How Many Students Need to be Replaced to Invalidate a Teacher's Evaluation Based on Value-added?

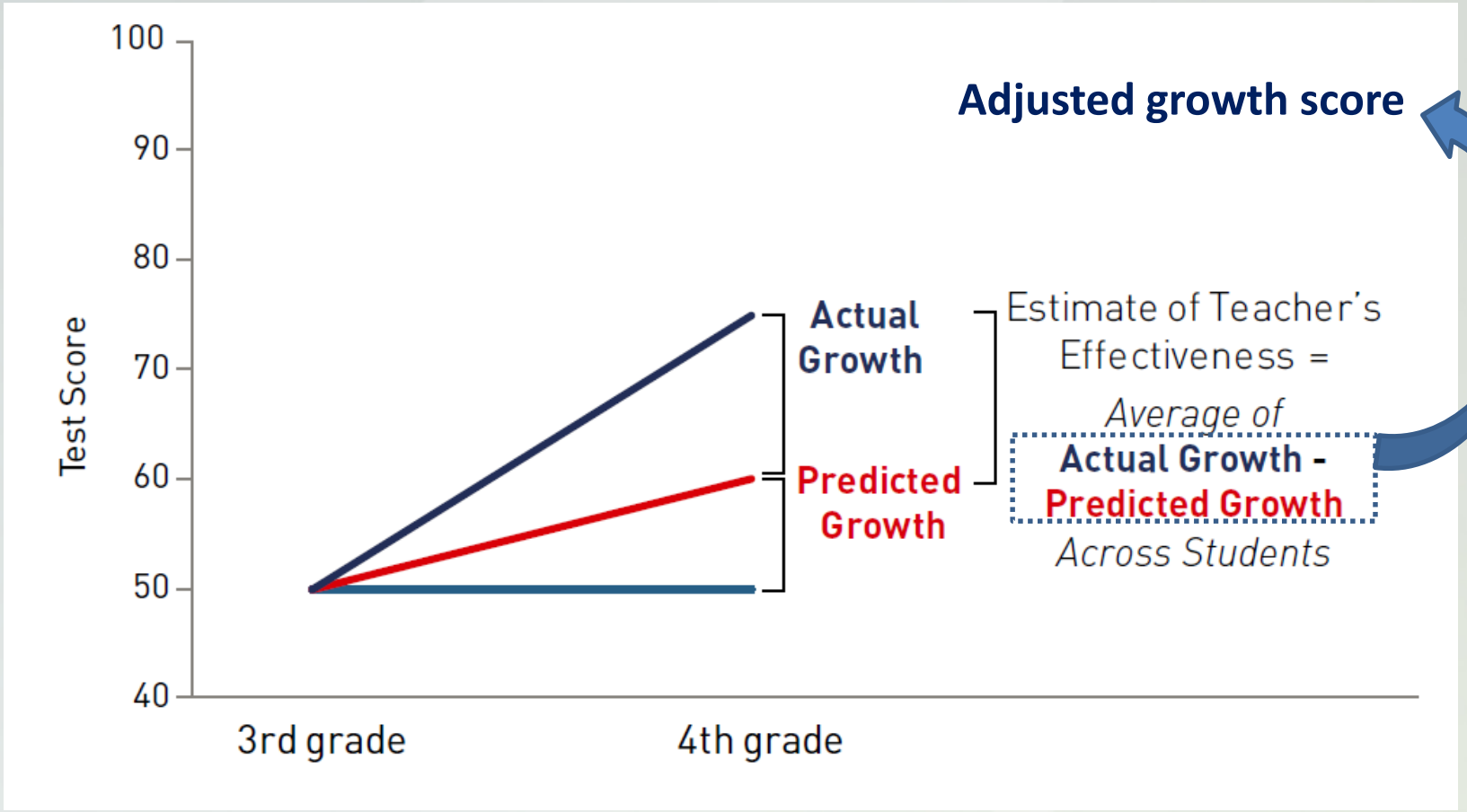
An Approach to Characterize the Uncertainty, Interpret and Make Use of Value-added

Qinyun Lin, Ken Frank
Michigan State University

Background

- No Child Left Behind and Race to the Top competition
- Some states have already incorporated value-added to inform high-stake decision-making for individual teachers
- It is essential to know the uncertainty of the measure for such high stakes decisions! Firing or promoting a teacher based on only an uncertain point estimate may be unfair or create a loss of investment or resources for a school.

What is value-added?



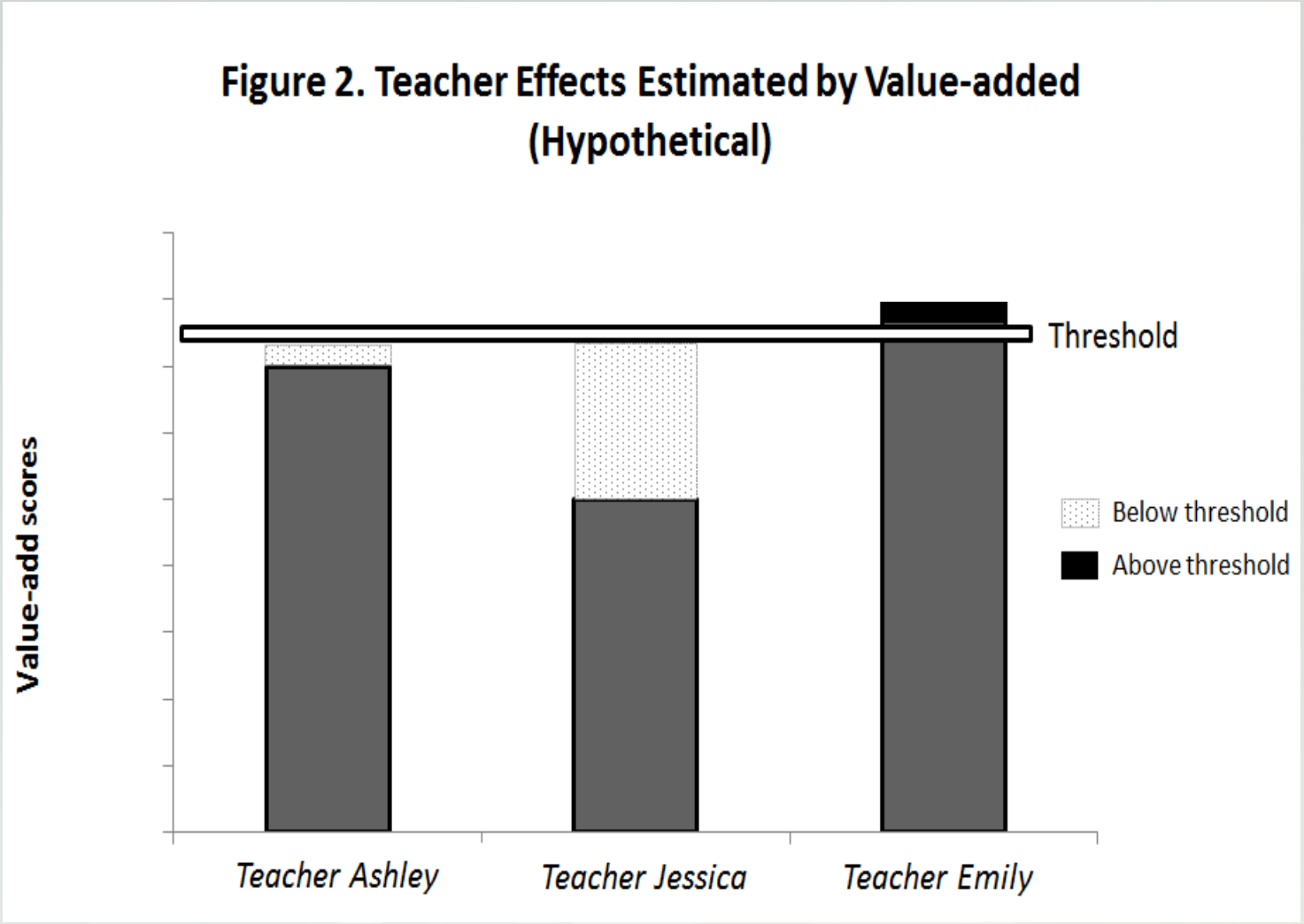
Value-Added Measures: How and Why the Strategic Data Project Uses Them to Study Teacher Effectiveness, Harvard University, 2011, p. 3. Available online: https://sdp.cepr.harvard.edu/files/cepr-sdp/files/sdp-va-memo_0.pdf.

Uncertainty of value-added

- Concerns about the validity and reliability of value added as a basis to inform high stake decisions:
  - Sorting based on pretests and non-random assignment of teachers to classrooms
  - Availability of test scores
    - It is recommended to have two years of prior tests in the model.
    - It is not uncommon for students (and even teachers) to change school.
  - Unreliability in test scores
    - Measurement errors
    - Different achievement measures

All these can cause inconsistency of value-added.

We summarized these sources of bias as they can lead to violation of the conditional random assignment assumption and thus bring uncertainty to the estimated teacher effect.



- *Dismiss both Ashley and Jessica?*
- *Send whom to professional development?*

Current approach to quantify the uncertainty

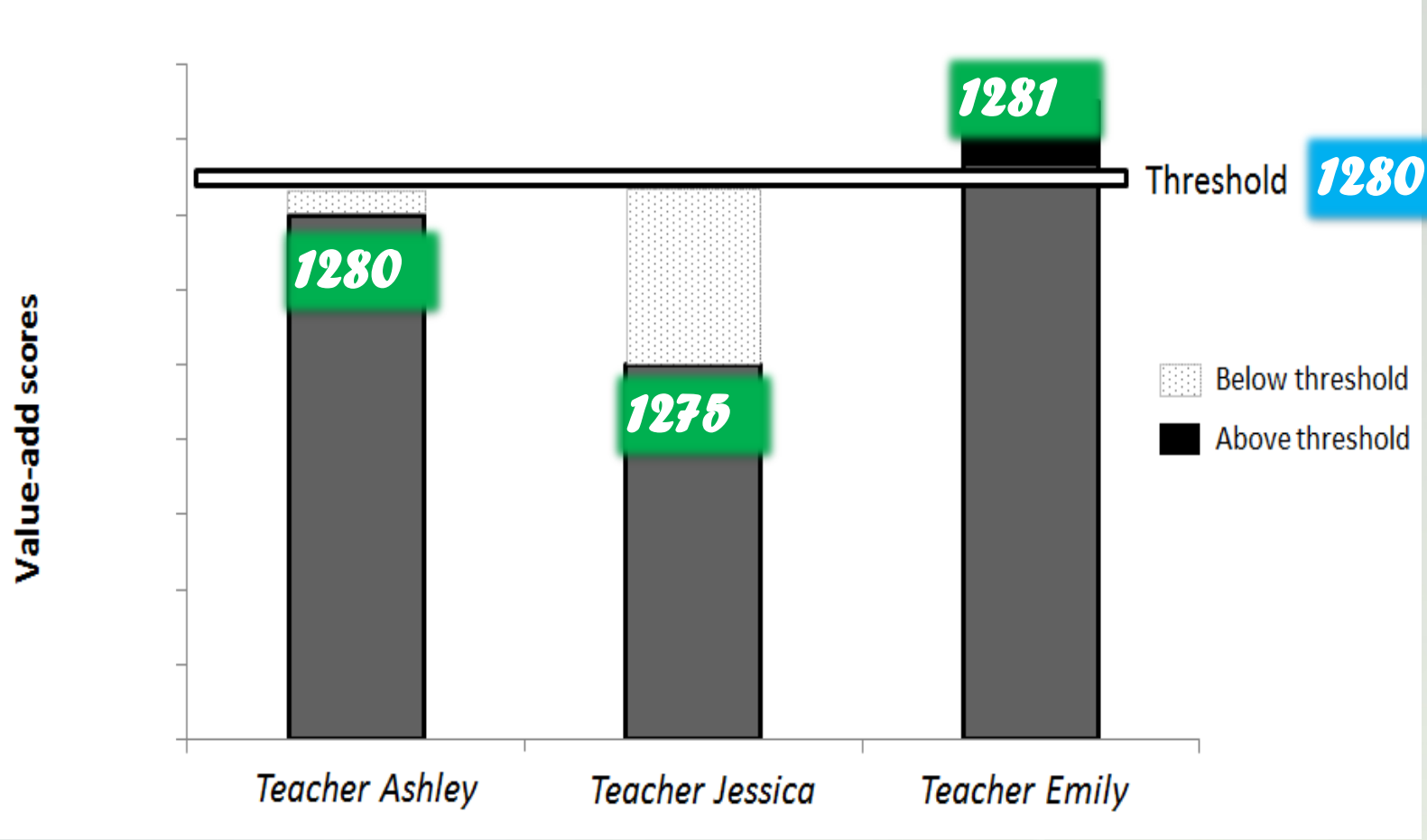
- standard errors or confidence interval
  - based on repeated sampling framework
  - difficult to understand and interpret for a decision regarding a single teacher
  - mainly concern about sampling error but not bias/inconsistency

The goal of this study is to provide an *intuitive* approach to quantify the uncertainty of value-added to inform decision-makings in practical education terms.

Theoretical Framework

- Replacement idea to quantify the robustness of an inference (Frank, Maroulis, Duong & Kelcey, 2013)
  - what percentage of the samples should be replaced with counterfactual no effect cases to invalidate the inference
- In this study, we ask: *how many students need to be changed to alter or invalidate the teacher evaluation based on value added?*

Hypothetical example of student replacement



Use the scale for Math, grade 3, in the 2017 M-STEP.

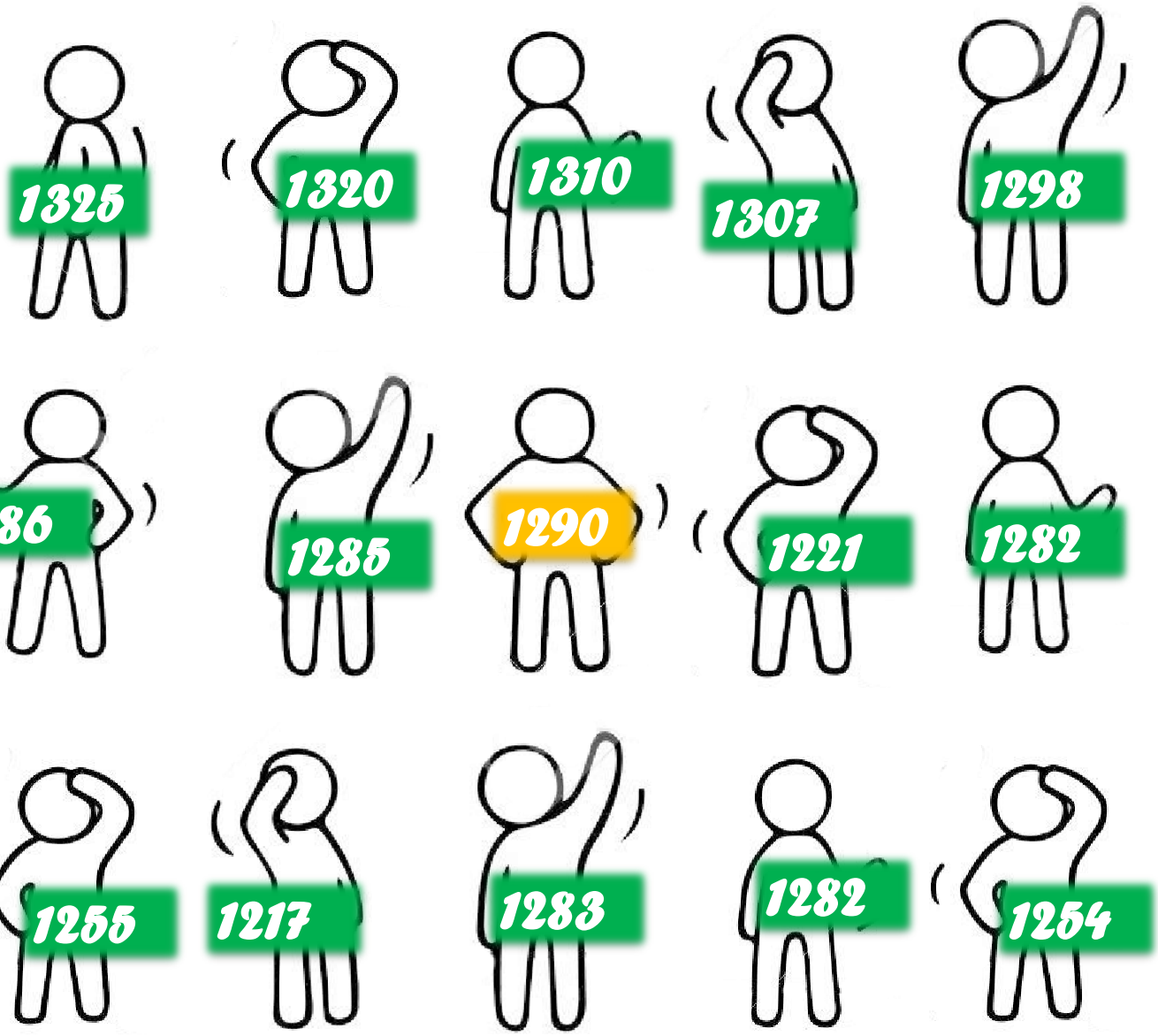
Motivate student replacement idea

- **Ashley:** I have a grade average effectiveness. I can achieve the threshold if I have more grade average students.
- **Administrator:** No. The below-threshold VAM indicates an inefficient effectiveness.
- The debate is about the uncertainty of VAM.
- *"How many grade average students need to be replaced to alter Ashley's evaluation?"*

Hypothetical example of student replacement - Ashley

After replacement Class average achieves 1281

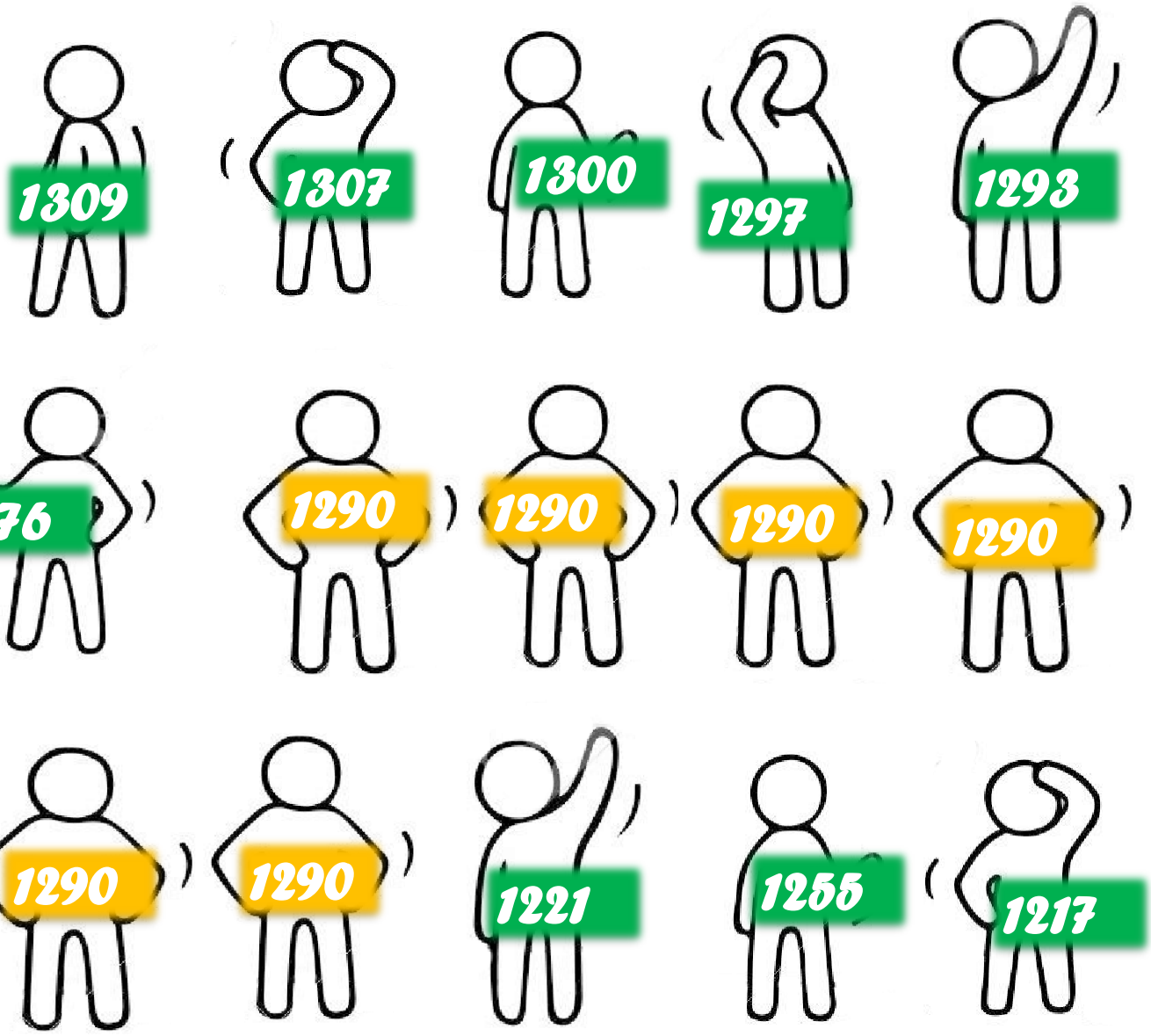
Class average = 1280



Hypothetical example of student replacement - Jessica

After replacement Class average improves to 1281

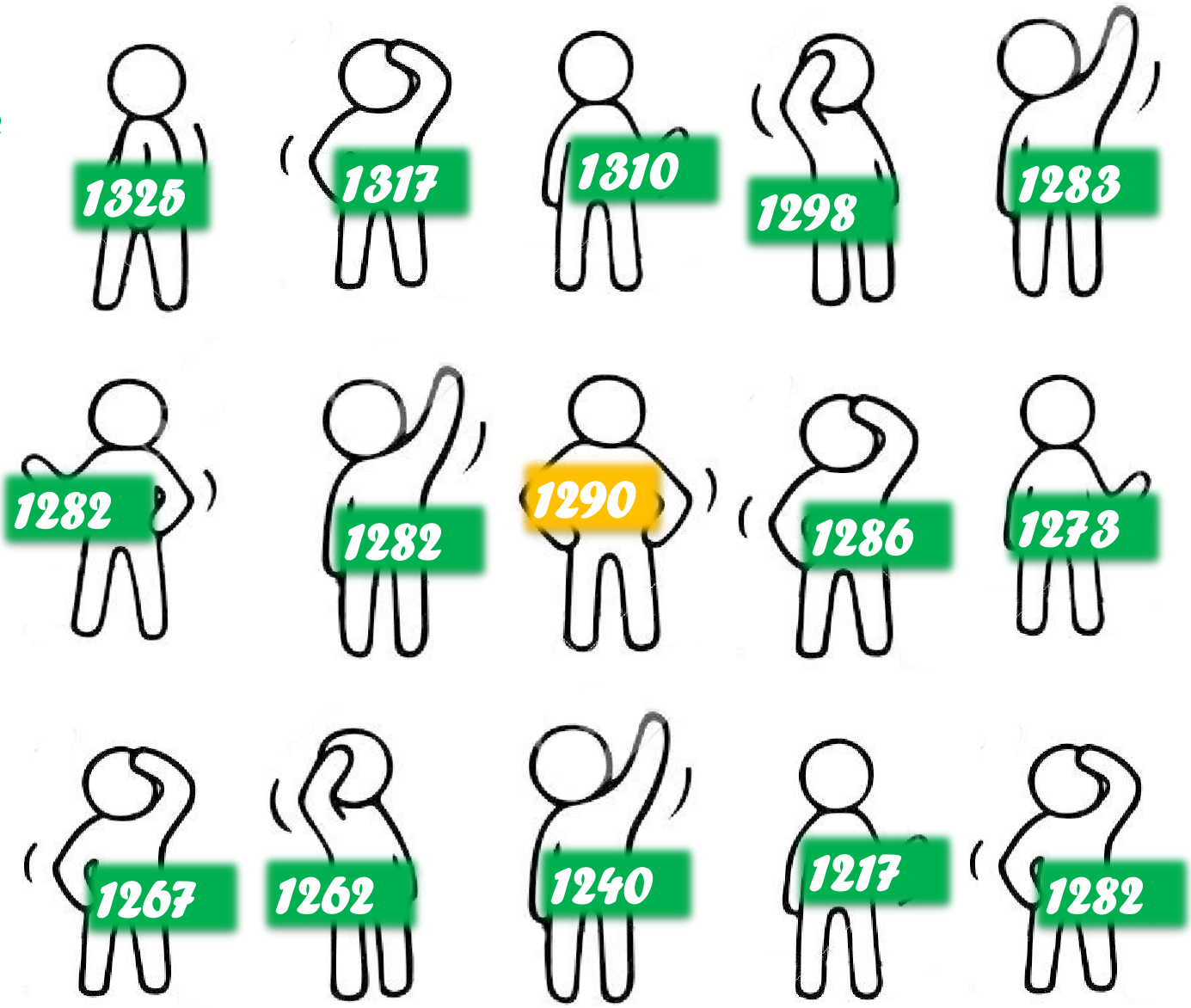
Class average = 1275



Hypothetical example of student replacement - Emily

Class average = 1282

After replacement Class average drops to 1281



Ashley: just by replacing one of her students, she can achieve the threshold.

Jessica: we need to replace six of her students to get her to the threshold.

Emily: just by replacing one of her students, she will drop to the threshold.

How can we interpret the number of students need to be replaced?

- *Uncertainty*
- *Distance to threshold*

Why this replacement idea (the mechanism of assigning students to teachers)?

- Easy to understand and interpret
- Directly related to the fundamental assumption of conditional random assignment
  - general framework for those specific sources of bias
- Recognize the agency of teachers, administrators and parents in authentic settings
- The replacement approach can be easily generalized for different contexts (next slide)

Generalize the student replacement idea

- *"How many grade average students need to be replaced to alter Ashley's evaluation?"*
  - Why grade average student? (choice of replacement student)
  - How to select students from Ashley's class to be replaced? (choice of students being replaced)

- Different contexts:
  - Concern about spillover effect?
  - Concern about violation of constant effect assumption?

Thank you very much!

Contact: linqinyu@msu.edu

If interested, please refer to the draft paper for more details.

Any suggestions and comments are very welcome!