Table 5 gives parameter estimates for the overall proposed model presented by Equations (1), (3), (4) and (5) and illustrated in Fig. 2 for the three types of missingness, assuming a two-latent class model for each of the latent variables “Access to Knowledge Sources” and “Response Propensity”.

The measurement model is quite robust under different missingness scenarios. This indicates that the contribution of the manifest variables in measuring the latent variable of interest is not affected by the type of missing data. On the contrary, the missingness model exhibits differences under the different types of missingness. The four missingness indicators are insignificant in defining the latent variable “Response Propensity” when data is MCAR. This is a logical result as the items were created to reflect a completely random pattern of missingness in the data, and thus the created dataset is a random subset of the data. The indicators do not measure “Response Propensity” in this case.

However, they are significant in measuring the latent variable when missing data is MAR or MNAR, as in both cases, those who have missing values are different from those who respond. However, while their contribution in measuring “Response Propensity” is almost equal for the four indicators in data MAR, it is higher for computers and smartphones than radio and telephone in the case of data MNAR.

The structural parameter ϕ, reflecting the relationship between “Response Propensity” and “Access to Knowledge Sources”, is given at the bottom of Table 5. As one would expect, this relationship is insignificant in case of data MCAR. On the contrary, there is a significant relationship in the case of data MAR and MNAR. The significant positive effect ϕ indicates that higher response levels are more likely to be found with higher levels of access to knowledge sources, even after controlling for covariates. The magnitude of the structural parameter is much higher in the case of data MNAR. A possible explanation of the significant effect indicating nonignorable missingness in the case of data that was originally created at random is that levels that are used in creating missingness are themselves confounded with certain levels of “Access to Knowledge Sources”, and thus there still exists kind of dependence of “Response Propensity” on “Access to Knowledge Sources”.

Table 6 shows the estimated conditional probabilities for the manifest items given class membership of the main latent variable “Access to Knowledge Sources”, and those of the missingness indicators given class membership of the latent variable “Response Propensity,” assuming that both latent variables are binary. The conditional probabilities are reported under different types of missingness. We may consider the first latent class of “Access to Knowledge Sources” to indicate “High access to knowledge sources” and the second to indicate “Low access to knowledge sources”, as the conditional probabilities of having any of the devices are consistently higher for the first-class than the second. The conditional probabilities resulting from the “Response Propensity” latent variable are not reliable in the case of data MCAR since the indicators are not significant in defining the “Response Propensity” latent variable (see Table V).

However, there is a clear pattern of higher estimated probabilities for the first class than the second, in data MAR or MNAR, although the differences are sometimes not too big. The first latent class may thus be labeled as “High response propensity” and the second latent class as “Low response propensity”.

TABLE VI
ITEM-RESPONSE CONDITIONAL PROBABILITIES FROM A TWO-CLASS LCM FOR "ACCESS TO KNOWLEDGE SOURCES" AND "RESPONSE PROPENSITY" LATENT VARIABLES UNDER THREE TYPES OF MISSINGNESS

	MCAR		MAR		MNAR	
"Access to Knowledge Sources"	1st class	2nd class	1st class	2nd class	1st class	2nd class
	Probability of a “Yes”					
Radio	0.472	0.223	0.466	0.221	0.517	0.250
Telephone	0.447	0.078	0.431	0.078	0.481	0.092
Computer	0.801	0.120	0.764	0.120	0.826	0.141
Smart phone	0.570	0.053	0.541	0.051	0.602	0.062
"Response Propensity"	1st class	2nd class	1st class	2nd class	1st class	2nd class
	Probability of being observed					
$r_{(Radio)}$	-	-	0.989	0.776	0.923	0.892
$r_{(Telephone)}$	-	-	0.990	0.780	0.935	0.888
$r_{(Computer)}$	-	-	0.988	0.772	0.996	0.866
$r_{(Smart\ phone)}$	-	-	0.990	0.771	0.964	0.877

Notes: The complement of the above probabilities indicates the probability of responding with a “No” to the corresponding item. The complement of the response indicator probabilities gives the probability of a “Missing” response.

From Table 7, and considering how the latent variable is defined, it can be concluded that the probability of having high access to knowledge sources is generally higher for more privileged people. That is to say; it is higher for males than females and people with a higher wealth index, of older age, and with higher levels of education. An unexpected result is that the probability of having high access to knowledge sources is higher for those living in rural areas compared to urban. However, it is not known whether the available media are used to access knowledge or mainly for entertainment and communication.

In our application, covariates effects on the missingness part of the model seem to be significant only in the case of data MAR. The sex covariate has a negative effect on “Response Propensity,” which means that females are more likely to respond. Place of residence and age positively affect “Response Propensity,” which means that younger people and people in urban areas are more likely to respond. For data MCAR, the “Response Propensity” latent variable is not well-defined due to randomness in creating the missingness in this case, so its relationship with “Access to Knowledge Sources” and covariates turned out to be insignificant. In the case of data MNAR, only place of residence significantly affects “Response Propensity”.

A possible explanation is that the effect of the other two covariates is already indirectly carried within the latent variable of interest “Access to Knowledge Sources”, since their effect on “Access to Knowledge Sources” is already highly significant.

TABLE VII

ESTIMATED COVARIATES EFFECTS FROM A TWO-CLASS LCM FOR "ACCESS TO KNOWLEDGE SOURCES" AND "RESPONSE PROPENSITY" LATENT VARIABLES IN CASE OF COMPLETE DATA AND UNDER DIFFERENT TYPES OF MISSINGNESS

		Complete data with covariates	MCAR	MAR	MNAR
Covariates effects on Z_a					
Place of residence (Rural)	β_1	-2.791***	-2.751***	-3.540***	-2.839***
Wealth index	β_2	-2.691***	-2.645***	-3.209***	-2.705***
Sex (Female)	β_3	0.376***	0.365***	0.408***	0.355***
Age	β_4	-0.043***	-0.045***	-0.047***	-0.044***
Educational level	β_5	-0.628***	-0.626***	-0.607***	-0.630***
Covariates effects on Z_r					
Place of residence (Rural)	γ_1		-0.047	2.548***	0.698**
Sex (Female)	γ_2		-0.109	-0.481***	0.305
Age	γ_3		0.001	0.018***	0.012

Note: *** denotes a p-value < $\mathbf{0.01}$ and ** denotes a p-value < $\mathbf{0.05}$.

IV. Conclusion

When multiple manifest variables are used as measures of a latent variable, it is quite often to have some missing values in the data due to item non-response. In this paper, we proposed to summarize item non-response by another latent variable that can be labeled as "Response Propensity". The missingness can thus be allowed to be non-random by allowing the "Response Propensity" latent variable to depend on the main latent variable of interest.

A model specification incorporating a missingness mechanism within a latent class model framework has been proposed to model multivariate binary data used as measures of a categorical latent variable. This model specification allows for nonignorable item non-response by letting the response propensity latent variable summarize the response indicators, depending on the latent variable of interest and covariates. Logistic regression equations are used to model relationships within the latent class model under the categorical nature of all manifest and latent variables in the model. Estimation of model parameters and goodness of fit measures use conventional methods that are usually used to fit latent variable models for multivariate data.

The proposed model has been applied to data from Egypt's Demographic and Health Survey 2014. Data missingness has been artificially created to generate three different types of missingness, MCAR, MAR and MNAR, to study the results of the model in each case. An important result of the model was that the measurement part defining the latent variable of interest, "Access to Knowledge Sources" is quite robust no matter how missingness was created. For data MAR and MNAR, the relationship between the "Response Propensity" latent variable and the "Access to Knowledge Sources" latent variable remains significant even after controlling for covariates.

Unlike other models already existing in the literature, such as Bacci and Bartolucci [22] and Beesley *et al* [32], the proposed model accounts for missingness and allows for this missingness to be non-random by depending on levels of the latent class of interest. The estimated probabilities of class membership of the "Response Propensity" latent variable are affected by class membership of the "Access to Knowledge Sources" latent variable making the missingness nonignorable. Lower levels of response were associated with lower levels of "Access to Knowledge Sources". This result confirms the importance of accommodating the missingness mechanism within the modeling of the data due to the systematic difference between respondents and nonrespondents. Covariates effects are also found to be robust on the measurement model; however, they are quite sensitive to the type of missingness in the missingness part of the model. We have used Bayesian estimation in Zakaria *et al.* [33] to fit the same model specification proposed in this article and to study the sensitivity of the results to different levels of missingness.

References

[1] G. Verbeke and G. Molenberghs, "Modeling Through Latent Variables," *Annual Review of Statistics and Its Application*, vol. 4, pp. 267–282, 2017.

[2] R. van Bork, M. Rhemtulla, L. J. Waldorp, J. Kruis, S. Rezvanifar, and D. Borsboom, "Latent Variable Models and Networks: Statistical Equivalence and Testability," *Multivariate Behavioral Research*, vol. 56, no. 2, pp. 175–198, 2021.

[3] D. J. Bartholomew, M. Knott, and I. Moustaki, *Latent Variable Models and Factor Analysis*, 3rd ed., Wiley series in probability and statistics, 2011.

[4] S. Jeon, J. Lee, J. C. Anthony, and C. H., "Latent class analysis for multiple discrete latent variables: a study on the association between violent behavior and drug-using behaviors," *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 24, pp. 911–925, 2017.

[5] J. W. Lee and H. Chung, "Latent class analysis with multiple latent group variables," *Communications for Statistical Applications and Methods*, vol. 24, pp. 173–191, 2017.

[6] K. J. Petersen, P. Qualter, and N. Humphrey, "The Application of Latent Class Analysis for Investigating Population Child Mental Health: A Systematic Review," *Frontiers in Psychology*, vol. 10, p. 1214, 2019.

[7] E. Kim, H. Chung, and S. Jeon, "Joint latent class analysis for longitudinal data: an application on adolescent emotional well-being," *Communications for Statistical Applications and Methods*, vol. 27, pp. 241–254, 2020.

[8] K. J. Petersen, N. Humphrey, and P. Qualter, "Latent Class Analysis of Mental Health in Middle Childhood: Evidence for the Dual-Factor Model," *School Mental Health*, vol. 12, pp. 786–800, 2020.

[9] A. Robitzsch, "Regularized Latent Class Analysis for Polytomous Item Responses: An Application to SPM-LS Data," *Journal of Intelligence*, vol. 8, p. 30, 2020.

[10] J. W. Lee and H. Chung, "A multivariate latent class profile analysis for longitudinal data with a latent group variable," *Communications for Statistical Applications and Methods*, vol. 27, pp. 15–35, 2020.

[11] J. H. M. Janssen, S. van Laar, M. J. de Rooij, J. Kuha, and Z. Bakk, "The Detection and Modeling of Direct Effects in Latent Class Analysis," *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 26, no. 2, pp. 280–290, 2019.

[12] Z. Bakk and J. Kuha, "Relating latent class membership to external variables: An overview," *British Journal of Mathematical and Statistical Psychology*, vol. 74, pp. 340–362, 2020.

[13] Z. Bakk and J. Kuha, “Two-step estimation of models between latent classes and external variables,” *Psychometrika*, vol. 83, no. 4, pp. 871–892, 2018.

[14] J. Kuha, S. Butt, M. Katsikatsou, and C. J. Skinner, “The Effect of Probing ‘Don’t Know’ Responses on Measurement Quality and Non-response in Surveys,” *Journal of the American Statistical Association*, vol. 113, no. 521, pp. 26–40, 2018.

[15] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, 2nd ed., John Wiley and Sons, 2002.

[16] H. Du, C. Enders, B. T. Keller, T. N. Bradbury, and B. R. Karney, “A Bayesian Latent Variable Selection Model for Nonignorable Missingness,” *Multivariate Behavioral Research*, 2021, doi: 10.1080/00273171.2021.1874259.

[17] N. Rose, M. von Davier, and B. Nagengast, “Modeling Omitted and Not-Reached Items in IRT Models,” *Psychometrika*, vol. 82, pp. 795–819, 2017.

[18] J. F. Cursio, R. J. Mermelstein, and D. Hedeker, “Latent trait shared-parameter mixed models for missing ecological momentary assessment data,” *Statistics in Medicine*, vol. 38, no. 4, pp. 660– 673, 2019.

[19] H. Jung, J. L. Schafer, and B. Seo, “A latent class selection model for nonignorably missing data,” *Computational Statistics and Data Analysis*, vol. 55, pp. 802–812, 2011.

[20] O. Harel and J. L. Schafer, “Partial and latent ignorability in missing-data problems,” *Biometrika*, vol. 96, no. 1, pp. 37–50, 2009.

[21] J. Kuha, M. Katsikatsou, and I. Moustaki, “Latent variable modelling with non-ignorable item non-response: multigroup response propensity models for cross-national analysis,” *Journal of the Royal Statistical Society, Series A*, vol. 181, pp. 1169–1192, 2018.

[22] S. Bacci and F. Bartolucci, “A Multidimensional Latent Class IRT Model for Non-Ignorable Missing Responses,” *Structural Equation Modeling: A Multidisciplinary Journal*, 2014, [Online]. Available: https://arxiv.org/abs/1410.4856.

[23] S. K. Sterba, “A Latent Transition Analysis Model for Latent-State-Dependent Nonignorable Missingness,” *Psychometrika*, vol. 81, no. 2, pp. 506–534, 2016.

[24] El-Zanaty, Associates, and I. International, “Egypt demographic and health survey 2014 (Data file and code book),” Ministry of Health and Population, Egypt, 2014. [Online]. Available: https://www.dhsprogram.com/data/available-datasets.cfm

[25] L. M. Collins and S. T. Lanza, *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. Wiley series in probability and statistics, 2010.

[26] L. K. Muthén and B. O. Muthén, *Mplus User’s Guide*, 8th ed., Los Angeles, CA: Muthén & Muthén, 1998-2017.

[27] L. A. Goodman, “Exploratory latent structure analysis using both identifiable and unidentifiable models,” *Biometrika*, vol. 61, pp. 215–231, 1974.

[28] G. Xu, “Identifiability of restricted latent class models with binary responses,” *The Annals of Statistics*, vol. 45, no. 2, pp. 675–707, 2017.

[29] W. Koo and H. Kim, “Bayesian nonparametric latent class model for longitudinal data,” *Statistical Methods in Medical Research*, vol. 29, no. 11, pp. 3381–3395, 2020.

[30] J. Tein, S. Coxe, and H. Cham, “Statistical Power to Detect the Correct Number of Classes in Latent Profile Analysis,” *Structural Equation Modeling*, vol. 20, no. 4, pp. 640–657, 2013.

[31] D. Gordon, S. Nandy, C. Pantazis, S. Pemberton, and P. Townsend, “The distribution of poverty in the developing world,” *University of Bristol, Centre for International Poverty Research, UK, 2003.*

[32] L. J. Beesley, J. M. G. Taylor, and R. J. A. Little, “Sequential imputation for models with latent variables assuming latent ignorability,” *Aust. N. Z. J. Stat.*, vol. 61, no. 2, pp. 213–233, 2019.

[33] S. Zakaria, M. S. Hafez, and A. M. Gad, “Bayesian Estimation of Latent Class Model for Survey Data Subject to Item Nonresponse,” *Pakistan Journal of Statistics and Operation Research*, vol. 15, no. 2, pp. 303–318, 2019.