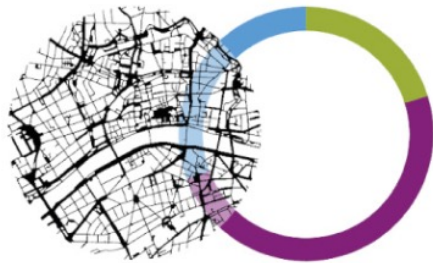# The Reproducibility in Reliability Generalization Meta-Analysis*

Research Synthesis & Big Data

18 - 21/05/2021
~~Frankfurt am Main~~
Virtual Conference

Pre-Conf MASEM Workshop
17/05

Carmen López-Ibáñez[1] & Julio Sánchez-Meca[1]

[1]Meta-analysis Unit, Department of Basic Psychology and Methodology, University of Murcia (Spain)

Reproducibility?

One of the principles of scientific method is the ability of an experiment to be reproducible.

Reproducibility can be defined as the ability to re-analyse the data following the procedures indicated in the original study and obtaining the same result (Maassen, 2020).

Artner et al. (2020) include in their definition that it should be verified by an independent researcher.

Reproducibility vs Replicability

It is important to differentiate between reproducibility and replicability.

It seems to exist some consensus about difference in the definitions.

Reproducibility involves using the same dataset as in the original study

Replicability refers to restarting the whole process, starting with coding the primary data.

Reproducibility in MA

Meta-Analysis is the main tool to synthesise quantitative evidence from multiple primary studies.

This statistical technique has different procedures that should vary depends on the primary data.

Because of that, it is important that the results of a MA can be reproduced by other researcher

Maassen et al. (2020) found that in 39% of the MAs, there were small discrepancies between the original MA results and the reproduced results. However, in most of them the differences were insignificant.
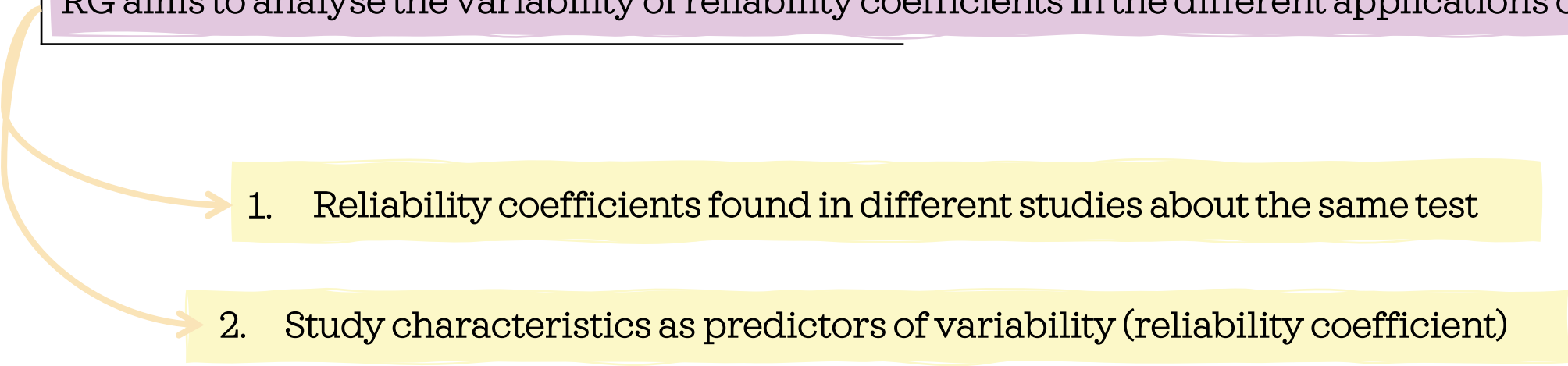
This is why we started to ask ourselves what was happening in a specific type of MA: the Reliability Generalization Meta-Analysis

Reliability Generalization MA

A specific type of MA. Main objective of the RG studies is to obtain an average reliability coefficient.

RG aims to analyse the variability of reliability coefficients in the different applications of a test.

1. Reliability coefficients found in different studies about the same test

2. Study characteristics as predictors of variability (reliability coefficient)

Current Study

We tried to examine how the reproducibility drama affects to this specific type of MA.

In order to do it, we wanted to prove whether the results of RG meta-analysis can be reproduced following the described procedure in each meta-analysis.

We collected all the RG MA that had provided the whole database with individual data and reanalysed the results.

Selection Criteria

1. Studies had to be written in English or Spanish.

2. Studies had to present an RG meta-analysis.

3. Studies had to focus on the measurement of a psychological construct.

4. Studies should report alpha coefficients.

5. Studies must provide dataset with individual alpha coefficients.

Search Strategies

Scopus, EBSCOHOST, & Google Scholar databases

MEDLINE
APA PscycInfo
Education Sourse
APA PsycArticles
Gender Studies Database
PSICODOC

Keywords

"Reliability Generalization"

"Meta-Analysis of Internal Consistency"

"Meta-Analysis of Alpha Coefficients"

Temporal Range: 1998-December 2020

Language Restrictions: studies should be written in English or Spanish.

Data Extraction

We have relied on the coding of the authors of each meta-analysis.

We have taken the data from the MAs, not from the primary studies.

Alpha coefficient

Lower Limit Confidence Interval

Upper Limit Confidence Interval

Index I2

Index heterogeneity Q

Statistical Model

Transformation for coefficients

Software

Data Extraction

We have relied on the coding of the authors of each meta-analysis.

We have taken the data from the MAs, not from the primary studies.

Alpha coefficient

Lower Limit Confidence Interval

Upper Limit Confidence Interval

Index I2

Index heterogeneity Q

Statistical Model

Transformation for coefficients

Software

From dataset we took sample size, reliability coefficients and number of items

Discrepancy Index

We calculated a discrepancy index (Artner et al., 2020)
to determine whether the results were reproduced.

$$Discrepancy\ Index\ (DI) = \left| \frac{T_{reproduced} - T_{reported}}{T_{reported}} \right|$$

We also transformed it in percentage

$$Percentage\ change = DI\ x100$$

We used a cutoff of 10% for discrepancy

Results

We obtained 170 reliability coefficients from 31 articles.

Not available data	
Alpha	5.29%
LL Confidence Interval	39.53%
UL Confidence Interval	39.53%
I2	56.4%
Q	55.23%
Statistical Model	4.07%
Transformation	7.56%
Software	32.56%

Results

We obtained 170 reliability coefficients from 31 articles.

Not available data	
Alpha	5.29%
LL Confidence Interval	39.53%
UL Confidence Interval	39.53%
I2	56.4%
Q	55.23%
Statistical Model	4.07%
Transformation	7.56%
Software	32.56%

Data not available to reproduce the analysis, but results reported in the article

Results

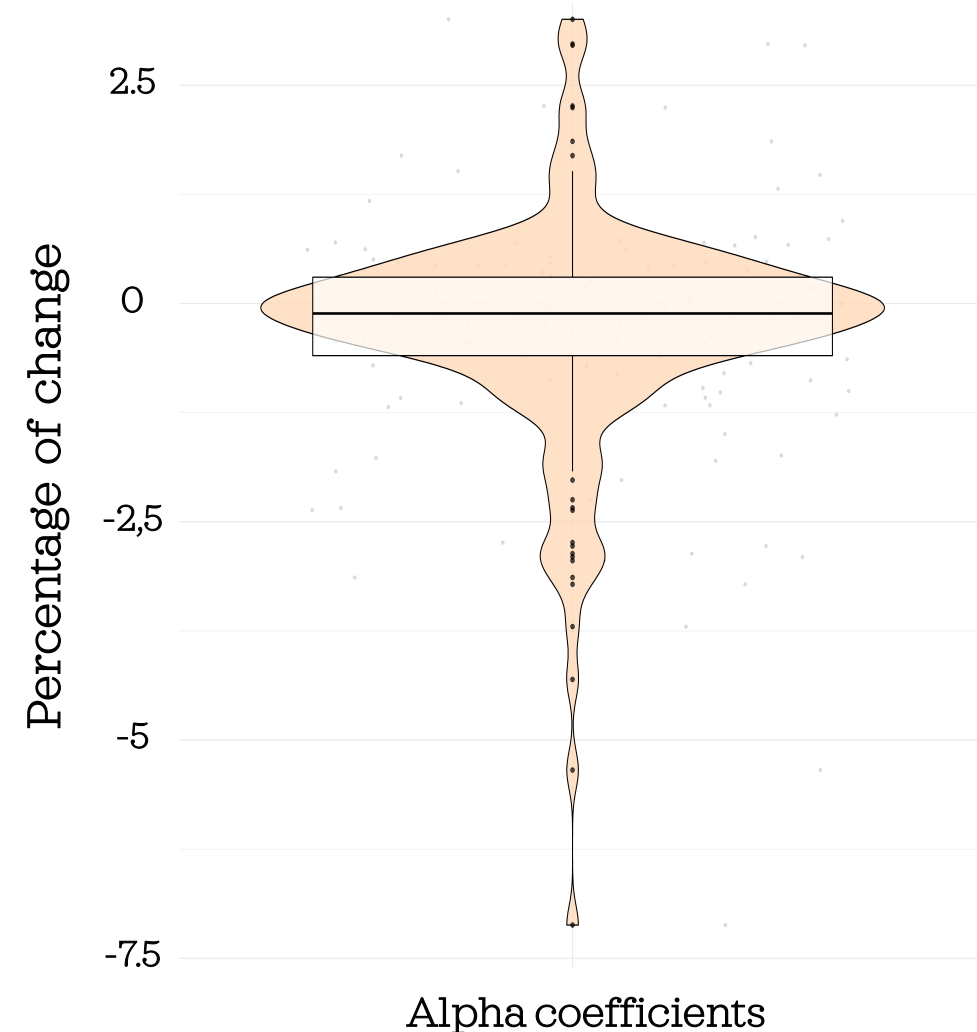We obtained 170 reliability coefficients from 31 articles.

Not available data	
Alpha	5.29%
LL Confidence Interval	39.53%
UL Confidence Interval	39.53%
I2	56.4%
Q	55.23%
Statistical Model	4.07%
Transformation	7.56%
Software	32.56%

More than 50% not reported

Discrepancy Index

Statistics	
N	161
Below 10%	100%
Below 5%	98,76%
Mean	0,83
SD	1,32
Min	0
Max	7,12
Range	10,38

Alpha coefficients

Discrepancy Index

Statistics

	LL		UL
N	93	N	93
Below 10%	98.92%	Below 10%	100%
Below 5%	94,62%	Below 5%	95.7%
Mean	1,48	Mean	1,02
SD	4,29	SD	1,46
Min	0	Min	0
Max	39,86	Max	6,42
Range	46,26	Range	12,4

Discrepancy Index

Differences

LL

Reported	Calculated
0.8	0.48
N=4	N=2

Dataset doesn't contain all data used in original analysis

Discrepancy Index

Statistics

	I^2
N	64
Below 10%	95.31%
Below 5%	90,63%
Mean	1,83
SD	5,26
Min	0
Max	29,64
Range	38,02

I^2 Index of Heterogeneity

Discrepancy Index

Heterogeneity

Differences	
I²	
Reported	Calculated
57,5	67,43
76	91,84
72	93,34

Original MA doesn't explicit heterogeneity estimator

I² Index of Heterogeneity

Discussion and Conclusion

Importantly, we only worked with 31 RG MA that provided the entire dataset.

From these studies, we reanalysed 161 reliability coefficients, 93 confidence intervals and 64 heterogeneity indices.

Discussion and Conclusion

We have reproduced the results in over 90% of cases in each condition.

The largest rate (100%) of reproducibility was obtained by calculating reliability coefficients.

The lowest rate (95,31%) of reproducibility was obtained by calculating the index I^2.

The largest range (46,26) was obtained by calculating confidence interval, in particular the lower limit.

Despite the positive data, we lost a lot of information because authors didn't report it.
Many articles didn't explicit the heterogeneity estimator applied in Random-Effects model.

Discussion and Conclusion

We have reproduced the results in over 90% of cases in each condition.

The largest rate (100%) of reproducibility was obtained by calculating reliability coefficients.

The lowest rate (95,31%) of reproducibility was obtained by calculating the index I^2.

The largest range (46,26) was obtained by calculating confidence interval, in particular the lower limit.

We would like to recall that more than 50% of the heterogeneity indices were not reported.

| I^2 | 56.4% |
| Q | 55.23% |

Identified Errors

1. Reported average alpha coefficient obtained thus Fixed-Effect model while reported confidence intervals belonged to Random Effects model.

2. Data are different from dataset and article.

3. Number of studies in the article is different from number of studies in dataset.

4. Dataset reported only one study for the calculation of MA

5. Confusing the procedure carried out

6. Incorrect number of items

7. Do not specify the number of items

8. Databases without some column names

9. Not all data in database

10. Different data in different parts of the article.

Future is now

We are currently working to improve this study

We are coding more data from original MAs to obtain more information on transparency in RG MA.

Nevertheless, we think that reproducibility data in this area are promising

Besides that , a guideline for the correct reporting of this type of study (REGEMA checklist) has recently been published (Sánchez-Meca et al., 2021).

Thank you for your attention

Carmen López-Ibáñez[1] & Julio Sánchez-Meca[1]

[1]Meta-analysis Unit, Department of Basic Psychology and Methodology, University of Murcia (Spain)

carmen.li@um.es