Supplementary Material S2

README – Reproducibility Instructions

1. **Data Description**

   The file *"Calculus1.csv"* contains the processed and anonymized dataset used in the analysis of:
   "Exploring Subgroup Heterogeneity in Online and In-Person Learning Outcomes Using Causal Forest."
   The dataset includes academic and demographic covariates, treatment assignment (online vs face-to-face), and course outcomes (all the used variables are explained in the variable dictionary document *"Variable Dictionary.xlsx"*). All direct identifiers were removed prior to analysis.

2. **Software Requirements**
   The analysis was conducted in R studio version 4.3.0 in macOS operating system.
   Required R packages: *"grf", "stddiff", "ggplot2", "metafor"*

3. **Analysis Script**
   The main analysis is contained in *"MainAnaysis codes.R"* which performs: data processing and recoding, Baseline covariate balance tests (Table 1), Average treatment effect estimation, Causal forest estimation, Subgroup CATE estimation, Forest plot and caterpillar plot generation.
   Clustering analysis can be found in *"Clustering analysis.R"* and standardized effects are found in *"srandardized analysis.R"*.

4. **Reproduction Steps**
   - Place *"Calculus1.csv"* and *"MainAnaysis codes.R"* in directory.
   - Open R or RStudio.
   - Set working directory to the file location.
   -Run the script from beginning to end (preferably line by line)
   All results (tables and figures) should be generated automatically.

5. **Output Files Generated**

   -Table1.csv
   - Subgroups.csv
   -Subgroup forest plot (displayed in R environment)

6. **Confidentiality Note**

The dataset provided is anonymized and contains no personally identifiable information. It is shared strictly for peer-review purposes in accordance with institutional data protection policies.