README

Replication Files for "Conscientiousness in the workplace: Evidence from a field experiment in West Africa"

Authors: Mathias Allemand (University of Zurich and University of Teacher Education Schaffhausen), Martina Kirchberger (Trinity College Dublin and CEPR), Sveta Milusheva (The World Bank), Carol Newman (Trinity College Dublin), Brent B Roberts (University of Illinois at Urbana-Champaign and University of Tübingen), Vincent Thorne (Paris School of Economics)

About this replication package: This package contains all data, code, and documentation necessary to replicate the results in "Conscientiousness in the workplace: Evidence from a field experiment in West Africa." The package includes raw survey data from three rounds, administrative records, and complete processing scripts that produce all tables and figures in the paper. This README provides comprehensive documentation of the data structure, variable definitions, and step-by-step replication instructions. The

Statement about Rights:

- I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.
- I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permission are documented in the LICENSE.txt file.

File Organization

- **codebook.html**: Complete codebook of the analysis dataset (`panel_analysis.dta`)

Data Files (`data/`)

Raw Data (`data/1_raw/`)

- **admin-data.dta**: Administrative data containing employee records (performance grades)
- **baseline_raw-deidentified.dta**: Baseline survey data (April 2019)
- **followup1_raw-deidentified.dta**: Mid-line survey data (January 2020)
- **followup2_raw-deidentified.dta**: End-line survey data (May 2020)

Processed Data (`data/2_processed/`)

Note: These files are created by the data preparation scripts

- **baseline_clean.dta**: Cleaned baseline survey data
- **followup1_clean.dta**: Cleaned mid-line survey data
- **followup2_clean.dta**: Cleaned end-line survey data
- **panel_appended.dta**: Appended panel with admin data
- **panel_analysis.dta**: Final analysis dataset with all variables

Scripts (`scripts/`)

- **run-all.do**: Master script that runs the entire replication (from data preparation to analysis) in proper sequence.

Data Preparation (`scripts/1_data-prep/`)

- **prepare-baseline.do**: Cleans and harmonizes baseline survey data. Starts with `baseline_raw-deidentified.dta`, produces `baseline_clean.dta`.
- **prepare-followup1.do**: Cleans and harmonizes mid-line survey data. Starts with `followup1_raw-deidentified.dta`, produces `followup1_clean.dta`.
- **prepare-followup2.do**: Cleans and harmonizes end-line survey data. Starts with `followup2_raw-deidentified.dta`, produces `followup2_clean.dta`.
- **append-rounds.do**: Appends all survey rounds into long format panel dataset. Starts with `baseline_raw-deidentified.dta`, `followup1_raw-deidentified.dta`, `followup2_raw-deidentified.dta` and `admin-data.dta`, and produces `panel_appended.dta`.
- **variables-creation.do**: Constructs analysis variables. Starts with `panel_appended.dta`, calls `variables-labels.do`, produces `panel_analysis.dta`.
- **variables-labels.do**: Applies variable labels, called from `variables-creation.do`.

Analysis (`scripts/2_analysis/`)

- **analysis.do**: Analysis script that produces all tables and figures in the paper.

Outputs (`outputs/`)

Note: This directory is created when running the replication files

- All tables (`.tex` and `.xlsx` formats) and figures (`.png` format) referenced in the paper

---

Software and System Requirements

Software: Stata 18.5 **Operating System:** Mac OS 15.5 **Required Stata Packages:**

- ietoolkit
- estout
- wyoung

Note: The master script automatically installs required packages if not present

Estimated Runtime: Approximately 5 minutes

Replication Instructions

1. **Setup**:

   - Download all replication files
   - Extract to your desired directory
   - Open `scripts/run-all.do` in Stata

2. **Set Path**:

   ○ In `run-all.do`, change the global root path (line 19):

   ```
   global root "path/to/replication-files"
   ```

3. **Run Replication**:

   ○ Execute `do run-all.do` in Stata
   ○ This will run all data preparation and analysis scripts in proper sequence
   ○ All outputs will be saved to the `outputs/` directory

Data Dictionary and Variable Definitions

Please refer to `codebook.html` for a complete codebook, including variable definitions and value labels.

Data Access and Provenance

Survey Data

Source: Primary data collection by research team **Collection Period:**

- Baseline: April 2019
- Mid-line: January 2020
- End-line: May 2020

The raw de-identified data from the surveys are available on the World Bank's Microdate Catalog: [TODO]

Administrative Data

Source: Personnel records of the multinational construction company which built the TER.

Access: Administrative data is included in the replication package as `admin-data.dta`.

Accessed on: 2024-09

Description: grade_2019_admin is the grade given by the supervisor and was taken into account by HR to determine salary increase in 2020.

Table and Figure Correspondence

The following tables and figures are produced by `analysis.do`:

Main Tables

- **Table 1**: Labor market outcomes (`table1_labour-market.tex`)
- **Table 2**: Earnings (`table2_earnings.tex`)

Appendix Tables — Additional tables

- **Table B1**: Balance tests — attrition vs non-attrition groups (`tableB1_balance-attrit_cohen.xlsx`)
- **Table B2**: Balance tests — treatment vs control groups (`tableB2_balance-treat_cohen.xlsx`)
- **Table B3**: Multiple hypothesis testing corrections (`tableB3_multiple-testing.xlsx`)
- **Table B4**: Occupations at baseline — retained workers (`tableB4_occupations-atbaseline-retained.xlsx`)
- **Table B5**: Jobs of those who left company by sector (`tableB5_jobs-left-company-sector.tex`)

Appendix Tables — Alternative earnings

- **Table C1**: Alternative earnings specifications (`tableC1_earnings-alternative.tex`)

Appendix Tables — Psychometric Traits

- **Table D1**: List of all conscientiousness items (produced manually in Excel)
- **Table D2**: Conscientiousness traits reliability (Cronbach's alpha) (`tableD2_consci-traits_reliability.xlsx`)
- **Table D3**: Treatment effects on organizational and industriousness traits (`tableD3_consci-orga-indus.tex`)
- **Table D4**: Treatment effects on responsibility and punctuality traits (`tableD4_consci-resp-punc.tex`)
- **Table D5**: Treatment effects on Big Five personality traits (`tableD5_big5.tex`)

Appendix Tables — COVID-19 Analysis

- **Table E1**: COVID-19 outcomes — bivariate (`tableE1_covid-outcomes-bivariate.tex`)
- **Table E2**: COVID-19 outcomes — with controls (`tableE2_covid-outcomes-controls.tex`)

Figures

- **Figure 1**: Timeline of intervention (produced manually using a vector graphics editor)
- **Figure 2**: Coefficient plot of treatment effects on conscientiousness items (`figure2_coefplot-conscien-items.png`)
- **Figure A1**: Detailed timeline of intervention (produced manually using a vector graphics editor)
- **Figure C1**: Distribution of earnings histogram (`figureC1_histogram-earnings.png`)

Note: Some tables are exported as Excel files (.xlsx) for easier customization, while others are exported as LaTeX files (.tex). All figures are saved as PNG files.

Notes

- All monetary values are in CFA francs
- Multiple hypothesis testing corrections applied using Westfall-Young step-down procedure
- **Output formats**: Some tables are exported as Excel files (.xlsx) for easier formatting and customization, while others are exported as LaTeX files (.tex). All figures are saved as PNG format.
- **Administrative data**: Company performance grades are included directly in the replication package

---

Contact Information

For questions about data access or replication, contact:

- **Corresponding Authors:** Martina Kirchberger (martina.kirchberger@tcd.ie) and Carol Newman (cnewman@tcd.ie)
- **Replication Support:** Vincent Thorne (vthorne@pm.me)