

# The effects of educational video-games on students motivation to math: A meta-analysis in K-12

List authors

List affiliations

## Analysis Report

*Edited by*

Claudio Zandonella Callegher\*

\*e-mail address

claudiozandonella@gmail.com

January 19, 2020

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Report sections . . . . .	1
<b>2</b>	<b>Statistical Approach</b>	<b>2</b>
2.1	Measure of effect size . . . . .	2
2.1.1	The pre- post- control group design . . . . .	2
2.1.2	The effect size estimate . . . . .	2
2.2	Multilevel meta-analysis . . . . .	3
2.3	Analysis plan . . . . .	5
2.4	Analysis reproducibility . . . . .	6
2.4.1	R-project structure . . . . .	7
2.4.2	Run the Analysis . . . . .	7
<b>3</b>	<b>Data Preparation</b>	<b>8</b>
3.1	Data presentation . . . . .	8
3.2	Data munging . . . . .	10
3.3	Compute $d_{ppc2}$ value and variance . . . . .	11
<b>4</b>	<b>Descriptive Statistics</b>	<b>13</b>
<b>5</b>	<b>Session Information</b>	<b>13</b>
	<b>Bibliography</b>	<b>14</b>

**List of Figures**

**List of Tables**

# 1 Introduction

In this report the statistical analyses of the article "*The effects of educational video-games on students motivation to math: A meta-analysis in K-12*" are presented. The aim of the meta-analysis was to synthesized results of the studies concerning the impact of educational video-games on students' motivation towards mathematics.

Here, we will focus only on the statistical analysis. For theoretical aspects, study selection, and results interpretation the reader can refer directly to the article.

## 1.1 Report sections

The analysis report is divided into different sections:

- **Section 2:** the statistical approach and the plan of analysis are presented.
- **Section 3:** the dataset is presented with a brief description of each variable and prepared for the analysis.
- **Section 4:** the descriptive statistics are presented.

## 2 Statistical Approach

In this section, the statistical approach and the plan of the meta-analysis are presented. First, we describe the measure of effect size used to evaluate effectiveness of the interventions. Subsequently, we discuss the reasons why multilevel meta-analysis is used to account for the dependence of multiple effect sizes within the same studies. Finally, the plan of the meta-analysis is summarized.

### 2.1 Measure of effect size

#### 2.1.1 The pre- post- control group design

Selected studies were characterized by a pre- post- control group design (PPC). In a PPC design, participants are randomly assigned to one (or more) experimental group and to one (or more) control group. Both groups are evaluated before (pre-test score) and after (post-test score) the experimental group is exposed to a treatment.

The PPC design allows to evaluate the efficacy of the treatment taking into account pre-existing differences between the two groups and concurrent factors or events other than the treatment that produce changes in the outcome variable.

The efficacy of the treatment can be evaluated considering the standardized mean change in both groups. The standardized mean change of each group is defined as the mean difference between post-test ( $\mu_{post}$ ) and pre-test ( $\mu_{pre}$ ) scores, divided by the standard deviation ( $\sigma$ ):

$$\delta = \frac{\mu_{post} - \mu_{pre}}{\sigma}. \quad (1)$$

Thus, assuming common standard deviation ( $\sigma$ ), the efficacy of the treatment can be defined as the difference in standardized mean change between the experimental group and the control group:

$$\Delta = \delta_{Eg} - \delta_{Cg} = \frac{(\mu_{Eg,post} - \mu_{Eg,pre}) - (\mu_{Cg,post} - \mu_{Cg,pre})}{\sigma}, \quad (2)$$

where  $Eg$  stands for experimental group and  $Cg$  stands for control group.

#### 2.1.2 The effect size estimate

To estimate the effect size, there are three alternatives that differ in the way the common standard deviation ( $\sigma$ ) is estimated. Morris (2008) discuss and evaluate the three alternatives:

1.  $d_{ppc1}$ : the effect size is computed using separate estimates of pre-test standard deviation for the experimental group ( $Eg$ ) and the control group ( $Cg$ ). That is,

$$d_{ppc1} = c_{Eg} \frac{(M_{Eg,post} - M_{Eg,pre})}{SD_{Eg,pre}} - c_{Cg} \frac{(M_{Cg,post} - M_{Cg,pre})}{SD_{Cg,pre}}, \quad (3)$$

where  $c_{Eg}$  and  $c_{Cg}$  are bias adjustments for small samples,  $M_{pre}$  and  $M_{post}$  are the mean scores in the pre-test and post-test, and  $SD_{pre}$  is the standard deviation in the pre test.

2.  $d_{ppc2}$ : the effect size is computed by pooling the data from the experimental and control group in the pre-test to estimate the population standard deviation. That is,

$$d_{ppc2} = c_P \left[ \frac{(M_{Eg,post} - M_{Eg,pre}) - (M_{Cg,post} - M_{Cg,pre})}{SD_{pre}} \right], \quad (4)$$

where  $SD_{pre}^1$  is the pooled standard deviation in the pre-test scores, and  $c_P^2$  is a bias adjustment for small sample size.

3.  $d_{ppc3}$ : the effect size is computed by pooling the data from the experimental and control group in both the pre-test and post-test to estimate the population standard deviation. That is,

$$d_{ppc3} = c_{PP} \left[ \frac{(M_{Eg,post} - M_{Eg,pre}) - (M_{Cg,post} - M_{Cg,pre})}{SD_{pre+post}} \right], \quad (5)$$

where  $SD_{pre+post}^3$  is the pooled standard deviation in the pre-test and post-test scores, and  $c_{PP}^4$  is a bias adjustment for small sample size.

Morris (2008) suggested the  $d_{ppc2}$  as the favorite effect size estimate in terms of bias, precision, and robustness to heterogeneity of variance. Given the assumption of homogeneity of variance in the two populations,  $d_{ppc2}$  allows a better estimate of the population standard deviation by pooling the data from the experimental and control groups in the pre-test. On the contrary,  $d_{ppc1}$  uses separate estimates of the sample standard deviation for experimental and control groups and  $d_{ppc3}$  includes in the pooled standard deviation also results from the post-test of the two groups. These solutions are not optimal as, in the first case, we do not take advantage of the assumption of homogeneity of variance, and in the second case post-test variances tend to be larger than pre-test variances given possible interaction between treatment effect and individual differences.

The formula to compute the variance of  $d_{ppc2}$  is provided by Morris (2008, see p.373 eq.25):

$$\sigma^2(d_{ppc2}) = 2(c_p^2)(1 - \rho) \left( \frac{n_{Eg} + n_{Cg}}{n_{Eg}n_{Cg}} \right) \left( \frac{n_{Eg} + n_{Cg} - 2}{n_{Eg} + n_{Cg} - 4} \right) \left( 1 + \frac{d_{ppc2}^2}{2(1 - \rho) \left( \frac{n_{Eg} + n_{Cg}}{n_{Eg}n_{Cg}} \right)} \right) - d_{ppc2}^2, \quad (6)$$

where  $\rho$  is the correlation between pre- and post-test scores, and  $n_{Eg}$  and  $n_{Cg}$  are the two groups sample size.

Following recommendations of the author, we used the  $d_{ppc2}$  to estimate the effect size of the studies included in the meta-analysis.

## 2.2 Multilevel meta-analysis

Traditional meta-analysis approaches are based on the assumption that each effect sizes is independent, so each study should contribute with only one effect size. However, dependency among effect sizes is very

---


$$\begin{aligned} {}^1SD_{pre} &= \sqrt{\frac{(n_{Eg}-1)SD_{Eg,pre}^2 + (n_{Cg}-1)SD_{Cg,pre}^2}{n_{Eg} + n_{Cg} - 2}} \\ {}^2c_P &= 1 - \frac{3}{4(n_{Eg} + n_{Cg} - 2) - 1} \\ {}^3SD_{pre+post} &= \sqrt{\frac{(n_{Eg}-1)SD_{Eg,pre}^2 + (n_{Cg}-1)SD_{Cg,pre}^2 + (n_{Eg}-1)SD_{Eg,post}^2 + (n_{Cg}-1)SD_{Cg,post}^2}{2(n_{Eg} + n_{Cg} - 2)}} \\ {}^4c_{PP} &= 1 - \frac{3}{4(2n_{Eg} + 2n_{Cg} - 4) - 1} \end{aligned}$$

common. Studies could evaluate the same subjects on multiple outcomes providing measures that are clearly not independent, or, even when outcomes are measured on independent samples, multiple effects within the same study are not independent. In fact, they share other common aspect as the design of the experiment, instruments used, treatment characteristics, research group, geographical region, etc., that may influence the effects to be more similar to each other.

In order to take into account the dependency between effect sizes, different approaches are proposed (Moeyaert et al., 2017; Pigott & Polanin, 2019):

1. **Averaging effect sizes.** When similar measures of the same underlying construct are provided, it is reasonable to use the average of the effect sizes to summarize the study results. Borenstein, Hedges, Higgins, and Rothstein (2009, see Part 5: Complex data structure) describe how to compute a composite effect size and variance taking according to the dependence structures of the effect size within the study.
2. **Robust variance estimation (RVE).** Larry V. Hedges, Tipton, and Johnson (2010) recommend to use the RVE when evaluating dependent effects in a meta-analysis. The overall effect size over studies is computed as a weighted mean of the observed effect size and the sampling variance estimate is obtained by means of an approximation of the covariance-matrix of each study. Larry V. Hedges et al. (2010) demonstrate that RVE meta-analysis with non-independent effect sizes produces valid results even in the case of misspecified covariance structure. As the number of the studies increase the results will converge to the correct values. Moreover, Tipton (2015) provided adjusted estimators to the RVE that increase the reliability of the results even when the number of studies is small.
3. **Multilevel meta-analysis.** All random-effects meta-analysis are multilevel as they assume a two-level structure where the observed effects are an estimation of each study "true" effect that is sampled from an overarching distribution of "true" effects. However, in this case we refer to three-level meta-analysis model, where participants are included at level 1, measured outcomes at level 2, and study at level 3. In this three level approach, the dependency between effect size within studies is automatically accounted for in the covariance matrix (Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2015).

Moeyaert et al. (2017) compared the three approaches and reported that both the RVA and the multilevel approach give unbiased parameter estimates, standard errors and variances. Whereas, averaging effect sizes is too conservative because in general standard errors are overestimated. Moreover, the RVA and the multilevel approach allow to include in the meta-analytic model predictors at the outcome level to account for differences between outcomes within the same studies.

In the present meta-analysis, we decided to adopt a multilevel approach as it allows us to specify the kind of dependency between effect size specifically for each study. Multilevel meta-analysis offers a great flexibility in accommodating different type of dependency among effect sizes within studies.

On the contrary, RVA approach allows only to decide between a correlated effects dependence structure or a hierarchical dependence structure for all the studies (Larry V. Hedges et al., 2010; Tipton, 2015). In the first case, within each study outcomes are assumed to be measured on the same underlying units. In the second case, multiple (independent) outcomes are measured within studies, but they are related by other aspects such as research organizations, research labs, or research groups.

## 2.3 Analysis plan

The analysis are conducted with R software (version 3, ., 6.1, ; R Core Team, 2018). First, we load and prepare the dataset for the analysis, computing the effect size estimate and variance for each outcome. As presented in Formula 6, values of the correlation between pre-test and post-test scores are needed to the compute  $d_{ppc2}$  variance. However, none of the studies reported this information. To overcome this issue,  $d_{ppc2}$  variance is computed with different correlation values and their influence on the meta-analysis results will be subsequently evaluated.

Next, descriptive statistics are presented to evaluate the main characteristics of the data included in the analysis.

Multilevel meta analysis is conduct using the restricted maximum likelihood method with the R-package *metafor* (Viechtbauer, 2010). Variance-covariance matrix is defined for each study according to the specific outcomes dependence structure. However, studies with multiple measure on the same subjects did not reported the correlation between outcomes needed to specify dependence structure. In these cases, variance-covariance matrix is computed with different correlation values and their influence on the meta-analysis results will be subsequently evaluated. Results are presented and heterogeneity between studies is assesed through inspection of forest plot and evaluation of the Q-statistic (Larry V Hedges & Olkin, 2014). Under the null hypothesis, the Q-statistic is distributed as a chi-square with degrees of freedom equal to the number of studies minus one. A significant chi-value indicates the presence of heterogeneity across studies. Moreover, to estimate the magnitude of the heterogeneity the  $I^2$  index is reported (i.e., the proportion of observed variance that reflects real and not random difference between studies effect sizes; Borenstein et al., 2009). High values of  $I^2$  suggest that differences between results are related to real differences across studies (i.e., different constructs or different study design). On the contrary, low values of  $I^2$  suggest that results across studies are similar and possible differences are related to random sampling.

To investigate robustness of the results, a sensitivity analysis is conducted to evaluate how the values of the correlation between pre-test and post-test scores and the values of the correlation between dependent outcomes influence the meta analysis results. Moreover, a *Leave-One-Out* (LOO) sensitivity analysis considering the study grouping level is run to assess the possible presence of influential studies. Substantial changes when a single study is removed would be interpreted as lack of homogeneity and unreliable results (Viechtbauer & Cheung, 2010).



Next, we present the alternative results obtained using the different meta-analytic approaches presented above, which are robust variance estimation (RVE) approach and averaging effect sizes approach. RVE is conducted using the R-package `robumeta` (Fisher & Tipton, 2015; Fisher, Tipton, & Zhipeng, 2017) considering a correlated effects dependence structure. Averaged effect sizes are obtained according to Borenstein et al. (2009) indications as implemented in the R-package `MAd` (Hoyt, 2014).

Subsequently, publication bias is assessed. To our knowledge, there are no specific methods designed to correctly evaluate publication bias in a multilevel meta-analysis. Funnel plot is presented as a graphical representation of the symmetry of the effects distribution. However, this is not optimal as studies with multiple effects are overrepresented in the plot respect to studies with only one effect. To overcome this issue, we conduct separately two Egger's regression tests to evaluate if effect sizes are associated to sample size or sampling variance of the estimated effect. In the absence of publication bias, the regression coefficient is expected to be zero (Egger, Smith, Schneider, & Minder, 1997; Lin & Chu, 2018). Moreover, the result of the rank correlation test is reported to examine whether the observed outcomes and the corresponding sampling variances are correlated (Begg & Mazumdar, 1994).

Alternatively, funnel plot with the *trim-and-fill* method (Duval & Tweedie, 2000; Rothstein, Sutton, & Borenstein, 2005) is presented considering the aggregated effects.

Finally, the role of possible moderators is examined using mixed-effects meta-regression models, the moderators are included as a fixed effects and are tested using Wald's chi-square (Viechtbauer, 2010). Considering the reduced number of studies and the unequal distribution among the different levels of the moderators, separate analyses are conducted for each moderator. Results have to be interpreted with caution as moderators are not evaluated together to understand the unique variance explained by each moderator.

## 2.4 Analysis reproducibility

To guarantee the reproducibility of the results, the whole analysis is structured within an *R-project* named `DMGC_Meta.Rproj` that is possible to download from this repository ([Todo: add repository link](#)).

The R-package `drake` (Landau, 2018) is used to manage the analysis workflow and to enhance the readability and transparency of the analysis. To know more about `drake` consider the official Git-hub page (<https://github.com/ropensci/drake>) or the user manual (<https://books.ropensci.org/drake/>). Summarizing, using `drake` the code of the analysis is organized into different scripts. The user defines the plan of the analysis where each step in the analysis is defined through functions. Functions can be appropriately defined to obtain desired targets (i.e., R-output with results of interests) and they are declared in another script. Subsequently, `drake` manages the whole analysis recognizing the dependency structure of the different targets. When any change is made to the code `drake` evaluates the analysis and updates the results. Using functions to define each step of the analysis allows to avoid "*copying and paste*" in the code, it makes debugging easier, and it facilitates the reading of the code.

Moreover, the R-package **renv** (Ushey, 2019) is used to manage the dependencies of the R-packages used in the analysis. The **renv** package allows to create an isolated, portable, and reproducible environment where the analyses are run. To know more about **renv** consider the official documentation (<https://rstudio.github.io/renv/articles/renv.html>).

Finally, git version control was used to track the changes during the analysis.

### 2.4.1 R-project structure

The R-project **DMGC\_Meta.Rproj** is organized into different folders. In the folder **Data/**, the raw datasets with the information regarding the studies selected in the literature review are stored.

In the folder **R/**, the R-scripts used in the analysis are stored. Using the **drake** package the analysis is organized into different R-scripts files:

- **Settings.R** contains the setting for the R sessions, including R-packages used.
- **Plan.R** contains the plan of the analysis. Where each target (i.e., R-output with results of interests) is defined through functions.
- **Function.R** contains the main functions used in **Plan.R** to obtain the targets of interest.
- **Auxiliary\_functions.R** contains other functions used in the analysis.
- **Analysis.R** is the script used to run the whole analysis.

### 2.4.2 Run the Analysis

In order to run the analysis follow these steps:

1. Make sure you have already the **renv** R-package installed in your lybrary. If not, run the command in R or R-studio `install.packages("renv")`
2. Open the R-project **DMGC\_Meta** by double-clicking the file **DMGC\_Meta.Rproj** you can find in the main directory. A new R-studio session should open and a similar message should appear in the console if **renv** was correctly installed:  

```
* Project '~/<your_path>/DMGC_Meta' loaded. [renv <version_number>]
```
3. Run the line `renv::restore()`, **renv** will ask the permission to install the R-packages used in the analysis, type y and return to confirm.
4. Open the file '**R/Analysis.R**' and run each line of the sections "Load", "Check", and "Make".
5. Now you can access the targets with the results using the functions `drake::load(<name_target>)` and `drake::read(<name_target>)`.

## 3 Data Preparation

In this section, first the raw dataset is presented, then the dataset is prepared for the analysis.

### 3.1 Data presentation

In the folder `Data/`, you can find the file `Dataset.csv` with all the informations about the selected studies. The dataset includes 43 effects sizes grouped within 19 different studies. Each line of the dataset is an effect and the characteristics of each effect are summarized in 25 variables:

1. **study** - Numerical variable that indicates the id number of the study. This id number is used in the analysis to refer to a specific study. Values range from 1 to 19.
2. **id** - Numerical variable that indicates the id number of the study used during the selection process of eligible studies.
3. **author** - Character string with the names of the authors.
4. **year** - Numerical variable that indicates the year of publication.<sup>5</sup>
5. **pub** - Character string that indicates if the study is published (**yes**) or unpublished/ grey-litterature (**no**).
6. **grade** - Numerical variable that indicates school grade of the participants. Primary school is indicated with 1 and secondary school is indicated with 2.
7. **weeks** - Numerical variable that indicates the length of the treatment in weeks.
8. **sessions** - Numerical variable that indicates the total number of sessions of the treatment.
9. **minutes** - Numerical variable that indicates average time in minutes of each session of the treatment.
10. **device** - Character string that indicates the device used during the treatment. Personal Computer (**pc**), console (**con**), or application on other device (**app**).
11. **n\_effect** - Numerical variable that identifies within each study the different effects. In each study the first effect is identified with 1, then the value increases if more effects are reported within the same study.
12. **id\_effect** - Numerical variable that identifies the unique effect. Values range from 1 to 43.
13. **dependence** - Character string that indicates the type of dependence among different effect sizes reported within the same study. Study with single effect size (**none**), multiple effects measured on the same participants (**multiple\_outcomes**), multiple comparison between different experimental groups and same control group (**multiple\_groups**), or independent effect sizes (**independent**).

---

<sup>5</sup>In the case of unpublished studies or grey-litterature the year refers to the realization of the study

14. **N** - Numerical variable that indicates the total number of participants included to compute the considered effect. This is not always equal to the total number of participants in the study, as multiple experimental or control groups could be included in the study.
15. **n\_cg** - Numerical variable that indicates the number of participants in the control group for the considered effect.
16. **n\_eg** - Numerical variable that indicates the number of participants in the experimental group for the considered effect.
17. **mot** - Numerical variable that indicates if the considered effect measured the motivation in terms of expectancy (1), in terms of value (2), or did not differentiate between the two aspects (NA).
18. **m\_t1\_cg** - Numerical variable that indicates the control group mean score in the pre-test.
19. **sd\_t1\_cg** - Numerical variable that indicates the control group standard deviation in the pre-test.
20. **m\_t2\_cg** - Numerical variable that indicates the control group mean score in the post-test.
21. **sd\_t2\_cg** - Numerical variable that indicates the control group standard deviation in the post-test.
22. **m\_t1\_eg** - Numerical variable that indicates the experimental group mean score in the pre-test.
23. **sd\_t1\_eg** - Numerical variable that indicates the experimental group standard deviation in the pre-test.
24. **m\_t2\_eg** - Numerical variable that indicates the experimental group mean score in the post-test.
25. **sd\_t2\_eg** - Numerical variable that indicates the experimental group standard deviation in the post-test.

The dataset is loaded in R and its structure is presented below.

```
data_raw <- read.csv("Data/Dataset.csv", sep = ";", header = T, stringsAsFactors = F)
```

```
str(data_raw)

## 'data.frame': 43 obs. of 25 variables:
## $ study      : int  1 2 2 2 3 3 3 4 5 6 ...
## $ id         : int  54 961 961 961 432 432 432 436 982 983 ...
## $ author     : chr   "Bai et al." "Ke" "Ke" "Ke" ...
## $ year       : int  2012 2006 2006 2006 2008 2008 2008 2010 2018 2015 ...
## $ pub        : chr   "yes" "no" "no" "no" ...
## $ grade      : int  2 1 1 1 1 1 1 2 1 1 ...
## $ weeks      : int  18 4 4 4 4 4 4 18 7 10 ...
## $ sessions   : int  NA 8 8 8 8 8 8 18 7 NA ...
## $ minutes    : int  NA 45 45 45 40 40 40 30 45 30 ...
## $ device     : chr   "pc" "pc" "pc" "pc" ...
## $ n_effect   : int  1 1 2 3 1 2 3 1 1 1 ...
## $ id_effect  : int  1 2 3 4 5 6 7 8 9 10 ...
## $ dependence: chr   "none" "independent" "independent" "independent" ...
```

```
## $ N      : int  437 115 114 129 76 77 79 193 80 1168 ...
## $ n_cg    : int  192 58 63 60 36 36 36 76 40 526 ...
## $ n_eg    : int  245 57 51 69 40 41 43 117 40 642 ...
## $ mot     : int  NA NA NA NA NA NA NA NA 1 2 ...
## $ m_t1_cg : num  75.6 156.2 149.4 145 77.5 ...
## $ sd_t1_cg : num  11.97 23.84 30.12 26.91 9.38 ...
## $ m_t2_cg : num  72.4 156.1 145.7 144.9 74.2 ...
## $ sd_t2_cg : num  11 26.5 28.9 28 12 ...
## $ m_t1_eg : num  71.2 150.4 142.1 152 77.7 ...
## $ sd_t1_eg : num  12.3 30.5 33.1 24.9 13.2 ...
## $ m_t2_eg : num  71.8 151.4 145.8 162 77.3 ...
## $ sd_t2_eg : num  13.5 29.7 30.2 23.9 15.6 ...
```

### 3.2 Data munging

Following **drake** guidelines, we define function for each step of the analysis. To organize the dataset we define the following function that allows to state which variables are categorical variables (factors) and uses more explicit labels in some variables. Moreover, we create a variable **autor\_y** to obtain a label in the format *"Author's (year)"* for each study and the variable **intensity** = **sessions\*minutes/weeks** as a rough measure of the weekly intensity of each treatment.

```
clean_data <- function(data){
  data%>%
    # Define which variable are factor
    mutate_at(vars("study","id","pub","dependence","grade","device","mot"), factor)%>%
    # Redefine factor labels for grade and device
    mutate(grade=recode_factor(grade,"1"="Primary","2"="Secondary"),
           device=recode_factor(device,"app"="App","con"="Console","pc"="PC"),
           # Obtain authors (year) label
           autor_y=paste0(author," (",year,")"),
           # Compute intensity of the intervention
           intensity = sessions*minutes/weeks)%>%
    select("study":"minutes","intensity",everything())
}
```

In Pareto, Arvemo, Dahl, Haake, and Gulz (2011) the gain scores (i.e., the mean difference between post-test and pre-test scores) are reported instead of the post-test scores. Thus, to obtain the post-test scores we define a function that computes the post-test scores. The mean is given by the sum between the mean pre-test score and the mean gain score ( $Mean_{post} = Mean_{pre} + Mean_{gain}$ ). The post-test Standard deviation is given by the square-root of the sum between the variance of the pre-test scores and the variance of the gain scores ( $SD_{post} = \sqrt{SD_{pre}^2 + SD_{gain}^2}$ ).

```
pareto_post <- function(data){
  # Control group gain
  m_gain_cg<-c(1.5,-.23) # Mean gain
  sd_gain_cg<-c(3.82,3.62) # Sd gain
```

```

# Experimental group gain
m_gain_eg<-c(.93,1.44) # Mean gain
sd_gain_eg<-c(3.09,3.52) # Sd gain

data %>%
  mutate_cond(study=="11", # Change only rows satisfying the condition
    # Control group
    m_t2_cg = m_t1_cg + m_gain_cg,
    sd_t2_cg = sqrt(sd_t1_cg^2+sd_gain_cg^2),
    #Experimental group
    m_t2_eg = m_t1_eg + m_gain_eg,
    sd_t2_eg = sqrt(sd_t1_eg^2+sd_gain_eg^2))
}

```

### 3.3 Compute $d_{ppc2}$ value and variance

To compute the  $d_{ppc2}$  value and variance, we use the functions reported in the metafor website (<http://www.metafor-project.org/doku.php/analyses:morris2008>). To compute the variance we need the correlation between pre-test and post-test scores. However, correlation are not reported so we try with different values high correlation ( $r_{high} = .8$ ), medium-high correlation ( $r_{mediumh} = .6$ ), medium-low correlation ( $r_{mediuml} = .4$ ), and low correlation ( $r_{low} = .2$ ).

```

compute_dppc2 <- function(data){

  data%>%
    mutate(
      # compute pooled standard deviation
      sd_pool= sqrt(((n_eg-1)*sd_t1_eg^2+(n_cg-1)*sd_t1_cg^2)/(n_eg+n_cg-2)),
      # Compute yi_dppc2
      yi_dppc2 = metafor:::cmicalc(n_eg+n_cg-2)*((m_t2_eg - m_t1_eg) - (m_t2_cg-m_t1_cg))
        /sd_pool,

      # add correlations pre-post
      r_high=.8, # high correlation
      r_mediumh=.6, # medium-high correlation
      r_mediuml=.4, # medium-low correlation
      r_low=.2)%>% # low correlation

      # Rearrange dataset from wide format to long format
      gather("r_high":"r_low", key="r_size",value="r_value", factor_key = TRUE)%>%
      # compute variance
      mutate(vi_dppc2=2*(1-r_value)*(1/n_eg + 1/n_cg)+ yi_dppc2^2/(2*(n_eg + n_cg)))%>%
      arrange(study,n_effect,r_size) # ordering dataset
}

```

Seamply looking at the effect obtained, we observe a strange value. In Hung, Huang, and Hwang (2014) values of  $yi\_dppc2$  for the second effect are extreainly huge 38.2. This value is impossible.

```
##   study  id      author year n_effect id_effect      dependence  N n_cg n_eg m_t1_cg
## 1    13 959 Hung et al. 2014      1      32 multiple_outcomes 46  23  23    3.3
## 2    13 959 Hung et al. 2014      2      33 multiple_outcomes 46  23  23    3.0
##   sd_t1_cg m_t2_cg sd_t2_cg m_t1_eg sd_t1_eg m_t2_eg sd_t2_eg   sd_pool  yi_dppc2
## 1    0.76   3.24   0.82   3.53   0.62   3.74   0.79 0.69354164 0.3826259
## 2    0.02   3.23   0.72   3.00   0.00   3.78   0.72 0.01414214 38.2235343
##      r_size r_value   vi_dppc2
## 1 r_mediumh    0.6 0.07115655
## 2 r_mediumh    0.6 15.95041926
```

Considering the measure in the pre-test we can understand the problem. The standard deviation in the pre-test are extremely low: 0.02 and 0. This leads to a low value for the pooled standard deviation and in turn an implausible high value of effect size. That is probably given by the fact that the measure used was not able to evaluate properly the variation among individuals in the pre-test, (maybe floor effect).

We do not include this effect in the meta-analysis and we remove it from the dataset. Thus the final dataset includes 19

```
rm_hung <- function(data) {
  data %>% filter(study != "13" | n_effect != 2)
}
```

## 4 Descriptive Statistics

## 5 Session Information

```
sessionInfo(package = NULL)

## R version 3.6.1 (2019-07-05)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Mojave 10.14.6
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] tidysselect_0.2.5 drake_7.8.0      ggplot2_3.2.1    kableExtra_1.1.0 knitr_1.26
##
## loaded via a namespace (and not attached):
## [1] Rcpp_1.0.3      highr_0.8        formatR_1.7      pillar_1.4.3
## [5] compiler_3.6.1 tools_3.6.1      zeallot_0.1.0    digest_0.6.23
## [9] gtable_0.3.0    evaluate_0.14    tibble_2.1.3     lifecycle_0.1.0
## [13] viridisLite_0.3.0 pkgconfig_2.0.3  rlang_0.4.2      igraph_1.2.4.2
## [17] filehash_2.4-2  rstudioapi_0.10  filelock_1.0.2   xfun_0.11
## [21] storrr_1.2.1    withr_2.1.2      dplyr_0.8.3      stringr_1.4.0
## [25] httr_1.4.1      xml2_1.2.2       vctrs_0.2.1      hms_0.5.2
## [29] grid_3.6.1      webshot_0.5.1    glue_1.3.1       R6_2.4.1
## [33] base64url_1.4   rmarkdown_1.16   txtq_0.2.0       purrr_0.3.3
## [37] readr_1.3.1     magrittr_1.5     codetools_0.2-16 backports_1.1.5
## [41] scales_1.1.0    htmltools_0.4.0  assertthat_0.2.1 rvest_0.3.4
## [45] tikzDevice_0.12.3 colorspace_1.4-1 stringi_1.4.3     lazyeval_0.2.2
## [49] munsell_0.5.0   crayon_1.3.4
```



## References

- Begg, C. B., & Mazumdar, M. (1994). Operating Characteristics of a Rank Correlation Test for Publication Bias. *Biometrics*, 50(4), 1088. doi:10.2307/2533446
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (Eds.). (2009). *Introduction to meta-analysis*. OCLC: ocn263294996. Chichester, U.K: John Wiley & Sons.
- Duval, S., & Tweedie, R. (2000). Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. *Biometrics*, 56(2), 455–463. doi:10.1111/j.0006-341X.2000.00455.x
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109), 629–634. doi:10.1136/bmj.315.7109.629
- Fisher, Z., & Tipton, E. (2015). Robumeta: An R-package for robust variance estimation in meta-analysis. *arXiv:1503.02220 [stat]*. arXiv: 1503.02220 [stat]
- Fisher, Z., Tipton, E., & Zhipeng, H. (2017). *Robumeta: Robust variance meta-regression*.
- Hedges, L. V. [Larry V.], Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1), 39–65. doi:10.1002/jrsm.5
- Hedges, L. V. [Larry V.], & Olkin, I. (2014). *Statistical methods for meta-analysis*. Academic press.
- Hoyt, A. C. D. R. W. T. (2014). *MAd: Meta-analysis with mean differences*.
- Hung, C.-M., Huang, I., & Hwang, G.-J. (2014). Effects of digital game-based learning on students' self-efficacy, motivation, anxiety, and achievements in learning mathematics. *Journal of Computers in Education*, 1(2-3), 151–166. doi:10.1007/s40692-014-0008-8
- Landau, W. M. (2018). The drake R package: A pipeline toolkit for reproducibility and high-performance computing. *Journal of Open Source Software*, 3(21).
- Lin, L., & Chu, H. (2018). Quantifying publication bias in meta-analysis: Quantifying Publication Bias. *Biometrics*, 74(3), 785–794. doi:10.1111/biom.12817
- Moeyaert, M., Ugille, M., Natasha Beretvas, S., Ferron, J., Bunuan, R., & Van den Noortgate, W. (2017). Methods for dealing with multiple outcomes in meta-analysis : a comparison between averaging effect sizes, robust variance estimation and multilevel meta-analysis. *International Journal of Social Research Methodology*, 20(6), 559–572. doi:10.1080/13645579.2016.1252189
- Morris, S. B. (2008). Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. *Organizational Research Methods*, 11(2), 364–386. doi:10.1177/1094428106291059
- Pareto, L., Arvemo, T., Dahl, Y., Haake, M., & Gulz, A. (2011). A Teachable-Agent Arithmetic Game's Effects on Mathematics Understanding, Attitude and Self-efficacy. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), *Artificial Intelligence in Education* (Vol. 6738, pp. 247–255). doi:10.1007/978-3-642-21869-9\_33
- Pigott, T. D., & Polanin, J. R. (2019). Methodological Guidance Papers: High-Quality Meta-Analysis in a Systematic Review. *Review of Educational Research*, 23. doi:10.3102/0034654319877153
- R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org/>. Vienna, Austria.
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Wiley.
- Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, 20(3), 375–393. doi:10.1037/met0000011
- Ushey, K. (2019). *Renv: Project environments*.
- Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2015). Meta-analysis of multiple outcomes: A multilevel approach. *Behavior Research Methods*, 47(4), 1274–1294. doi:10.3758/s13428-014-0527-2
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48.
- Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2), 112–125. doi:10.1002/jrsm.11