The Majority Rules Procedure: Aiming to Develop Alternative Approaches to Missing Data for Latent Profile Analysis

Dr. Alexander V Struck Jannini, University of Cincinnati

Alexander Struck Jannini is an Assistant Professor - Educator in the Department of Engineering and Computing Education at the University of Cincinnati. His main research interests are motivation in engineering students, engineering graduate student culture, and using person-centered approaches in understanding student mindsets.

Dr. Siqing Wei, Youngstown State University - Rayen School of Engineering

Dr. Siqing Wei is an Assistant Professor at the Youngstown State University. He received a B.S. and M.S. in Electrical Engineering and a Ph.D. in Engineering Education program at Purdue University as a triple boiler. He was a postdoc fellow at the University of Cincinnati under the supervision of Dr. David Reeping. His research interests span several major research topics, which are teamwork, AI in education, cultural diversity, and international and Asian/ Asian American student experiences. He utilizes innovative and cutting-edge methods, such as person-centered approaches, NLP, ML, and Social Relation Models. He studies and promotes multicultural teaming experiences to promote an inclusive and welcoming learning space for all to thrive in engineering. Particularly, he aims to help students improve intercultural competency and teamwork competency through interventions, counseling, pedagogy, and mentoring. Siqing received the Outstanding Graduate Student Research Award in 2024 from Purdue College of Engineering, the Bilsland Dissertation fellow in the 2023-24 academic year, and the 2024 FIE New Faculty Fellow Award.

The Majority Rules Procedure: Aiming to Develop Alternative Approaches to Missing Data for Latent Profile Analysis

Abstract

In this method full paper, we present the Majority Rules Procedure, a modification of the existing Majority Vote Procedure, for handling missing data based on multiple imputation in Latent Profile Analysis (LPA) studies. While full-information maximum likelihood (FIML) is considered standard practice for addressing missing data in LPA studies, recent research has identified limitations in its performance. The main insights from this work are that multiple imputation methods, such as the Majority Vote Procedure, are more effective in determining latent profiles within datasets. However, a major limitation still exists with the Majority Vote Procedure; namely, the procedure does not identify nor provide a single dataset for use in subsequent analysis. This limitation hinders follow-up analyses, such as correlating profile membership with critical variables not used in the LPA. While pooling may be an effective method for linear regression, this method is conceptually problematic for LPA, as it raises the question of what a student's pooled profile membership represents. To address this issue, we proposed the Majority Rules Procedure, which not only identifies the most frequent latent profile result from multiple imputations but also designates a single representative dataset for follow-up analysis. This dataset is determined based on multiple fit indices that consider validity, reliability, and parsimony, including Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), sample-size adjusted Bayesian Criterion (SABIC), entropy, and profile size. We tested the effectiveness of this method using a dataset we have previously used for an LPA study. This dataset contains a full dataset from 1630 first-year engineering students related to their cultural value orientations. After introducing varying levels of artificial missingness, we conducted the Majority Rules Procedure, comparing our LPA results across the different missing-data conditions. Overall, we found that several limitations for this procedure exist, including that high missingness conditions can lead to differing LPA outcomes and that profile separation may be an important factor that limits the performance of imputation techniques. We hope that this work helps to promote others to tackle important questions regarding imputation techniques in person-centered approaches.

Keywords: Person-Centered Approaches; Latent Profile Analysis; Missing Data; Imputation

Introduction

Dealing with missing data is a fundamental issue with quantitative analysis. When collecting data using surveys, there are many factors that can lead to incomplete datasets, such as failures to complete, human error when responding to the survey, and attrition with the study [1]. With person-centered approaches to quantitative analysis, specifically latent profile analysis (LPA), the full-information maximum likelihood (FIML) method is considered the gold standard for dealing with missing data [2], [3]. A limitation of the FIML that should be noted is that the LPA program for R, tidyLPA [4], cannot perform this technique. The most commonly used software for LPA, MPlus [5], requires a 745 USD purchase for the software from a university [6]. This pricing can be a barrier to researchers who may be interested in conducting an LPA but may not be able to purchase the software necessary to run FIML to deal with missing data.

Not only is the FIML strategy currently limited to specific software packages, but recent research also suggests that it is not always the best way to deal with missing data for an LPA. Specifically, Waldman [3] demonstrated that multiple imputation methods (MI) can account for nonresponse errors better than FIML when the profile separation is considerably high (i.e., entropy is greater than 0.74) and when all profiles are of reasonable size (i.e., 10% or more of the sample is within each profile). Waldman also noted that two multiple imputation techniques, specifically the Majority Vote and pooling, performed the best as MI techniques [3].

While these multiple imputation methods may work, there are still some issues with current approaches that need to be addressed. Using a pooling method, for example, leads to a conceptual issue; specifically, what does a student's pooled profile membership mean? For Majority Vote, in which the number of profiles is determined based on the mode of best performance parameters, the issue is that no singular dataset is used for determining the profile placement of the students. This gap suggests that exploration and innovation are necessary to develop MI methods that are relevant and valid for person-centered education research.

Current Study and Research Questions

This work is meant to be a preliminary proof of concept showing that this Majority Rules procedure is a viable option for dealing with missing data for latent profile analysis. Our goal for this paper is to advance the methodological rigor of LPA by introducing new techniques for dealing with missing data, and to welcome discussions about this topic within person-centered quantitative methodology. We hope that this work helps to establish conversation and invite others who are interested in this topic to discuss possible avenues for collaboration. With this in mind, we consider the following as our Research Questions:

1) Does the Majority Rules Procedure return similar profiles to the full dataset?
2) What differences exist between running the Majority Rules Procedure when using either the dataset with the best fit parameters or a dataset at random as the exemplary dataset?

Data Collection and Handling

Dataset

The dataset was collected from a cohort of first-year engineering students at a large Midwestern university as part of a course requirement, along with a teamwork evaluation, during the Fall 2022 semester. While this dataset was originally collected for a different research purpose, the high response rate and completeness of the data are particularly useful as an empirical dataset for pilot testing new methods, especially for the research project presented here. The sample contains 1630 full responses to the Personal Cultural Orientation (PCO) instrument, which assesses an individual's orientation from a cultural perspective [7]. Eight constructs being collected, each with four items, including independence (IND), interdependence (INT), power (POW), social inequality (IEQ), risk aversion (RSK), ambiguity intolerance (AMB), masculinity (MAS), and gender equality (GEQ). Interested readers can learn more about these constructs from the cited reference from Sharma [7].

Introducing Missingness

The "Random Missingness" program in R was used to introduce missingness to the full dataset, simulating missing completely at random based on a binomial process [8]. Missingness was added at levels of 8%, 20%, and 32%, which are considered low, medium, and large amounts of missingness, where multiple imputation methods are still appropriate [9]. These three datasets were therefore labeled *Low Missingness, Medium Missingness,* and *High Missingness*, respectively.

Multiple Imputation

For the imputation process, we chose to use predictive mean matching (PMM). PMM was chosen due to one author's familiarity with the process [10]. Other potential imputation candidate methods within MI framework to be explored in the future might include regression-based imputation, hot-deck, and random forest. Briefly, the imputation method was run using the "mice" program in R [11], specifying the method as "pmm" and the maximum iterations to be 100. A dataset was generated based on the percent of missing data, as is considered good practice for multiple imputation [12]. Therefore, 8 datasets were developed for the *Low Missingness* condition, 20 for the *Medium Missingness* Condition, and 32 for the *High Missingness* condition.

The "Majority Rules" Procedure

Latent Profile Analysis

Once missingness was applied to the dataset, the "Majority Rules" procedure was enacted. Briefly, an LPA was conducted for each dataset that was generated for all three missingness

conditions. The LPAs were run in R using the tidyLPA program [4]. We followed the same conditions that were used in our previous work with this dataset [13]. Briefly, all LPAs were developed with variances and covariances being considered equal. We ran the program to develop multiple models, assuming 1 through 9 profiles.

Fit Statistics

Several fit statistics were collected to compare across profiles. Specifically, we collected the Akaike Information Criterion (AIC; [14]), Bayesian Information Criterion (BIC; [15]), and the sample-size adjusted Bayesian Information Criterion (SABIC; [16]). Lower AIC, BIC, and SABIC scores indicate a better fit.

Entropy was also collected and considered when making comparisons between models. Higher entropy scores indicate that subjects have a higher likelihood of belonging to the profiles that they are assigned to, with a value of 0.80 or higher being considered optimal [2]. In our initial work, we omitted any profile model with an entropy lower than 0.80. In this work, multiple models had no profiles with an entropy score that met this criterion, and so we modified our method to prioritize the highest entropy scores.

We also analyzed the results of the bootstrapped likelihood ratio test (BLRT; [17]). The BLRT reports a p-value, and a p-value below 0.05 indicates that the new model with a higher number of profiles is significantly different than the previous model with the lower number of profiles. As a holistic measure, we also considered the minimum number of subjects in a profile (n_{min}). In our original work, we aimed for n_{min} values above 2 percent. However, there were several instances where the n_{min} was always below 2 for each profile. Therefore, we prioritized profiles with n_{min} values that were at least 1.5 percent, which still aligns with reported best practices [2].

Determining Profiles and the Exemplary Dataset

This procedure follows a similar method known as the “Majority Vote” procedure, as described in Waldman’s thesis [3]. When each of the LPA’s are conducted, the fit statistics are compared across models with different profiles. The model with the best fit statistics is considered the one with the appropriate number of profiles. This analysis is completed for each dataset. Once these analyses were complete, the mode of the appropriate number of profiles was found, and this was considered the correct number of profiles for the data. This is normally where the “Majority Vote” procedure ends.

We then considered two options for determining the exemplary dataset with the “Majority Vote” mode as the number of profiles to use for additional analyses. First, we considered the one dataset that also had the best fit parameters (considered the “Best Fit” condition). We also chose one of the datasets at random (considered the “Random Fit” condition).

Comparisons

We made comparisons between the Best Fit and Random Fit conditions with the results from our previous study. First, we compared the profiles developed and their relative sizes. Second, we report the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) for each fit condition. To determine the difference in class assignment for each sample point, we also calculated the Rand Index (RI) and Adjusted Rand Index (ARI). The Rand Index is a number between 0 and 1, with 1 indicating perfect matching between the two conditions and 0 meaning no matching [18]. For the Adjusted Rand Index, the number returned can range from -1 to 1: A value of -1 indicates that the profile placement is worse than chance, 0 indicates that the profile placement is no better than random placement, and 1 indicating perfect placement between the predicted and actual [19], [20].

Results

Determining Profiles

As mentioned above, the first step was to determine the mode of the number of profiles that fit the data for each missingness condition. In our initial dataset containing no missing values, the number of profiles was determined to be 5 [13]. For this study, only the Low Missingness condition returned the same result. In the Medium and Large Missingness conditions, the ideal number of profiles that was determined the most times was 3. A total of 6 datasets resulted in 3 profiles for the Medium Missingness condition, and a total 10 datasets for the Large Missingness condition. A summary of these findings is provided in Table 1. The best-fit parameters for each dataset of each condition are also provided in Appendix A. Considering that the Medium and High Missingness conditions did not return the same number of profiles as the full dataset, they are excluded from the remainder of the analysis. We discuss this deviation within the Medium and High Missingness conditions in the Discussion section.

Table 1.

Summary of Findings for Determining the Appropriate Number of Profiles Based on Missingness Conditions.

	Low Missingness	Medium Missingness	High Missingness
Total Number of Datasets generated?	8	20	32
Mode of Best Fit Profile	5	3	3
Number of Datasets that Returned Best Fit Profile	5	6	10

Best Fit Versus Random Fit Conditions

For the Low Missingness condition, we extracted the dataset with the best fit parameters (Best Fit) and another dataset that returned the mode at random (Random Fit). We then analyzed the profiles and named them based on their cultural value scores. We then compared the profiles, their overall descriptions, and the size of each profile. The average scores for each cultural value for each profile in each condition are provided in Figures 1 through 3. A summary of these comparisons is provided in Table 2. The averages and standard deviations for the profiles in the Full Dataset, Best Fit, and Random Fit conditions are provided in Appendix B.

Overall, there are similar naming conventions and patterns within each of the groups. For example, Classes 2, 3, and 4 follow similar trends within each group, and Class 3 has the same naming convention for each group. There is also considerable overlap in the naming and trends between the Best and Random Fit groups, with the main difference being a Moderate – All group in the Best Fit and the Moderate High – All group in the Random Fit. We also note the similar trend regarding Class 3 being the dominant class in each condition, containing the highest number of students in the Full Dataset, Best Fit, and Random Fit conditions.

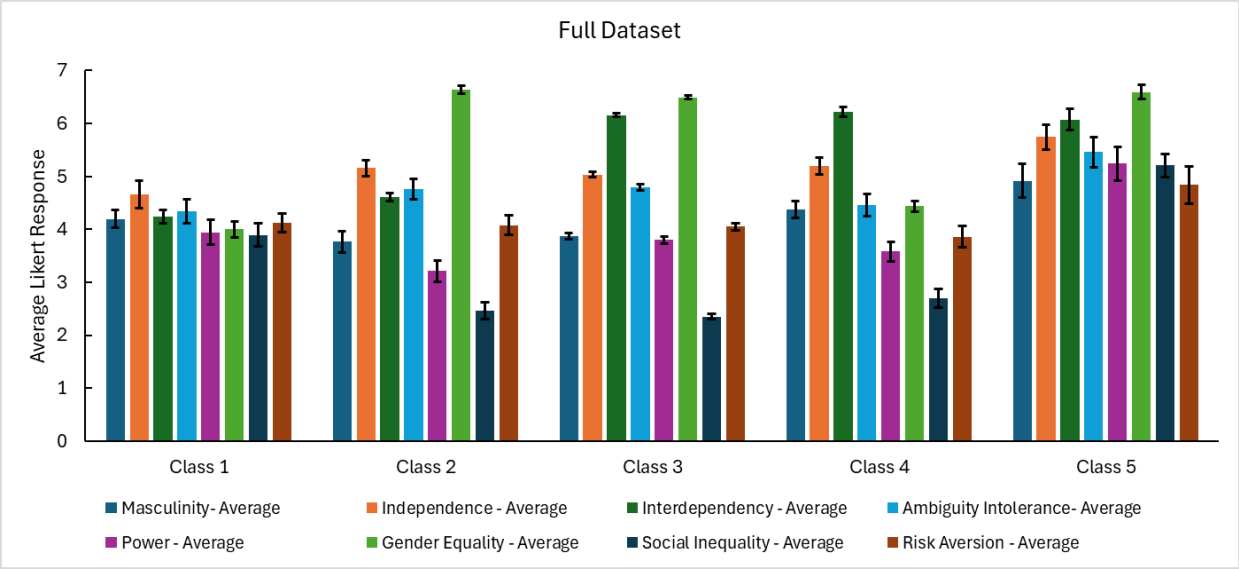

Figure 1. Profiles based on the full dataset. Class 1 is Moderate – All. Class 2 is High Gender Equality – Low Power and Social Inequality. Class 3 is High Gender Equality and Interdependency – Low Social Inequality. Class 4 is High Interdependency – Low Social Inequality. Class 5 is Moderate High – All. Error bars are 95% Confidence Intervals.

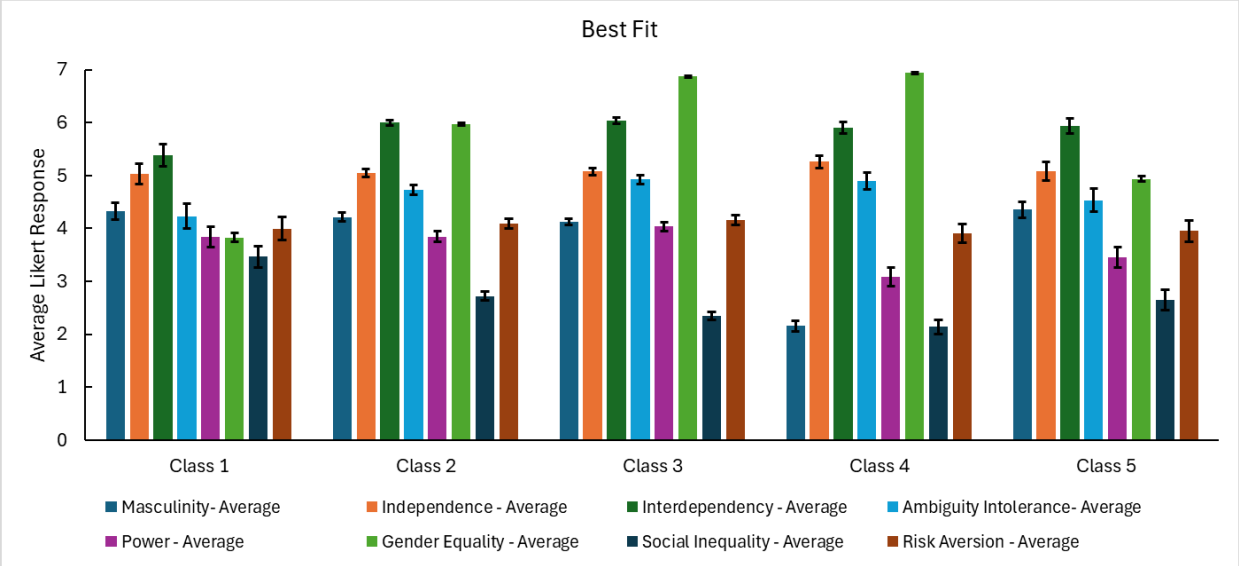

Figure 2. Profiles based on the Best Fit Condition. Class 1 is Moderate – All. Class 2 is Moderate High Interdependency and Gender Equality – Moderate Masculinity and Independence. Class 3 is High Gender Equality and Interdependency – Low Social Inequality. Class 4 is High Gender Equality – Low Masculinity and Social Inequality. Class 5 is Moderate Gender Equality, Independence, and Masculinity. Error bars are 95% Confidence Intervals.

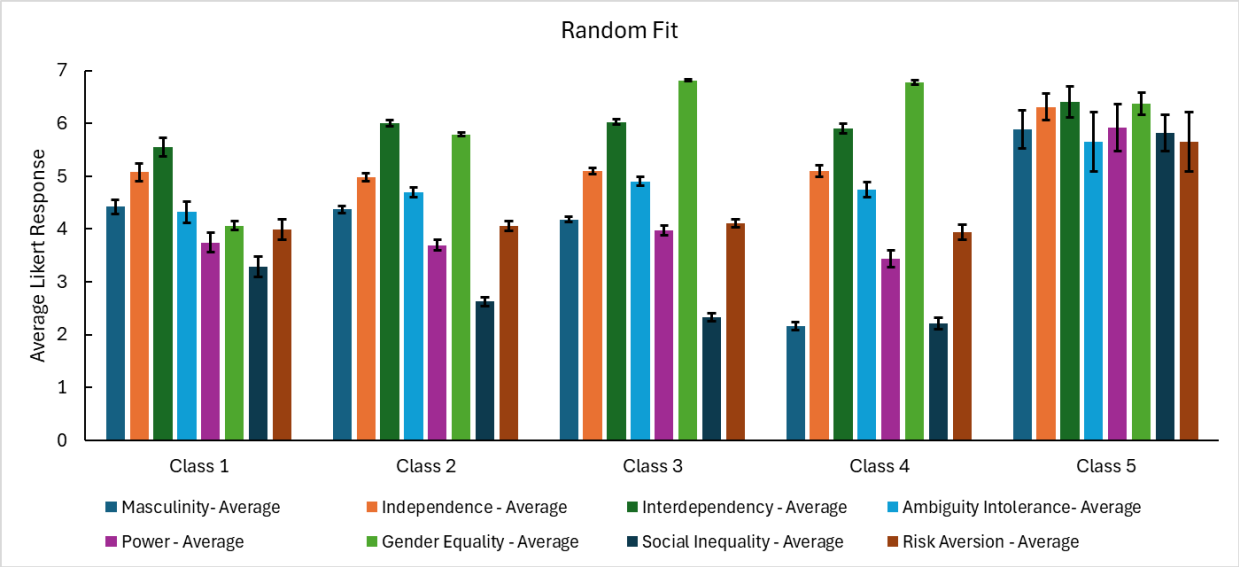

Figure 3. Profiles based on the Random Fit Condition. Class 1 is Moderate Gender Equality, Independence, and Masculinity. Class 2 is Moderate High Interdependency and Gender Equality – Moderate Masculinity and Independence. Class 3 is High Gender Equality and Interdependency – Low Social Inequality. Class 4 is High Gender Equality – Low Masculinity and Social Inequality. Class 5 is Moderate High - All. Error bars are 95% Confidence Intervals.

Table 2.

Summary of the Profiles for the Full, Best Fit, and Random Fit Conditions.

	Full Dataset	Best Fit Scenario	Random Fit Scenario
Profile 1 Name	Moderate - All	Moderate - All	Moderate Gender Equality, Independence, and Masculinity
Profile 1 Size	51	102	143
Profile 1 Description	Students reported moderate scores in all cultural dimensions	Students reported moderate scores in all cultural dimensions	Students only reported moderate gender equality, independence, and masculinity
Profile 2 Name	High Gender Equality - Low Power and Social Inequality	Moderate High Interdependency and Gender Equality - Moderate Masculinity and Independence	Moderate High Interdependency and Gender Equality - Moderate Masculinity and Independence
Profile 2 Size	149	538	475
Profile 2 Description	Students reported higher scores in gender equality, but lower in power and social inequality	Students reported moderately high scores for interdependence and gender equality, and moderately for masculinity and independence	Students reported moderately high scores for interdependence and gender equality, and moderately for masculinity and independence
Profile 3 Name	High Gender Equality and Interdependency - Low Social Inequality	High Gender Equality and Interdependency - Low Social Inequality	High Gender Equality and Interdependency - Low Social Inequality
Profile 3 Size	1240	682	718
Profile 3 Description	Students reported higher scores in gender equality and interdependence, but lower in social inequality	Students reported higher scores in gender equality and interdependence, but lower in social inequality	Students reported higher scores in gender equality and interdependence, but lower in social inequality
Profile 4 Name	High Interdependency - Low Social Inequality	High Gender Equality - Low Masculinity and Social Inequality	High Gender Equality - Low Masculinity and Social Inequality
Profile 4 Size	138	196	274
Profile 4 Description	Students reported higher interdependence and lower social inequality	Students reported higher gender equality, low masculinity, and low social inequality	Students reported higher gender equality, low masculinity, and low social inequality
Profile 5 Name	Moderate High - All	Moderate Gender Equality, Independence, and Masculinity	Moderate High - All
Profile 5 Size	52	112	20
Profile 5 Description	Students reported moderately high scores in all cultural dimensions	Students only reported moderate gender equality, independence, and masculinity	Students reported moderately high scores in all cultural dimensions

We also compared the MAE, RMSE, RI, and ARI between the Best Fit and Random Fit Conditions, as shown in Table 3. For the Best Fit Condition, the MAE, the average difference between the actual and the predicted score, was 1.52. The Random Fit Condition had similar results, with an MAE of 1.51. The RMSE results were the same for both conditions, which is an indication that there is little difference between the two conditions. This indication is reinforced when looking at both the RI and the ARI values, which are identical and differ by only 0.002, respectively.

Table 3.

Summary of Error Statistics for Best and Random Fit Conditions.

	Best Fit	Random Fit
Mean Absolute Error (MAE)	1.52	1.51
Root Mean Squared Error (RMSE)	2.02	2.02
Rand Index (RI)	0.540	0.540
Adjusted Rand Index (ARI)	0.143	0.141

Discussion

Overall, there are several key findings from this preliminary study. First, the Majority Rules procedure resulted in a different number of profiles than the full data set when there was medium (20%) or high (32%) missingness. In each of these cases, 3 was returned the most times as the ideal number of profiles, while the full dataset had 5 as the ideal. When reviewing the LPA for the full dataset, there is an argument that 3 could also be the ideal number of profiles if the analyst prioritized entropy above all other fit parameters. During our analysis of the datasets for the Medium and Large Missingness Conditions, entropy was prioritized, as many datasets had entropy values that were well below our original target (0.80). Considering this, we propose three recommendations: First, the Majority Rules procedure seems ideal for when there is low missingness (i.e., less than 20%) within the data. Second, the Majority Rules procedure does use MI, and is therefore limited if there is not high separation between the profiles and there are small profiles. Third, when using the Majority Rules Procedure, researchers should prioritize high entropy to ensure high separation between the profiles as well as high profile sizes, aiming for entropy values above 0.80 and profiles that contain at least 10% of the total sample.

We note that the Majority Rules Procedure ended up with profiles of roughly similar characteristics. The overall trends of the profiles matched in several instances, as shown in Figures 1, 2, and 3. The major difference we see between each of them is that the distribution of students within the profile does vary. In the full dataset, a vast majority of the students are placed in Class 3 (1240 students, roughly 76%). While Class 3 also included the most students for the Best Fit and Random Fit Conditions, it only contained 42% and 44% of the students, respectively. This change in the distributions explains the RI and ARI values, which show that there is little profile recovery. An ARI of 0.143 and 0.141 shows that the profile classification of

the students in the Best Fit and Random Fit conditions is only slightly better than chance. A possible reason for this low ARI is also due to the classes looking very similar within the conditions. When viewing the different profiles, we see similar trends in the construct scores across profiles. Since there is low separation between the profiles, this could exacerbate profile membership discrepancies and should be considered when using the Majority Rules procedure.

When comparing the Best Fit and the Random Fit conditions, we note that there is little difference between the LPA results, especially when compared to the actual dataset. For both conditions, all model statistics were similar. These findings suggest that there is no difference between using the best fit statistics and a random data set. This finding is not surprising considering the methodology, in which the best fit statistics for each dataset are kept and compared to find the mode. Therefore, we see no justification for either picking the dataset with the best fit parameters, or a random dataset. Thus, we recommend interested researchers apply both methods and then compare the results to determine the most reasonable profiles to work with if they are different.

It is important to also note that there are limitations to this study, and our results should be interpreted with care. First, we acknowledge that there is low separation between our profiles, which can lead to issues with profile retention. An LPA with greater profile separation may result in improved profile retention and show better results for our procedure. Second, there is the possibility that 3 profiles may be a better fit to our data, considering how the entropy was highest for that condition with the full dataset. Running this study again but with 3 profiles instead of 5 may result in better outcomes and notes the importance of how LPA assessment can introduce bias into the analysis. Lastly, we recognize that this is a preliminary study that may lack more robust methods to determine the effectiveness of our imputation method. Our goal with this work, however, was to note that there may be other options for handling missing data and begin the conversation about possible new methods that may be appropriate for this field.

Conclusion

The Majority Rules procedure was developed after a review of the current state of multiple imputation options for person-centered approaches. The goal of the procedure is to not only return the appropriate profiles that are within a population, but to determine a specific dataset that can be used for additional analyses. We performed a baseline test of this procedure under multiple missing data conditions, finding that more than 20% missing data leads to difficulty in returning the appropriate number of profiles. We also found that using either the best fit parameters or a random dataset that contains the determined number of profiles yields similar results regarding individual imputation error and profile retention. While there are limitations to this procedure, we are hopeful that others will be interested in our findings and will be willing to explore this area of research moving forward, including studies looking at 3 profiles instead of 5 and exploring the 8% to 20% missing data range for further analysis of this method.

References

[1] A. D. Woods *et al.*, "Best practices for addressing missing data through multiple imputation," *Infant Child Dev.*, vol. 33, no. 1, pp. 1–37, 2024, doi: 10.1002/icd.2407.

[2] D. Spurk, A. Hirschi, M. Wang, D. Valero, and S. Kauffeld, "Latent profile analysis: A review and 'how to' guide of its application within vocational behavior research," *J. Vocat. Behav.*, vol. 120, no. January 2019, p. 103445, 2020, doi: 10.1016/j.jvb.2020.103445.

[3] M. R. Waldman, "Multiple Imputation Methods for Latent Profile Analysis in Education and the Behavioral Sciences," Harvard University, 2020.

[4] J. M. Rosenburg, P. N. Beymer, D. J. Anderson, C. J. Van Lissa, and J. A. Schmidt, "tidyLPA: An R package to easily carry out latent profile analysis (LPA) using open-source or commercial software," *J. Open Source Softw.*, vol. 3, no. 30, p. 978, 2018, doi: https://doi.org/10.21105/joss.00978.

[5] L. Muthen and B. Muthen, *Mplus user's guide*, 8th ed. Muthen & Muthen, 2017.

[6] Muthen & Muthen, "University Pricing." Accessed: Mar. 23, 2026. [Online]. Available: https://www.statmodel.com/pricing.shtml

[7] P. Sharma, "Measuring personal cultural orientations: Scale development and validation," *J. Acad. Mark. Sci.*, vol. 38, no. 6, pp. 787–806, 2010, doi: 10.1007/s11747-009-0184-7.

[8] M. J. Lajeunesse, "Random generation of missingness in a data frame."

[9] J. C. Jakobsen, C. Gluud, J. Wetterslev, and P. Winkel, "When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts," *BMC Med. Res. Methodol.*, vol. 17, no. 162, pp. 1–10, 2017, doi: 10.1186/s12874-017-0442-1.

[10] A. V. Struck Jannini, "Developing motivational profiles of first-year engineering students using latent profile analysis," Purdue University, 2024.

[11] P. Royston, "Multiple imputation of missing values: Update of ice," *Stata J.*, vol. 5, no. 4, pp. 527–536, 2005, doi: 10.1177/1536867x0500500404.

[12] I. R. White, P. Royston, and A. M. Wood, "Multiple imputation using chained equations: Issues and guidance for practice," *Stat. Med.*, vol. 30, no. 4, pp. 377–399, 2011, doi: 10.1002/sim.4067.

[13] S. Wei, A. Struck Jannini, and D. Reeping, "WIP: Characterizing Personal Cultural Orientations of First-Year Engineering Students by Latent Profile Analysis: A Person-Centered Approach," in *ASEE Annual Conference and Exposition*, Montreal, 2025, pp. 1–10. doi: 10.18260/1-2--57384.

[14] H. Akaike, “A new look at the statistical model identification,” *IEEE Trans. Automat. Contr.*, vol. AC-19, no. 6, pp. 716–723, 1974.

[15] G. Schwarz, “Estimating the dimension of a model,” *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, 1978.

[16] S. L. Sclove, “Application of model-selection criteria to some problems in multivariate analysis,” *Psychometrika*, vol. 52, no. 3, pp. 333–343, 1987, doi: 10.1007/BF02294360.

[17] G. McLachlan and D. Peel, *Finite mixture models*, Wiley Seri. John Wiley & Sons, Inc., 2000. doi: http://dx.doi.org/10.1002/0471721182.

[18] M. J. Warrens and H. van der Hoef, “Understanding the Rand Index,” in *Advanced Studies in Classification and Data Science*, T. Imaizumi, A. OKada, S. Miyamoto, F. Sakaori, Y. Yamamoto, and M. Vichi, Eds., Springer, Singapore, 2020, pp. 301–313. doi: 10.1007/978-981-15-3311-2_24.

[19] J. M. Santos and M. Embrechts, “On the use of the Adjusted Rand Index as a metric for evaluating supervised classification,” in *Artificial Neural Networks – ICANN 2009*, vol. 5769, C. Alippi, M. Polycarpou, C. Panayiotou, and G. Ellinas, Eds., Springer Berlin, 2009, pp. 175–184. doi: 10.1007/978-3-642-04277-5_18.

[20] L. Hubert and P. Arabie, “Comparing partitions,” *J. Classif.*, vol. 2, no. 1, pp. 193–218, 1985, doi: 10.1007/BF01908075.

Appendix A.

The Majority Rules Results for Each Missingness Condition

Table A.1

Majority Rules Results for the Low Missingness Condition

Dataset #	*Determined Profile Numbers*	*AIC*	*BIC*	*SABIC*	*Entropy*	*n_min*	*BLRT_p*
1	5	34442.52	34874.22	34620.08	0.837335	0.062577	0.009901
2	5	34680.00	35111.71	34857.56	0.773981	0.01227	0.009901
3	7	34345.04	34873.89	34562.56	0.830032	0.010429	0.009901
4	2	34945.16	35231.17	35062.8	0.943734	0.111043	0.009901
5	5	34697.89	35129.6	34875.45	0.734353	0.034969	0.009901
6	6	34560.91	35041.18	34758.44	0.760767	0.030675	0.009901
7	5	34632.98	35064.69	34810.54	0.769556	0.018405	0.009901
8	5	34676.78	35108.48	34854.34	0.814753	0.030675	0.009901

Note: AIC – Akaike Information Criterion. BIC – Bayesian Information Criterion. SABIC - Sample-size Adjusted Bayesian Information Criterion. n_min – Percentage of the total sample size of the smallest profile. BLRT_p – p-value returned from the Bootstrapped Likelihood Ratio Test.

Table A.2

Majority Rules Results for the Medium Missingness Condition

Dataset #	*Determined Profile Numbers*	*AIC*	*BIC*	*SABIC*	*Entropy*	*n_min*	*BLRT_p*
1	3	34787.61	35122.18	34925.22	0.786449	0.119018	0.009901
2	7	34672.64	35201.48	34890.16	0.702841	0.01227	0.009901
3	3	34816.2	35150.77	34953.81	0.885333	0.04908	0.009901
4	3	34871.24	35205.81	35008.85	0.801399	0.12454	0.009901
5	6	34555.82	35036.1	34753.36	0.751506	0.019632	0.009901
6	7	34286.88	34815.72	34504.39	0.845545	0.02454	0.009901
7	3	34964.15	35298.72	35101.76	0.784711	0.119632	0.009901
8	8	34157.1	34734.51	34394.59	0.839314	0.018405	0.009901
9	7	34570.95	35099.79	34788.46	0.710551	0.03681	0.009901
10	2	34898.96	35184.97	35016.6	0.939172	0.121472	0.009901
11	8	34331.93	34909.34	34569.42	0.778347	0.01227	0.009901
12	9	34443.27	35069.24	34700.73	0.704243	0.025153	0.009901
13	7	34326.94	34855.78	34544.46	0.778114	0.01227	0.009901
14	3	34934.47	35269.04	35072.08	0.67281	0.122699	0.009901
15	5	34761.38	35193.09	34938.94	0.736937	0.02638	0.009901
16	5	34803.4	35235.11	34980.96	0.720205	0.027607	0.009901
17	3	34841.1	35175.68	34978.71	0.799534	0.120859	0.009901
18	5	34741.96	35173.67	34919.52	0.720144	0.045399	0.009901
19	8	34335.68	34913.08	34573.16	0.770302	0.020859	0.009901
20	5	34698.05	35129.76	34875.61	0.762285	0.029448	0.009901

Note: AIC – Akaike Information Criterion. BIC – Bayesian Information Criterion. SABIC - Sample-size Adjusted Bayesian Information Criterion. n_min – Percentage of the total sample size of the smallest profile. BLRT_p – p-value returned from the Bootstrapped Likelihood Ratio Test.

Table A.3

Majority Rules Results for the High Missingness Condition

Set #	*Determined Profile Numbers*	*AIC*	*BIC*	*SABIC*	*Entropy*	*n_min*	*BLRT_p*
1	3	34858.93	35193.5	34996.53	0.879416	0.062577	0.009901
2	7	34517.29	35046.13	34734.8	0.761655	0.004908	0.009901
3	6	34516.69	34996.97	34714.23	0.738369	0.042945	0.009901
4	4	34775.29	35158.43	34932.88	0.761739	0.095706	0.009901
5	5	34810.24	35241.95	34987.81	0.730016	0.045399	0.009901
6	3	34718.79	35053.36	34856.4	0.810181	0.11227	0.009901
7	4	34678.88	35062.02	34836.47	0.691181	0.057669	0.009901
8	3	34956.62	35291.2	35094.23	0.751459	0.105521	0.009901
9	4	34818.45	35201.59	34976.03	0.78396	0.058896	0.009901
10	9	34366.35	34992.33	34623.81	0.80665	0.017178	0.009901
11	6	34567.98	35048.26	34765.52	0.726734	0.035583	0.009901
12	5	34620.04	35051.75	34797.6	0.736391	0.05092	0.009901
13	6	34668.27	35148.55	34865.81	0.741005	0.03681	0.009901
14	3	34850.89	35185.46	34988.5	0.794861	0.100613	0.009901
15	5	34716.7	35148.4	34894.26	0.690702	0.050307	0.009901
16	3	34778.59	35113.16	34916.2	0.899022	0.056442	0.009901
17	6	34791.19	35271.47	34988.73	0.691063	0.030675	0.009901
18	3	34920.21	35254.79	35057.82	0.762481	0.103681	0.009901
19	4	34894.43	35277.57	35052.01	0.782653	0.031288	0.009901
20	7	34441.51	34970.35	34659.02	0.690849	0.043558	0.009901
21	3	34905.65	35240.22	35043.26	0.811409	0.096933	0.009901
22	4	34687.63	35070.77	34845.22	0.818758	0.057669	0.009901
23	3	34882.85	35217.43	35020.46	0.800523	0.097546	0.009901
24	2	34966.01	35252.01	35083.64	0.951428	0.093252	0.009901
25	5	34816.97	35248.68	34994.53	0.780321	0.046012	0.009901
26	7	34499.47	35028.31	34716.98	0.730878	0.030675	0.009901
27	2	34876.48	35162.49	34994.12	0.945803	0.106135	0.009901
28	2	34876.48	35162.49	34994.12	0.945803	0.106135	0.009901
29	4	34783.86	35167	34941.44	0.755371	0.074233	0.009901
30	3	34935.21	35269.79	35072.82	0.787098	0.10184	0.009901
31	2	34921.3	35207.31	35038.94	0.944856	0.095706	0.009901
32	3	34886.61	35221.18	35024.22	0.814745	0.102454	0.009901

Note: AIC – Akaike Information Criterion. BIC – Bayesian Information Criterion. SABIC - Sample-size Adjusted Bayesian Information Criterion. n_min – Percentage of the total sample size of the smallest profile. BLRT_p – p-value returned from the Bootstrapped Likelihood Ratio Test.

Appendix B.

Profile Specifications for Each Fit Condition

Table B.1

Profile Specifications for the Full Dataset.

	Masculinity Score	Independence Score	Interdependency Score	Ambiguity Intolerance Score	Power Score	Gender Equality Score	Social Inequality Score	Risk Aversion Score
Class 1	4.20 (0.61)	4.66 (0.94)	4.24 (0.45)	4.34 (0.82)	3.94 (0.85)	4.00 (0.55)	3.90 (0.80)	4.12 (0.66)
Class 2	3.77 (1.27)	5.16 (0.97)	4.61 (0.44)	4.77 (1.20)	3.21 (1.24)	6.63 (0.48)	2.46 (1.02)	4.08 (1.13)
Class 3	3.87 (1.10)	5.03 (0.86)	6.16 (0.53)	4.80 (1.13)	3.80 (1.19)	6.49 (0.53)	2.34 (0.88)	4.05 (1.13)
Class 4	4.37 (0.99)	5.20 (0.94)	6.22 (0.58)	4.46 (1.26)	3.58 (1.09)	4.43 (0.63)	2.70 (1.06)	3.86 (1.22)
Class 5	4.92 (1.16)	5.74 (0.88)	6.07 (0.74)	5.46 (1.06)	5.24 (1.15)	6.59 (0.47)	5.21 (0.80)	4.84 (1.30)

Note: Values in parentheses are standard deviations. Class 1 is Moderate – All. Class 2 is High Gender Equality – Low Power and Social Inequality. Class 3 is High Gender Equality and Interdependency – Low Social Inequality. Class 4 is High Interdependency – Low Social Inequality. Class 5 is Moderate High – All.

Table B.2

Profile Specifications for Best Fit Condition.

	Masculinity Score	Independence Score	Interdependency Score	Ambiguity Intolerance Score	Power Score	Gender Equality Score	Social Inequality Score	Risk Aversion Score
Class 1	4.33 (0.83)	5.03 (1.00)	5.38 (1.09)	4.23 (1.19)	3.84 (1.03)	3.83 (0.46)	3.47 (1.03)	4.00 (1.12)
Class 2	4.22 (1.00)	5.05 (0.86)	6.00 (0.63)	4.73 (1.08)	3.84 (1.19)	5.97 (0.23)	2.72 (1.00)	4.10 (1.11)
Class 3	4.16 (0.86)	5.07 (0.89)	6.03 (0.71)	4.92 (1.15)	4.03 (1.19)	6.87 (0.19)	2.35 (1.04)	4.16 (1.16)
Class 4	2.16 (0.71)	5.26 (0.85)	5.90 (0.78)	4.90 (1.18)	3.09 (1.23)	6.94 (0.13)	2.14 (0.97)	3.91 (1.22)
Class 5	4.36 (0.82)	5.08 (0.93)	5.93 (0.77)	4.53 (1.18)	3.45 (1.05)	4.94 (0.26)	2.66 (1.03)	3.95 (1.08)

Note: Values in parentheses are standard deviations. Class 1 is Moderate – All. Class 2 is Moderate High Interdependency and Gender Equality – Moderate Masculinity and Independence. Class 3 is High Gender Equality and Interdependency – Low Social Inequality. Class 4 is High Gender Equality – Low Masculinity and Social Inequality. Class 5 is Moderate Gender Equality, Independence, and Masculinity.

Table B.3

Profile Specifications for Random Fit Condition.

	Masculinity Score	Independence Score	Interdependency Score	Ambiguity Intolerance Score	Power Score	Gender Equality Score	Social Inequality Score	Risk Aversion Score
Class 1	4.42 (0.83)	5.07 (1.00)	5.55 (1.06)	4.32 (1.21)	3.75 (1.12)	4.06 (0.55)	3.28 (1.15)	3.99 (1.14)
Class 2	4.37 (0.81)	4.97 (0.85)	5.99 (0.65)	4.70 (1.04)	3.69 (1.12)	5.79 (0.35)	2.63 (0.92)	4.06 (1.07)
Class 3	4.18 (0.73)	5.10 (0.86)	6.02 (0.70)	4.90 (1.13)	3.97 (1.18)	6.81 (0.26)	2.33 (0.95)	4.11 (1.12)
Class 4	2.16 (0.62)	5.10 (0.92)	5.91 (0.77)	4.75 (1.22)	3.44 (1.37)	6.78 (0.36)	2.21 (0.94)	3.94 (1.17)
Class 5	5.89 (0.82)	6.31 (0.56)	6.41 (0.67)	5.65 (1.28)	5.92 (1.01)	6.38 (0.48)	5.81 (0.79)	5.65 (1.29)

Note: Values in parentheses are standard deviations. Class 1 is Moderate Gender Equality, Independence, and Masculinity. Class 2 is Moderate High Interdependency and Gender Equality – Moderate Masculinity and Independence. Class 3 is High Gender Equality and Interdependency – Low Social Inequality. Class 4 is High Gender Equality – Low Masculinity and Social Inequality. Class 5 is Moderate High - All.