



Leibniz Institute for  
Economic Research

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

September 13, 2024

MAER-Net Colloquium

Lise Masselus, RWI & U Passau

Christina Petrik, U Passau

Jörg Ankel-Peters, RWI & U Passau

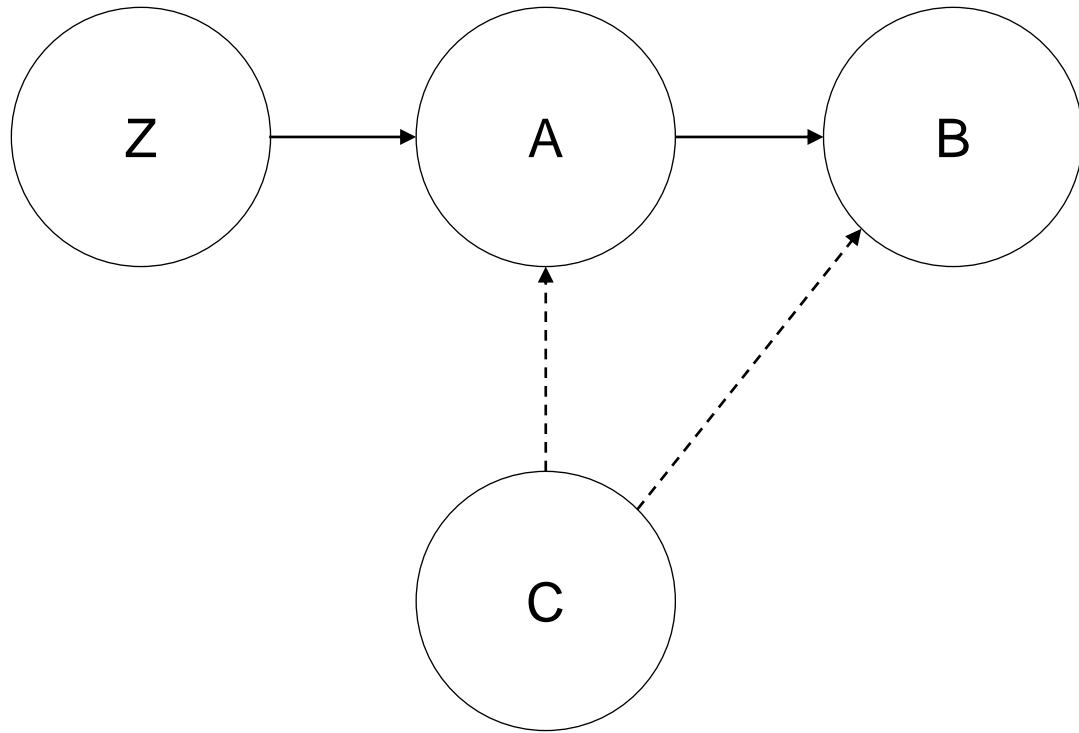


# Motivation: beyond internal validity

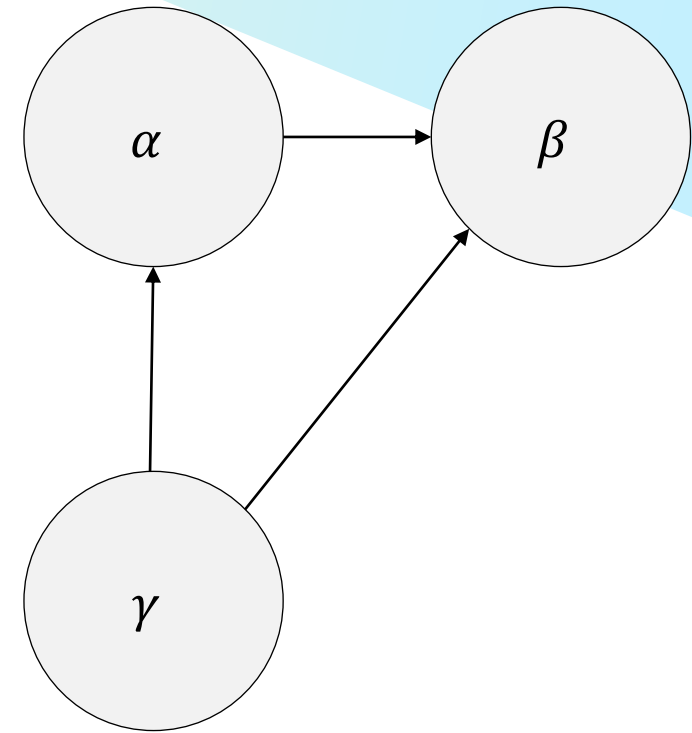
- ▶ Randomized Controlled Trials (RCTs) gained a prominent role in the credibility revolution due to their high internal validity (Angrist & Pischke, 2010)
- ▶ Critics argue that this high internal validity often comes at the expense of limited generalizability (e.g. Cartwright, 2002; Deaton & Cartwright, 2018)
  - ▶ Often results do not scale beyond a particular setting (Bold et al., 2018; Gechter, 2024; Kerwin & Thornton, 2021; List, 2024; Usmani et al., 2022)
- ▶ This is due to low external validity and low construct validity
- ▶ Construct validity: deals with how the operationalization of a treatment corresponds to the broader construct it intends to speak to (Cook & Campbell, 1979; Esterling et al., 2023; Findley et al., 2021; Nadel & Pritchett, 2016; Shadish et al., 2002)
  - ▶ Construct validity of the **cause** -> **our paper**
  - ▶ Construct validity of the outcome

# Conceptual Framework

Operational level

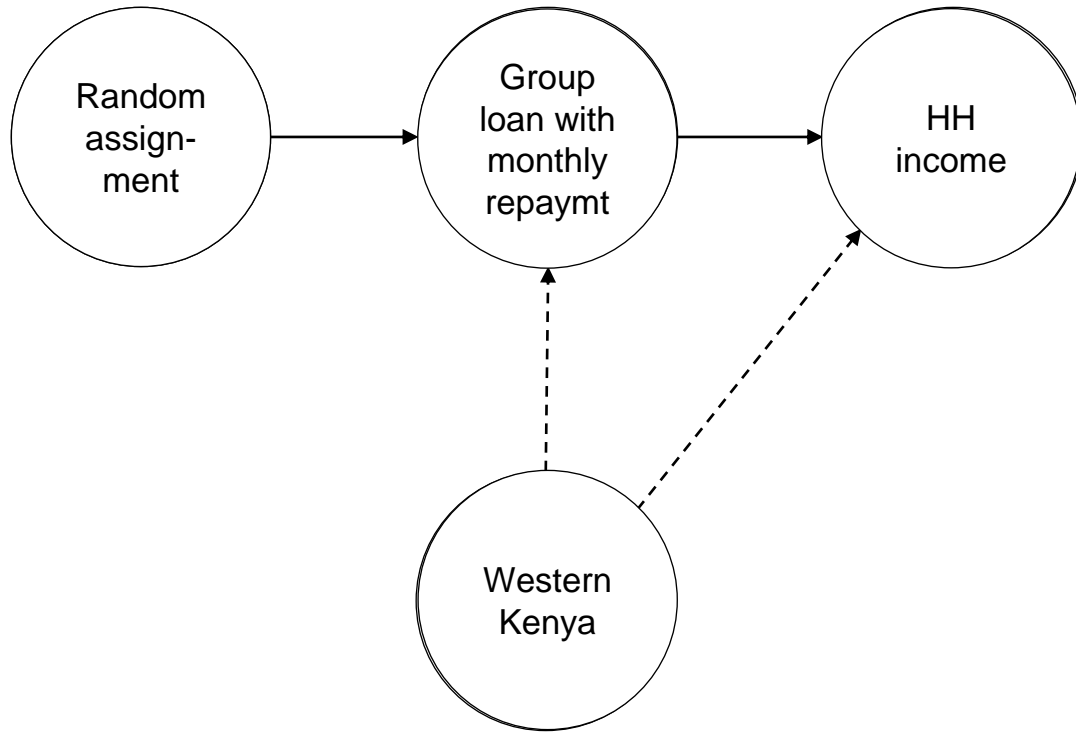


Level of inference

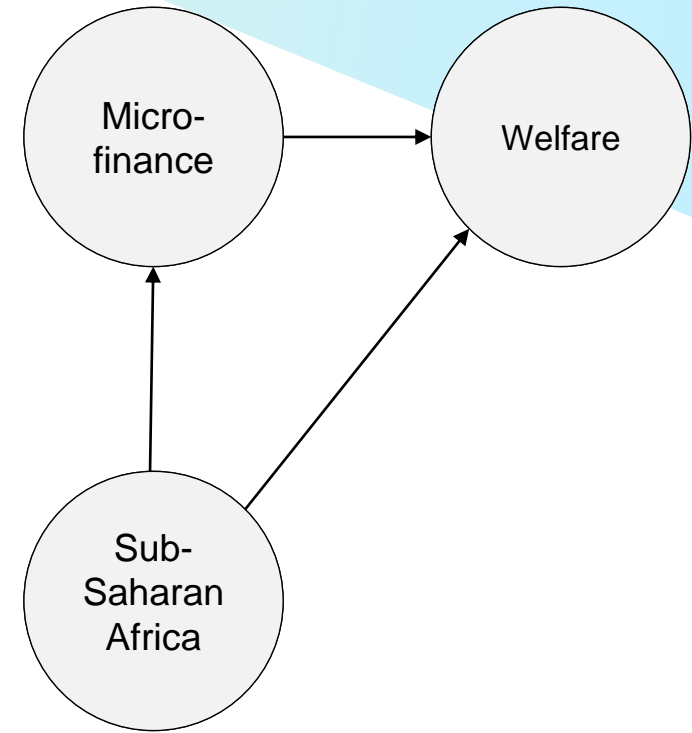


# Conceptual Framework

## Operational level



## Level of inference



# The pointillist painting and its link to construct validity

- ▶ **Prevailing epistemic model: the “pointillist painting”**
  - ▶ Inducing generalized causal claims based on multiple RCTs

## II. How Lessons Are Drawn: Microcredit

The first flaw in the strawman is its misunderstanding of how RCT advances science. RCT researchers do not come to sweeping conclusions about the potential impact of a program based on any single experiment. **Instead, each experiment is like a dot on a pointillist painting:** on its own it does not mean much, but the accumulation of experimental results eventually paints a picture that helps make sense of the world, and guide policy. It is the accretion of results that makes sense and justifies the whole enterprise.

- ▶ Assumption: external and construct validity ensue automatically from carrying out enough RCTs (Angrist & Pischke 2010; Banerjee & Duflo, 2009; Duflo, 2020)
- ▶ How many dots are needed to see the painting?
- ▶ **Key message: construct validity increases the number of required dots considerably, making the “pointillist painting” very hard to compile**

# What we do in this paper – short summary

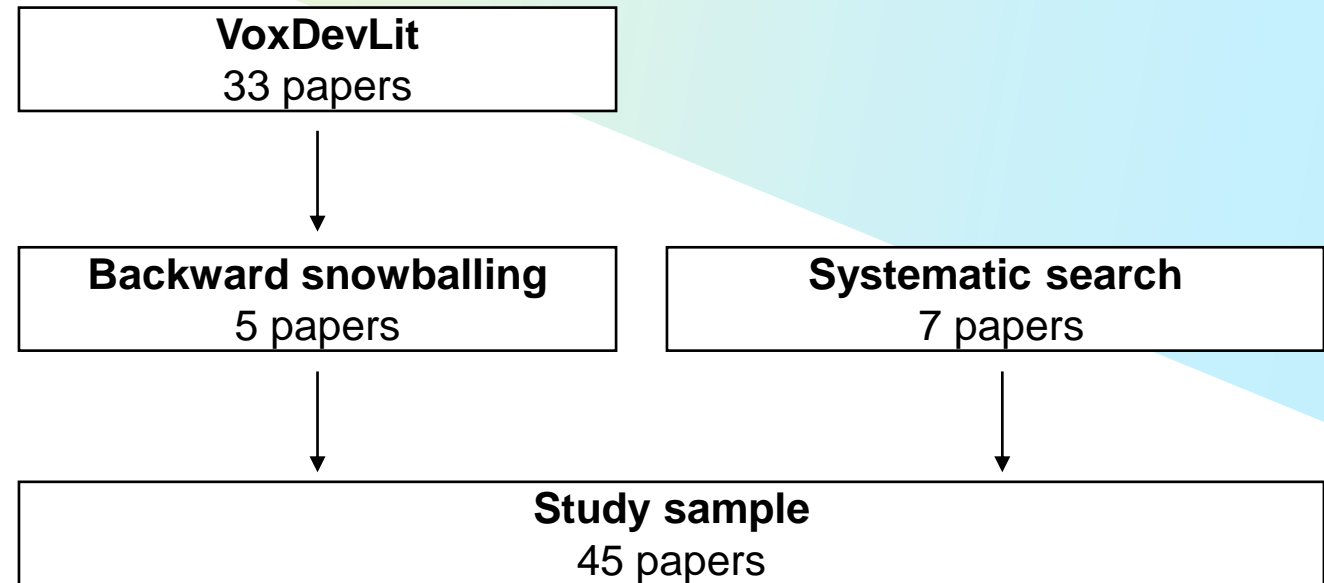
- ▶ Review the microfinance literature, focusing on RCTs
- ▶ ...to better understand, empirically, how important construct validity is creating generalizable and policy-relevant knowledge
- ▶ Based on 45 RCTs, we (conservatively) estimate the universe of potential operationalizations (“the design space”)
- ▶ We demonstrate that variations in treatment operationalization may matter for the observed impact
- ▶ **This makes the pointillist painting very hard to compile**
- ▶ We then show that there is little awareness for construct validity: One RCT operationalizes specific MF treatment(s), while the causal claim in the paper often refers to the monolithic, general construct MF.

# Why Microfinance?

- ▶ Flagship topic of RCT movement (Duflo, 2020)
- ▶ A special issue in AEJ:AE, bundling the first six RCTs, laid the cornerstone
  - ▶ “Collective spanning of contextual, social, and market environments, [lends] external validity to their results, especially given the consistent pattern of findings” (Banerjee et al., 2015, p. 2)
- ▶ Many RCTs have followed
- ▶ **Where if not here, should the pointillist painting take shape?**
- ▶ We build on previous work on generalizability of microfinance literature
  - ▶ Push-back on AEJ:AE and microfinance in general, mostly related to external validity (Morduch, 2020; Wydick, 2015; Allcott, 2015; Bédécarrats et al., 2020)
  - ▶ Bayesian Hierarchical Modelling to assess external validity (Meager, 2019; Vivaldi, 2020)
- ▶ We focus systematically on the operationalization of the treatment, in a larger sample
- ▶ MF as example, but we believe our framework and findings are transferable to other identification strategies and interventions

# Data

- ▶ Three sources
  - ▶ aim to cover the well-known and most influential papers
- ▶ Inclusion criteria
  - ▶ 1. RCT
  - ▶ 2. Microfinance intervention
  - ▶ 3. LMIC country
  - ▶ 4. Paper looks at welfare impact on borrowers
  - ▶ 5. Econ journals or WP series
- ▶ Two readers independently extracted key information from each paper



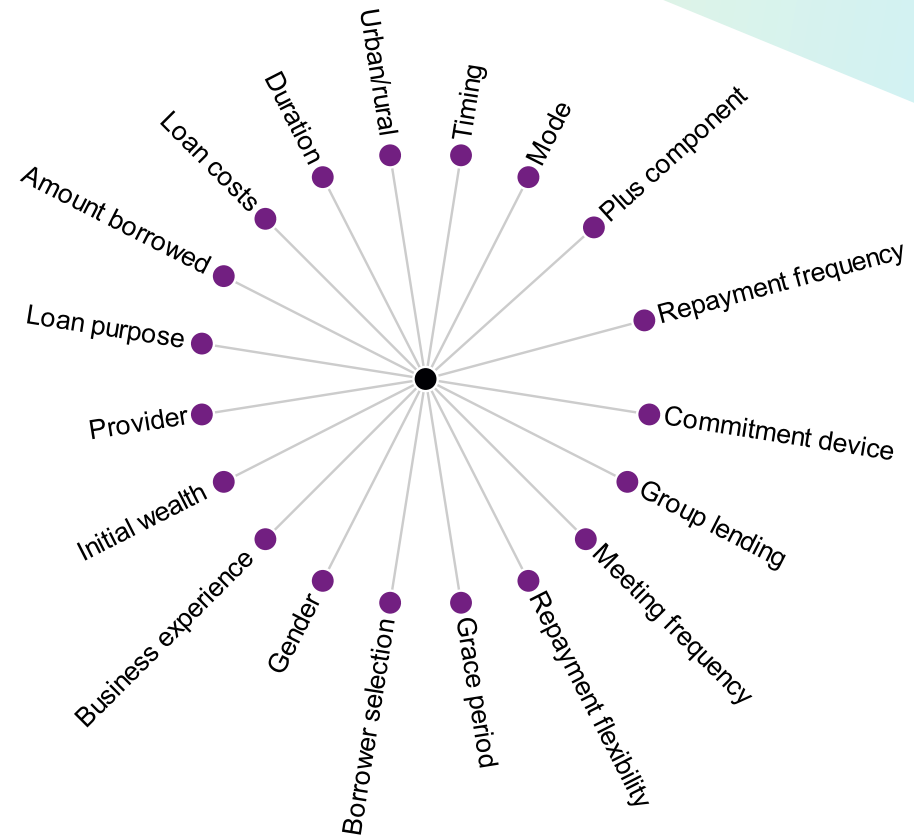


## RQ (1): Mapping the design space

- ▶ We use treatment variations and design features made salient by authors to conservatively map the design space
- ▶ Design space = the universe of potential operationalizations
- ▶ We extract key design element from each paper
  - ▶ 1. Design element is feature that is referred to in title or abstract
  - ▶ 2. Design element is feature that is randomly varied in study (25 papers)
- ▶ Together, these features form the *design space* of a microfinance intervention
  
- ▶ 45 papers result in **21 unique elements**

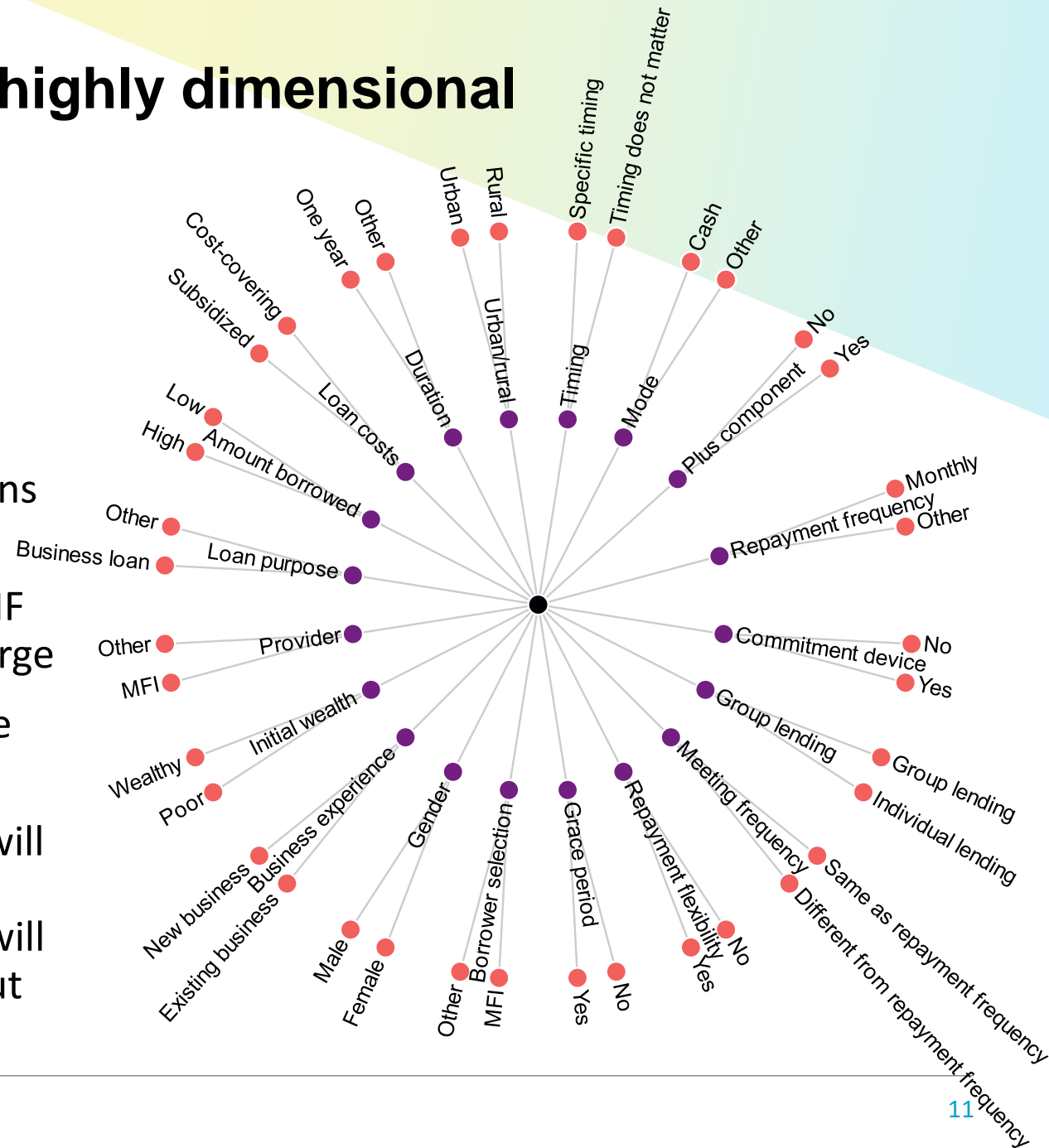
# The design space is highly dimensional

- ▶ We identify 21 unique design elements
- ▶ Each design element (purple dots)...



# The design space is highly dimensional

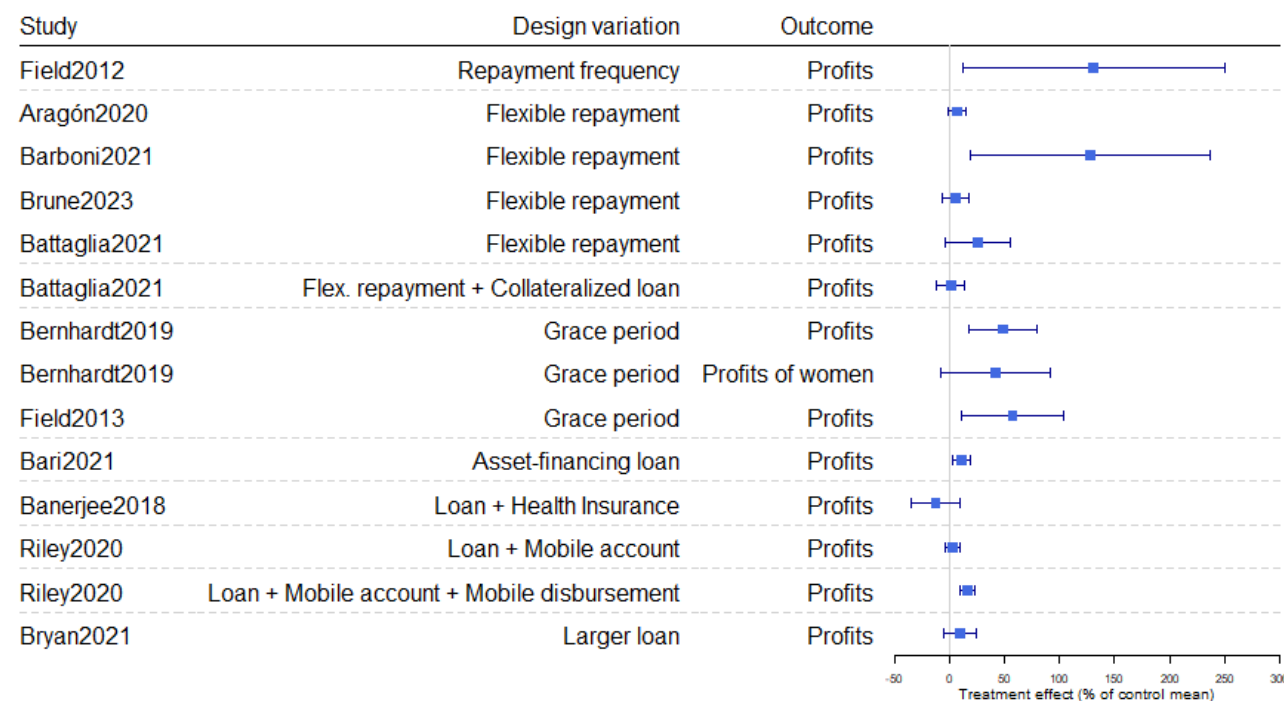
- ▶ We identify 21 unique design elements
- ▶ Each design element (purple dots)...
- ▶ ... consists of multiple dimensions (red dots)
- ▶ The potential total number of MF interventions easily becomes large
- ▶ But branches could easily diffuse more
- ▶ Evidently, not all combinations will matter, but generalization from one specific operationalization will need to make assumptions about stability of effects



## RQ (2): Do variations in the treatment operationalization matter?

- ▶ Comparisons across RCTs would confound difference in context with variations in treatment operationalization
- ▶ We look at papers that compare design variations within one RCT
- ▶ For household income and business profits: heterogeneity in the degree to which changing one design element matters
  - ▶ Around half of the estimated effects are statistically significant at the 5% level
  - ▶ Some papers reveal massive effects

### Design variation effects on business profits

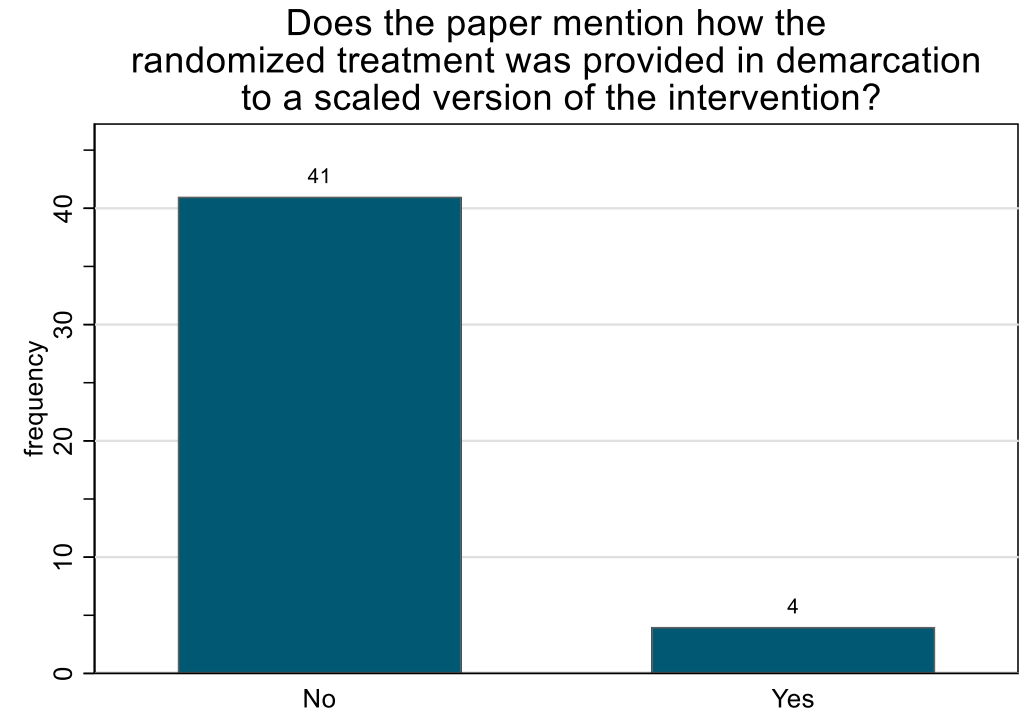


## RQ (3): Do papers mention researcher special care?

- ▶ We now assess each paper with respect to researcher special care as an important construct validity feature
- ▶ RCTs evaluate interventions that may not be implemented exactly the same in a non-controlled setting
- ▶ E.g. Researchers are involved in the treatment provision or the nature of organizations that collaborate with researchers differs from organizations that do not participate in an RCT
- ▶ Therefore, program effectiveness observed in RCTs could be different than when the evaluated program is implemented outside a RCT
- ▶ Researcher special care could be considered as a design element, and should be reflected in the broader construct
- ▶ Implemented by two proxy questions

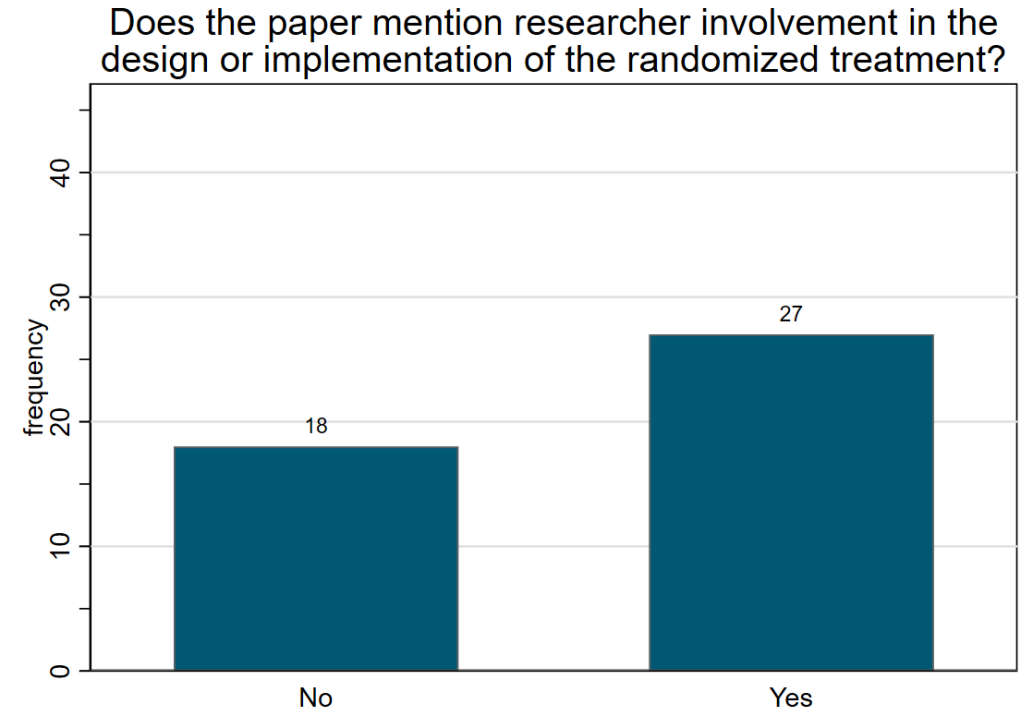
## RQ (3): Do papers mention researcher special care?

- ▶ Does the paper demarcate if the treatment provision differs from a scaled version of the treatment?
  - E.g. discussion of challenges that occur when intervention would scale up, statements about intervention without influence from researchers or statements about a version of the intervention implemented by another entity.
- ▶ 9% of papers mention how their treatment was provided in demarcation to a potential scaled version of the intervention.
- ▶ We coded this question as “yes” as soon as the paper makes a brief statement about whether the implementation in their RCTs might differ from the potential scaled version, irrespective of the discussion’s comprehensiveness or plausibility



## RQ (3): Do papers mention researcher special care?

- ▶ Researcher involvement in the delivery of the treatment determines how effectively the treatment is delivered and may influence treatment effects
- ▶ 60% of the papers mention researcher involvement in the design or implementation of the intervention
- ▶ 36% of the papers mention involvement in the implementation



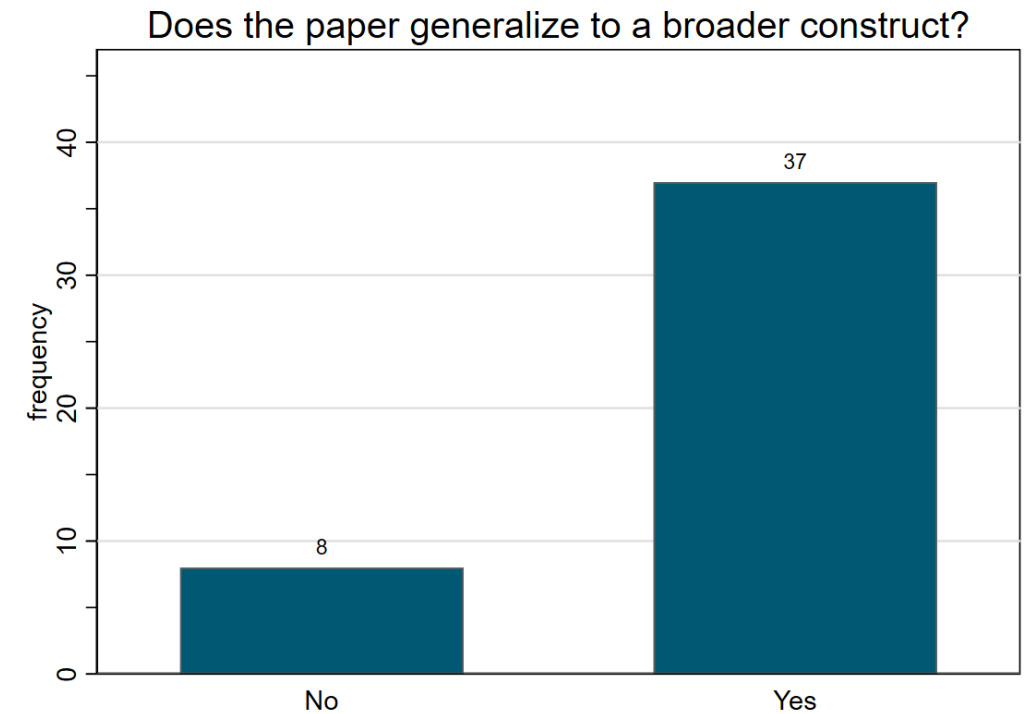
## RQ (4): Do papers generalize?

- ▶ **Construct validity hinges on the inference made and the evidence provided:** bolder claims decrease construct validity, more cautious claims increase construct validity
- ▶ We assess whether papers generalize and whether the inferential leaps are discussed
- ▶ Our ambition for this question is to capture whether papers are cautious in drawing inference beyond their operationalized treatment, and hence, showcase a sensitivity for construct validity
- ▶ Coding in favor of the paper



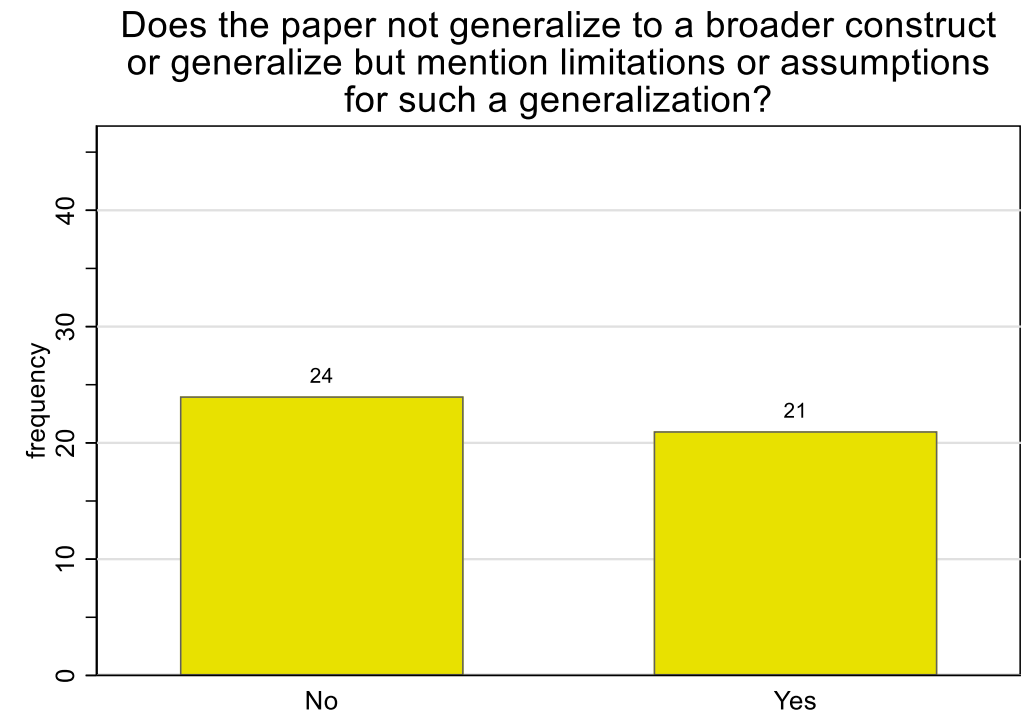
## RQ (4): Do papers generalize?

- ▶ 82% of papers generalize to a construct broader than the operationalized treatment



## RQ (4): Do papers generalize?

- ▶ 82% of papers generalize to a construct broader than the operationalized treatment
  - ▶ Among those papers, 65% does not mention potential limitations for the generalization
- ▶ 47% either do not generalize or mention limitation or assumptions for their generalization
- ▶ **So more than half of the papers in our sample do not show sensitivity for construct validity in how they report and interpret findings**
- ▶ **Problematic for how priors are formed among researchers, but also for experts and decision makers**



# Conclusion

- ▶ Main purpose: sensitize for the concept of construct validity
- ▶ For the case of microfinance, we have demonstrated that
  - ▶ 1. The number of potential operationalizations is very large. Each operationalization underrepresents the construct “microfinance”
  - ▶ 2. Design variations can be consequential for treatment effects
  - ▶ 3. there is not much sensitivity for the relevance of design features, such as researcher special care
  - ▶ **4. ...yet many papers generalize without sensitivity for construct validity**

# Conclusion

- ▶ The need to establish the different types of validity arises with the claims of the paper
  - ▶ Claim causality: establish internal validity
  - ▶ Claim generalizability to other populations: establish external validity
  - ▶ Claim generalizability to a broader construct: establish construct validity
- ▶ Implications:
  - ▶ More **transparent reporting** of treatment operationalization and construct
  - ▶ Focus more on **local learning**: less generalizing from individual papers automatically increases construct validity
  - ▶ For **global learning** more work needs to be done to improve the construct validity of literatures
    - ▶ Meta-analyses with sensitivity for external and construct validity can help
    - ▶ As well as coordinated replication effects (e.g. Metaketa)
    - ▶ But also: more foundation in theory and a broader methodological toolkit

# Thank you!

Please reach out: [Masselus@rwi-essen.de](mailto:Masselus@rwi-essen.de)

**Working paper:** Masselus, L., Petrik, C., & Ankel-Peters, J. (2024, August 13).  
Lost in the Design Space? Construct Validity in the Microfinance Literature.  
<https://doi.org/10.31219/osf.io/nwp8k>



