

# Responsible Research Assessment

Evaluating research quality beyond *h*-index and JIF  
with the RESQUE framework  
(**RE**search **QU**ality **E**valuation)

*Anne Gärtner & Felix Schönbrodt*

*Alp Kaan Aksu, Maximilian Ernst, Maximilian Frank, Nele Freyer, Mario Gollwitzer,  
Malika Ihle, Jens Lange, Dorothee Mischkowski, Philipp Musfeld, Aaron Peikert, Le Vy  
Phan, Manfred Schmitt, Anne M. Scheel, Anna-Lena Schubert, Ulf Steinberg, Daniel  
Leising*

# What is the goal of research assessment in academic hiring?

- To find a colleague who ...
  - brings a lot of third-party funding to the university?
  - boosts the university's THE ranking by having a lot of papers, which are also highly cited?
  - excels in teaching and mentoring?
  - contributes to scientific progress and credible knowledge?
  - is a nice and agreeable person who makes no trouble in the department?

# What is scientific progress, and can we predict it?

- See COBIT maturity model: „[...] accounting for the fact that a **well-developed process naturally leads to products of a certain quality**“  
(zu Castell, 2024)
- “**Quality** is primarily an activity-oriented concept, concerning the skill and competence in the performance of some task. **Progress** is a result-oriented concept, concerning the success of a product relative to some goal.
  - All acceptable work in science has to fulfill certain standards of quality. But it seems that there are no necessary connections between quality and progress in science. Sometimes very well-qualified research projects fail to produce important new results, while less competent but more lucky works lead to success.
- Nevertheless, the skillful use of the methods of science will make progress highly probable. *Hence, the best practical strategy in promoting scientific progress is to support high-quality research.*” (Niiniluoto, 2019, p. 6)
- Overall, very low interrater reliability on quality assessment of publications/grant proposals
  - But: Higher agreement on the low end of the scale (Cicchetti, 1991)

# Bewertungskriterien wissenschaftlicher Leistung

„If we want science with impact, we need to reward **behaviour that is consistent with good research practices**, not impact factors.“

*Ivan Oransky*  
(founder of Retraction Watch)



„1. Recognise the **diversity of contributions** to research.

- „practices that contribute to robustness, openness, transparency,
- „Inappropriate uses of journal- and publication-based metrics in research assessment should be abandoned. In particular, this means **moving away from using metrics like the Journal Impact Factor (JIF), Article Influence Score (AIS) and h-index** as proxies for quality and impact.“



„das ganze Spektrum wissenschaftlicher Publikationsformen gleichwertig in [...] Lebensläufen abzubilden“

„Dies können beispielsweise Artikel auf Preprint-Servern, Datensätze oder Softwarepakete sein.“

„Angaben zu quantitativen Metriken wie Impact-Faktoren und h-Indizes im Lebenslauf oder Antrag werden nicht benötigt und sollen in der Begutachtung keine Berücksichtigung finden.“

# Bewertungskriterien wissenschaftlicher Leistung

- Wachsende **Unzufriedenheit mit den klassischen Metriken** zur Bewertung von Forschungsleistungen (JIF; h-Index, Anzahl Publikationen, Summe Drittmittel etc.)
- Um hier Alternativen zu entwickeln, wurden u.a. die **San Francisco Declaration on Research Assessment** (DORA: <https://sfdora.org>) formuliert und die **Coalition for Advancing Research Assessment** (CoARA: <https://coara.eu>) gegründet.
- Die DGPs ist der CoARA kürzlich beigetreten und ermuntert die Institute, dies ebenfalls zu tun.
  - > 600 institutionelle Mitglieder, inkl. European Research Council, DFG, DGPs, Leibniz-Gemeinschaft, Wissenschaftsrat
  - „German National Chapter“
- In der DGPs hat sich eine **Arbeitsgruppe** mit DORA und CoARA und deren Umsetzung befasst (verankert an DGPs-Kommission „Open Science“).

# Neue Bewertungskriterien – der Prozess

2022



Ten Steps Towards  
better Personality  
Science

Quality rating  
scheme for  
publications



19 commentaries

The Hong Kong Principles for assessing  
researchers: Fostering research integrity



2023



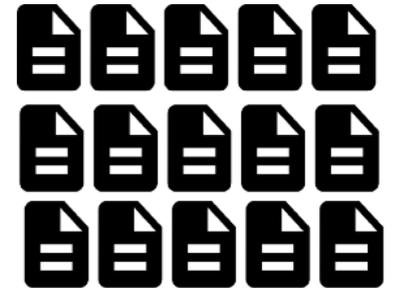
RRA 1 (Principles)



RRA 2 (Implementation)



Empfehlungen zur Bewertung  
wissenschaftlicher  
Leistungen bei  
Berufungsverfahren in der  
Psychologie



15 commentaries



6 commentaries

- ✓ Community-proofed
- ✓ Collaboratively improved with multiple breakout groups focusing on specific indicators (e.g., software, theory development, ...)
- ✓ Development backed by DGPs
- ✓ Ready-to-use templates, adaptable to local needs
- ✓ Efficiency & reduced committee workload by automation

## RESQUE Framework:

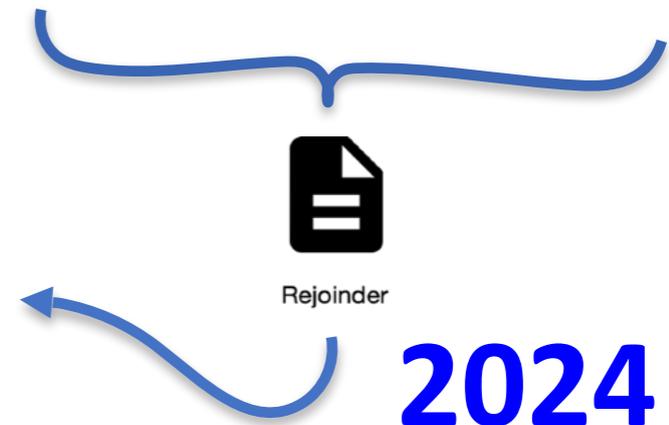
- 4 principles
- 2-phase assessment
- How-tos for hiring committees

## Rating schemes:

- RESQUE-Publication
- RESQUE-Data
- RESQUE-Software

## Tools:

- Interactive web form
- R-scripts for profile visualization



# Zum Nachlesen ...

## Diskussionsforum

### Empfehlungen zur Bewertung wissenschaftlicher Leistungen bei Berufungsverfahren in der Psychologie

Anne Gärtner<sup>1</sup>, Daniel Leising<sup>1</sup> und Felix D. Schönbrodt<sup>2</sup>

<sup>1</sup>Technische Universität Dresden, Deutschland

<sup>2</sup>Ludwig-Maximilians-Universität München, Deutschland

### Responsible Research Assessment I: Implementing DORA for hiring and promotion in psychology

AUTHORS

Felix Schönbrodt, Anne Gärtner, Maximilian Frank, Mario Gollwitzer, Malika Ihle, Dorothee Mischkowski, Le Vy Phan, Manfred Schmitt, Anne M. Scheel, Anna-Lena Schubert, Ulf Steinberg, Daniel Leising

AUTHOR ASSERTIONS

Conflict of Interest: No ▾

Public Data: Not applicable ▾

Preregistration: Not applicable ▾

### Responsible Research Assessment II: A specific proposal for hiring and promotion in psychology

AUTHORS

Anne Gärtner, Daniel Leising, Felix Schönbrodt

AUTHOR ASSERTIONS

Conflict of Interest: No ▾

Public Data: Not applicable ▾

Preregistration: Not applicable ▾

*Psychologische Rundschau*  
(Diskussionsforum mit 6 Kommentaren)

*Meta-Psychology*  
(Special Issue mit 15 Kommentaren)

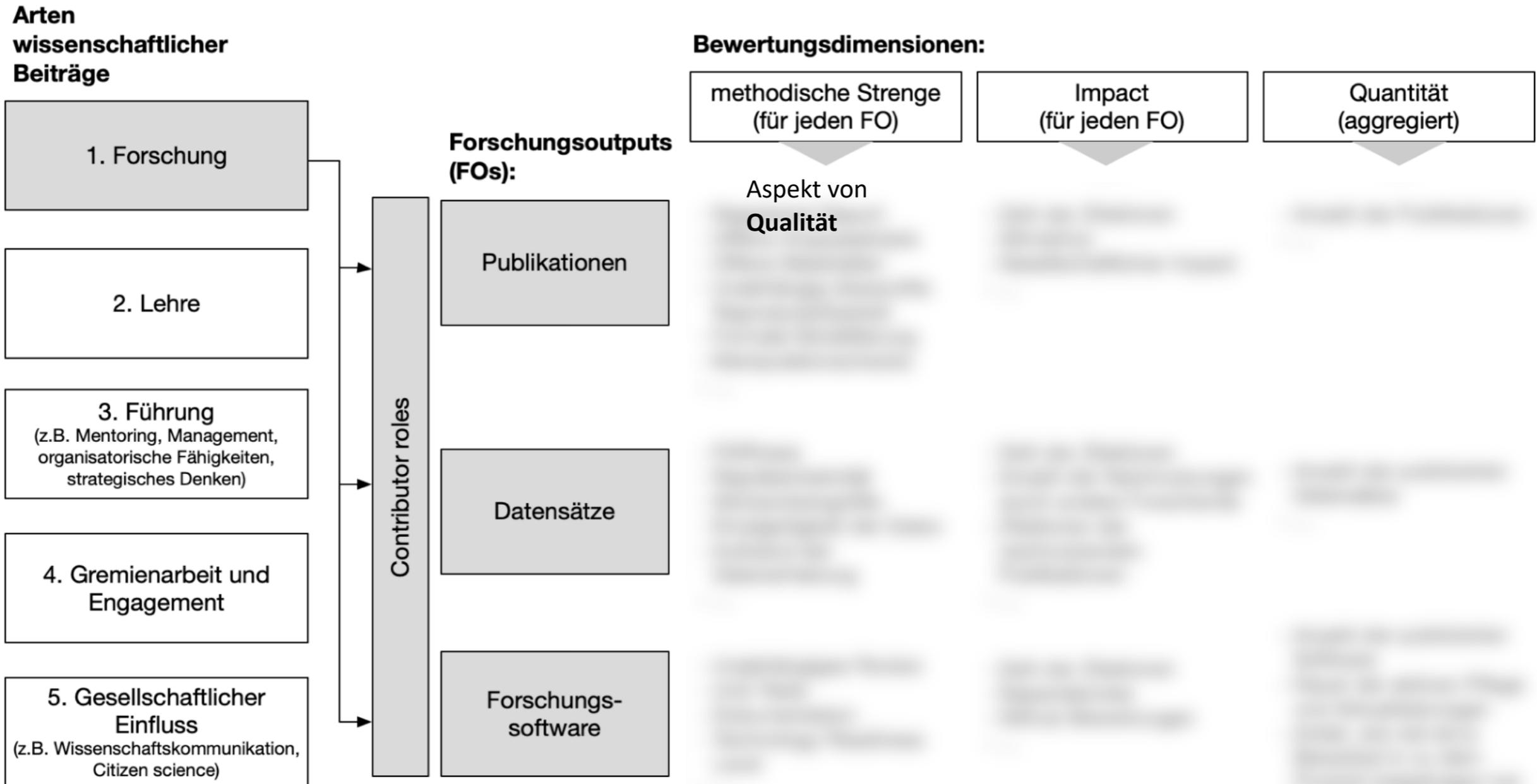
**+ Rejoinder (coming spring 2024)**

siehe Gärtner, A., Leising, D., & Schönbrodt, F. D. (2023). Empfehlungen zur Bewertung wissenschaftlicher Leistungen bei Berufungsverfahren in der Psychologie. *Psychologische Rundschau*, 74(3), 166–174. <https://doi.org/10.1026/0033-3042/a000630>; sowie <https://doi.org/10.31234/osf.io/rgh5b> und <https://doi.org/10.31234/osf.io/5yexm>

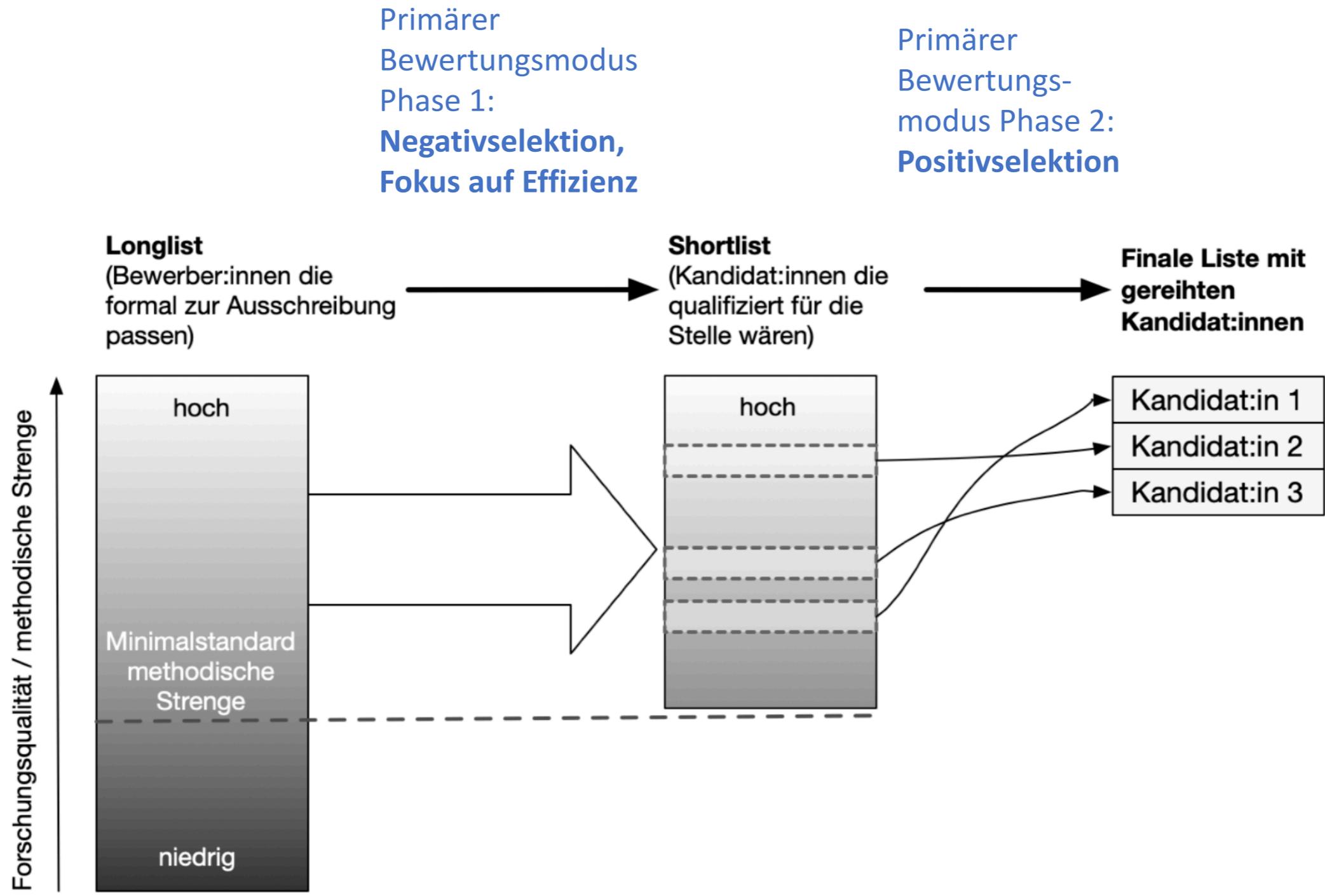
# Quality over quantity

- „Quality“ is multidimensional:
  - basic aspects (methodological rigor)
  - elusive and complex aspects (innovation, creativity, ingenuity)
- Rigor - as one part of quality - can be measured (quite) objectively:  
Whether research has been **skillfully executed according to standards of good scientific practice within the field.**
- Quality cannot be *reduced* to rigor!
  - Rigor is *not a sufficient* condition for high-quality research –  
but it can be seen as a *necessary* condition for valid knowledge.

# Research Quality Evaluation Scheme (RESQUE)



# RESQUE: Zwei Selektionsphasen



siehe Gärtner, A., Leising, D., & Schönbrodt, F. D. (2023). Empfehlungen zur Bewertung wissenschaftlicher Leistungen bei Berufungsverfahren in der Psychologie. *Psychologische Rundschau*, 74(3), 166–174. <https://doi.org/10.1026/0033-3042/a000630>; sowie <https://doi.org/10.31234/osf.io/rgh5b> und <https://doi.org/10.31234/osf.io/5yexm>

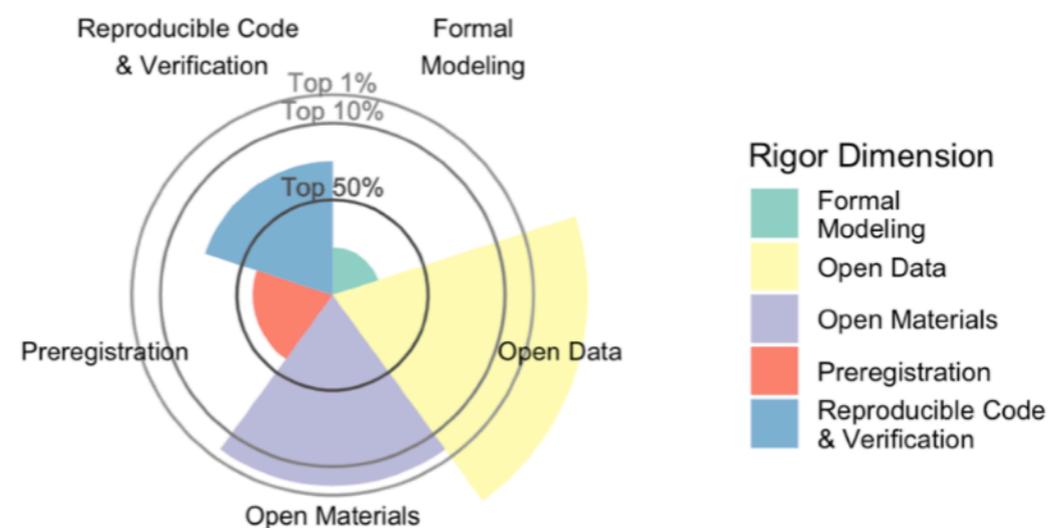
# Scoring-Prinzipien in Phase 1

- Grundprinzipien:
  - Es gibt viele Arten, Wissenschaft zu machen. Aber was man macht, sollte man gut machen.
    - Bsp: Man muss nicht theoriegeleitet arbeiten - aber wenn man es tut, sollte es gut sein.
  - In Phase 1 messen wir „fast and frugal indicators“, die möglichst objektiv erfassbar sind.
  - Das Fokus liegt auf basalen *Hygienefaktoren*. (Wir versuchen erst gar nicht, Innovation etc. zu messen).
  - Man sollte nicht bestraft werden, wenn man Punkte prinzipiell nicht erhalten kann
- Relative Rigor Score
  - „POMP“: Percentage of maximum points
  - Wenn ein Indikator begründbar „not applicable“ ist, dann wird der Maximalscore entsprechend reduziert und man kann immer noch theoretisch 100% erreichen.

# RESQUE Tool: Profil für Bewerber:innen

## Contributorship profile (CRediT roles)

Based on 10 submitted publications, this is the self-reported contributorship profile:



## Research impact: Highly popular publications

TITLE	3 YEAR CITATION COUNT	OVERALL CITATION COUNT	POPULARITY	CANDIDATES' CREDIT INVOLVEMENT	CANDIDATES' CREDIT MAIN ROLES
🔥 Bayes factor design analysis: Planning for compelling evidence	63	300	Top 0.1%	Very High	<b>Lead:</b> Conceptualization, Formal Analysis, Visualization, Writing: Original draft, Writing: Review & Editing <b>Equal:</b> Methodology
🔥 Replicability, Robustness, and Reproducibility in Psychological Science	105	105	Top 1%	Medium	<b>Equal:</b> Data Curation, Formal Analysis

# Demo:

<https://nicebread.github.io/RESQUE/web/>

[Note: The RESQUE app is in beta stage and might change in the near future. If you want to use it in practice, please [contact](#) us.]

**RESQUE**

10 of 10 slots used, you can't add more publications.

**Author / Metadata**  
Schönbrodt

**Publication 1**  
Machine Learning and Risk Assessment: Random Forest Does Not Outperform Logistic Regression in the Prediction of Sexual Recidivism

**Publication 2**  
Big little lies: a compendium and simulation of  $p$ -hacking strategies

**Publication 3**  
Replicability, Robustness, and Reproducibility in Psychological

**Author / Metadata**

**Last name of applicant**

**First name**

**Year of PhD / Dr.**

**Years that are subtracted from active academic life (e.g. due to child care). You can enter fractional years, e.g. 0.5.**

**ORCID identifier (as full link, e.g. <https://orcid.org/0000-0002-1825-0097>)**

**Who enters the data?**

I am the applicant & a co-author of these research outputs

# Arbeitsauftrag

- Reicht dieses core-set an Indikatoren aus, um wesentliche Qualitätsmerkmale in Ihrem Fachgebiet messbar zu machen?
  - Welche Indikatoren passen vielleicht nicht so gut?
  - Welche Indikatoren fehlen?
- Möglichkeit, ein fachgruppenspezifisches Erweiterungsset an Indikatoren zu entwickeln und in die RESQUE-Plattform einzubinden.
  - Falls ja: Wenn möglich bis 30.6.2024.
  - Diskussion der Ergebnisse auf dem DGPs-Kongress in Wien

# RESQUE: Umsetzung vor Ort

- Diskussion vor Ort initiieren:
  - Wie wird wissenschaftliche Leistung gemessen? In ...
    - Leistungsorientierter Mittelvergabe (LOM)
    - Tenure-track-Verfahren und Habilitationen
    - Berufungskommissionen
  - Kritische Reflektion über (möglicherweise) bisher genutzte Metriken (z.B. JIF, *h*-Index, Anzahl Publikationen)
  - → **Wir bieten Impulsvorträge an, um diese Diskussion zu starten**
- „Menü-Ansatz“: Die Elemente von RESQUE auswählen, die vor Ort passen; eigene Gewichtung festlegen
- Umsetzung in Berufungskommissionen
  - Zeit sparen durch technische Umsetzung: Eingabemaske für Bewerber:innen + übersichtliches und vergleichendes Profil
  - → **Wir unterstützen gerne bei der Umsetzung & Evaluation**

„Wir“ = DGPs-Kommission Open Science (Kontakt: [felix.schoenbrodt@psy.lmu.de](mailto:felix.schoenbrodt@psy.lmu.de)), Anne Gärtner, Daniel Leising