

Lost in the Design Space? Construct Validity in the Microfinance Literature

Lise Masselus^{1 2}, Christina Petrik², Jörg Ankel-Peters^{1 2}

¹RWI Essen, ²University of Passau

Leibniz Open Science Day 2025

October 27, 2025

Motivation

- Randomized controlled trials (RCTs) are valued due to their high internal validity (Angrist & Pischke, 2010)
- Generalizability of an RCT's results is potentially hampered by external validity (Findley et al., 2021; Pritchett & Sandefur, 2015, Peters et al., 2018; Vivalt, 2020) and **construct validity** (Esterling et al., 2023)

Motivation

- Randomized controlled trials (RCTs) are valued due to their high internal validity (Angrist & Pischke, 2010)
- Generalizability of an RCT's results is potentially hampered by external validity (Findley et al., 2021; Pritchett & Sandefur, 2015, Peters et al., 2018; Vivalt, 2020) and **construct validity** (Esterling et al., 2023)
- **Construct validity**: Link between the operationalization and the theoretical construct of interest
 - ▶ Construct validity of the **cause**
 - ▶ Construct validity of the outcome

Motivation

- Randomized controlled trials (RCTs) are valued due to their high internal validity (Angrist & Pischke, 2010)
- Generalizability of an RCT's results is potentially hampered by external validity (Findley et al., 2021; Pritchett & Sandefur, 2015, Peters et al., 2018; Vivalt, 2020) and **construct validity** (Esterling et al., 2023)
- **Construct validity**: Link between the operationalization and the theoretical construct of interest
 - ▶ Construct validity of the **cause**
 - ▶ Construct validity of the outcome
- Without construct (and external) validity, the implications of an RCT's results for theory or policy are unclear, even if results are internally valid

Contribution

- Generalizability

- ▶ External validity concerns are on the agenda (Banerjee et al., 2017; Peters et al., 2018; Ogden, 2020; Reidpath et al., 2022)
- ▶ Replicability and scalability problems of RCTs (Al-Ubaydli et al., 2017; Ankel-Peters et al., 2023; Camerer et al., 2016; Maniadis et al., 2017)
- ▶ Nascent literature on construct validity concerns (Esterling et al., 2023; Nadel & Pritchett, 2016; Pritchett & Sandefur, 2015)

- Microfinance

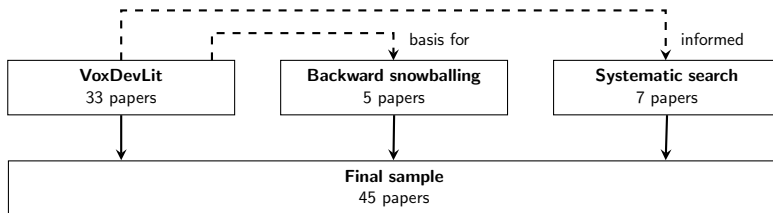
- ▶ Better-managed MFIs in RCTs and researcher involvement (Alcott, 2015; Bédécarrats et al., 2020)
- ▶ Interventions (beyond MF) implemented by NGOs and academics tend to find larger treatment effects (Vivalt, 2020)
- ▶ Borrowers' gender, loan size and interest rates are more predictive of heterogeneity in effect sizes across studies than study protocol differences (Meager, 2019)

Inclusion criteria:

1. Randomized controlled trial with MF loan treatment
2. Lower- or Middle-Income Country
3. Study at least one welfare impact (beyond take-up)
4. Journals or working paper series from economics

Inclusion criteria:

1. Randomized controlled trial with MF loan treatment
2. Lower- or Middle-Income Country
3. Study at least one welfare impact (beyond take-up)
4. Journals or working paper series from economics



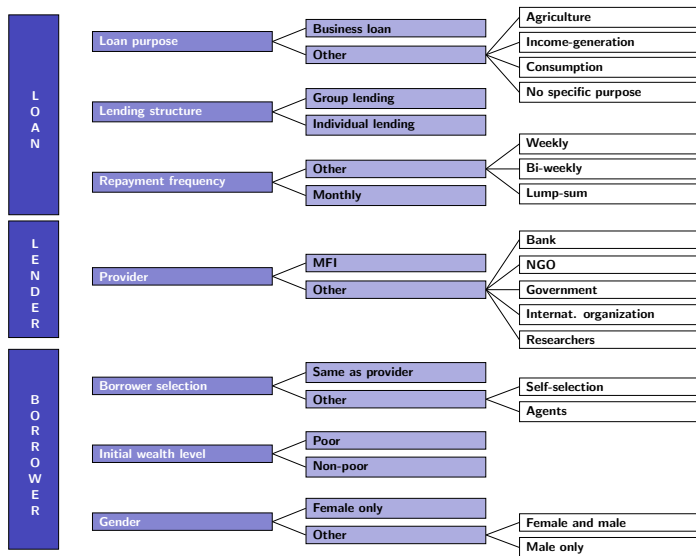
► Detailed selection

► Sample descriptives

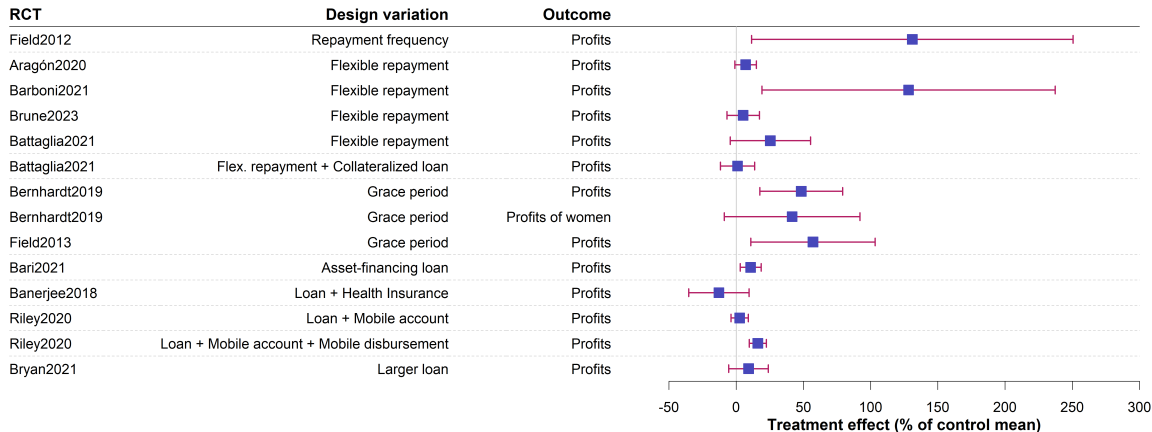
Mapping the design space

- Eliciting design elements
 - ▶ Design element is feature that is **randomly varied** in the RCT
 - ▶ Design element is feature that is referred to in **title or abstract**
- This systematic approach yields 21 design elements that can take on different dimensions
- Conservatively coding all design elements as binary yields over **two million potential combinations**

Mapping the design space (small selection)



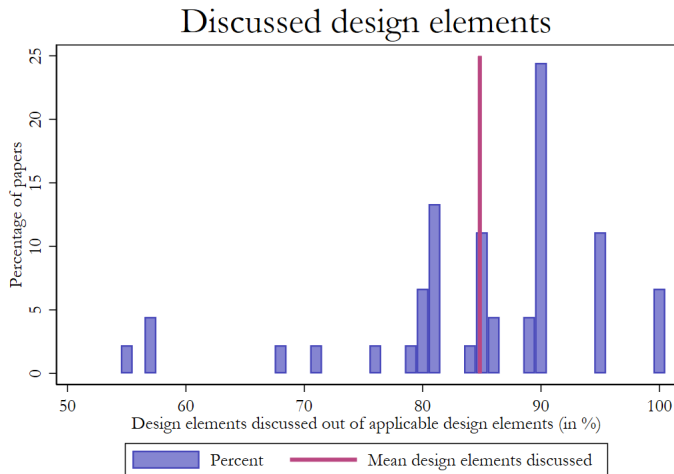
Design variations can matter



► HH income

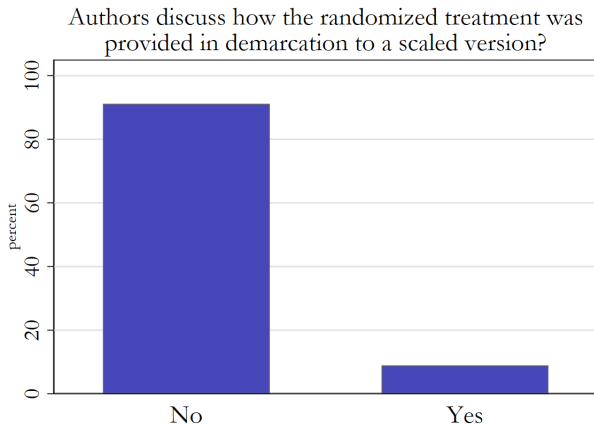
Assessing construct validity

- Complete information on all design elements from 7% of papers
- On average, 85% of the applicable design elements are discussed



Researcher special care

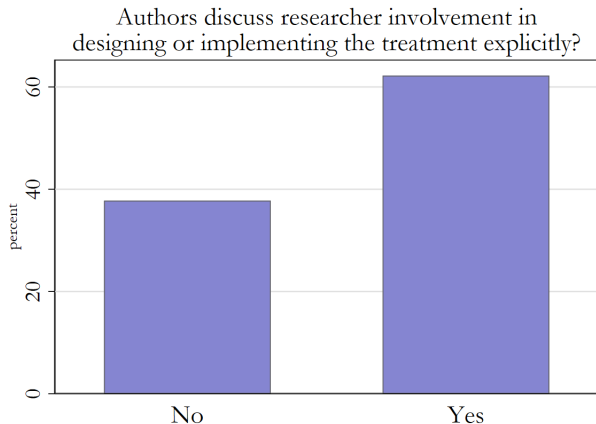
- Replicated or scaled version treatment may be provided differently than in RCT due to researcher presence
- Only 9% of papers discuss the provision of the treatment in the RCT compared to a replicated or scaled version of the treatment



► Extraction protocol

Researcher special care

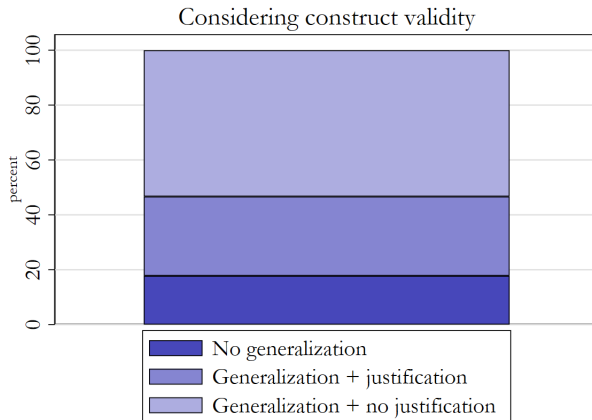
- Researcher involvement or non-involvement in the design or implementation of the treatment is discussed in 62% of papers
- Only 26% of papers discuss both



► Extraction protocol

Construct validity is not on the agenda

- 18% of papers do not generalize beyond the operationalized treatment
- 29% of papers generalize, but provide assumptions/justifications
- 53% of papers generalize and do not provide assumptions/justifications, which **disregards construct validity**



Conclusion

- Greater awareness of construct validity is needed for inference and policy-oriented learning
- Using microfinance RCTs as an example, we demonstrate the large design space and that design variations can be consequential for treatment effects
- Yet, reporting on all design elements and researcher special care is not done at large and generalizing beyond the operationalized treatment is common
- Three practical responses:
 - ▶ **Transparent and explicit reporting:** design elements, researcher special care, and assumptions for generalizations
 - ▶ **Epistemic humility:** limit generalizations and emphasize local learning
 - ▶ **Better conceptual understanding and linking:** framing RCTs as conceptual replications, coordinated replications, construct-sensitive meta-analyses, and using qualitative insights and expertise more systematically

▶ Checklist

Appendix

Appendix: Epistemic approaches to construct validity

- **Deductive approach**

- ▶ Deducing generalized causal claims based on a single RCT (Cartwright, 2010, 2011; Esterling et al. 2023)
- ▶ Make explicit assumptions about construct validity to draw generalized causal claims

- **"Pointillist painting" approach**

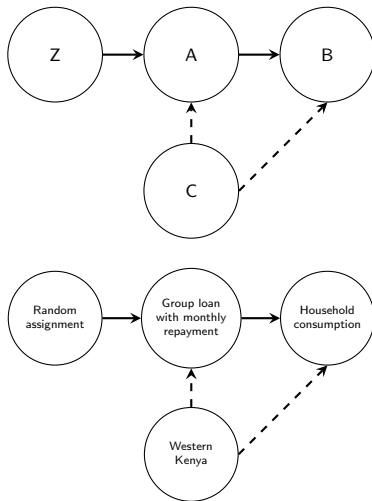
- ▶ Inducing generalized causal claims based on multiple RCTs
- ▶ External and construct validity ensue automatically from carrying out enough RCTs (Banerjee & Duflo, 2009; Duflo, 2020)

- **"Local tinkering" approach**

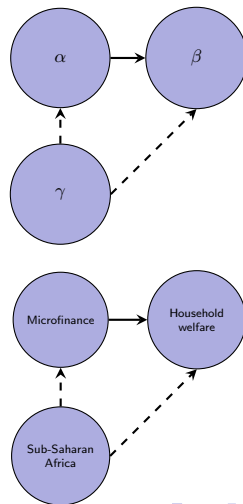
- ▶ RCTs to go from proof of concept to a scalable policy (Banerjee et al., 2017; Samii, 2020)
- ▶ Approach does not strive for generalized causal claims

Conceptual framework

Operational level

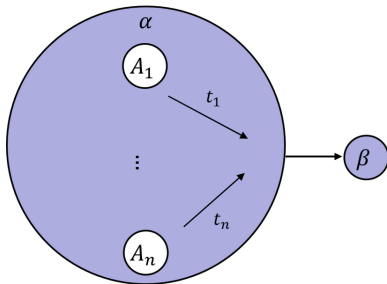


Conceptual level

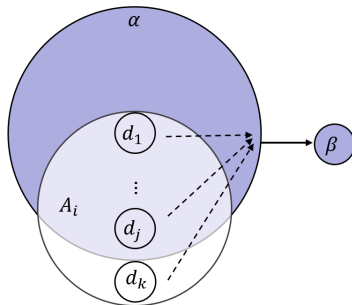


Threats to construct validity

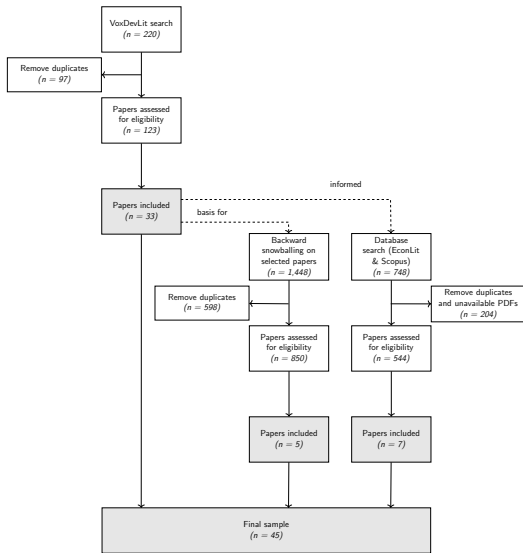
Multiple operationalizations



Researcher special care



Appendix: Sample selection



Appendix: Sample descriptives

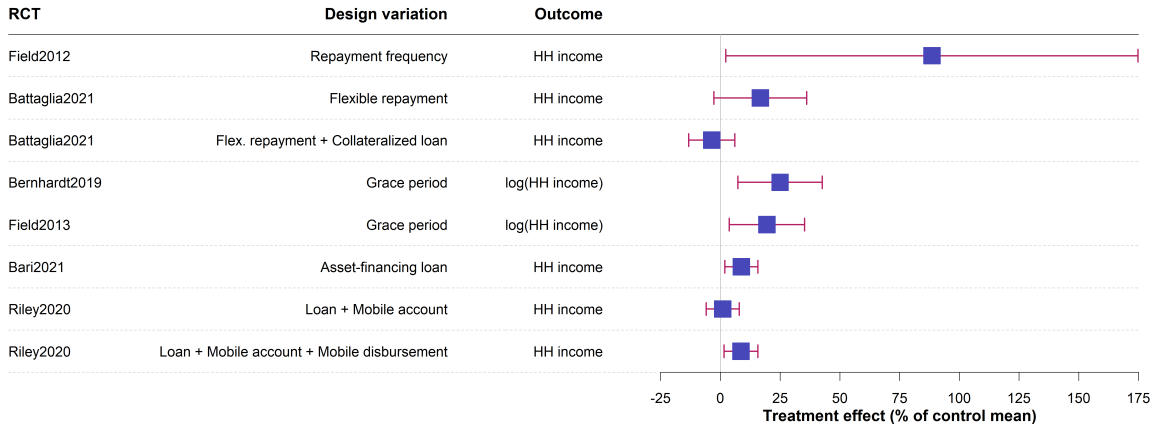
Publication status	Frequency	Percent
Working paper	9	20
Published paper	36	80
Total	45	100

Journals	Frequency	Percent
American Economic Journal: Applied Economics	7	19.44
American Economic Review	4	11.11
American Economic Review: Insights	1	2.78
American Journal of Agricultural Economics	1	2.78
Econometrica	2	5.56
Economica	1	2.78
Economic Development and Cultural Change	1	2.78
Economic Journal	1	2.78
Journal of Development Economics	5	13.89
Journal of Policy Analysis and Management	1	2.78
Journal of the European Economic Association	1	2.78
PLoS ONE	1	2.78
Review of Development Economics	1	2.78
Review of Economic Studies	3	8.33
Science	1	3.45
Springer Briefs in Economics	1	2.78
The Quarterly Journal of Economics	1	2.78
The Review of Financial Studies	1	2.78
World Development	2	5.56
Total	36	100

Appendix: Sample descriptives

Country where the study took place	Frequency	Percent
Bangladesh	5	11.11
Bosnia and Herzegovina	1	2.22
China	1	2.22
Colombia	1	2.22
Egypt	2	4.44
Ethiopia	2	4.44
India	13	28.89
Indonesia	1	2.22
Kenya	3	6.67
Mali	1	2.22
Mexico	1	2.22
Mongolia	1	2.22
Morocco	1	2.22
Pakistan	4	8.89
Philippines	2	4.44
South Africa	1	2.22
Tanzania	2	4.44
Uganda	2	4.44
Zambia	1	2.22
Total	45	100

Appendix: Design variation effects - HH income



- ◀ Design variations can matter

Appendix: Extraction protocol - Special care

1. Does the paper mention how the randomized treatment was provided in demarcation to a scaled version of the intervention?

- the program effectiveness observed in RCTs could be higher than when the evaluated program is implemented outside an RCT $\Rightarrow$ e.g., because the researchers were involved in the treatment provision or because organizations that collaborate with researchers differ from organizations that do not participate in an RCT
- Yes, if
 - ▶ the paper makes a statement on how the treatment compares to a replicated or scaled version of the program or to a potential real-world construct (regardless of personal judgement) $\Rightarrow$ statements should refer to the nature of the treatment, so what happens on the provision-side
 - ▶ the paper makes a statement about the specific sample problem, if the demarcation between the implemented treatment and a scaled treatment is made explicit
- No, if
 - ▶ the statement is about Hawthorne effects, John Henry effects and general equilibrium effects (external validity)
 - ▶ the paper only makes a cost-benefit analysis

◀ Researcher involvement

Appendix: Extraction protocol - Researcher involvement

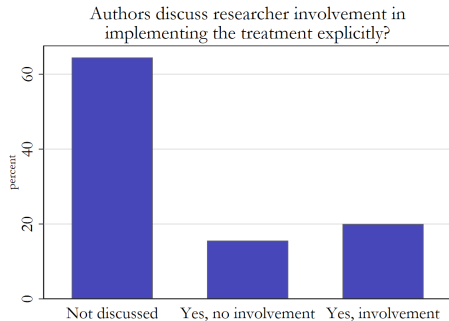
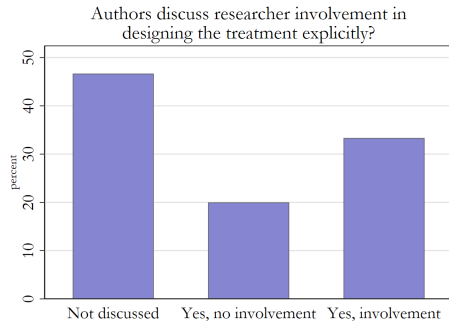
2a. Does the paper discuss researcher involvement in the design of the randomized treatment?

- Yes, researcher involvement
 - ▶ as soon as it is recognized that the researchers were involved in at least one component of the design (or one of the treatment arms)
- Yes, no researcher involvement
 - ▶ explicit statement that the partner organization designed the treatment
 - ▶ authors explicitly state that the intervention was already existing before the RCT was implemented

2b. Does the paper discuss researcher involvement in the implementation of the randomized treatment?

- Yes, researcher involvement
 - ▶ explicitly stated that the researchers were involved in at least one component of the implementation of the treatment (or one of the treatment arms)
- Yes, no researcher involvement
 - ▶ explicit statement that the researchers were not involved
 - ▶ explicitly highlight that the partner organization was responsible for the implementation of the treatment (excluding statements that are only made about subcomponents)

Appendix: Researcher involvement in design and implementation



◀ Researcher involvement

Appendix: Extraction protocol - Generalization

Does the paper generalize to a broader construct?

- to capture whether papers are cautious in drawing inference beyond their operationalized treatment
- Yes, if
 - ▶ authors draw generalized conclusions or recommendations towards microfinance providers, other researchers or policy makers, beyond the operationalized treatment (e.g., “MFIs should/should not”, “our results suggest that policy makers should”)
 - ▶ general statements about the effectiveness of a construct that is broader than the operationalized treatment
- No, if
 - ▶ the authors do not generalize or make a policy recommendation at all or not beyond the operationalized treatment

◀ Generalization

Appendix: Extraction protocol - Construct validity discussion

Does the paper discuss potential limitations of generalizability to a broader construct?

- capture whether papers show some sensitivity towards construct validity by mentioning potential limitations to the generalizability to a broader construct, beyond the operationalized treatment
- Yes, if
 - ▶ the authors discuss some limitations (regardless of personal judgement)
 - ▶ the authors explicitly state that the results have to be interpreted in light of the intervention that was studied, or if authors label their intervention as a “proof-of-concept” or “exploratory study” for a particular MF intervention
 - ▶ the authors say that they do not want to draw any generalizable conclusion or make policy recommendations
- No, if
 - ▶ the authors do not discuss limitations
 - ▶ the authors discuss limitations with respect to external validity only

◀ Construct validity discussion

Appendix: Check list I

- **Construct:**

- ▶ Does the paper provide comprehensive information about the details of the intervention, that would allow other researchers/implementers to implement an intervention that is sufficiently similar to the intervention under evaluation? (Authors can be guided by the design elements that are already studied in the literature.)

- **Researcher special care:**

- ▶ Does the paper discuss how the randomized treatment was provided in demarcation to a scaled version of the intervention?
- ▶ Does the paper describe to what extent the researchers were involved in designing the intervention?
- ▶ Does the paper describe who implemented the intervention, and to what extent the researchers were involved in the implementation?

Appendix: Check list II

- **Hawthorne and John Henry effects:**

- ▶ Does the paper describe whether the participants in the treatment and control group are aware of being part of an experiment?

- **Generalizability:**

- ▶ Does the paper determine the scope of generalization regarding the intervention design that is being studied?
- ▶ Does the paper lay out the assumptions on which base this generalization would be valid?

◀ Conclusion

References

- Allcott, H. (2015). Site selection bias in program evaluation. *The Quarterly Journal of Economics*, 130(3), 1117–1165.
- Al-Ubaydli, O., List, J. A., & Suskind, D. L. (2017). What can we learn from experiments? Understanding the threats to the scalability of experimental results. *American Economic Review Papers and Proceedings*, 107(5), 282–86.
- Angrist, J. D., & Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24(2), 3–30.
- Ankel-Peters, J., Fiala, N., & Neubauer, F. (2023). Do economists replicate? *Journal of Economic Behavior & Organization*, 212, 219–232.
- Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., & Walton, M. (2017). From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives*, 31(4), 73–102.
- Bédécarrats, F., Guérin, I., & Roubaud, F. (2020). Microcredit RCTs in development: Miracle or mirage? *Oxford Academic*.
- Camerer, C.F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., & Wu, H. (2016). Evaluating Replicability of Laboratory Experiments in Economics. *Science* 351, 1433–1436.
- Esterling, K., Brady, D., & Schwitzgebel, E. (2023). The necessity of construct and external validity for generalized causal claims. Institute for Replication (I4R), s.l., No. 18.
- Findley, M. G., Kikuta, K., & Denly, M. (2021). External validity. *Annual Review of Political Science*, 24(1), 365–393.
- Maniadis, Z., Tufano, F., & List, J. A. (2017). To replicate or not to replicate? Exploring reproducibility in economics through the lens of a model and a pilot study. *American Economic Review*, 107(5), 282–86.
- Maniadis, Z., Tufano, F., & List, J. A. (2014). One swallow doesn't make a summer: New evidence on anchoring effects. *American Economic Review*, 104(1), 277–90.

- Meager, R. (2019). Understanding the average impact of microcredit expansions: A bayesian hierarchical analysis of seven randomized experiments. *American Economic Journal: Applied Economics*, 11(1), 57–91.
- Nadel, S., & Pritchett, L. (2016). Searching for the devil in the details: Learning about development program design. *SSRN Electronic Journal*.
- Ogden, T. (2020). RCTs in development economics, their critics and their evolution. *Randomized Control Trials in the Field of Development: A Critical Perspective*, 126-151.
- Peters, J., Langbein, J., & Roberts, G. (2018). Generalization in the tropics: Development policy, randomized controlled trials, and external validity. *The World Bank Research Observer*, 33(1), 34–64.
- Pritchett, L., & Sandefur, J. (2015). Learning from experiments when context matters. *American Economic Review*, 105(5), 471–475.
- Reidpath, D. D., Allotey, P., Barker, S. F., Clasen, T., French, M., Leder, K., Ramirez-Lovering, D., Rhule, E. L. M., & Siri, J. (2022). Implementing “from here to there”: A case study of conceptual and practical challenges in implementation science. *Social Science & Medicine*, 301, 114959.
- Vivalt, E. (2020). How much can we generalize from impact evaluations? *Journal of the European Economic Association*, 18(6), 3045–3089.