data Imputation

DISCUSSION

This study shows just how much of an effect the preferred imputation methods have on the expression of quality in educational measurement. Specifically the findings related to the psychometrics of the user's data when compared to the population of interest from Mainstream Schools (Zargar et al. 2022). The evidence of data distortion from the Multiple Imputation and Bayesian Estimation are seen to be less biased as the RMSE for the value of the data are seen to be less diverging (Pereira et al. 2020). These methods have their missing observations correctly imputed while preserving the true structure of the data in the process.

In other words, the methods did not simply fill in the gaps. They improved the overall understanding of the picture (Afkanpour et al. 2024). About 45% of the missing data mechanisms being MAR makes the use of sophisticated imputation methods that much more important within this context as well (Li et al. 2020). As MAR assumes that the probability of missingness is related to the observed data, it follows that not every case can be used for missing data handling techniques like multiple imputation. MAR is a common and reasonable assumption, however (Tahir et al., 2022). Ways such as traditional methods, for instance, listwise deletion and mean imputation, did not solve the problem with patterns of missing data. What they did was inflate error rates, bias estimates, and compromise validity across a number of assessment outcomes (Li et al., 2020). Imputing data affects the retrieval of IRT values. Using listwise and mean imputation will create bias in engineering estimates of item difficulty. This has significant implications due to the effect of item difficulty on the psychometric definition of an assessment (what is measured and the quality of measurement) (Tiwaskar et al., 2024).

From an educational standpoint, it is important to conduct calibration and keep students' assessment accuracy in mind (González-Vidal et al., 2020). As for the accuracy of the measurements, the results also attest to the effectiveness of sophisticated strategies to cope with absence (Sun et al., 2023). When several imputations and Bayesian methods were used in the tests, internal consistency was shown to be at a higher level, as per the Cronbach's alpha coefficients. The tests showed higher consistency and reliability (Thomas & Rajabi, 2021). On the contrary, traditional approaches showed worse results because of the loss of variance in the data. The performance of the machine learning models was such that it placed them as winners of some categories, demonstrating that they are reliable and efficient methods. In spite of that, the best statistical models showed marginally less bias and better RMSE scores (Noghrehchi et al., 2021).

The purpose of this research is to identify suitable techniques for handling missing data. The most fitting techniques must be chosen to correspond to the data set characteristics and the purpose of the assessment (Pindiyan & Pramila, 2024). However, the random forest and k-NN techniques, especially in machine learning, remain the most accurate in predictions when dealing with large, complex data sets (Altamimi et al., 2024). It is not purely methodological in nature when educational researchers and practitioners working with incomplete data, especially

in mainstream schools, choose to employ robust imputation techniques. It is a methodological choice that promotes fairness, equity, and 'evidence-based' decision making (Thomas & Rajabi, 2021).

CONCLUSION

The purpose of this empirical study is to explain the data missingness techniques used in educational assessments to improve policy evaluations. This empirical study has demonstrated the importance of missing data techniques in educational assessments. Merging data trends and longitudinal missing data capture in educational assessments is addressed mostly in this study. This study is also concerned with the relative importance of comparing missing at the data assessment techniques. This focus has the potential to bias educational assessments made on mainstream school populations. Within this study, Bayesian estimation and multiple imputation are deemed the most bias free, and the estimates are deemed less reliable, and that estimation is consistent with the psychometric model. This study is particularly concerned with missing data due to challenges like learning disabilities, fatigue, or disengagement. This study also picks missing data imputation techniques.

7. Recommendations for Future Research

1. Future studies should explore the development and testing of hybrid imputation techniques that combine the strengths of statistical and machine-learning methods to further enhance accuracy and efficiency.
2. Longitudinal research is needed to assess the impact of imputation techniques on student growth measures over time, especially in adaptive testing environments.
3. Future studies should be conducted with diverse special education populations to determine the generalizability of imputation performance across different types of disabilities and educational contexts.
4. Practical guidelines and user-friendly tools should be developed for educators and researchers to implement robust imputation methods in real-time assessment systems.

REFERENCES

1. Afkanpour, M., Hosseinzadeh, E., & Tabesh, H. (2024). Identify the most appropriate imputation method for handling missing values in clinical structured datasets: a systematic review. BMC Medical Research Methodology, 24(1), 188.
2. Altamimi, A., Alarfaj, A. A., Umer, M., Alabdulqader, E. A., Alsubai, S., Kim, T.-h., & Ashraf, I. (2024). An automated approach to predict diabetic patients using KNN imputation and effective data mining techniques. BMC Medical Research Methodology, 24(1), 221.
3. Antholzer, S., Haltmeier, M., & Schwab, J. (2019). Deep learning for photoacoustic tomography from sparse data. Inverse problems in science and engineering, 27(7), 987-1005.
4. Austin, P. C., White, I. R., Lee, D. S., & van Buuren, S. (2021). Missing data in clinical research: a tutorial on multiple imputation. Canadian Journal of Cardiology, 37(9), 1322-1331.
5. Blazek, K., van Zwieten, A., Saglimbene, V., & Teixeira-Pinto, A. (2021). A practical guide to multiple imputation of missing data in nephrology. Kidney International, 99(1), 68-74.
6. Bulut, O., & Kim, D. (2021). The use of data imputation when investigating dimensionality in sparse data from computerized adaptive tests. Journal of Applied Testing Technology.
7. Carpenter, J. R., & Smuk, M. (2021). Missing data: A statistical framework for practice. Biometrical Journal, 63(5), 915-947.
8. Dai, S. (2021). Handling missing responses in psychometrics: Methods and software. Psych, 3(4), 673-693.
9. de Chiusole, D., Stefanutti, L., Anselmi, P., & Robusto, E. (2015). Modeling missing data in knowledge space theory. Psychological Methods, 20(4), 506.
10. Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., & Tabona, O. (2021). A survey on missing data in machine learning. Journal of Big data, 8, 1-37.
11. Enders, C. K. (2022). Applied missing data analysis. Guilford Publications.
12. Epskamp, S. (2020). Psychometric network models from time-series and panel data. Psychometrika, 85(1), 206-231.
13. Epskamp, S., Maris, G., Waldorp, L. J., & Borsboom, D. (2018). Network psychometrics. The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development, 953-986.
14. Feng, R., Grana, D., & Balling, N. (2021). Imputation of missing well log data by random forest and its uncertainty analysis. Computers & Geosciences, 152, 104763.
15. González-Vidal, A., Rathore, P., Rao, A. S., Mendoza-Bernal, J., Palaniswami, M., & Skarmeta-Gómez, A. F. (2020). Missing data imputation with bayesian maximum entropy for internet of things applications. IEEE Internet of Things Journal, 8(21), 16108-16120.
16. Graham, B., & Van der Maaten, L. (2017). Submanifold sparse convolutional networks. arXiv preprint arXiv:1706.01307.
17. Guerra-Urzola, R., Van Deun, K., Vera, J. C., & Sijtsma, K. (2021). A guide for sparse pca: Model comparison and applications. Psychometrika, 86(4), 893-919.
18. Heine, J. H., & Tarnai, C. (2015). Pairwise Rasch model item parameter recovery under sparse data conditions. Psychological Test and Assessment Modeling, 57(1), 3-36.

19. Hong, S., & Lynn, H. S. (2020). Accuracy of random-forest-based imputation of missing data in the presence of non-normality, non-linearity, and interaction. BMC Medical Research Methodology, 20, 1-12.
20. Huo, Y., de la Torre, J., Mun, E.-Y., Kim, S.-Y., Ray, A. E., Jiao, Y., & White, H. R. (2015). A hierarchical multi-unidimensional IRT approach for analyzing sparse, multi-group data for integrative data analysis. Psychometrika, 80, 834-855.
21. Jaritz, M., De Charette, R., Wirbel, E., Perrotton, X., & Nashashibi, F. (2018). Sparse and dense data with cnns: Depth completion and semantic segmentation. 2018 International Conference on 3D Vision (3DV),
22. Jena, M., & Dehuri, S. (2022). An integrated novel framework for coping missing values imputation and classification. IEEE Access, 10, 69373-69387.
23. Kline, R. B. (2023). Principles and practice of structural equation modeling. Guilford publications.
24. Kwak, S. K., & Kim, J. H. (2017). Statistical data preparation: management of missing values and outliers. Korean journal of anesthesiology, 70(4), 407-411.
25. Lang, K. M., & Little, T. D. (2018). Principled missing data treatments. Prevention science, 19(3), 284-294.
26. Lei, P.-W., & Shiverdecker, L. K. (2020). Performance of estimators for confirmatory factor analysis of ordinal variables with missing data. Structural Equation Modeling: A Multidisciplinary Journal, 27(4), 584-601.
27. Levy, R., & Mislevy, R. J. (2017). Bayesian psychometric modeling. Chapman and Hall/CRC.
28. Li, X., Wang, Y., & Ruiz, R. (2020). A survey on sparse learning models for feature selection. IEEE transactions on cybernetics, 52(3), 1642-1660.
29. Ma, J., Cheng, J. C., Jiang, F., Chen, W., Wang, M., & Zhai, C. (2020). A bi-directional missing data imputation scheme based on LSTM and transfer learning for building energy data. Energy and Buildings, 216, 109941.
30. Malan, L., Smuts, C. M., Baumgartner, J., & Ricci, C. (2020). Missing data imputation via the expectation-maximization algorithm can improve principal component analysis aimed at deriving biomarker profiles and dietary patterns. Nutrition Research, 75, 67-76.
31. McNeish, D. (2017a). Exploratory factor analysis with small samples and missing data. Journal of personality assessment, 99(6), 637-652.
32. McNeish, D. (2017b). Missing data methods for arbitrary missingness with small samples. Journal of Applied Statistics, 44(1), 24-39.
33. Mohan, K., & Pearl, J. (2021). Graphical models for processing missing data. Journal of the American Statistical Association, 116(534), 1023-1037.
34. Morin, A. J., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. Structural Equation Modeling: A Multidisciplinary Journal, 23(1), 116-139.
35. Newgard, C. D., & Lewis, R. J. (2015). Missing data: how to best account for what is not known. Jama, 314(9), 940-941.
36. Noghrehchi, F., Stoklosa, J., Penev, S., & Warton, D. I. (2021). Selecting the model for multiple imputation of missing data: Just use an IC! Statistics in Medicine, 40(10), 2467-2497.
37. Patra, B. K., Launonen, R., Ollikainen, V., & Nandi, S. (2015). A new similarity measure using Bhattacharyya coefficient for collaborative filtering in sparse data. Knowledge-Based Systems, 82, 163-177.
38. Pedersen, A. B., Mikkelsen, E. M., Cronin-Fenton, D., Kristensen, N. R., Pham, T. M., Pedersen, L., & Petersen, I. (2017). Missing data and multiple imputation in clinical epidemiological research. Clinical epidemiology, 157-166.
39. Pepinsky, T. B. (2018). A note on listwise deletion versus multiple imputation. Political Analysis, 26(4), 480-488.
40. Pereira, E., Ali, M., Wu, L., & Abourizk, S. (2020). Distributed simulation–based analytics approach for enhancing safety management systems in industrial construction. Journal of construction engineering and management, 146(1), 04019091.
41. Pindiyan, A. V., & Pramila, R. (2024). Machine Learning-Based Imputation Techniques Analysis and Study. 2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT),
42. Sun, Y., Li, J., Xu, Y., Zhang, T., & Wang, X. (2023). Deep learning versus conventional methods for missing data imputation: A review and comparative study. Expert Systems with Applications, 227, 120201.
43. Syahrizal, M., Aripin, S., Utomo, D. P., Mesran, M., Sarwandi, S., & Hasibuan, N. A. (2024). The application of the K-NN imputation method for handling missing values in a dataset. AIP Conference Proceedings,
44. Tahir, H. B., Haque, M. M., Yasmin, S., & King, M. (2022). A simulation-based empirical bayes approach: Incorporating unobserved heterogeneity in the before-after evaluation of engineering treatments. Accident Analysis & Prevention, 165, 106527.
45. Tang, F., & Ishwaran, H. (2017). Random forest missing data algorithms. Statistical Analysis and Data Mining: The ASA Data Science Journal, 10(6), 363-377.
46. Thomas, T., & Rajabi, E. (2021). A systematic review of machine learning-based missing value imputation techniques. Data Technologies and Applications, 55(4), 558-585.
47. Tiwaskar, S., Rashid, M., & Gokhale, P. (2024). Impact of machine learning-based imputation techniques on medical datasets-a comparative analysis. Multimedia Tools and Applications, 1-21.
48. Wang, J. S., & Aronow, P. M. (2023). Listwise deletion in high dimensions. Political Analysis, 31(1), 149-155.

49. Wang, X., Zhang, R., Sun, Y., & Qi, J. (2019). Doubly robust joint learning for recommendation on data missing not at random. International Conference on Machine Learning,
50. Wang, Z., Wang, L., Tan, Y., & Yuan, J. (2021). Fault detection based on Bayesian network and missing data imputation for building energy systems. Applied Thermal Engineering, 182, 116051.
51. Wei, R., Wang, J., Su, M., Jia, E., Chen, S., Chen, T., & Ni, Y. (2018). Missing value imputation approach for mass spectrometry-based metabolomics data. Scientific Reports, 8(1), 663.
52. Wu, W., Jia, F., & Enders, C. (2015). A comparison of imputation strategies for ordinal missing data on Likert scale variables. Multivariate behavioral research, 50(5), 484-503.
53. Xiao, J., & Bulut, O. (2020). Evaluating the performances of missing data handling methods in ability estimation from sparse data. Educational and psychological Measurement, 80(5), 932-954.
54. Zargar, S., Yao, Y., & Tu, Q. (2022). A review of inventory modeling methods for missing data in life cycle assessment. Journal of Industrial Ecology, 26(5), 1676-1689.