

Meta analysis coding title

Supplementary materials

Chiara Montuori* Filippo Gambarota[†] Gianmarco Altoè[‡] Barbara Arfè[§]

Contents

1	Introduction	2
1.1	Meta-analysis	2
1.2	Papers	2
2	Dataset description	2
2.1	Pre-processing steps	3
2.2	Multiple effects for the same outcome	3
3	Effect size computation	4
4	Meta-analysis models	5
4.1	Univariate vs multivariate model	5
4.2	Random-effect vs fixed-effect model	5
4.3	Modelling functions	6
5	Correlations inputation	7
6	Meta-analysis table	7
7	Multiverse analysis	8
	References	15

*Department of Developmental and Social Psychology, chiara.montuori@phd.unipd.it

[†]Department of Developmental and Social Psychology, filippo.gambarota@phd.unipd.it

[‡]Department of Developmental and Social Psychology, gianmarco.altoe@unipd.it

[§]Department of Developmental and Social Psychology, barbara.arfe@unipd.it

1 Introduction

This document explain in details the statistical approach of the main paper. All the analysis are conducted with R (R Core Team, 2021) and all scripts are available on the Open Science Framework repository.

- Section 2 present the full dataset description with all extracted variables. Despite only a subset of these variables are used to compute the meta-analysis model, we reported all extracted study-level information.
- Section 3 describe the computation of the effect size measure and the sampling variance.
- Section 4 describe the main meta-analysis model presented in the paper together with modeling alternatives
- Section 5 describe the correlations imputation that is necessary to compute the meta-analysis model
- Section 7 present the sensitivity analysis for different correlations values and different meta-analytic models

1.1 Meta-analysis

The meta-analysis is the statistical procedure to combine information from already conducted studies. The core aspect of a meta-analysis is to weight each included effect according to the amount of information (i.e., precision) that the study provide. In order to compute a meta-analysis model we need:

- The effect size measure
- The effect size variance (i.e., the inverse of the precision)
- The meta-analysis model

1.2 Papers

Almost all published papers reported enough information to calculate the effect size. Di Lieto and colleagues (2020) provided us raw data in order to calculate all relevant parameters. They use a design with 3 time points (T0 = pre, T1 = post and T2 = follow-up) for both the experimental and control group. Furthermore they measured several different outcomes measures. In order to include the paper we performed these pre-processing steps:

- We selected only the T0 and T1 in order to have a single pre-post measure
- We included all participants that have both time points (T0 and T1) in at least one outcome measure. For example, if a child has T0 and T1 for the outcome x but only T0 for the outcome y we keep the child in the analysis. This create a situation where different outcomes have different sample sizes but maximize the amount of available information.

2 Dataset description

The main dataset can be found on the online OSF repository under `data/raw/meta_table_cleaned.csv`. The following list describes the meaning of each column:

- **paper**: Unique id for every paper
- **paper_id**: Unique id for every paper with authors and year
- **author**: Paper's authors
- **year**: Publication year
- **Participants**: Participants' characteristics: ST (typical development), SA (atypical development)
- **n_EX**: Sample size for the experimental group
- **n_CT**: Sample size for the control group
- **male_EX**: Number of males for the experimental group
- **female_EX**: Number of females for the experimental group

- **male_CT**: Number of males for the control group
- **female_CT**: Number of females for the control group
- **Training_length**: Length of the training (in weeks)
- **Minute_Training_length**: Length of a single training session (in minutes)
- **Training_Mode**: Type of training activity: CV (Virtual Coding), ER (Educational Robotics)
- **Training_CV_Type**: Types of virtual coding activities: structured and unstructured
- **M_SES_EX**: Mean socio-economic status (SES) for the experimental group
- **SD_SES_EX**: Standard deviation of socio-economic status (SES) for the experimental group
- **M_SES_CT**: Mean socio-economic status (SES) for the control group
- **SD_SES_CT**: Standard deviation of socio-economic status (SES) for the control group
- **M_age_CT**: Mean age for the control group
- **M_age_EX**: Mean age for the experimental group
- **outcome**: The outcome measure/test used
- **M_pre_EX**: Mean pre-training score for the experimental group
- **M_post_EX**: Mean post-training score for the experimental group
- **SD_pre_EX**: Standard deviation pre-training score for the experimental group
- **SD_post_EX**: Standard deviation post-training score for the experimental group
- **M_pre_CT**: Mean pre-training score for the control group
- **M_post_CT**: Mean post-training score for the control group
- **SD_pre_CT**: Standard deviation pre-training score for the control group
- **SD_post_CT**: Standard deviation post-training score for the control group
- **outcome2**: The recoded outcome variable
- **Flip_NoFlip**: Whether the effect size need to be flipped in order to have the same direction (i.e., positive values, the training improve performance)
- **eff_size**: The computed $dpcc_2$
- **eff_size_var**: The computed $dpcc_2$ variance
- **eff_size_se**: The computed $dpcc_2$ standard error

2.1 Pre-processing steps

The raw dataset is available on the online Open Science Framework repository. We performed few minimal pre-processing steps:

- renaming relevant columns
- re-coding the **outcome** variable into the **outcome2** variable
- separating the Arfè et al. (2019) paper into two distinct papers (see below)
- converting outcomes names in English

2.1.1 Arfè et al. (2019)

Arfè et al. (2019) is the only paper that contain multiple sub-studies creating a multilevel structure. In order to reduce the dataset complexity and given that the two sub-studies refers to a different pool of subjects, we decided to consider this studies as two separate papers, creating Arfè et al. (2019a) and Arfè et al. (2019b).

2.2 Multiple effects for the same outcome

We decided to recode the **outcome** into **outcome2** in order to have a small set of outcomes according to the underlying psychological construct. For example, if the test x and the test y are both considered a Working Memory measure we decided to recode both outcomes as “Working Memory”. This reduce the dataset complexity but create a situation where a single paper has multiple effects belonging to the same outcome. We decided to aggregate multiple effects of the same **outcome2** using the approach suggested by Borenstein et al. (2009, pp. 225–233) implemented using the `metafor::aggregate.esalc()` function (<https://wwiechthb.github.io/metafor/reference/aggregate.esalc.html>). To note, we use a slightly different

approach compared to Borenstein et al.(2009, pp. 225–233) computing an *inverse-variance weighted average* instead of an *un-weighted average*. Essentially, the weighted average combine multiple effect sizes taking into account the precision (i.e., the inverse of the variance).

We used the `metafor::aggregate.escalc()` as follows:

```
aggregate_effects <- function(data, rho, split_by = paper, weighted = TRUE){
  split_by <- rlang::enexpr(split_by)
  data <- metafor::escalc(yi = eff_size, vi = eff_size_var,
    var.names = c("eff_size", "eff_size_var"),
    data = data)
  dat_split <- split(data, pull(data, !!split_by))
  dat_split <- purrr::map_dfr(dat_split, function(x) metafor::aggregate.escalc(x, cluster = outcome2,
    bind_rows(dat_split) %>% tibble()
  })
```

This is a wrapper of the `metafor::aggregate.escalc()` function that split the dataset according to the paper and then aggregate within each outcome2.

3 Effect size computation

All included studies used a Pretest-Posttest-Control Group (PPC) design where the treated group is compared with an age-matched control group before and after a certain treatment. Morris (2008) described different effect size indexes for the PPC design. We decided to use the d_{ppc2} index because it provide an unbiased estimation of the true effect size and the sampling variance can be analytically computed. The d_{ppc2} is calculated as indicated in Equation (1).

$$d_{ppc2} = c_p \frac{(M_{T,post} - M_{T,pre}) - (M_{C,post} - M_{C,pre})}{SD_{pooled,pre}} \quad (1)$$

The numerator describe the actual mean difference between treated and control group of the respective *pre-post* scores. The difference is standardized using only a pooled standard deviation of pre-test scores. The variance is calculated in Equation (2):

$$\sigma^2(d_{ppc2}) = c_p^2(1 - \rho) \left(\frac{n_t + n_c}{n_t n_c} \right) \left(\frac{n_t + n_c - 2}{n_t + n_c - 4} \right) \left(\frac{1 + \Delta^2}{2(1 - \rho) \left(\frac{n_t + n_c}{n_t n_c} \right)} \right) \quad (2)$$

Both the d_{ppc2} and the variance are adjusted using a small sample correction factor c_p calculated in Equation (3)

$$c_p = 1 - \frac{3}{4(2n_t + 2n_c - 4) - 1} \quad (3)$$

Positive values of the effect size means that the treatment increased the performance of the experimental group. However for some measures, higher values corresponds to worse performance (e.g., number of errors). In this case we computed the effect size as described in the previous section and then we flipped the sign obtaining the same interpretation regardless the original measure.

For the actual computation we wrote two functions following the documentation of the `metafor` package: (<http://www.metafor-project.org/doku.php/analyses:morris2008>), `get_dpcc2()`:

```

get_dpcc2 <- function(mt_pre, mt_post, mc_pre, mc_post,
                     st_pre, st_post, sc_pre, sc_post,
                     nt, nc){

  mdiff <- (mt_post - mt_pre) - (mc_post - mc_pre)
  poolvar <- sqrt((((nt - 1) * st_pre^2) + ((nc - 1) * sc_pre^2)) / (nt + nc - 2))
  cp <- 1 - (3/(4*(nt + nc - 2) - 1))

  dpcc2 <- (mdiff / poolvar) * cp

  return(dpcc2)
}

and get_dpcc2_var():
get_dpcc2_var <- function(dpcc2, nt, nc, rho){

  dpcc2_var <- 2*(1-rho) * (1/nt + 1/nc) + dpcc2^2 / (2*(nt + nc))

  return(dpcc2_var)
}

```

4 Meta-analysis models

4.1 Univariate vs multivariate model

When multiple variables are collected from the same pool of participants (e.g., multiple outcomes studies) and/or there are multiple experiments within the same paper, effects size cannot be considered independent (Cheung, 2014, 2019). Instead of pooling multiple effects into a single measure or keeping only a single effect per study, we computed a multivariate meta-analysis model taking into account the dependency (i.e., the correlation) between multiple effects size calculated on the same pool of participants. In order to compute the multivariate model we need to include or impute the variance-covariance matrix that describe how multiple effects correlate within each study. The section 5 explain our imputation approach.

4.2 Random-effect vs fixed-effect model

Another important choice for the meta-analysis is between the fixed-effect and the random-effect model. Under the fixed-effect model we make the assumption of a single *true* effect size that is estimated by a sample of studies. For this reason, the variability observed in the empirical meta-analysis is only caused by the sampling error.

On the other side, the *random-effect* model assume a *distribution* of *true* effects. The model estimate the average effect (the *mean* of the distribution) and the between-study heterogeneity τ^2 (i.e., the *variance* of the distribution). The critical assumption is that the observed heterogeneity is composed by sampling variability (as the *fixed-effect* model) and *true* between-study variability. The latter can be caused by different experimental designs, participants features or other study-level variables and can be explained using moderators (i.e., meta-regression).

The goal of the *random-effect* model is to generalize the meta-analytic findings at the population level while the *fixed-effect* model is focused on a specific pool of studies.

We decided to use a *fixed-effect* model for several reasons. Under the *random-effect* framework, the multivariate model estimate the average effect and the variability (τ^2) for each included outcome. Firstly, one of the main reason to estimate τ^2 (i.e., the focus of the *random-effect* model) is to explain the observed heterogeneity with study-level variables. Given our limited pool of studies, we did not included a meta-regression analysis. Crucially, with a limited number of studies the τ^2 estimation can be strongly biased (Veroniki et al., 2016) influencing also the pooled effect estimation (Borenstein et al., 2009, pp. 73–75). Finally, the *fixed-effect* is less complex in terms of model parameters beacuse we need to estimate only the average effect of each included outcome, taking into account the statistical dependency (see 4.1).

4.3 Modelling functions

We tested each model parameter (i.e., average effect for a specific outcome) using the Wald *z-test* with $\alpha = 0.05$.

For the *univariate fixed-effect* model we use the `metafor::rma()` function:

```
fit_uni_fixed <- function(data){
  rma(yi = eff_size,
      vi = eff_size_var,
      method = "FE",
      data = data)
}
```

For the *univariate random-effect* model we use the `metafor::rma()` function and the REML estimator for τ^2 :

```
fit_uni_random <- function(data){
  rma(yi = eff_size,
      vi = eff_size_var,
      method = "REML",
      data = data)
}
```

For the *multivariate fixed-effect* model we use the `metafor::rma.mv()` function:

```
fit_multi_fixed <- function(data, cov_matrix){
  rma.mv(
    yi = eff_size,
    V = cov_matrix,
    mods = ~ 0 + outcome2,
    data = data)
}
```

In this case we used `outcome2` as moderator with a *cell-mean parametrization* (Schad et al., 2020) (i.e., removing the intercept `0 + outcome2`). In this way model parameters and statistical tests correspond directly to the average effect for each outcome.

For the *multivariate random-effect* model we use the `metafor::rma.mv()` function:

```
fit_multi_random <- function(data, cov_matrix){
  rma.mv(
    yi = eff_size,
    V = cov_matrix,
    mods = ~ 0 + outcome2,
    random = ~ outcome2|paper_id,
    struct = "UN",
```

```

    method = "ML",
    data = data)
}

```

As explained in the previous section, this model estimate also a τ^2 for each outcome (`random = ~ outcome2|paper_id`).

4.3.1 Variance-covariance matrix

For the *multivariate* model is necessary to include a block variance-covariance matrix. Essentially, each study has n different outcomes and we need a $n \times n$ variance-covariance matrix where the diagonal is the d_{pcc2} variance for a specific outcome and off-diagonal elements are the covariance between pairs of outcomes. The covariance between two outcomes is calculated as the product of the correlation and the variances:

$$Cov_{O_1, O_2} = \rho_{(O_1, O_2)} \sigma_{O_1} \sigma_{O_2} \quad (4)$$

Combining all study-level matrices we obtain a full block-variance-covariance matrix to use within the `rma.mv()` function. The `metafor::vcalc()` function allow to automatically create the full matrix specifying the assumed correlation and a clustering variable:

```

get_block_cov_matrix <- function(data, rho){
  vcalc(eff_size_var, cluster = paper_id, obs = effect_id, data = data, rho = rho)
}

```

5 Correlations inputation

As explained in previous sections, we decided to use a *fixed-effect multivariate* model. In order to compute the model we need to include 3 correlations measures:

- For the effect size computation, we need the correlation between pre-post (*pre-post correlation*) scores for both groups
- For the aggregation of multiple effects within the same paper (*aggregation correlation*) (see Section 2.2) we need the correlation between the effects
- For the actual multivariate model we need the full variance-covariance matrix of different outcomes. To create the matrix we need to include the correlation between outcomes within each study (*multivariate correlation*).

Given that correlations are rarely reported in published papers, we decided to impute these values and assessing the impact with a multiverse-like approach (Steege et al., 2016). In particular we used:

- A $\rho_{pre-post}$ of 0.5, 0.7 and 0.9
- A ρ_{agg} of 0.3, 0.5, 0.7
- A ρ_{multi} of 0.3, 0.5 and 0.7

6 Meta-analysis table

Table 1 describe all included papers with pre-post scores for the experimental and control group along with the computed effect size. Furthermore, the table is organized grouping the affects according to the considered outcome in order to highlight the multivariate data structure.

Table 1: Pre-post scores (M = mean, SD = standard deviation, N = sample size) for the control and experimental group. The last two columns represent the computed effect size and the variance. The effects are organized according to the outcome value and the corresponding paper.

		Experimental Group					Control Group					Effect size	
Outcome	Paper	N	M _{pre}	M _{post}	SD _{pre}	SD _{post}	N	M _{pre}	M _{post}	SD _{pre}	SD _{post}	d _{ppc2}	σ^2 d _{ppc2}
Cognitive Flexibility Acc.	Di Lieto et al. (2020a)	96	7.22	8.39	2.06	1.69	91	7.19	8.27	2.10	1.66	0.04	0.01
	Di Lieto et al. (2020b)	18	5.72	6.61	2.37	2.12	18	7.05	7.05	2.15	2.41	0.38	0.07
Inhibition Acc.	Arfè et al. (2019a)	44	4.63	1.79	4.45	2.01	36	5.76	4.16	4.36	3.72	0.21	0.02
	Arfè et al. (2019b)	19	4.03	1.58	3.80	1.81	19	3.74	2.82	2.79	2.27	0.46	0.05
	Arfè et al. (2020)	88	5.71	1.97	6.01	2.23	91	5.66	4.12	0.26	1.50	0.44	0.01
	Di Lieto et al. (2020a)	96	5.72	3.39	4.62	3.42	91	4.61	3.02	3.53	3.13	0.14	0.01
	Di Lieto et al. (2020b)	18	8.16	6.50	7.64	6.14	18	6.58	5.38	4.39	6.66	0.06	0.05
Planning Acc.	Arfè et al. (2019a)	44	7.13	9.84	2.46	2.45	36	5.75	7.57	2.91	2.80	0.30	0.02
	Arfè et al. (2019b)	19	8.39	11.42	3.51	2.60	19	8.92	8.55	3.23	3.29	0.99	0.06
	Arfè et al. (2020)	88	6.02	9.59	2.88	2.39	91	4.95	7.05	1.34	1.18	0.61	0.01
Problem Solving	Akcaoglu & Koehler (2014)	20	-1.11	-0.07	1.23	1.11	24	-1.17	-1.29	0.91	1.15	1.05	0.05
	Arfè et al. (2019a)	44	4.31	6.12	1.46	1.06	36	3.09	3.68	1.60	1.92	0.79	0.03
	Arfè et al. (2019b)	19	6.05	7.16	1.08	0.96	19	5.58	5.21	1.17	1.08	1.29	0.08
	Arfè et al. (2020)	88	3.90	5.86	1.45	0.87	91	3.57	3.74	0.08	0.08	1.75	0.02
	Erol & Cirak (2021)	16	93.25	99.12	16.32	12.39	18	89.11	86.72	10.79	10.88	0.59	0.08
	La Paglia et al. (2017)	30	54.19	58.09	7.52	6.89	30	56.29	56.86	11.58	12.22	0.34	0.04
	Nam et al. (2010)	30	9.55	13.36	3.50	3.55	30	10.09	11.50	3.16	2.84	0.71	0.04
	Pardamean et al. (2011)	43	9.98	12.05	3.08	2.90	42	9.35	10.76	2.02	3.23	0.26	0.02
	WonNam et al. (2019)	25	25.72	37.16	6.29	3.88	28	26.79	29.75	5.07	6.33	1.47	0.07
Working Memory Acc.	Di Lieto et al. (2020a)	96	3.34	4.93	1.92	2.31	91	3.15	4.23	1.76	1.90	0.26	0.01
	Di Lieto et al. (2020b)	18	2.36	3.34	1.25	1.94	18	2.11	3.42	1.14	2.01	-0.21	0.04

7 Multiverse analysis

We assessed the impact of our modelling assumptions using a sensitivity analysis approach. In particular we considered:

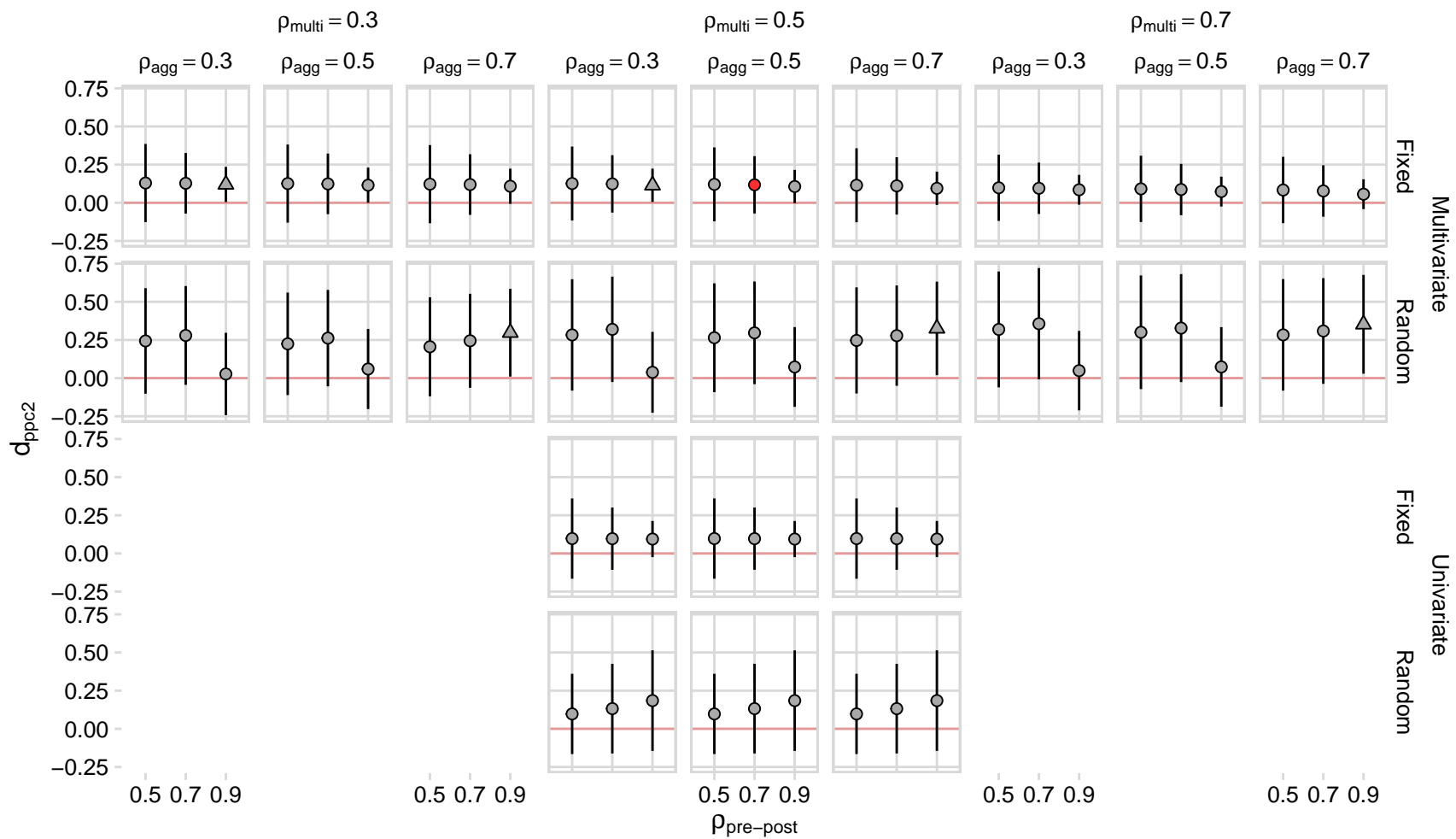
- 4 meta-analysis models:
 - *univariate fixed-effect*

- *univariate random-effect*
- *multivariate fixed-effect*
- *multivariate random-effect*
- All correlations combinations ($3\rho_{pre-post} \times 3\rho_{aggregation} \times 3\rho_{multivariate}$)

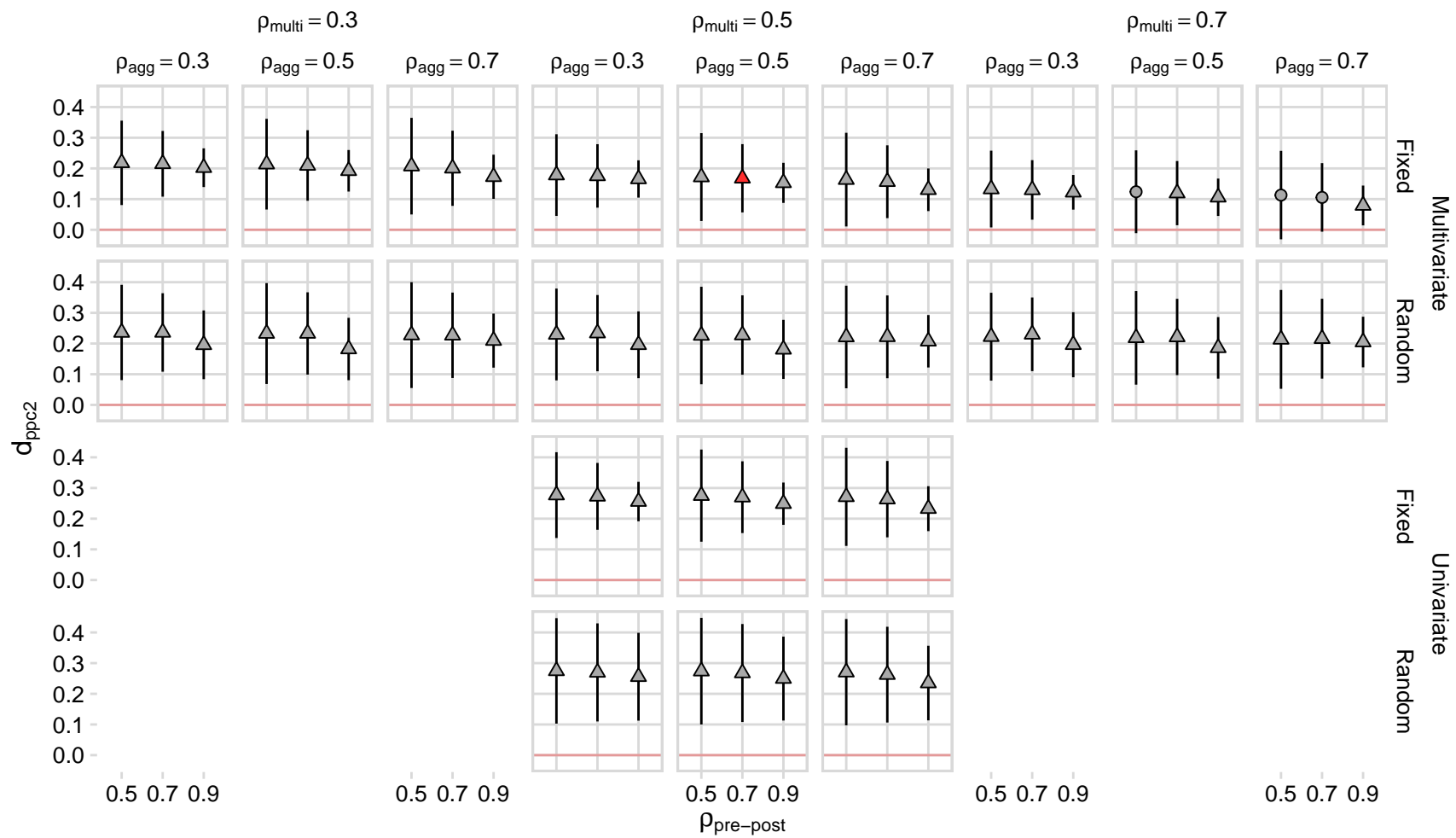
In this way we computed 72 meta-analysis directly assessing the impact on the parameters estimation and p-values¹. All plots in the following pages depicted the sensitivity analysis for each outcome. Triangles indicates parameters with $p < 0.05$ while points indicates parameters with $p > 0.05$. Red shapes indicates the correlations/model combination included in the paper. Clearly, *random effect models* (both *univariate* and *multivariate*) is associated with wider confidence intervals (i.e., less precise estimation) compared to *fixed effect models*. To note, given that the ρ_{multi} is not relevant for *univariate* models, the lower part of the figure is repeated.

¹For *multivariate* models (*fixed* and *random*) we have 3x3x3 correlation combinations while for *univariate* models (*fixed* and *random*) only 3x3 ($\rho_{pre-post}$ and ρ_{agg})

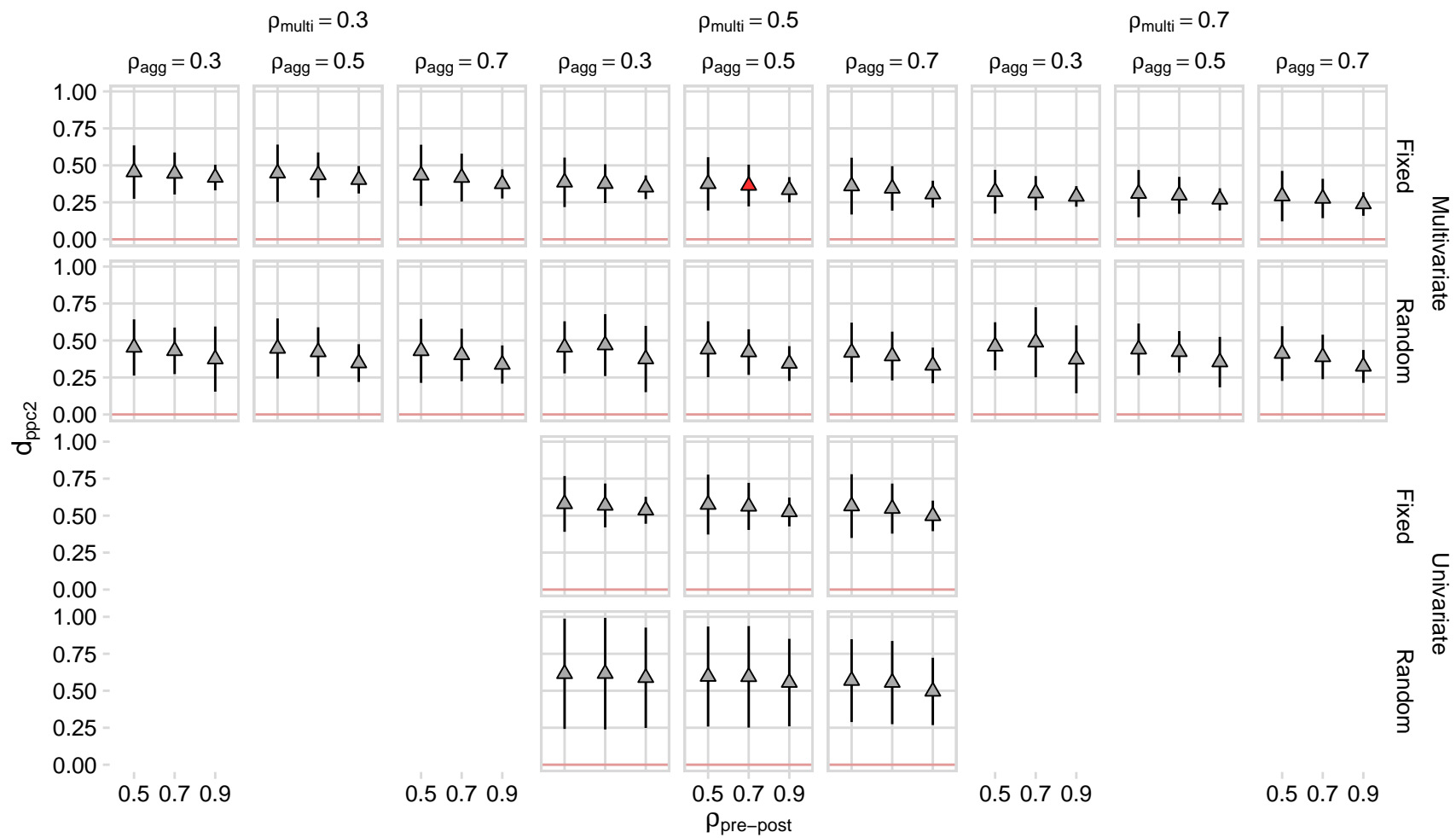
Cognitive Flexibility Acc.



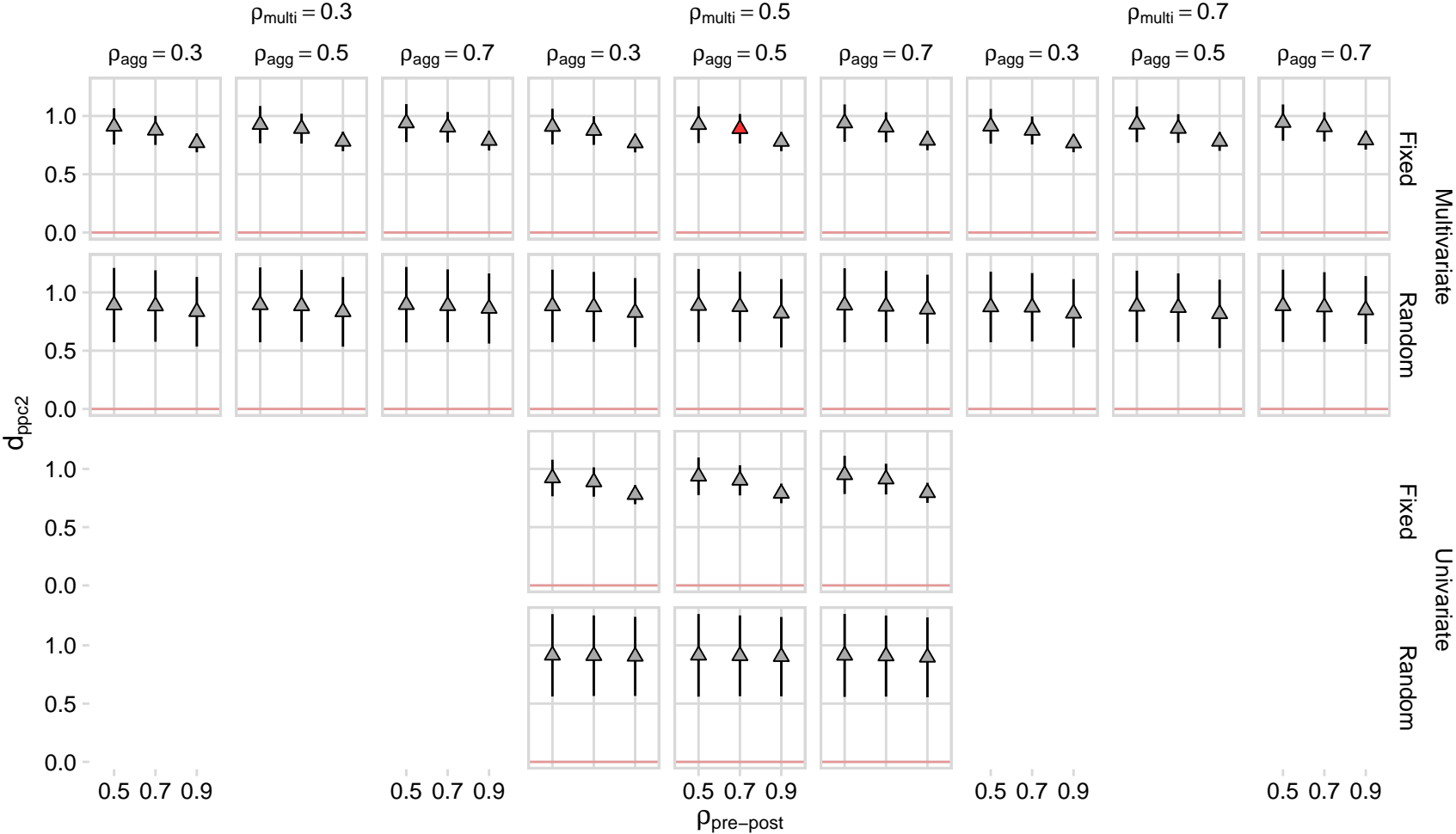
Inhibition Acc.



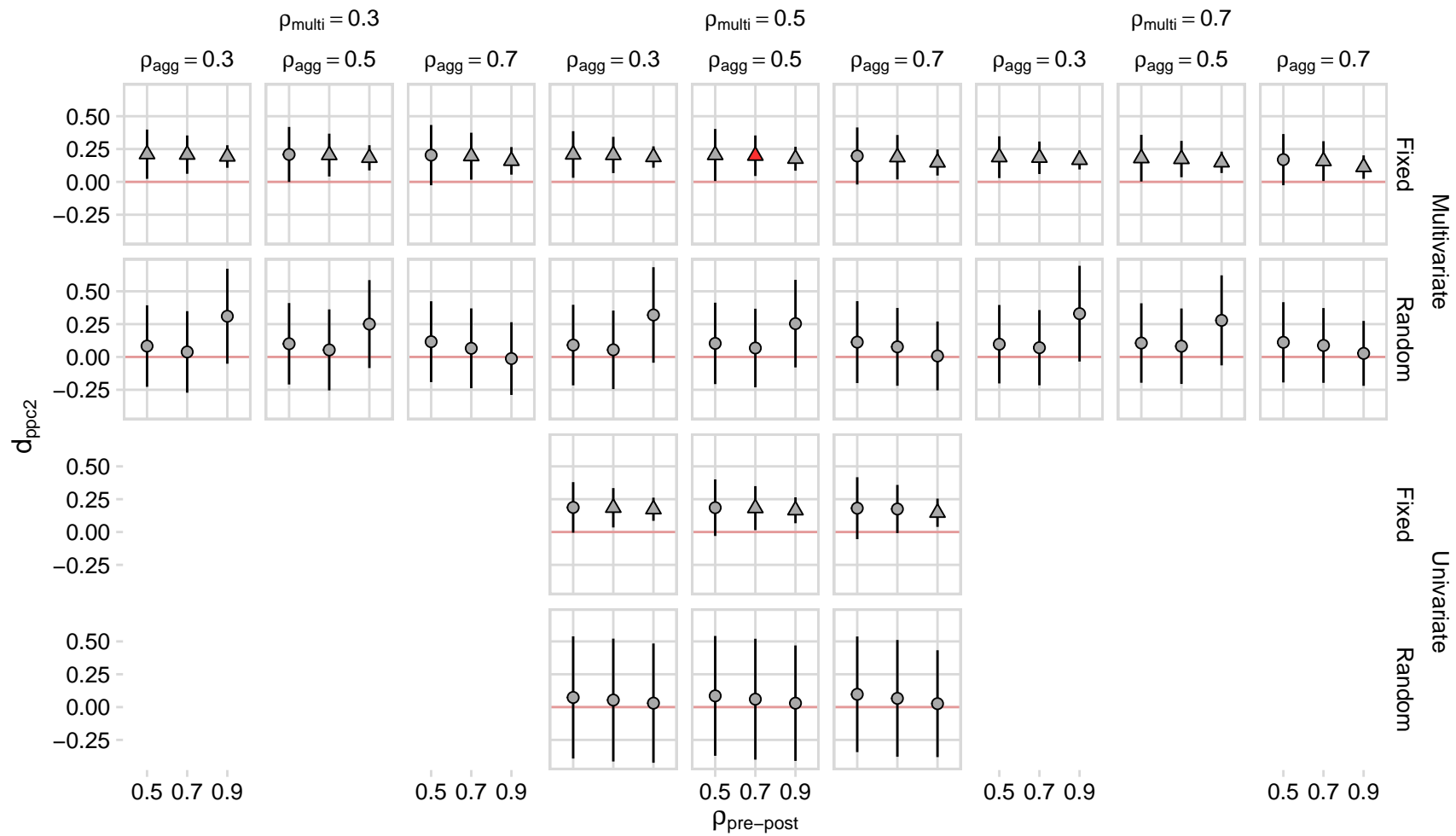
Planning Acc.



Problem Solving



Working Memory Acc.



References

- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. <https://doi.org/10.1002/9780470743386>
- Cheung, M. W.-L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychol. Methods*, 19(2), 211–229. <https://doi.org/10.1037/a0032968>
- Cheung, M. W.-L. (2019). A guide to conducting a Meta-Analysis with Non-Independent effect sizes. *Neuropsychol. Rev.*, 29(4), 387–396. <https://doi.org/10.1007/s11065-019-09415-6>
- Di Lieto, M. C., Castro, E., Pecini, C., Inguaggiato, E., Cecchi, F., Dario, P., Cioni, G., & Sgandurra, G. (2020). Improving executive functions at school in children with special needs by educational robotics. *Front. Psychol.*, 10, 2813. <https://doi.org/10.3389/fpsyg.2019.02813>
- Morris, S. B. (2008). Estimating effect sizes from Pretest-Posttest-Control group designs. *Organizational Research Methods*, 11(2), 364–386. <https://doi.org/10.1177/1094428106291059>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *J. Mem. Lang.*, 110, 104038. <https://doi.org/10.1016/j.jml.2019.104038>
- Steenen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspect. Psychol. Sci.*, 11(5), 702–712. <https://doi.org/10.1177/1745691616658637>
- Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuss, O., Higgins, J. P. T., Langan, D., & Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res. Synth. Methods*, 7(1), 55–79. <https://doi.org/10.1002/jrsm.1164>