

Reproducibility Package

Project: Socio-Emotional Skills

1. Overview

This reproducibility package contains the data processing, analysis, and visualization code used in the paper: **"Labor Market Returns to Socio-Emotional Traits and the Impacts of Training Programs: A Meta-Analysis"**

Collaborators

- **Diego Angel-Urdinola**¹
- **Maurice Kugler**²
- **Narae Lee**¹
- **Angelo Santos**¹²
- **Olzhas Zhorayev**¹²

¹ World Bank

² George Mason University

Software

- **Python**
- **Stata**

Objective

The objective of this project is to study:

1. **Labor market returns to socio-emotional skills (SES)**, and
2. **The effectiveness of SES training programs**,

using a systematic evidence synthesis and meta-analytic framework.

The package is designed to be run end-to-end using a **Python master script**, which calls Stata in batch mode to process the data and then generates tables using Stata and figures using Python.

Date of Production

December 2025

2. Data Availability

Data Availability Statement

- ☐ All data are publicly available.

- ☐ Some data cannot be made publicly available.
- ☒ **No data can be made publicly available.**

Data Files Used

This package relies on two cleaned data files:

1. **Data_clean.dta**

- Content: Harmonized wage-return estimates for socio-emotional and cognitive skills
- Format: Stata **.dta**
- Used for: Wage-returns meta-analysis

2. **Training_data_clean.xlsx**

- Content: Harmonized meta-analysis results on socio-emotional skill training impacts
- Format: Excel **.xlsx**
- Used for: Training-impact meta-meta-analysis

Source:

World Bank Development Data Hub (*forthcoming*). The data were collected by the project team. For detailed information on data collection please refer to the **Database Construction** section and **Appendix A** of the paper.

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.

Data Access Note

All data used in this project are available upon request to the authors. The data are **included in a private GitHub repository for authorized replicators**.

3. Instructions for Replicators

Prerequisites

To replicate the results, users must:

1. Obtain access to the required data files (**Data_clean.dta**, **Training_data_clean.xlsx**)
2. Place the files in the **Data/** folder
3. Install the required software (see Section 5)

Running the Package

This project uses **Python as the master runner**.

- Step 1. Open a terminal in the project root directory
- Step 2. Set the path to your local Stata executable (example for Windows):

```
$env:STATA_EXE="C:\Program Files\Stata18\StataMP-64.exe"
```

- Step 3. Run the full pipeline:

```
python Code/00-Master-run-all.py
```

This will:

- run all Stata do-files in batch mode
- run all Python scripts
- generate intermediate Excel outputs in Temp/
- generate all tables in Tables/
- generate all figures in Figures/

4. List of Exhibits

Reproducibility Coverage

- ☐ All numbers provided in text in the paper
- ☐ All tables and figures in the paper
- ☒ Selected tables and figures in the paper (as explained below)

Outputs Generated by This Package

Exhibit	Output Filename	Script	Note
Table 2	Table2_DescriptiveStats_Income.xlsx	6.Descriptive-Stat-Income.do	Found in the Tables Folder
Table 3	Table3_DescriptiveStats_Training.xlsx	10.Descriptive-Stat-Training.do	Found in the Tables Folder
Table A3	TableA3_DescriptiveStats_Region.xlsx	14.Descriptive-Stat-Region.do	Found in the Tables Folder
Figure 1	Figure1.WageReturns.png	7.Wage-Returns-RE.py	Found in the Figures Folder
Figure 2	Figure2.Heterogeneity.png	8.Heterogeneity.py	Found in the Figures Folder
Figure 3	Figure3.FunnelPlot.png	9.Funnel-Plot.do	Found in the Figures Folder

Exhibit	Output Filename	Script	Note
Figure 4	Figure4.TrainingImpact.png	11.Training-Impact.py	Found in the Figures Folder
Figure 5	Figure5.TrainingHeterogeneity.png	12.Training-Heterogeneity.py	Found in the Figures Folder
Figure 6	Figure6.FunnelPlotTraining.png	13.Funnel-Plot-Training.do	Found in the Figures Folder
Figure A7	FigureA7.WageReturnsREMain.png	15.Wage-Returns-RE-Main.py	Found in the Figures Folder
Figure A8	FigureA8.WageReturnsFEAll.png	17.Wage-Returns-FE-All.py	Found in the Figures Folder
Figure A9	FigureA9.WageReturnsREBig5.png	17.Wage-Returns-RE-Big5.py	Found in the Figures Folder

5. Requirements

Computational Requirements

- Operating system: Windows (tested)
- Internet access: not required once data are available
- Storage: minimal (temporary Excel and figure outputs)

Software Requirements

- **Stata 18** (tested)
 - Required for meta commands and data format compatibility
 - Earlier versions (Stata 16-17) may work but are not guaranteed due to differences in the meta command behavior
- **Python ≥ 3.9**

Python Packages

- pandas
- numpy
- matplotlib
- openpyxl

Install via:

```
pip install pandas numpy matplotlib openpyxl
```

Runtime

- Full pipeline runtime: approximately 2.5 minutes on a tested laptop
 - Memory usage: modest
-

6. Code Description

Master Script

- **00-Master-run-all.py** is a Python master script that:
 - sets project paths
 - calls Stata do-files in batch mode
 - runs all Python scripts

Stata Setup

- **Setup.do** defines project directories and ensures all Stata scripts run reproducibly when called from Python.

Stata Data Preparation and Analysis Files

- **1.Pool-Estimates-RE-All.do** prepares data and runs meta-analysis for wage returns using random effects.
- **2.Pool-Estimates-Training.do** prepares data and runs meta-meta-analysis for training impacts using random effects.
- **3.Pool-Estimates-RE-Main.do** prepares data and runs meta-analysis for wage returns using random effects (restricted to main specifications in the analyzed papers).
- **4.Pool-Estimates-FE-All.do** prepares data and runs meta-analysis for wage returns using fixed effects.
- **5.Pool-Estimates-RE-Big5.do** prepares data and runs meta-analysis for wage returns using random effects (restricted to papers controlling for all Big Five personality traits).

Stata Table Code

- **6.Descriptive-Stat-Income.do**, **10.Descriptive-Stat-Training.do**, and **14.Descriptive-Stat-Region.do** generate descriptive statistics tables.

Stata Figure Code

- **9.Funnel-Plot.do** and **13.Funnel-Plot-Training.do** generate funnel plots.

Python Figure Scripts

- **7.Wage>Returns-RE.py**, **8.Heterogeneity.py**, **11.Training-Impact.py**, **12.Training-Heterogeneity.py**, **15.Wage>Returns-RE-Main.py**, **16.Wage>Returns-FE-All.py**, and **17.Wage>Returns-RE-Big5.py** generate figures.
-

7. Folder Structure

```
Code/
├─ 00-Master-run-all.py
├─ Paths.py
├─ Setup.do
├─ 1.Pool-Estimates-RE-All.do
├─ 2.Pool-Estimates-Training.do
├─ 3.Pool-Estimates-RE-Main.do
├─ 4.Pool-Estimates-FE-All.do
├─ 5.Pool-Estimates-RE-Big5.do
├─ 6.Descriptive-Stat-Income.do
├─ 7.Wage>Returns-RE.py
├─ 8.Heterogeneity.py
├─ 9.Funnel-Plot.do
├─ 10.Descriptive-Stat-Training.do
├─ 11.Training-Impact.py
├─ 12.Training-Heterogeneity.py
├─ 13.Funnel-Plot-Training.do
├─ 14.Descriptive-Stat-Region.do
├─ 15.Wage>Returns-RE-Main.py
├─ 16.Wage>Returns-FE-All.py
└─ 17.Wage>Returns-RE-Big5.py
Data/
├─ Data_clean.dta
└─ Training_data_clean.xlsx
Figures/
└─ Generated figures
Tables/
└─ Generated tables
Temp/
└─ Intermediate Excel outputs
README.md
README brief.txt
Manuscript.pdf
```