What Impacts Should We Expect from Tutoring at Scale? Exploring Meta-Analytic Generalizability

Matthew A. Kraft, Brown University

Beth E. Schueler, University of Virginia

Grace T. Falken, Brown University

Funded by the National Science Foundation and William T. Grant Foundation

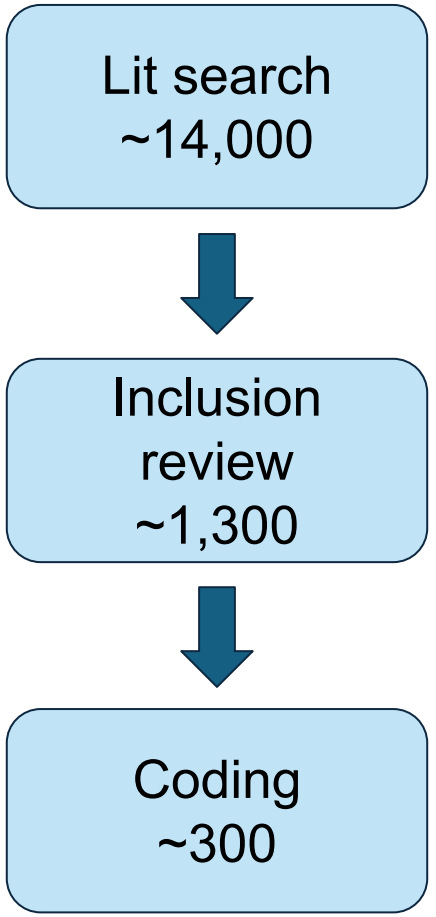
Research support from: Virginia Lovison, Luyan Nguyen, Sam Lynch, Hannah Sexton, and teams at Brown and UVA

Motivation and overview

- National calls to scale high-dosage tutoring
  - …but suggestive evidence that programs taken to scale have fallen short of expectations (e.g., Carbonari et al., 2022)
- We update the tutoring review literature to ask:
  - What are realistic expectations for tutoring effects on achievement when implemented at scale as part of the U.S. K-12 public school system?
- We find:
  1) Tutoring impact expectations are somewhat inflated for programs at scale
  2) More generalizable estimates are still meaningful in size
  3) A selected bundle of program design features appear to help preserve impacts at scale

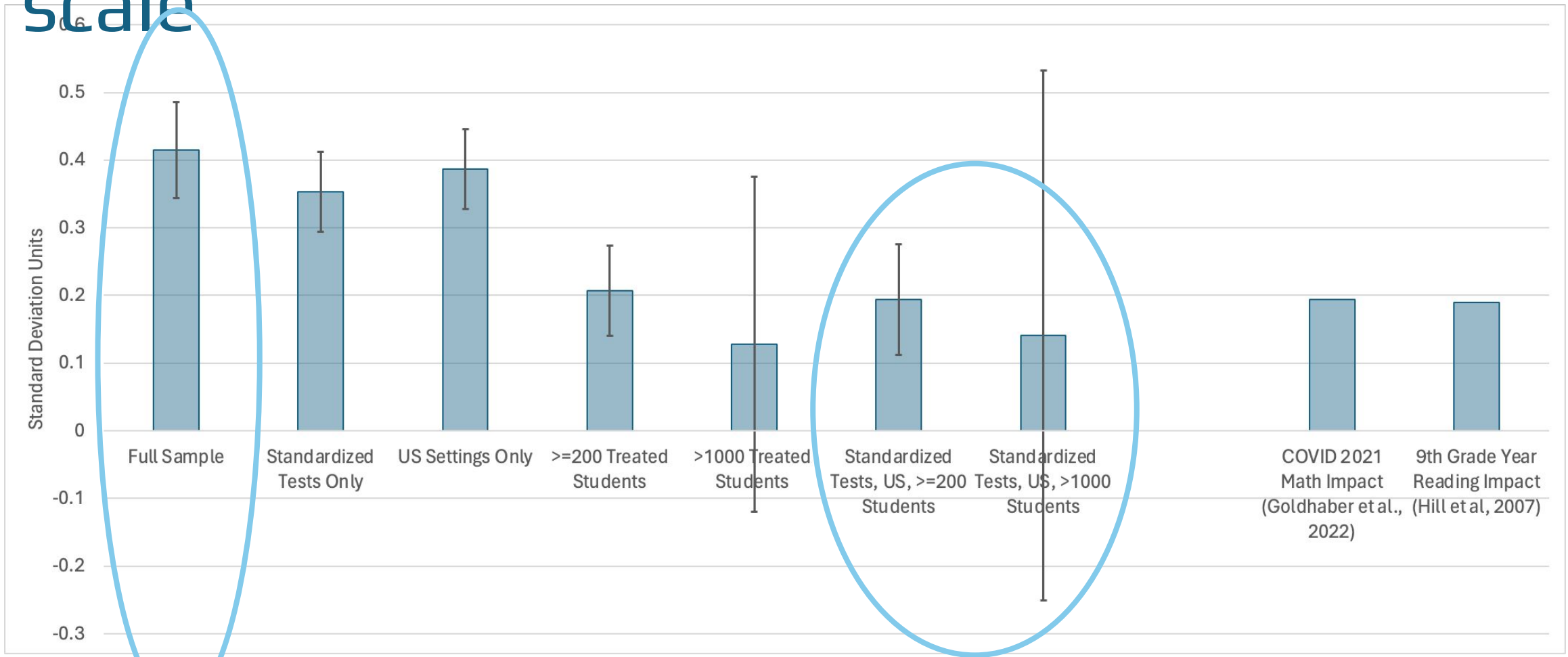
Methods

- Largest meta-analysis of tutoring impact evaluations (~300 RCT studies, 2,515 estimates)
- Use larger set of estimates to examine variation relevant to scaling tutoring in U.S. public schools
- Code studies with: human tutors, ratios, academic content, reading or math test outcomes, K-12, OECD countries, RCTs, sample of >20 students
- Calculate standardized Hedge's *g* effect sizes and estimate pooled effects with robust variance estimation accounting for nesting of estimates in studies

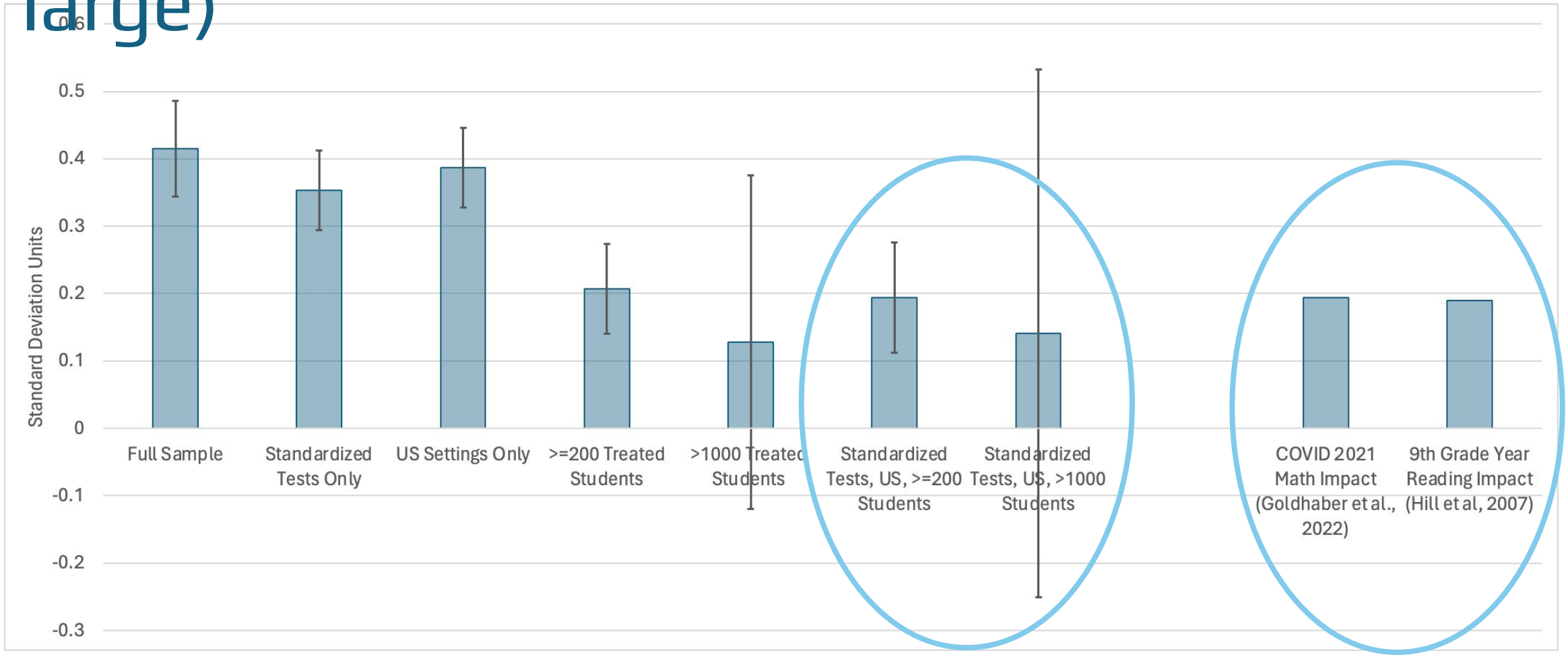


What tutoring effects can we realistically expect?

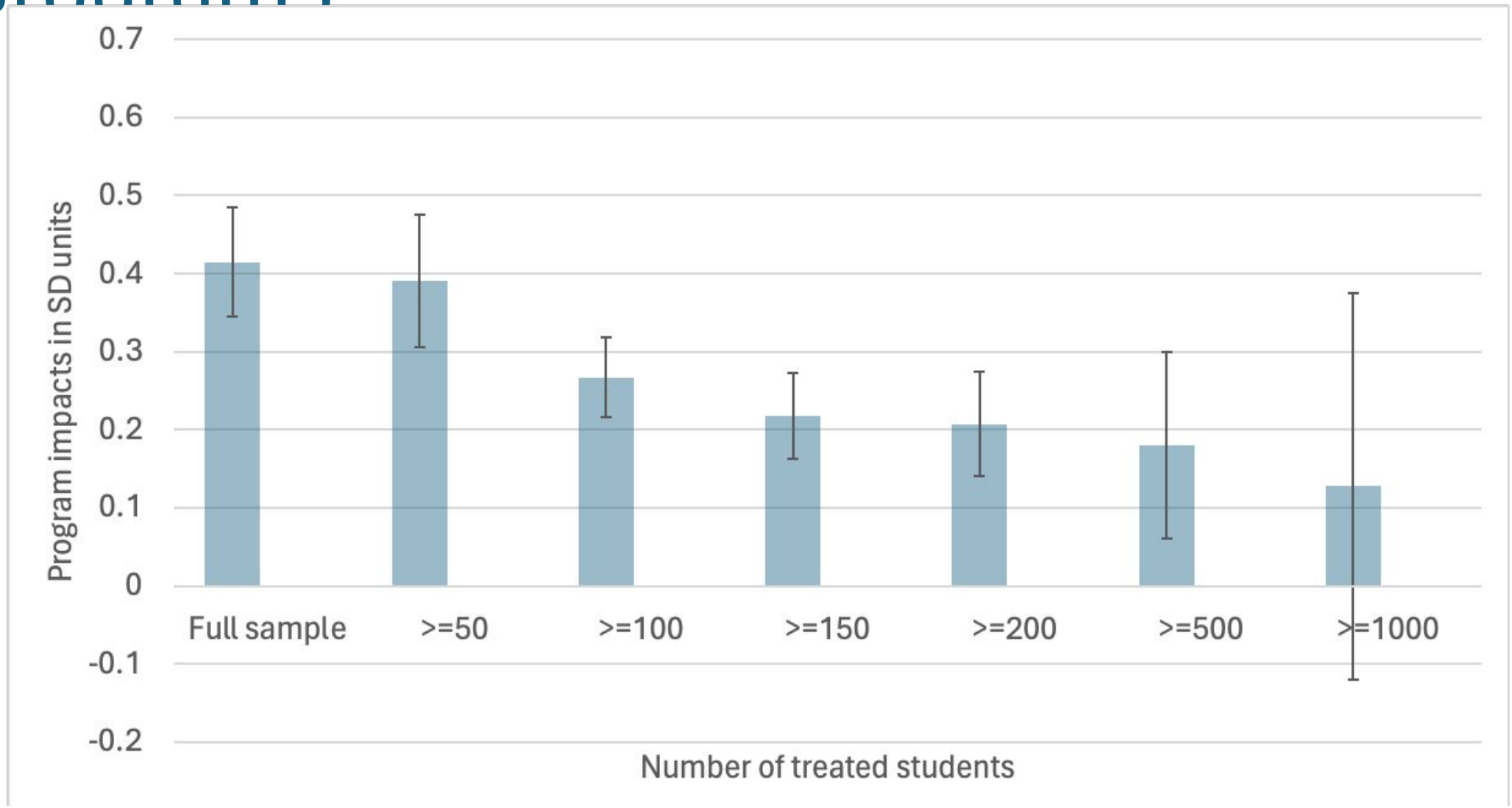
Expectations for tutoring effects are higher than effects for programs implemented at scale



Tutoring effects for programs at scale are still quite meaningful (i.e., moderate to large)



Tutoring effects are smaller for larger programs



Potential hypotheses for declining effects

- Publication bias could differentially inflate effects for smaller programs

No evidence of this

- Implementation quality could change as tutoring scales
  - Less effective tutors hired as programs expand
  - Implementation quality / fidelity declines as programs expand
  - Student attendance declines as programs expand

- Average targeted student could benefit less as tutoring scales
  - Differential benefits across students / less effective targeting as programs expand

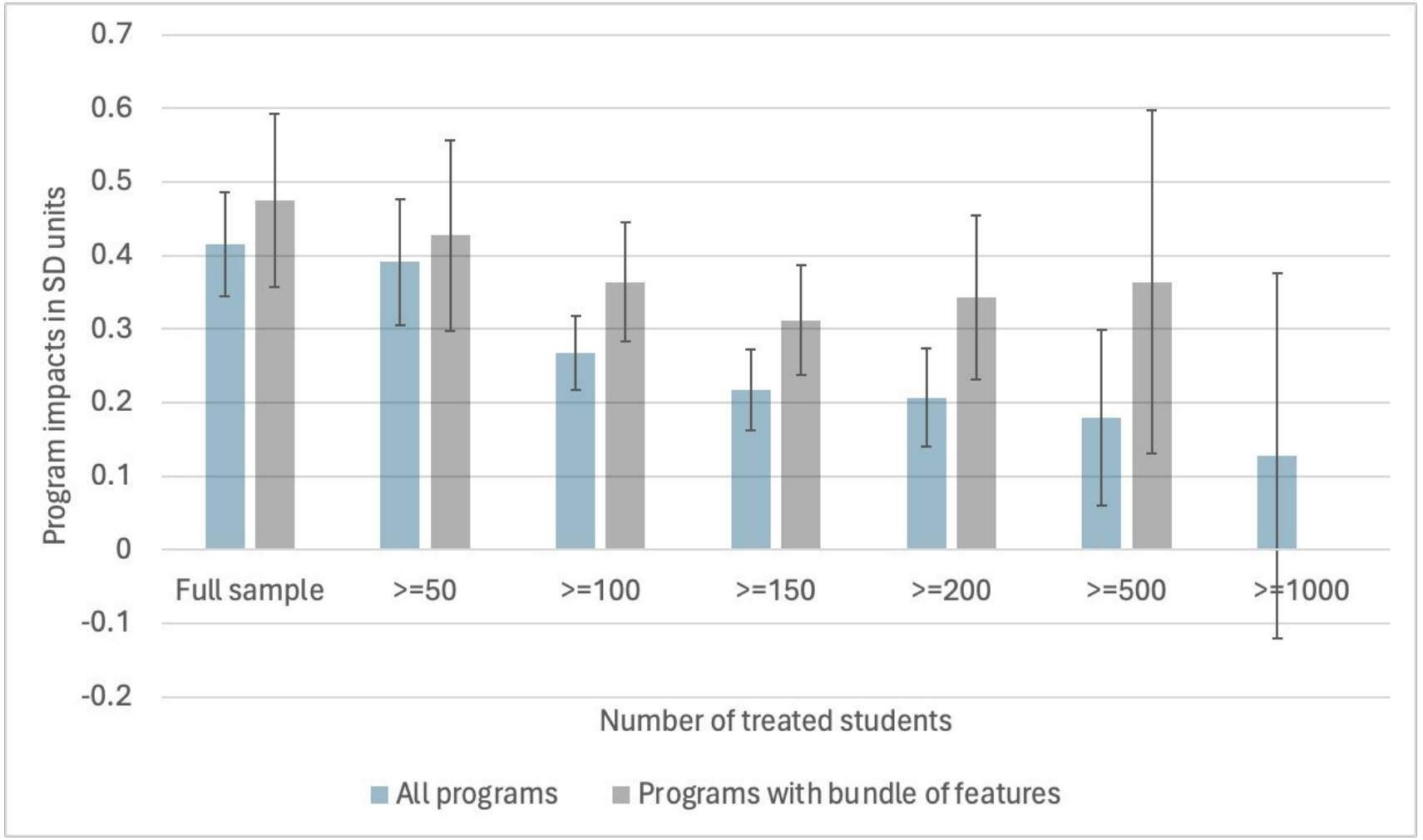
Cannot test these possibilities

- Program design could change as tutoring scales
  - Lower dosage or frequency
  - Higher student-tutor ratios
  - Less qualified tutors
  - Less tutor support

Programs with bundle of features can scale

Limit to programs with bundle of hypothesized effective features:

- 15+ hours
- Meet 3+/week
- In-person
- At school
- During school
- <=3:1 ratio
- Curricula



Robinson, Kraft, Loeb & Schueler (2021). Design Principles for Accelerating Student Learning With High-Impact Tutoring. EdResearch for Action Brief No. 8. https://edresearchforaction.com/research-briefs/accelerating-student-learning-with-high-dosage-tutoring/

Big picture takeaways

- Policymakers and researchers may have somewhat inflated expectations for effects of tutoring at scale (0.4 vs. 0.15-0.2 SD)

- Gold-standard evidence suggests we can still expect meaningful effects for a range of program types across subjects and grades

- Impacts are smaller for larger programs but a bundle of program design features appears to help maintain effectiveness at scale

Thank you
& stay in touch

Beth Schueler: beth.schueler@virginia.edu

Matt Kraft: mkraft@brown.edu

Grace Falken: gfalken@uw.edu