Open Science And Reproducibility

ALEXANDRA SARAFOGLOU

MYRTHE VEENMAN

Open Science and Reproducibility

Outline

- Confirmatory versus Exploratory Research
- Practical Tools to Increase Transparency
- Preregistration and Analysis Blinding
- Teaching Good Research Practices: Student Perspective

Confirmatory versus Exploratory Analysis

Disclaimer

Slides on Confirmatory versus Exploratory Research were taken from Eric-Jan Wagenmakers TeaP talk on the reproducibility crisis and future prospects of transparent research practices.

His talk is available at:

https://www.youtube.com/watch?v=LxRHkk2MBWw&t=365s

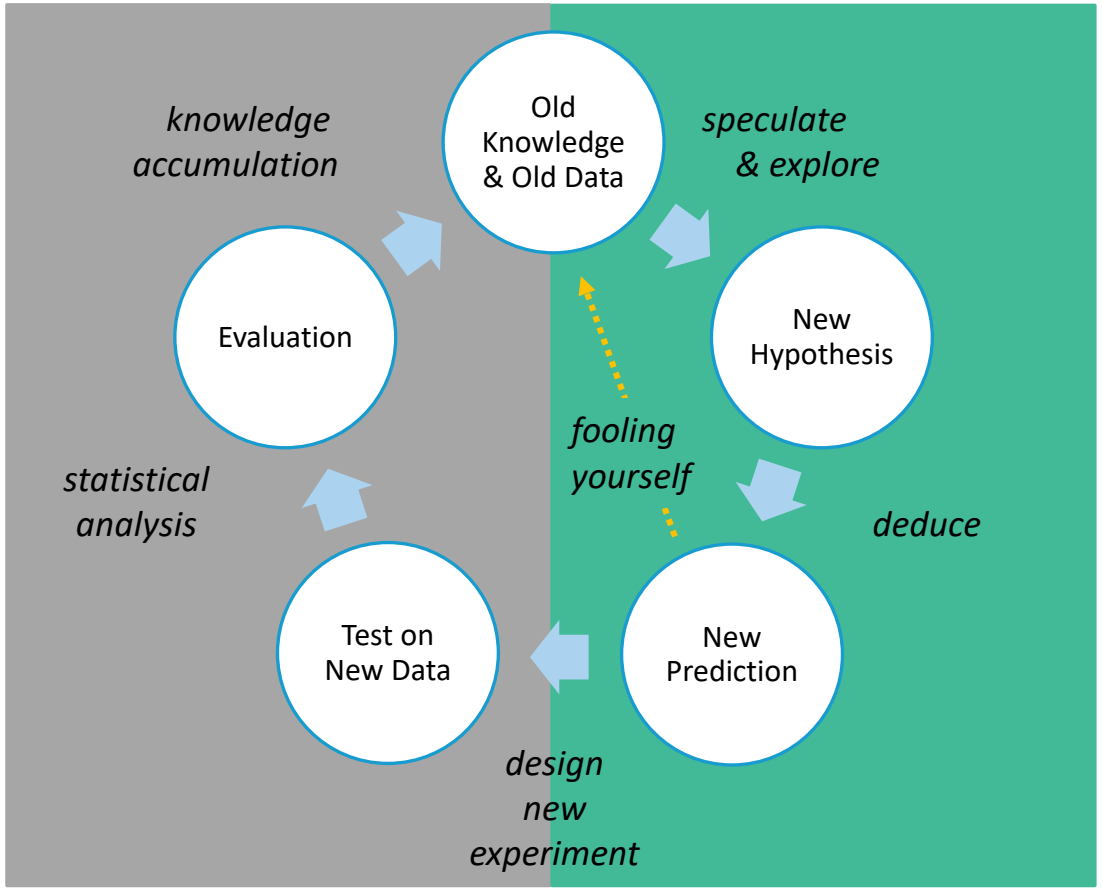
Confirmatory versus Exploratory Analysis

Using the Data Twice

- There is a conceptual distinction between **hypothesis-generating** and **hypothesis-testing** research (De Groot 1956/2014; Reichenbach, 1938)
- When the data inspire a hypothesis, you cannot use the same data to test this hypothesis

Confirmatory versus Exploratory Analysis

The Empirical Cycle



Poppers disconfirmability test

The Statistical Context of Justification – Confirmatory Research

The Creative Context of Discovery – Exploratory Research

Confirmatory versus Exploratory Analysis

Why multiple testing matters

- Exploratory analyses introduce a multiple comparisons problem which increases the Type-I-error rate
- Maximum allowable probability to make a Type-I-error is α (typically $\alpha = 0.05$)
- $P(\text{not making an error}) = 1 - \alpha$
- $P(\text{not making an error in } m \text{ tests}) = (1 - \alpha)^m$
- $P(\text{making at least one error in } m \text{ tests}) = 1 - (1 - \alpha)^m$

Confirmatory versus Exploratory Analysis

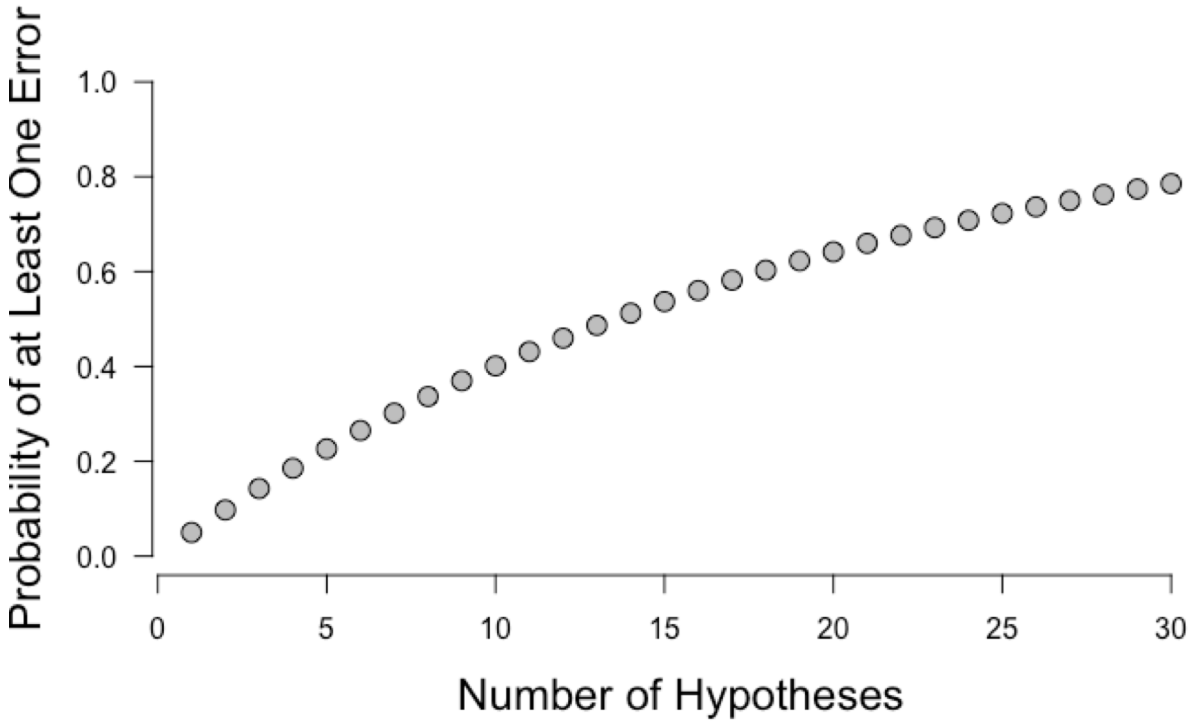
Why multiple testing matters

- Trying out m = 5 different analyses:

$$1 - (1 - 0.05)^5 = 0.23$$

Probability to make at least 1

Type-I error in 5 tests is 23%.



Confirmatory versus Exploratory Analysis

Why multiple testing matters

Simmons, Nelson, Simonsohn (2011)

Table 1. Likelihood of Obtaining a False-Positive Result

	Significance level		
Researcher degrees of freedom	$p < .1$	$p < .05$	$p < .01$
Situation A: two dependent variables ($r = .50$)	17.8%	9.5%	2.2%
Situation B: addition of 10 more observations per cell	14.5%	7.7%	1.6%
Situation C: controlling for gender or interaction of gender with treatment	21.6%	11.7%	2.7%
Situation D: dropping (or not dropping) one of three conditions	23.2%	12.6%	2.8%
Combine Situations A and B	26.0%	14.4%	3.3%
Combine Situations A, B, and C	50.9%	30.9%	8.4%
Combine Situations A, B, C, and D	81.5%	60.7%	21.5%

Confirmatory versus Exploratory Analysis

Why multiple testing matters

- If the tests are informed by the data, they lose their predictive interpretation, and, with it, their statistical validity
- Exploratory analyses introduce a multiple comparisons problem which increases the Type-I-error rate

→ **Statistical tests are only appropriate for purely confirmatory research efforts**

Confirmatory versus Exploratory Analysis

Solutions

1. Be transparent about your research methods and study design
2. Specify your hypotheses and your analysis plan before you start data collection
3. Exchange experiences with peers (e.g., through an Open Science Community)
4. Start small: incorporate good research practices in your student projects and teaching

Confirmatory versus Exploratory Analysis

Solutions

1. **Be transparent about your research methods and study design**
2. Specify your hypotheses and your analysis plan before you start data collection
3. Exchange experiences with peers (e.g., through an Open Science Community)
4. Start small: incorporate good research practices in your student projects and teaching

Transparency

Transparency Checklist

- Created by open science advocates and editors from leading psychology journals
- https://eltedecisionlab.shinyapps.io/TransparencyChecklist/

Submitted Manuscript: Confidential

Title: A Consensus-Based Transparency Checklist for Social and Behavioral Researchers

Authors:

B. Aczel[1*], B. Szaszi[1], A. Sarafoglou[2], Z. Kekecs[1], Š. Kucharský[2], D. Benjamin[3], C. D. Chambers[4], A. Fisher[2], A. Gelman[5] , M. A. Gernsbacher[6], J. P. Ioannidis[7], E. Johnson[5], K. Jonas[8], S. Kousta[9], S. O. Lilienfeld[10,11], D. S. Lindsay[12], C. C Morey[4], M. Monafò[13], B. R. Newell[14], H. Pashler[15] , D. R. Shanks[16], D. J. Simons[17], J. M. Wicherts[18], D. Albarracin[17] , N. D. Anderson[19], J. Antonakis[20], H. Arkes[21], M. D. Back[22], G. C. Banks[23], C. Beevers[24], A. A. Bennett[25], W. Bleidorn[26], T. W. Boyer[27], C. Cacciari[28], A. S. Carter[29], J. Cesario[30], C. Clifton[31], R.M. Conroy[33], M. Cortese[34], F. Cosci[35], N. Cowan[36], J. Crawford[37], E. A. Crone[38], J. Curtin[6], R. Engle[39], S. Farrell[40], P. Fearon[16], M. Fichman[41], W. Frankenhuis[42], A. M. Freund[43], M. G. Gaskell[44], R. Giner-Sorolla[45], D. P. Green[5], R. L. Greene[46], L. L. Harlow[47], F. Hoces de la Guardia[48], D. Isaacowitz[49], J. Kolodner[50], D. Lieberman[51], G. D. Logan[52], W. B. Mendes[53], L. Moersdorf[43], B. Nyhan[54], J. Pollack[55], C. Sullivan[56], S. Vazire[26], E-J. Wagenmakers[2]

Confirmatory versus Exploratory Analysis

Solutions

1. Be transparent about your research methods and study design
2. **Specify your hypotheses and your analysis plan before you start data collection**
3. Exchange experiences with peers (e.g., through an Open Science Community)
4. Start small: incorporate good research practices in your student projects and teaching

Preregistration

Solutions

- Specify your **hypotheses** and your **analysis plan** before you start data collection
  - Preregistration (Registered Reports; over 200 participating journals!)

CENTER FOR OPEN SCIENCE

https://cos.io/

PREREGISTERED

Preregistration

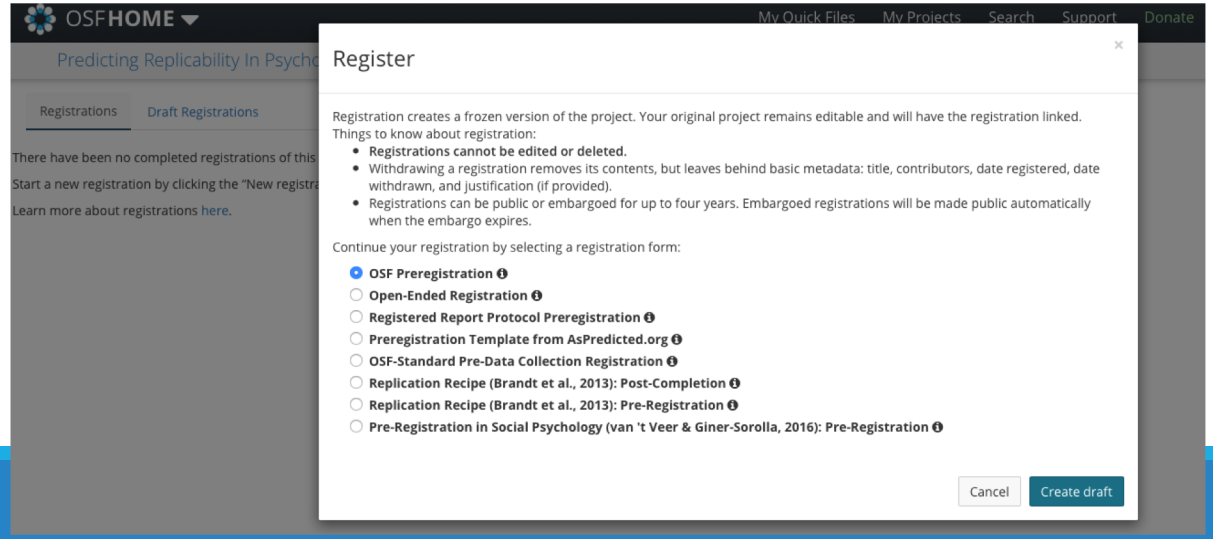
Solutions

- Specify your **hypotheses** and your **analysis plan** before you start data collection
  - Preregistration (Registered Reports; over 200 participating journals!)

Preregistration

Challenges

- Specify your **hypotheses** and your **analysis plan** before you start data collection
  - Preregistration (Registered Reports)
- Problem: Not clear what should be included in the preregistration



OSF

https://osf.io/

PREREGISTERED

Preregistration

Challenges

- Specify your **hypotheses** and your **analysis plan** before you start data collection
  - Preregistration (Registered Reports)
- Problem: Hypotheses and analysis plan are often too vague

Preregistration

Challenges

- Specify your **hypotheses** and your **analysis plan** before you start data collection

1. Hypotheses and Analysis Plans are Ambiguous
   - Example: “Are soccer referees more likely to give red cards to players with dark skin than to players with light skin?”

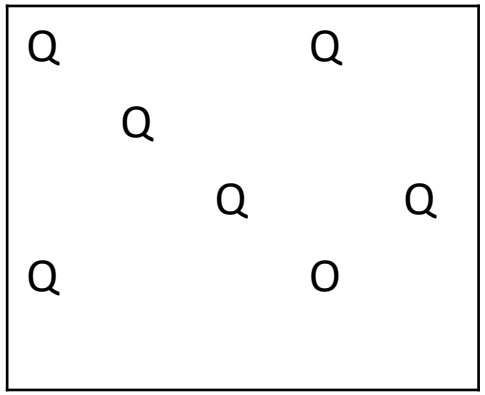
     - Chi-square test? Logistic regression?
     - How do you measure ‘dark’ and ‘light’ skin? Dichotomous factor or multiple levels?
     - Would you control for the referees skin color?
     - Would you treat each red-card decision as an independent observation?
     - Would you try to control for the seniority of the referee?

Preregistration

Challenges

- Specify your **hypotheses** and your **analysis plan** before you start data collection

2. Crucial Information is Missing
   - Are you testing assumptions? What will you do if assumptions are violated?
   - How do you measure your dependent variable? (Example: “classical music improves attention”)



Preregistration

Challenges

- Specify your **hypotheses** and your **analysis plan** before you start data collection

3. Analysis Plans Have Not Been Tested
   - Create dummy data! (or use data from an similar study/pilot study)
   - Try out your analysis (does your analysis plan actually work?)
   - Ask for feedback for this analysis

Preregistration

Challenges

- Specify your **hypotheses** and your **analysis plan** before you start data collection

4. Time pressure and Unexpected Features of the Data

Preregistration

Challenges: Unexpected Features of the Data

Dutilh et al. (2017)
Preregistration of an impossible analysis

Reproducibility project:
Cancer Biology

Horrigan (2017)
Spontaneous tumor regressions

Arid, Kandela & Mantis (2017)
Unexpected early deaths in control group

Blinded Analysis

Blinded Analysis

- Common practice in (Astro)physics
- Allows you to make data-dependent choices without introducing bias

Blind analysis: Hide results to seek the truth

Robert MacCoun & Saul Perlmutter

07 October 2015

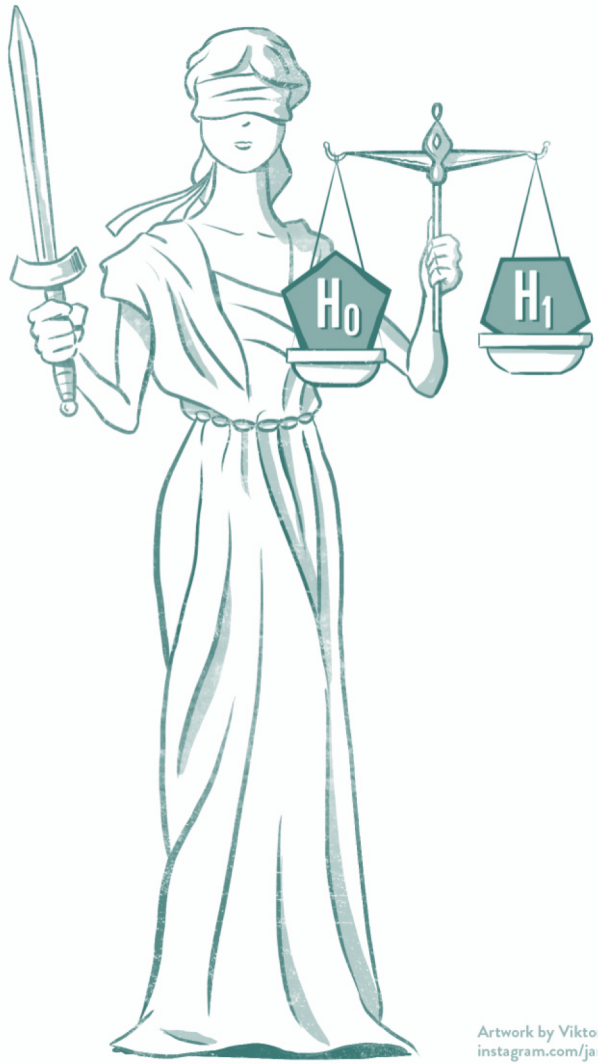
More fields should, like particle physics, adopt blind analysis to thwart bias, urge Robert MacCoun and Saul Perlmutter.

PDF | Rights & Permissions

Subject terms: Research management

Illustration by Dale Edwin Murray

MacCoun & Perlmutter (2015)



Blinded Analysis

How does Blinding work?

1. Experimenter collects data

2. Data is blinded

Blinded Analysis

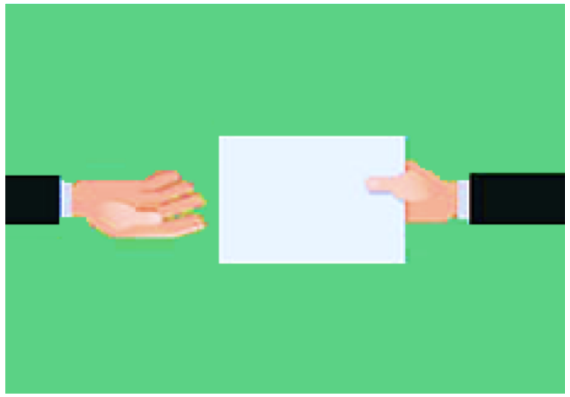
How does Blinding work?

3. Experimenter gives the *hypotheses* and *blinded data* to analyst

4. Analyst decides on analysis plan

Blinded Analysis

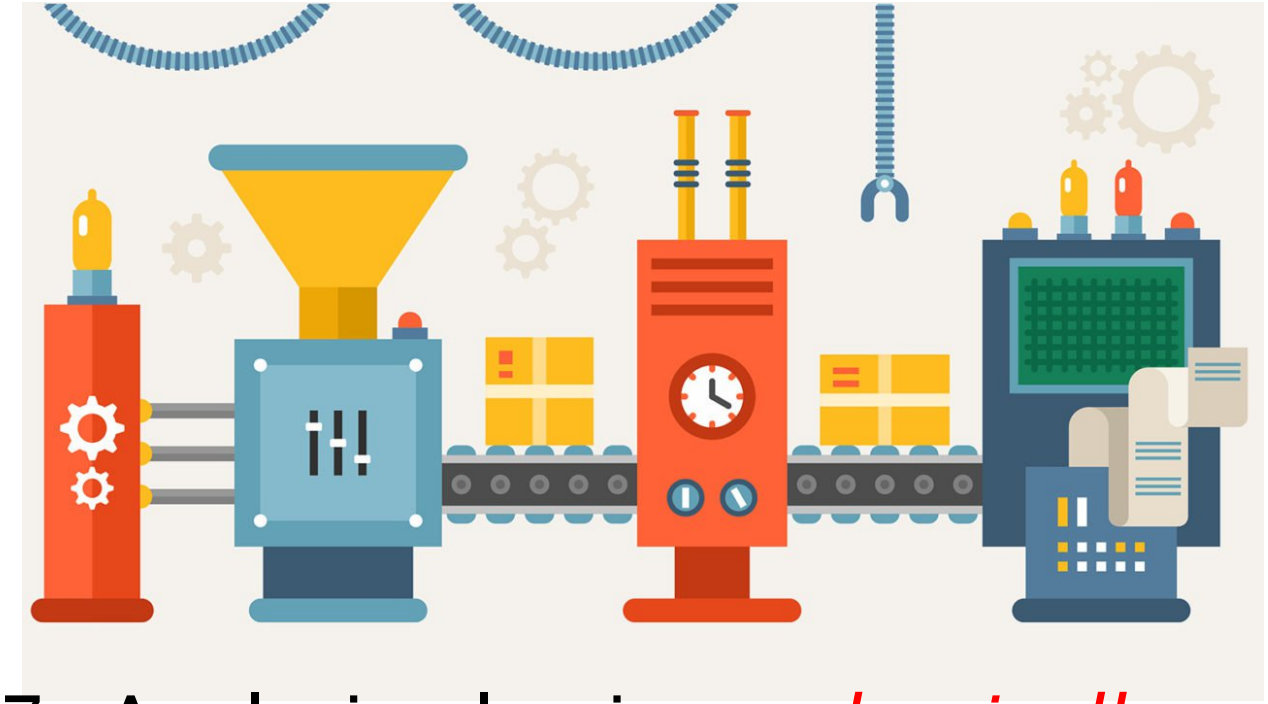
How does Blinding work?



5. Analyst gives *analysis plan* to experimenter

6. The blind is lifted

Blinded Analysis

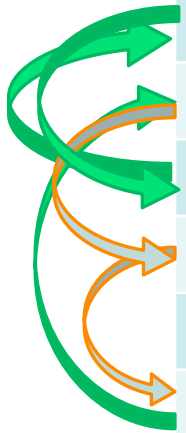


7. Analysis plan is *mechanically* conducted on the original data

Blinded Analysis

Different Research Designs Demand Different Blinding Methods

e.g., Regression



# Facebook Friends	Amygdala Size	Gender	Age	Overal Brain volume
190	0.655	male	28	1238
232	0.680	male	35	1386
84	0.647	female	18	1176
311	0.698	male	20	1172
138	0.699	female	27	1185
356	0.732	female	34	970

Blinded Analysis

Different Research Designs Demand Different Blinding Methods

e.g., Regression

# Facebook Friends	Amygdala Size	Gender	Age	Overall Brain volume
84	0.655	male	28	1238
356	0.680	male	35	1386
190	0.647	female	18	1176
232	0.698	male	20	1172
138	0.699	female	27	1185
311	0.732	female	34	970

Blinded Analysis

Different Research Designs Demand Different Blinding Methods

e.g., ANOVA

Education	Gender: male	Gender: female
high	13.7 (SD 1)	14.5 (SD 1.2)
low	11.8 (SD 2)	9.46 (SD 2.5)

Blinded Analysis

Different Research Designs Demand Different Blinding Methods

e.g., ANOVA

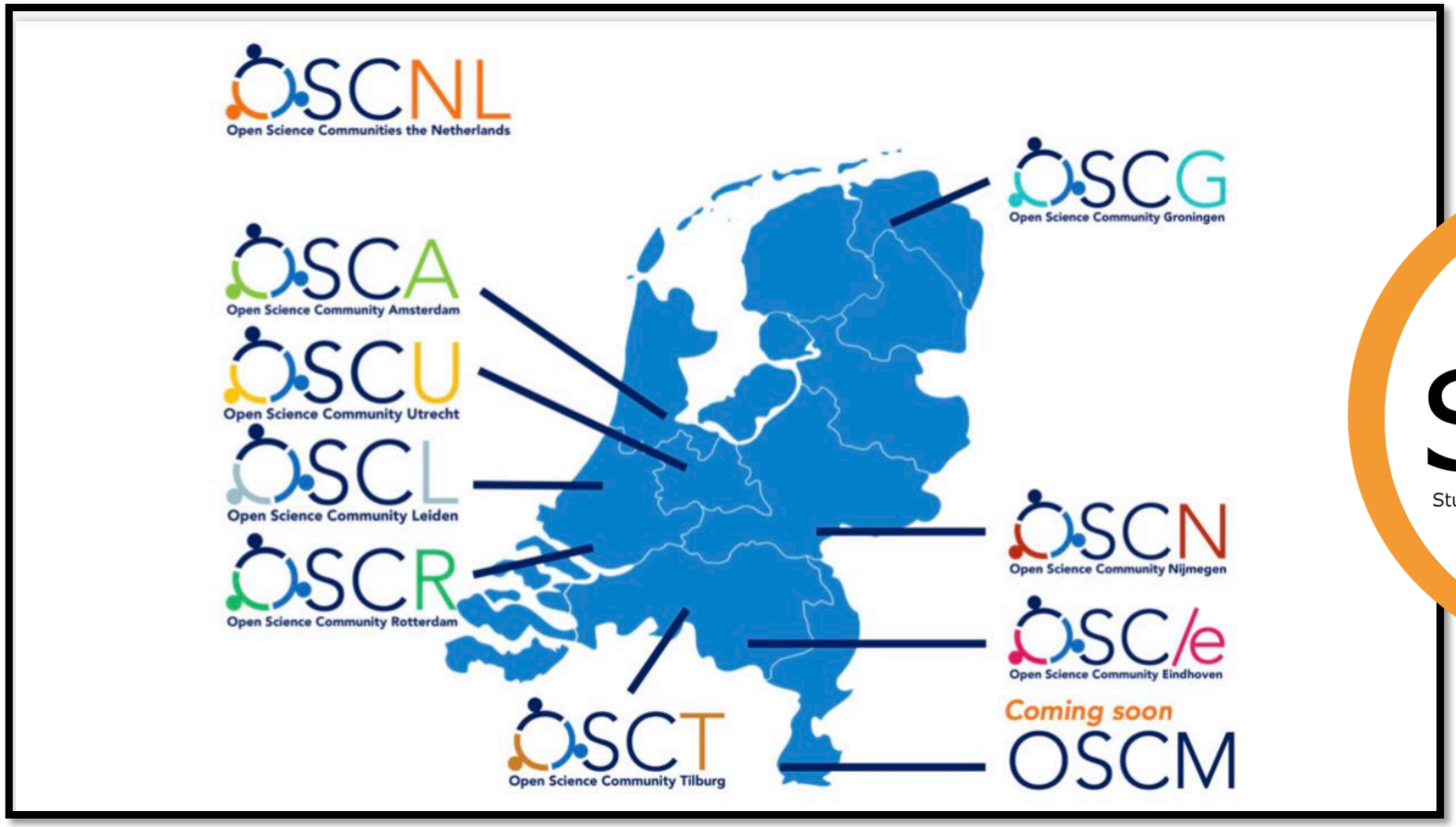
		Gender	
		?	?
Education	?	0 (SD 1)	0 (SD 1.2)
	?	0 (SD 2)	0 (SD 2.5)

Confirmatory versus Exploratory Analysis

Solutions

1. Be transparent about your research methods and study design
2. Specify your hypotheses and your analysis plan before you start data collection
3. **Exchange experiences with peers (e.g., through an Open Science Community)**
4. Start small: incorporate good research practices in your student projects and teaching

Open Science Communities in the Netherlands



SIOS
Student Initiative for Open Science

Use Available Resources

Further Topics

- Open Data and Open Access
- JASP Data Documentation Guidelines
- Helpdesks at University Libraries
  - (in Heerlen: Gjalt-Jorn Peters)

OPEN ACCESS

Confirmatory versus Exploratory Analysis

Solutions

1. Be transparent about your research methods and study design
2. Specify your hypotheses and your analysis plan before you start data collection
3. Exchange experiences with peers (e.g., through an Open Science Community)
4. **Start small: incorporate good research practices in your student projects and teaching**

Teaching Good Research Practices

TEACHING GOOD RESEARCH PRACTICES 1

Psychology Learning and Teaching

Teaching Good Research Practices: Protocol of a Research Master Course

Alexandra Sarafoglou[1], Suzanne Hoogeveen[2], Dora Matzke[1], Eric-Jan Wagenmakers[1]

[1] Department of Psychology, Psychological Methods, University of Amsterdam, The Netherlands

[2] Department of Psychology, Social Psychology, University of Amsterdam, The Netherlands

SIOS
Student Initiative for Open Science

Sarafoglou A., Hoogeveen S., Matzke D., & Wagenmakers, E.-J. (in press). Teaching Good Research Practices: Protocol of a Research Master Course. Preprint available on PsyArXiv: https://psyarxiv.com/gvesh/

SIOS
Student Initiative for Open Science

Open Science: The Student Experience

Myrthe Veenman

Open Science:
The Student Experience

Outline

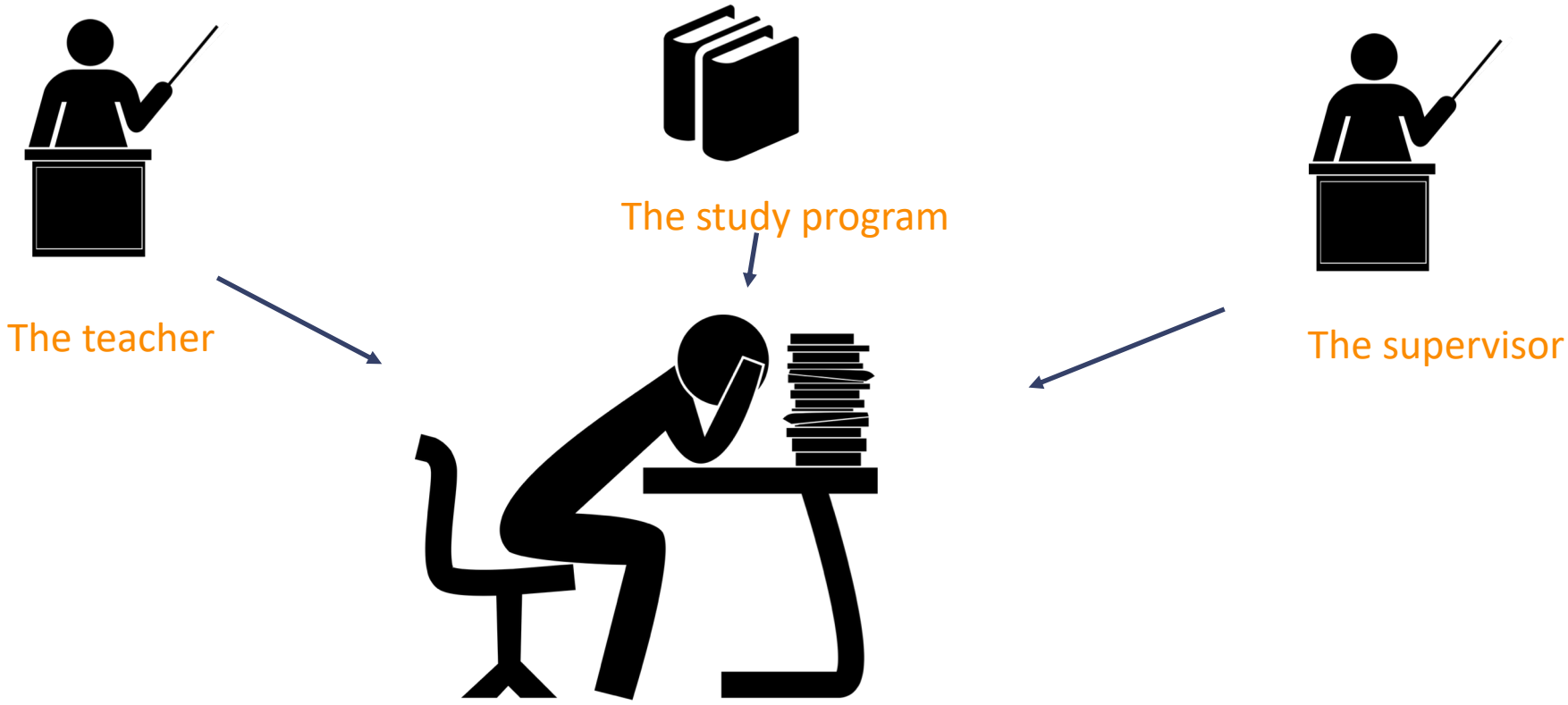
- Why?
- Supervisor
- Teacher
- Study Program
- Advisor
- Challenges
- Improvements

The Importance of Open Science in The Study Program

Why?

- Why not?
  - If universities teach students how to do research, then teach them to do research the right way
- Researchers of the future
- Change the research culture
  - Publish, publish, publish

Open Science:
The Student Experience



The Importance of the Supervisor

Open Science Practices in Internship or Thesis

- Preregistration
  - Benefits
    - Confirmatory versus Exploratory
    - Feedback from supervisor before analysis
    - Analysis plan ready before data collection

The Importance of the Supervisor

Methodology Shop

- Free methodological advice
- Students come by with their analysis problems when it is already too late
- Come before conducting your analysis to make an analysis plan
  - Preregistration avoids that students still stress at the end of their project about their analyses

The Importance of the Supervisor

Open Science Practices in Internship or Thesis

- Preregistration
  - OSF
  - Proposal
- Open Data
  - Shared data
- Open Code
- Open Material
  - Provided access to material

The Importance of the Teacher

Open Science Practices in Courses

- Preregistration
  - Predata report
  - Either applied or discussed

The Importance of the Teacher

Other Open Science Practices in Courses

- Share R Code/SPSS syntax
  - Hand in your code for assignments/projects
- Discussed Exploratory versus Confirmatory
- Open Material
  - Material for the course was online
  - Encouraged to put own assignments/projects online

Open Science in The Study Program

Course: Good Research Practices

- Students really appreciated the course
  - They would recommend it to other universities
  - Some suggested instead to apply the course content in existing courses

Open Science in The Study Program

Programming or Not?

- Make your analysis reproducible
- Different programs that allow for that
  - R
  - JASP
- University of Glasgow and University of Tübingen teach Psychology students in the Bachelor program R instead of SPSS
  - Good responses from Glasgow students

The Importance of an Advisor

Who to Ask for Help?

- Researchers
- Professors
- Supervisor
- Open Science fans
- However, most students did not know who to ask exactly
  - Which researchers/professors?
  - No general university advisor for students or not aware of the existence

Challenges

- Time constraints (especially for preregistration)

Challenges

Time Constraints

- Example of Bachelor Project

1. **Week 1 – 8**
   a) Literature research
   b) Study design
   c) Ethical approval, prepare materials, book lab etc.
   d) Introduction, Methods, and Analysis Plan
   e) Create Dummy Data, Analyze Dummy Data, Adjust Analysis Plan (R Code)
   f) One round of Feedback

2. **Week 9 – 12**
   a) Data collection
   b) Incorporate feedback for Introduction and Methods part

3. **Week 13 – 16**
   a) Analyze Data
   b) Results section and Discussion
   c) Reflection report
   d) Prepare presentation

Challenges

- Time constraints (especially for preregistration)
  - Different experiences
    - Time was not a problem for some
    - Others felt the time pressure
    - Some experienced it as time saving in the end
- Lack of experience
  - Never done before themself
  - Supervisor could not help

How Can We Improve the Experience for Students?

How Can We Improve the Experience for Students?

Make It Practical!

How Can We Improve the Experience for Students?

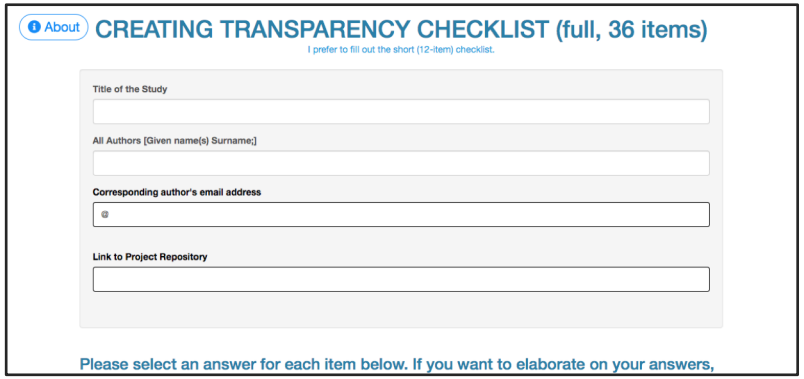
Make It Practical!

- Tutorials, step-by-step instructions
- Data Management
  - Guidelines how to upload/share data (to OSF)
  - Code styles
- Guidelines on how to share experiments
- Examples
- Let students do it

How Can We Improve the Experience for Students?

Make It Practical!

- Tutorials, step-by-step instructions
- Data Management
  - Guidelines how to upload/share data (to OSF) and create a README file
  - Code styles
- Guidelines on how to share experiments
- Examples
- Let students do it



https://eltedecisionlab.shinyapps.io/TransparencyChecklist/

How Can We Improve the Experience for Students?

Other Suggestions

- An Open Science advisor per university/department
- Listing pro Open Science journals

How Can We Improve the Experience for Students?

Other Suggestions

- An Open Science advisor per university/department
- Listing pro Open Science journals

OSCA
Open Science Community Amsterdam

Journals that implement Open Science badges:
https://cos.io/our-services/open-science-badges/

How Can Teachers Improve the Experience for Students?

Suggestions for Courses

- Provide open access to course material
- Implement preregistration more (e.g., predata report for final project)
- Implement Open Data + data management
- Apply Open Science tools, for instance the OSF

What Would Students Like to Learn More About?

Suggestions for Study Programs

- Latest techniques in Open Science
- Developments in Open Science
  - For example a seminar once a month
- How to apply Open Science in your study
- How Open Science can improve your scientific career
- Concrete steps for junior researchers

What Would Students Like to Learn More About?

Suggestions for Study Programs

- Latest techniques in Open Science
- Developments in Open Science
  - For example a seminar once a month
- How to apply Open Science in your study
- How Open Science can improve your scientific career
- Concrete steps for junior researchers

Open Science and Reproducibility

Summary

- Confirmatory versus Exploratory Research
- Practical Tools to Increase Transparency
- Preregistration and Analysis Blinding
- Teaching Good Research Practices: Student Perspective