



Where are Pakistan's Out-of-School Children? Mapping accessibility of schooling using satellite data and machine learning methods with an application to the 2022 floods

Third Submission: RR_PAK_2025_493

*Maria Reyes Retana
reproducibility@worldbank.org
March 18, 2026*

This review verifies the reproducibility of the exhibits included in the paper " *Where are Pakistan's Out-of-School Children? Mapping accessibility of schooling using satellite data and machine learning methods with an application to the 2022 floods*".

Contents in this review:

1. Main findings
2. List of exhibits and reproducibility status
3. Reproduction Environment

Main findings

- **Every exhibit in the paper has been reproduced accurately.**
- The code was successfully executed on a new computer after:
 1. Following the steps in the README file.
- The output demonstrates consistent stability across multiple runs. Specifically, executing the code two times consecutively yielded identical results.
- The code takes approximately 19 hours to run.
- We conducted our reproducibility analysis based on the paper shared by the authors by mail on January 22nd, 2026.
- **Verification Process and Data Handling:**
 - The reproducibility package relies on 3 types of data: open, limited-access, and restricted data.
 - Open data is included in the public reproducibility package.
 - Reviewers used limited-access and restricted data provided directly by the authors to conduct the reproducibility verification, and this is not included in the public reproducibility package.
 - *data_hash_report.csv* lists the SHA256 hashes of all files in the Data folder to support data integrity checks. Users who have access to limited or restricted data can use this file to verify that the data has not been altered.
- **Reproducibility Summary:**

- **Data:** Some data is restricted and has not been included in the reproducibility package. For more details, please refer to the README file.
- **Code:** All code files (from cleaning to analysis) are included in the reproducibility package.
- **Outputs:** Some outputs are not generated by code (created manually in Excel, ArcGIS, etc) but instructions for producing these outputs are included in the README file.
- **Reproducibility verification:** Reviewers used data provided directly by the authors to conduct the reproducibility verification, and this is not included in the public reproducibility package. The reviewers did not verify if publicly available data matches the data provided by the authors.
- **Dependencies environment:** The reviewers created a new environment for dependencies using the latest versions available for each dependency at the moment of the review.

List of exhibits and reproducibility status

Results in the Main Section of the Paper

- **Figure 1** Does not show analysis results.
- **Figure 2** Does not show analysis results.
- **Figure 3** Does not show analysis results.
- **Figure 4a** Reproduced.
- **Figure 4b** Reproduced.
- **Figure 5** Reproduced; requires manual steps. File oosc_pk_all_all.csv. The file used to create these figures is identical to the one used by the authors. The underlying data were shared with the World Bank's cartography unit, which produced these figures. Since the same underlying data were used, these figures are considered reproducible.
- **Table 1** Does not show analysis results.
- **Table 2** Reproduced, requires manual steps. Columns 1-4 verified against table2.csv. Column 5 Count of private schools in 2023/24 Pakistan Education Statistics is directly obtained from Pakistan Education Statistics 2023-24 Report. Column 5 was not verified. Format is different, but this does not affect reproducibility.
- **Table 3** Reproduced, requires manual steps. Verified row one from table3.csv and row 2 from table2.csv. Format is different, but this does not affect reproducibility.
- **Table 4** Reproduced with manual steps. Verified row private using table4.csv. For public enrolment figures, sources are in the paper. Format is different, but this does not affect reproducibility.
- **Table 5** Reproduced, requires manual steps. Format is different, but this does not affect reproducibility.
- **Table 6** Reproduced, requires manual steps. Verified columns 1-3 against table6.csv and column 4 from Pakistan Education Statistics Report 2021-22, page 99.

- **Table 7 Reproduced, requires manual steps.** The information for the first panel (Khyber Pakhtunkhwa, Punjab, Sindh, and Balochistan) is not included in table7.csv; the values must be manually compiled from:

For **column 3**, private_school_analysis_all_all_whitespace_summary.json.gz. For **column 2**, the values come from spatial_accessibility/results/all_province_variants.

To review **column 1**, private_school_analysis_all_all_potential.json.

- **Table 8 Reproduced, requires manual steps.** The results are not included in the code outputs and must be manually compiled.

For **column 3**, values were obtained from private_school_analysis_all_all_whitespace_summary.json and reproduced. For **column 2**, values were obtained from spatial_accessibility/results/all_province_variants and reproduced.

For **column 1**, the reproduced results yielded values of 0 for all entries (0, 0, 0) and therefore this could not be verified. **Action:** Review the functionality of this portion of the code (catchment)

- **Table 9 Reproduced, requires manual steps.**

The first general panel (Azad Jammu and Kashmir, Balochistan, Khyber Pakhtunkhwa, Punjab, and Sindh) is missing from table9.csv. These results, therefore, need to be manually compiled as follows:

Accessibility indicators: For **column 1**, values were obtained from government_all_all_accessibility.json.

For **column 2**, values were obtained from private_school_analysis_all_all_accessibility.json.

Whitespace indicators: For **column 1**, values were obtained from government_all_all_whitespace_summary.json.gz.

For **column 2**, values were obtained from private_school_analysis_all_all_whitespace_summary.json.

- **Table 10 Reproduced, requires manual steps.** Verified against table10.csv, share is manually calculated.
- **Table 11 Reproduced, requires manual steps.** Verified against table11.csv, totals are manually calculated.
- **Table 12 Reproduced, requires manual steps.** Verified against table12.csv, differences are manually calculated. Numbers for Pakistan, Provinces and most affected districts are not in the file table12.csv
- **Table 13 Reproduced, requires manual steps.** Verified against table13.csv. Order is different, but this does not jeopardize reproducibility.
- **Table 14 Reproduced, requires manual steps.** Table is reproduced by running the code Survey/-ground_truth_tables.do

Results in the Annex

- **Annex Figure 1** Does not show analysis results.
- **Annex Figure 2** Does not show analysis results.
- **Annex 3 Figure 3 Reproduced**
- **Annex 3 Fig 4 Reproduced**

- **Table A5.1** Reproduced, requires manual steps. Table is reproduced by running the code Survey/-ground_truth_tables.do.
- **Table A5.2** Reproduced, requires manual steps. Table is reproduced by running the code Survey/-ground_truth_tables.do.
- **Table A5.3** Reproduced, requires manual steps. Table is reproduced by running the code Survey/-ground_truth_tables.do.
- **Appendix 6. Table 1** Does not show analysis results.
- **Appendix 6. Table 2** Does not show analysis results.

Reproduction Environment

- Paper exhibits were reproduced in a computer with the following specifications:
 - OS: Windows 11 Enterprise
 - Processor INTEL(R) XEON(R) PLATINUM 8562Y+, 2800 Mhz, 8 Core(s), 8 Logical Processor(s)
 - Memory available: 106 GB
 - Software version: Python 3.9.11, Stata 18.5 MP, QGIS 3.38.1