

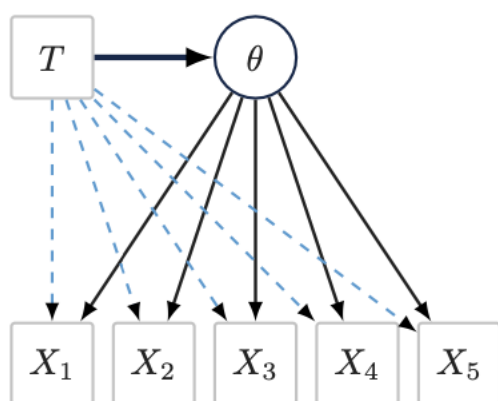
# How Sensitive Are Treatment Effects to the Specific Outcome Measure Used?

#2400864: Statistical Methods and Sensitivity Analyses: How the Choice of Outcome Measures Affects the Conclusions of STEM Education Research Studies (<https://ecrhub.org/ecr-projects?id=2400864>)

**Question:** Why does the outcome measure matter so much?

| Same Intervention       |             |
|-------------------------|-------------|
| Outcome measure         | Effect Size |
| Researcher developed    | 0.30 SD     |
| Independently developed | 0.05 SD     |

**Answer:** Item-specific effects (net of impact) do not generalize to other measures of the same construct that use different items



**Methods Developed**

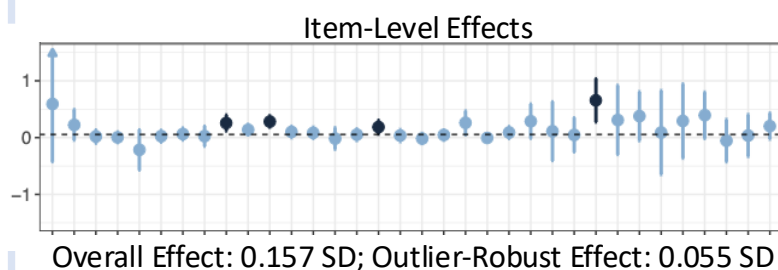
Item Response Theory

+

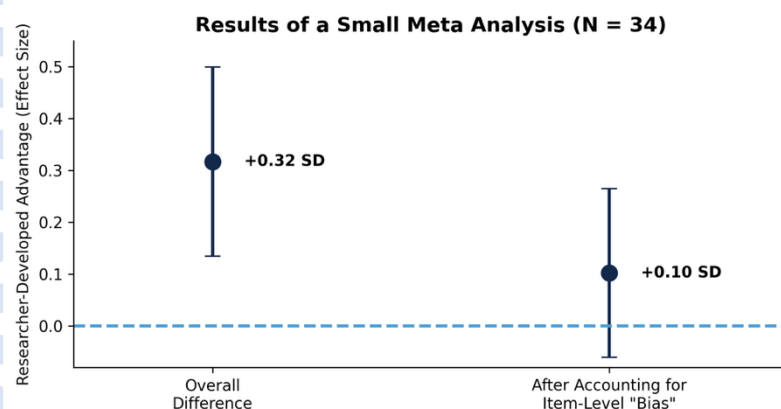
Program Evaluation

- Sensitivity analysis for treatment effects
- Robust estimator of impact
- Tests for item-level heterogeneity
- Open-source software

Reported effects can be driven by only a few items



Item-level “bias” accounts 68% of discrepancy between Researcher vs Independent measures



**Challenges Overcome**

Item-Level Data

- Data no longer available from NCES
- Used [itemresponsewarehouse.org](https://itemresponsewarehouse.org)
- But limited access to assessment materials

Methods paper



Empirical paper



R package



Project Website



✉ [peter.halpin@unc.edu](mailto:peter.halpin@unc.edu)

@peterhalpin

