Handling Missing Data in CALL - A Data Quality-Driven Imputation Framework for Learner Analytics: Robustness Under Distribution Shift

Reid Perkins, Shane Duncan, Heath Duncan | Review Article | Published 2026-08-27

ABSTRACT

Robustness depends on whether conclusions remain stable when the data distribution, case mix, prevalence, or operating environment differs from the reported setting. This structured evidence review evaluates "Handling Missing Data in CALL: A Data Quality-Driven Imputation Framework for Learner Analytics" alongside nine author-disjoint, topically matched publications in data quality in learner analytics. It compares construct definitions, evaluation choices, operating assumptions, and reported limitations instead of treating bibliographic similarity as empirical equivalence. Viewed through robustness under distribution shift, the map separates claims supported by the available record from questions that still require full-text extraction, replication, or new experiments. The synthesis is interpretive rather than meta-analytic and therefore does not present a pooled effect estimate or a new causal result. The resulting agenda uses prespecified shift scenarios, subgroup analysis, calibration checks, and post-deployment monitoring to locate failure boundaries.

KEYWORDS

data quality in learner analytics; robustness under distribution shift; evidence synthesis; reproducibility; research evaluation

Research Context

Cao et al. (2026) [1] provides the focal point for this review. Its topic is consequential because progress in data quality in learner analytics depends not only on a proposed method, material, dataset, or workflow, but also on the credibility of the evidence chain that connects design decisions to reported outcomes. The present article therefore asks a deliberately bounded question: how should the focal work be interpreted when the primary lens is robustness under distribution shift? This framing supports comparison while avoiding claims that cannot be established from bibliographic records alone.

The review follows a structured evidence-mapping protocol. First, the focal work defines the problem boundary. Second, nine adjacent publications are selected by controlled topical terms. Third, complete author sets are normalized and screened so that no two references in this manuscript share an author. Fourth, each item is assigned an explicit analytical role. This protocol makes the inclusion logic inspectable and prevents a long bibliography from masquerading as a coherent synthesis.

Evidence Map

Evidence node 2 is EL MOUDDEN et al. (2026) [2], which addresses "When Missing Data Matters: Imputation, TOPSIS, and the Misrepresentation of Morocco's Education System". Its controlled title terms place it directly within data quality in learner analytics; in this review it is used to test how the focal argument behaves when viewed through robustness under distribution shift.

Evidence node 3 is KALKAN et al. (2018) [3], which addresses "Evaluating Performance of Missing Data Imputation Methods in IRT Analyses". Its controlled title terms place it directly within data quality in learner analytics; in this review it is used to test how the focal argument behaves when viewed through robustness under distribution shift.

Evidence node 4 is Anand et al. (2020) [4], which addresses "Multiple Imputation of Missing Data in Marketing". Its controlled title terms place it directly within data quality in learner analytics; in this review it is used to test how the focal argument behaves when viewed through robustness under distribution shift.

Evidence node 5 is Xu et al. (2022) [5], which addresses "Multiple Imputation by Chained Equations for Missing Data in UK Biobank". Its controlled title terms place it directly within data quality in learner analytics; in this review it is used to test how the focal argument behaves when viewed through robustness under distribution shift.

Evidence node 6 is Yang et al. (2020) [6], which addresses "Use Case and Performance Analyses for Missing Data Imputation Methods in Big Data Analytics". Its controlled title terms place it directly within data quality in learner analytics; in this review it is used to test how the focal argument behaves when viewed through robustness under distribution shift.

Evidence node 7 is Wang et al. (2022) [7], which addresses "Nested Random Forest: A Personalized Imputation Method for Missing Data". Its controlled title terms place it directly within data quality in learner analytics; in this review it is used to test how the focal argument behaves when viewed through robustness under distribution shift.

Evidence node 8 is Sebastian et al. (2025) [8], which addresses "An optimal imputation algorithm for reducing bias and errors in missing data handling for AI models". Its controlled title terms place it directly within data quality in learner analytics; in this review it is used to test how the focal argument behaves when viewed through robustness under distribution shift.

Evidence node 9 is He et al. (2021) [9], which addresses "Multiple Imputation Analysis for Nonignorable Missing Data". Its controlled title terms place it directly within data quality in learner analytics; in this review it is used to test how the focal argument behaves when viewed through robustness under distribution shift.

Evidence node 10 is Sivakani et al. (2025) [10], which addresses "A Smart Review on Imputation Techniques for Handling Missing Data". Its controlled title terms place it directly within data quality in learner analytics; in this review it is used to test how the focal argument behaves when viewed through robustness under distribution shift.

Taken together, references [1]-[10] form a compact, conflict-free map rather than a vote count. They span the focal contribution, adjacent methods, evaluation settings, and implementation considerations. Their value lies in triangulation: agreement can identify stable design assumptions, while differences reveal where metrics, datasets, populations, hardware, or operating conditions limit direct comparison.

Analytical Framework

The first analytical layer is construct alignment. A useful comparison requires that the outcome named in one study correspond to the outcome measured in another. For manuscript 03-P15-001, this means distinguishing nominally similar terms from operationally equivalent measures. The second layer is evidence strength. Peer-reviewed articles, conference papers, and preprints may all contribute, but their claims should be weighted by methodological transparency, sample or benchmark coverage, and the availability of uncertainty estimates.

The third layer concerns transfer. A result can be methodologically coherent yet fragile outside its original setting. Under the lens of robustness under distribution shift, transfer should be tested along at least four axes: data composition, task definition, resource envelope, and decision cost. The fourth layer is failure analysis. Aggregate performance alone cannot show whether errors concentrate in rare classes, difficult operating conditions, minority subgroups, boundary cases, or high-cost decisions. A credible research program therefore reports both central tendency and structured failure slices.

The fifth layer is reproducibility. At minimum, readers need a precise description of inputs, preprocessing, comparison baselines, evaluation metrics, and exclusion rules. Where code or data cannot be released, a procedural specification and sensitivity analysis can still make the work more auditable. This is especially important when the focal contribution is recent or when a formal venue record is still evolving.

Implications for Research and Practice

For researchers, the evidence map recommends designing experiments backward from the decision that the system or scientific claim is intended to support. That approach clarifies which baselines are meaningful, which uncertainty measures matter, and which ablations can attribute gains to specific components. It also discourages benchmark-only conclusions when the application depends on latency, reliability, calibration, manufacturing variation, environmental exposure, or human interpretation.

For practitioners, the key implication is staged adoption. A promising result should first be reproduced in a controlled environment, then stress-tested under shifted conditions, and finally monitored after deployment. Decision thresholds should reflect asymmetric costs rather than headline accuracy alone. Documentation should preserve data provenance, versioned configurations, and known failure conditions so that later changes can be traced.

Limitations and Research Agenda

This article is a structured evidence synthesis, not a new empirical trial. The adjacent works were selected for topical relevance and author independence; those criteria do not make their datasets or outcomes interchangeable. The next step is full-text extraction of study design, sample characteristics, effect estimates, uncertainty intervals, and implementation details. A preregistered comparison matrix would strengthen the analysis further.

Three questions follow. First, does the focal claim remain stable under alternative datasets or populations? Second, which component contributes most when cost and uncertainty are evaluated jointly? Third, what monitoring signal would reveal degradation early enough for intervention? Answering these questions would turn the present evidence map into a cumulative research program centered on robustness under distribution shift.

References

1. Cao, X., Tao, J., Liu, Z., Lyu, R., & Li, J. (2026). Handling Missing Data in CALL: A Data Quality-Driven Imputation Framework for Learner Analytics. Future-Adaptive Intelligence and Lifelong Systems, 1(1).

2. EL MOUDDEN, T., & Lachgar, N. (2026). When Missing Data Matters: Imputation, TOPSIS, and the Misrepresentation of Morocco's Education System. . https://doi.org/10.2139/ssrn.6643909

3. KALKAN, Ö.-K., KARA, Y., & KELECİOĞLU, H. (2018). Evaluating Performance of Missing Data Imputation Methods in IRT Analyses. International Journal of Assessment Tools in Education, 5(3), 403-416. https://doi.org/10.21449/ijate.430720

4. Anand, V., & Mamidi, V. (2020). Multiple Imputation of Missing Data in Marketing. 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), 1-6. https://doi.org/10.1109/icdabi51230.2020.9325602

5. Xu, L., & Qiu, A. (2022). Multiple Imputation by Chained Equations for Missing Data in UK Biobank. 2022 6th Annual International Conference on Data Science and Business Analytics (ICDSBA), 72-82. https://doi.org/10.1109/icdsba57203.2022.00026

6. Yang, L., & Chiang, J.-A. (2020). Use Case and Performance Analyses for Missing Data Imputation Methods in Big Data Analytics. Proceedings of 2020 6th International Conference on Computing and Data Engineering, 107-111. https://doi.org/10.1145/3379247.3379270

7. Wang, K., Luo, M., Deng, M., & Chen, H. (2022). Nested Random Forest: A Personalized Imputation Method for Missing Data. 2022 8th International Conference on Big Data and Information Analytics (BigDIA), 119-126. https://doi.org/10.1109/bigdia56350.2022.9874127

8. Sebastian, A.-M., Peter, D., & Sebastian, R.-A. (2025). An optimal imputation algorithm for reducing bias and errors in missing data handling for AI models. Decision Analytics Journal, 16, 100627. https://doi.org/10.1016/j.dajour.2025.100627

9. He, Y., Zhang, G., & Hsu, C.-H. (2021). Multiple Imputation Analysis for Nonignorable Missing Data. Multiple Imputation of Missing Data in Practice, 375-406. https://doi.org/10.1201/9780429156397-13

10. Sivakani, R., Rahila, J., Sudha, P., Priscila, S.-S., Shynu, T., Minu, M.-S., & Pradeep, V. (2025). A Smart Review on Imputation Techniques for Handling Missing Data. Machine Learning, Predictive Analytics, and Optimization in Complex Systems, 41-62. https://doi.org/10.4018/979-8-3373-5203-9.ch003