**README for materials associated with Hart et al., 2024 *Psychological Bulletin***

The materials detailed below have been posted to LDbase and can be used to replicate the analyses presented in Hart et al., 2024 published in *Psychological Bulletin*.

Contact Emma Hart (erh2169@tc.columbia.edu) or Tyler Watts (tww2108@tc.columbia.edu) with any questions.

***Primary Datasets***

There are two versions of the dataset: MERF\_soc\_cog\_wide.dta and MERF\_soc\_cog\_long.dta.

MERF\_soc\_cog\_wide.dta is organized in wide format. This is the primary dataset used for the analyses in the manuscript. Each observation represents one unit of analysis: an aligned group in which intervention impacts were reported for a construct measured consistently (i.e., consistent measure, subscale, and reporter) at post-test and at least one follow-up assessment, organized within follow-up assessment bins (i.e., 6- to 12-months follow-up, 1- to 2-years follow-up, etc). Statistical details (effect size, standard errors, sample size, etc.) and descriptive details (study, construct, measure, reporter, whether the outcome is social-emotional or cognitive) are indicated for each aligned group. This version of the data contains all variables necessary for running the main regression models in R.

MERF\_soc\_cog\_long.dta is organized in long format. Each observation represents one intervention impact estimate from one assessment wave, either post-test or follow-up, for all of the aligned groups included in our sample. This version of the data was used in the Stata do file for creating study-level descriptive information and creating figures. It does not have all of the variables necessary for the regression models (which the wide version of the data contains).

***Analytic Syntax***

There are two analytic syntax files for this manuscript: MERF\_soc\_cog\_syntax1.R and MERF\_soc\_cog\_syntax2.dta.

MERF\_soc\_cog\_syntax1.R is an R syntax file that contains the code for the main and supplemental meta-regression models. This is the file to look at for replicating all meta-regression models (both those used to estimate descriptive information and to estimate predictive associations).

MERF\_soc\_cog\_syntax2.dta is a Stata syntax file that contains the code for creating all of the figures and Table 1 and 2. This is the file to look at for re-creating figures and some study-level descriptives (note: meta-analytic estimates of average effects are estimated in the R file).

***Supplemental Datasets***

There are three datasets provided for supplemental models that involve different aligned groups. They only contain the variables necessary to run the analyses.

MERF\_soc\_cog\_supp\_altgroups1.dta is organized in wide format. This is a supplemental dataset associated with Table S9 column 1 in Hart et al., 2024 for alternate grouping approach #1 where measure, subscale, and reporter were allowed to vary across waves.

MERF\_soc\_cog\_supp\_altgroups2.dta is organized in wide format. This is a supplemental dataset associated with Table S9 column 2 in Hart et al., 2024 for alternate grouping approach #2 where the analytic sample was comprised of all social-emotional and cognitive outcomes averaged together at each wave (i.e., no specific construct, measure, subscale, reporter alignment).

MERF\_soc\_cog\_supp\_estdrop.dta is organized in wide format. This is a supplemental dataset associated with Table S5 column 7 in Hart et al., 2024. The aligned groups were formed after dropping all effects that required more “estimation” (see manuscript for more details).

***Additional Files***

1. MERF\_soc\_cog\_codebook.csv contains a full list of variables included in each dataset.
2. MERF\_protocol contains detailed information about the creation of the MERF dataset (i.e., inclusion/exclusion process, coding, etc.).

***FAQs***

Why does the intervention impact estimate reported in the dataset look different than what is reported in the paper?

* Estimates were averaged in several cases:
	+ First, if multiple papers that we coded reported intervention impacts for the same outcome (with aligned construct, measure, subscale, reporter) at the same assessment wave (e.g., post-test, 12-mo follow-up, etc.), these impact estimates were averaged for the purposes of our analysis.
	+ Second, if there were multiple estimates of impacts from the same follow-up window (i.e., 6- to 12-months follow-up, 1- to 2-years follow-up, etc.) derived from one or multiple papers, these were averaged to create a single estimated treatment impact. For example, if a paper reported impacts at 6 and 9 months follow-up for the same outcome, these impacts would be averaged to form a single estimate for analysis in 6- to 12-month follow-up models.
	+ Additionally, there were cases when data were reported for multiple treatment groups formed via random assignment within a study. We opted to leave effect sizes presented separately by treatment group when possible since the effects reflected experimental treatment impacts. However, there were some instances when effect sizes were reported for each treatment group separately at earlier assessment waves (e.g., pre-test, post-test, 6-12-month follow-up), and in aggregated form at later assessment waves (e.g., 3-year follow-up). In these cases, treatment-specific effect sizes, standard errors, and p-values were averaged to form an average treatment effect that could be investigated in alignment with the effect sizes from later assessment waves.
	+ Finally, there were cases when data were reported separately for different sub-samples within a study (e.g., for boys and girls, for “low-risk” and “high-risk” participants, etc.). For these cases, we derived a main treatment effect by averaging the effect size estimates for each group, weighted by the group sample size. The same weighted averaging was used for standard errors and p-values. Critically, if the treatment effect was only reported for one sub-sample (e.g., only boys, only “low-risk” participants, etc.), then the effects were dropped from the meta-analysis so that each estimate represented a main treatment impact of the original random assignment to treatment or control.

Why is this treatment impact estimate not in the sample?

* If there was not enough information reported to calculate an intervention effect and standard error, the outcome is not included in our sample.
* As detailed in the MERF protocol, to avoid duplication issues, we opted to code in composite measures, rather than subscales, when possible (see protocol for more details).

What is the difference between “studyid” and “studyname\_txname”?

* Within a particular study, there can be multiple treatment groups that were formed through randomization. Studyid refers to the larger study, and studyname\_txname refers to the specific treatment-control group contrast formed through randomization. If there is only one treatment group within a study, then studyname\_txname will indicate the study name. If there are multiple treatment groups, then studyname\_txname will indicate the study name followed by the treatment group name.