Replicability crisis ... is not the only one: Other current challenges/opportunities to enhance science

Gianmarco Altoè

Department of Developmental and Social Psychology
University of Padova

19 September 2023, Padova

Workshop
Replicability crisis in science?

Introduction: What the talk is about

You may say there's a replicability crisis ...

Introduction: What the talk is about

You may say there's a replicability crisis ...

but it's not the only one!

Premise (I): Exploratory and Confirmatory Research

EXPLORATORY RESEARCH

- Explore your data
- Use results to generate hypotheses

E.g. *Exploratory Factor Analysis*

CONFIRMATORY RESEARCH

- Pre-plan and formalize hypotheses
- Test hypotheses

E.g. *Confirmatory Factor Analysis*

Premise (I): Exploratory and Confirmatory Research

EXPLORATORY RESEARCH

- Explore your data
- Use results to generate hypotheses

E.g. *Exploratory Factor Analysis*

CONFIRMATORY RESEARCH

- Pre-plan and formalize hypotheses
- Test hypotheses

E.g. *Confirmatory Factor Analysis*

Exploratory and Confirmatory research are complementary

Premise (I): Exploratory and Confirmatory Research

EXPLORATORY RESEARCH

- Explore your data
- Use results to generate hypotheses

E.g. *Exploratory Factor Analysis*

CONFIRMATORY RESEARCH

- Pre-plan and formalize hypotheses
- Test hypotheses

E.g. *Confirmatory Factor Analysis*

Usually: Exploratory Research $\Longrightarrow$ Confirmatory Research

Premise (I): Exploratory and Confirmatory Research

EXPLORATORY RESEARCH

- Explore your data
- Use results to generate hypotheses

E.g. *Exploratory Factor Analysis*

CONFIRMATORY RESEARCH

- Pre-plan and formalize hypotheses
- Test hypotheses

E.g. *Confirmatory Factor Analysis*

- **Both kinds of research are important**
- **As J.W. Tukey stated:**
  "Finding the question is often more important than finding the answer"

Premise (I): Exploratory and Confirmatory Research

EXPLORATORY RESEARCH

- Explore your data
- Use results to generate hypotheses

E.g. *Exploratory Factor Analysis*

CONFIRMATORY RESEARCH

- Pre-plan and formalize hypotheses
- Test hypotheses

E.g. *Confirmatory Factor Analysis*

With regard to replication crisis

- **Bad Science:**
  Exploratory Findings, Confirmatory Conclusions
- **Good Science:**
  Confirmatory Conclusions,
  Require Pre-planned Hypotheses

(Wagenmakers et al., 2012)

Premise (I): Exploratory and Confirmatory Research

CONFIRMATORY RESEARCH

- Pre-plan and formalize hypotheses
- Test hypotheses

IN THIS TALK,
I WILL FOCUS ON CONFIRMATORY RESEARCH

Premise (II): Psychology as a Motivating Case

- Replicability crisis started examining psychological papers
  (Ioannidis, 2005. Why most published research findings are false.)

- Psychological Science was among the first to introduce potential remedies (e.g., Open Science, Replication Studies and Pre-registration)

- In the past 5 years, at least two more crises have emerged in psychological research.
  These crises are pervasive across all scientific disciplines

In theory, how to conduct confirmatory research is clear

1. Based on theory, formulate and formalize your hypotheses
2. Identify the relevant variables and how to measure them
3. Plan your research design, sample size and statistical analysis

4. Collect data
5. Test your hypotheses
6. Interpret and share results

... but in practice, a lot of bad things can happen!

... but in practice, a lot of bad things can happen!

Questionable Research Practices (QRPs) ... for all tastes
(Loewenstein & Prelec, 2012)

- Failing to report all of a study's variables that are relevant for a finding
- Failing to report all of a study's conditions
- Deciding whether to exclude data after looking at the impact of doing so on the desired results
- Stop collecting data earlier than planned because an expected finding was already obtained
- Collect more data than planned in order to obtain the expected results
- Cherry picking: presenting the results of a study that best support the hypothesis instead of reporting all the findings
- HARKing: hypothesizing after the results are known
- Falsifying data (FRAUD)

One remedy to move toward a more rigorous science

Pre-registration via Registered Reports

'Because the study is accepted in advance,
the incentives for authors change from
producing the most beautiful story to the most accurate one"
(Chambers, 2018)

What about other recent crises?

Let's focus on three main aspects of confirmatory research:

1. Formalization of hypotheses and associated pre-planned statistical analysis
2. Operationalization of the variables of interest
3. Interpretation of (statistical) findings

What about other recent crises?

For each of these aspects, there is evidence of a crisis:

1. Formalization of hypotheses and associated pre-planned statistical analysis
2. Operationalization of the variables of interest
3. Interpretation of (statistical) findings

What about other recent crises?

For each of these aspects, there is evidence of a crisis:

1. Formalization of hypotheses and associated pre-planned statistical analysis

2. Operationalization of the variables of interest

3. Interpretation of (statistical) findings

The replication crisis (Ioannidis, 2005)

What about other recent crises?

For each of these aspects, there is evidence of a crisis:

1. Formalization of hypotheses and associated pre-planned statistical analysis

2. Operationalization of the variables of interest

   The validation crisis (Schimmack, 2021)

3. Interpretation of (statistical) findings

   The replication crisis (Ioannidis, 2005)

What about other recent crises?

For each of these aspects, there is evidence of a crisis:

1. Formalization of hypotheses and associated pre-planned statistical analysis

   The theory crisis (Eronen & Bringmann, 2021)

2. Operationalization of the variables of interest

   The validation crisis (Schimmack, 2021)

3. Interpretation of (statistical) findings

   The replication crisis (Ioannidis, 2005)

What about other recent crises?

For each of these aspects, there is evidence of a crisis:

1. Formalization of hypotheses and associated pre-planned statistical analysis

   The theory crisis (Eronen & Bringmann, 2021)

2. Operationalization of the variables of interest

   The validation crisis (Schimmack, 2020)

3. Interpretation of (statistical) findings

   The replication crisis (Ioannidis, 2005)

NOTE: These crises are not new. Since 1950, there have been many articles about them. However, these articles have had little impact on the scientific community

The validation (or measurement) crisis

Meta-Psychology, 2021, vol 5, MP.2019.1645
https://doi.org/10.15626/MP.2019.1645
Article type: Commentary

Open data: N/A
Open materials: N/A
Open and reproducible analysis: N/A
Open reviews and editorial process: Yes
Preregistration: N/A

Edited by: Daniel Lakens
Reviewed by: J. Gottfried, M. Wolf
Analysis reproduced by: N/A
All supplementary files can be accessed at OSF:
https://osf.io/afusj/

The Validation Crisis in Psychology

Ulrich Schimmack
Department of Psychology, University of Toronto Mississauga

Abstract

Cronbach and Meehl (1955) introduced the concept of construct validity and described how researchers can demonstrate that their measures have construct validity. Although the term construct validity is widely used, few researchers follow Cronbach and Meehl's recommendation to quantify construct validity with the help of nomological networks. As a result, the construct validity of many popular measures in psychology is unknown. I call for rigorous tests of construct validity that follow Cronbach and Meehl's recommendations to improve psychology as a science. Without valid measures even replicable results are uninformative. I suggest that a proper program of validation research requires a multi-method approach and causal modeling of correlations with structural equation models. Construct validity should be quantified to enable cost-benefit analyses and to replace existing measures with better measures that have superior construct validity.

Keywords: Measurement, Construct Validity, Convergent Validity, Discriminant Validity, Structural Equation Modeling; Nomological Networks

The validation (or measurement) crisis

Advances in Methods and Practices in Psychological Science
Volume 3, Issue 4, December 2020, Pages 456-465

https://doi.org/10.1177/2515245920952393

General Article

Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them

Jessica Kay Flake [1] **and Eiko I. Fried** [2]

Abstract

In this article, we define questionable measurement practices (QMPs) as decisions researchers make that raise doubts about the validity of the measures, and ultimately the validity of study conclusions. Doubts arise for a host of reasons, including a lack of transparency, ignorance, negligence, or misrepresentation of the evidence. We describe the scope of the problem and focus on how transparency is a part of the solution. A lack of measurement transparency makes it impossible to evaluate potential threats to internal, external, statistical-conclusion, and construct validity. We demonstrate that psychology is plagued by a *measurement schmeasurement* attitude: QMPs are common, hide a stunning source of researcher degrees of freedom, and pose a serious threat to cumulative psychological science, but are largely ignored. We address these challenges by providing a set of questions that researchers and consumers of scientific research can consider to identify and avoid QMPs. Transparent answers to these measurement questions promote rigorous research, allow for thorough evaluations of a study's inferences, and are necessary for meaningful replication studies.

Keywords

construct validity, validity, psychometrics, replication, *p*-hacking

The validation (or measurement) crisis

(psychometrically) invalid measurements

invalid and non replicable results

The validation (or measurement) crisis

AN EXAMPLE

A recent methodological review of 100 studies (both experimental and correlational) published in 3 leading journals of Psychology during 2008 shows that:

- Of 193 scales used, only 29% were reported with a citation to the associated validation study
- Surprising evidence that issues related to validity were mostly neglected
- Strong evidence of poor psychometric properties of the measures used

(Flake et al., 2022)

The theory crisis

aps
ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Perspectives on Psychological Science
2021, Vol. 16(4) 779–788


Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1745691620970586
www.psychologicalscience.org/PPS
SAGE

The Theory Crisis in Psychology: How to Move Forward

Markus I. Eronen[1] **and Laura F. Bringmann**[2]
[1]Department of Theoretical Philosophy, and [2]Department of Psychometrics and Statistics, University of Groningen

Abstract
Meehl argued in 1978 that theories in psychology come and go, with little cumulative progress. We believe that this assessment still holds, as also evidenced by increasingly common claims that psychology is facing a "theory crisis" and that psychologists should invest more in theory building. In this article, we argue that the root cause of the theory crisis is that developing good psychological theories is extremely difficult and that understanding the reasons why it is so difficult is crucial for moving forward in the theory crisis. We discuss three key reasons based on philosophy of science for why developing good psychological theories is so hard: the relative lack of robust phenomena that impose constraints on possible theories, problems of validity of psychological constructs, and obstacles to discovering causal relationships between psychological variables. We conclude with recommendations on how to move past the theory crisis.

Keywords
theory, phenomena, robustness, validity, causation

The theory crisis

PSYCHOLOGICAL INQUIRY
2020, VOL. 31, NO. 4, 271–288
https://doi.org/10.1080/1047840X.2020.1853461

TARGET ARTICLE

Lack of Theory Building and Testing Impedes Progress in The Factor and Network Literature

Eiko I. Fried

Department of Clinical Psychology, Leiden University, Leiden, Netherlands

ABSTRACT

The applied social science literature using factor and network models continues to grow rapidly. Most work reads like an exercise in model fitting, and falls short of theory building and testing in three ways. First, statistical and theoretical models are conflated, leading to invalid inferences such as the existence of psychological constructs based on factor models, or recommendations for clinical interventions based on network models. I demonstrate this inferential gap in a simulation: excellent model fit does little to corroborate a theory, regardless of quality or quantity of data. Second, researchers fail to explicate theories about psychological constructs, but use implicit causal beliefs to guide inferences. These latent theories have led to problematic best practices. Third, explicated theories are often weak theories: imprecise descriptions vulnerable to hidden assumptions and unknowns. Such theories do not offer precise predictions, and it is often unclear whether statistical effects actually corroborate weak theories or not. I demonstrate that these three challenges are common and harmful, and impede theory formation, failure, and reform. Matching theoretical and statistical models is necessary to bring data to bear on theories, and a renewed focus on theoretical psychology and formalizing theories offers a way forward.

KEYWORDS

Factor model; formal theory; network model; open science; replicability; statistical equivalence; theory

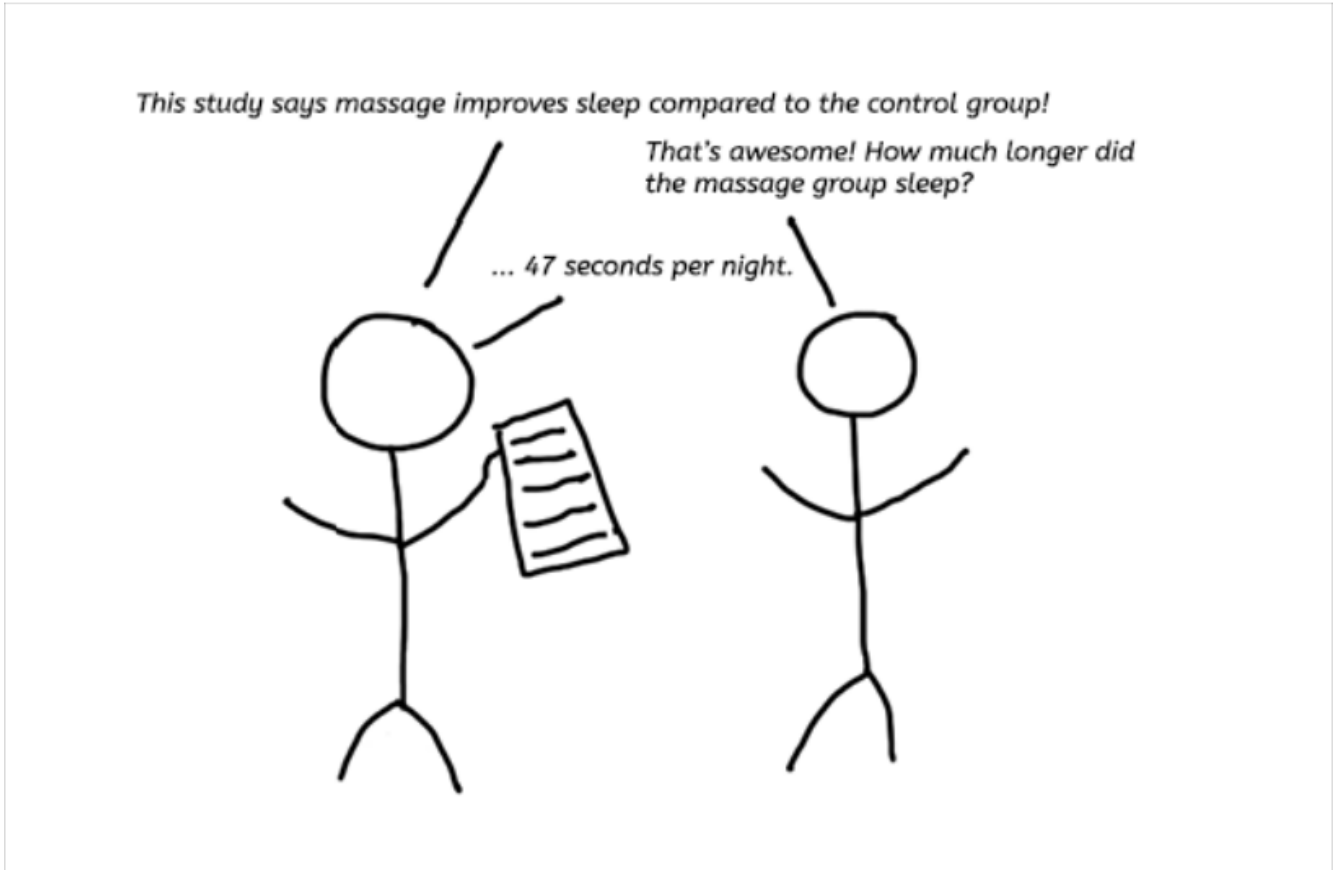
The theory crisis

THE MAIN INGREDIENTS

- Vague and not testable theories
- Lack of mathematical formalization of hypotheses
- Mismatch between theoretical and statistical models

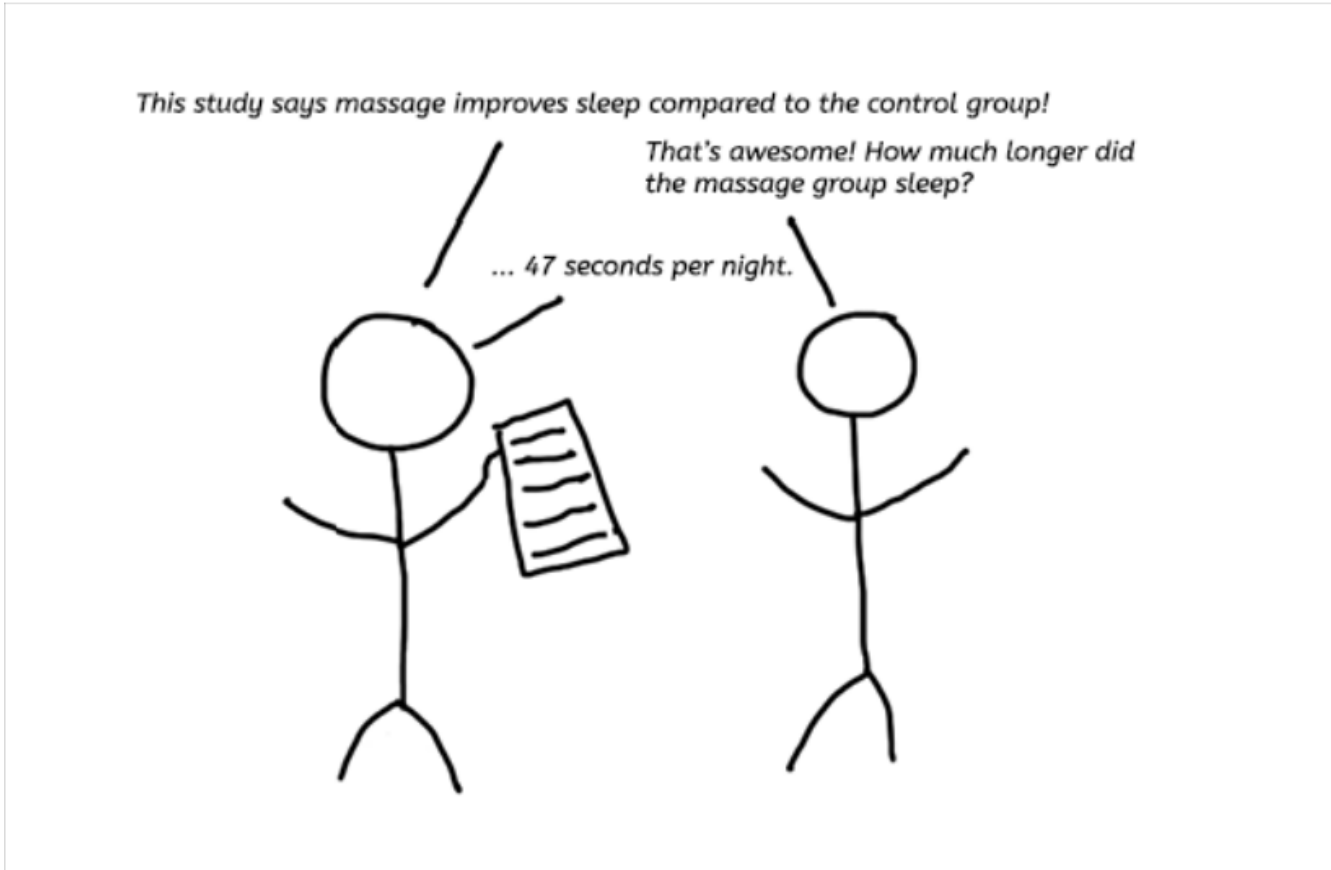
The theory crisis

EXAMPLE 1



The theory crisis

EXAMPLE 1

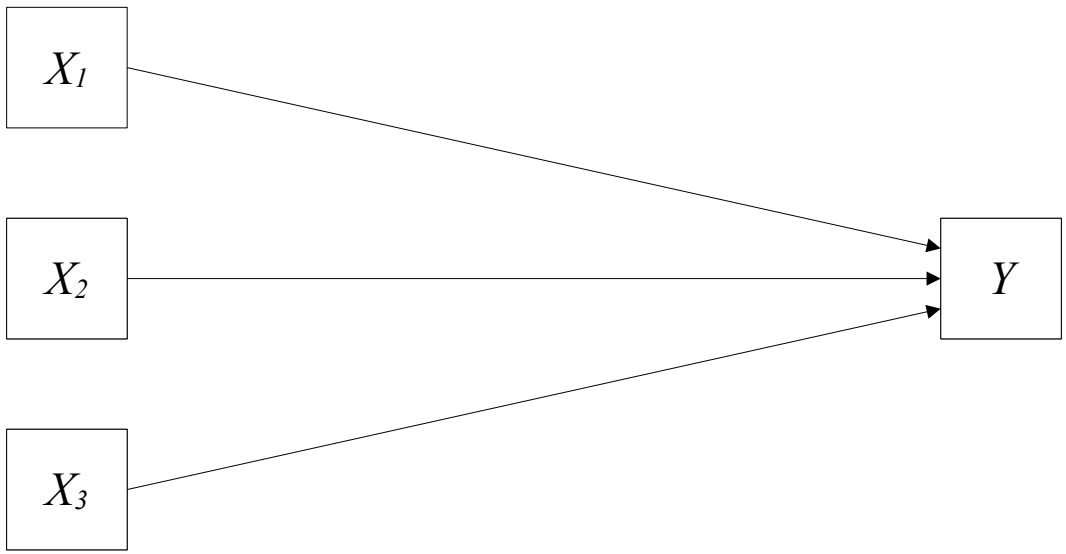


- bad hypothesis: the treatment has an effect
- sound hypothesis: the treatment has a specified clinically-relevant effect

The theory crisis

EXAMPLE 2

:-(

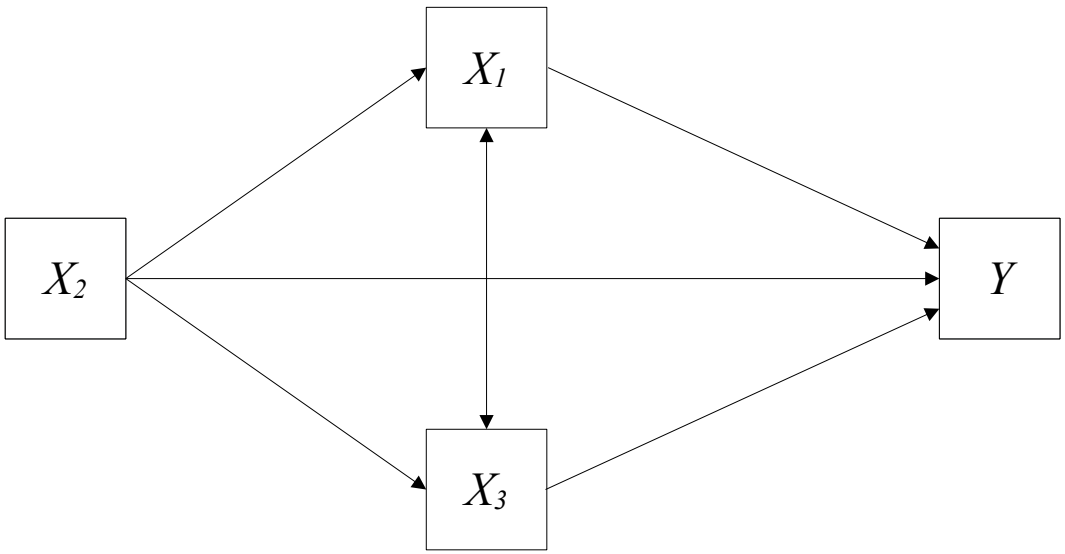


- poor formalization of the *theoretical model*
- linear regression as an inadequate *statistical model*

The theory crisis

EXAMPLE 2

:-)



- good formalization of the *theoretical model*
- Structural Equation Models as an adequate *statistical tool*

Statistical remedies/proposals

Statisticians should know, at least, the basics of:

- Philosophy of Science
- Measurement Theory

Hot statistical approaches:

- Structural Equation Models
- Causal Inference
- Multiverse Analysis

Statisticians should be trained to work with other scientists

Multidisciplinary remedies/proposals

Scientists should know, at least, the basics of:

- Statistical Reasoning
  (how to deal with data, variation and chance)

Experts in the study phenomenon and statisticians must:

- COLLABORATE from the beginning,
  and **BEFORE** data are collected
  (see Ronald Fisher, 1938)

Conclusion

Let's replace Crisis with Renaissance!

gianmarco.altoe@unipd.it

https://psicostat.dpss.psy.unipd.it/