The Generalizability Framework

Alison Fahey

Associate Director of Policy, J-PAL Global

J-PAL
ABDUL LATIF JAMEEL POVERTY ACTION LAB

Course Overview

1. Introduction to Randomized Evaluations in the Humanitarian Sector and Why Randomize?

2. Research Ethics in Humanitarian Settings

3. The Generalizability Framework

4. Academic-NGO Partnerships in the Humanitarian Space

"We keep running into the same problem from place to place to place. … The solutions, in a sense, can be the same. **You learn something general**, and from this general finding, you can **extract a lesson that policymakers will then tailor to each individual context**."

Esther Duflo, interview after the announcement of the 2019 Prize in Economic Sciences
https://bit.ly/2WI37Bk

TELEPHONE INTERVIEW
ESTHER DUFLO

THE SVERIGES RIKSBANK PRIZE IN ECONOMIC SCIENCES IN MEMORY OF ALFRED NOBEL 2019

Illustration: Niklas Elmehed

Example:
Teaching at the Right Level, India

A remedial learning tutoring program helped primary school students in India acquire basic reading skills.

Banerjee et al. 2016

8 + 14 − 7

7 x 4

If 3x − 10 = 24, then x = *?*
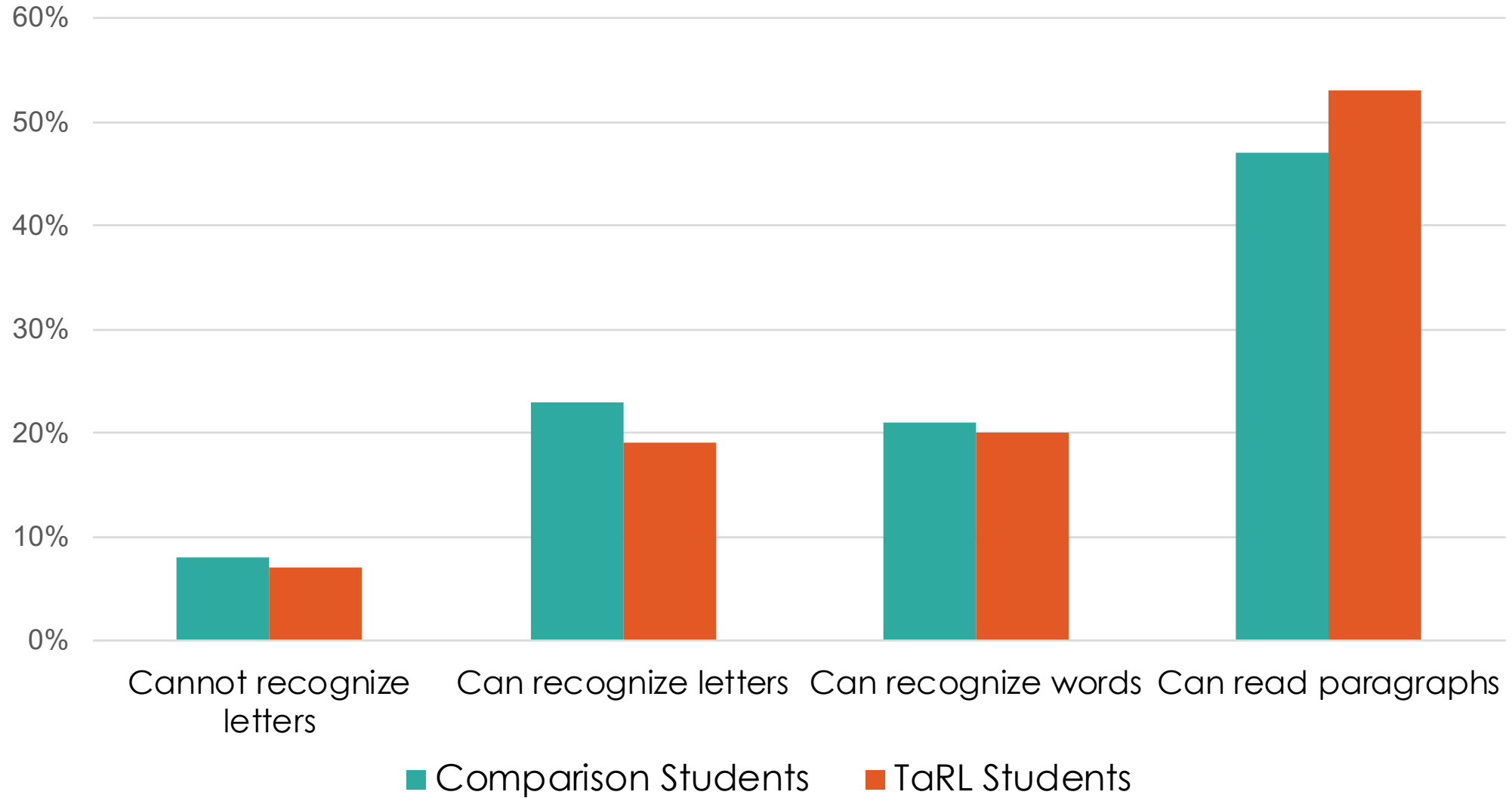
For all *a* and *b*,
$6a^2b^3 − 3a^2b$ is equivalent to
which of the expressions?

Randomized evaluation: Teaching at the Right Level

- Study by Abhijit Banerjee, Esther Duflo, Michael Walton, and Rukmini Banerji

- Location: Haryana, India

- 400 schools, 9,000 students

- Pratham (NGO) staff trained local teachers in the Teaching at the Right Level approach

  - Problem:

    - Classes are organized according to age, rather than learning level

  - Solution:

    - Group students by learning level based on test scores

    - Tailor lessons according to each group's learning level

Student Competency in Hindi

Bar chart legend: Comparison Students (teal), TaRL Students (orange)

Category	Comparison Students	TaRL Students
Cannot recognize letters	8%	7%
Can recognize letters	23%	19%
Can recognize words	21%	20%
Can read paragraphs	47%	53%

Breakout Discussion I

Breakout Discussion I – 10 minutes

- Housekeeping:
  - Please turn your video on
  - When speaking, turn your audio on as well
- Agenda
  - Introduction to the Generalizability Framework

Suppose you work for the Ministry of Education in your current country of residence …

- Would you implement TaRL in your country?
  - Why or why not?
- What additional information would you ideally have to make your decision?
- What data would you use to determine:
  - If students are performing below grade level?
  - If teachers are teaching at only one level for all students?
  - Which students to focus on with a Teaching at the Right Level program?

End of Breakout I

The challenge

- Dramatic rise in the number of rigorous impact evaluations in high, low, and middle income countries in last 20 years

- Unlikely to be rigorous evaluation of the program policy makers wants to introduce in exactly same location

- So can we learn anything at all from the evidence that exists?

A narrow view on applying evidence might say no

- A study can only inform policy in the location where the study was conducted.

- And within our specific location, we can only use whatever evidence was generated here.

- If we wanted to scale up a program, it should always be tested with a randomized evaluation in our location first.

- Or before scaling up a program, an identical version of the program or policy be replicated and tested a specific number of times

The generalizability framework

- Instead of focusing on place, focus on people

  - What lessons can we learn about **general behavior**

- Evidence from single study just one part of the puzzle

- Combine theory, descriptive evidence, and results of rigorous impact evaluations to consider:

  - What are the local conditions which shape the study's context? What did the local capacity to implement the program look like?

  - And to answer: Might results from one country be likely to replicate in another? When do/don't we need more evaluation?

- For more details, see Mary Ann Bates and Rachel Glennerster, "The Generalizability Framework," *Stanford Social Innovation Review*, 2017. https://ssir.org/articles/entry/the_generalizability_puzzle

Applying the Generalizability Framework

Three examples

1. Scaling immunization incentives
2. Teaching at the right level
3. Promoting refugee acceptance through sport

Scaling immunization incentives

A parent receives a kilogram of lentils at a vaccination clinic in Rajasthan, India.

- Seva Mandir program to increase immunization rates in rural Rajasthan, tested with RCT

    – Banerjee, Duflo, Glennerster, Kothari, 2010

- Fixing **supply**: regular monthly immunization camps with nurse present without fail

- Building **demand**: 1kg lentils for every vaccination, set of plates on completed immunization schedule

Percentage of Children Aged 1-3 Years Who Have Completed A Course of Immunizations
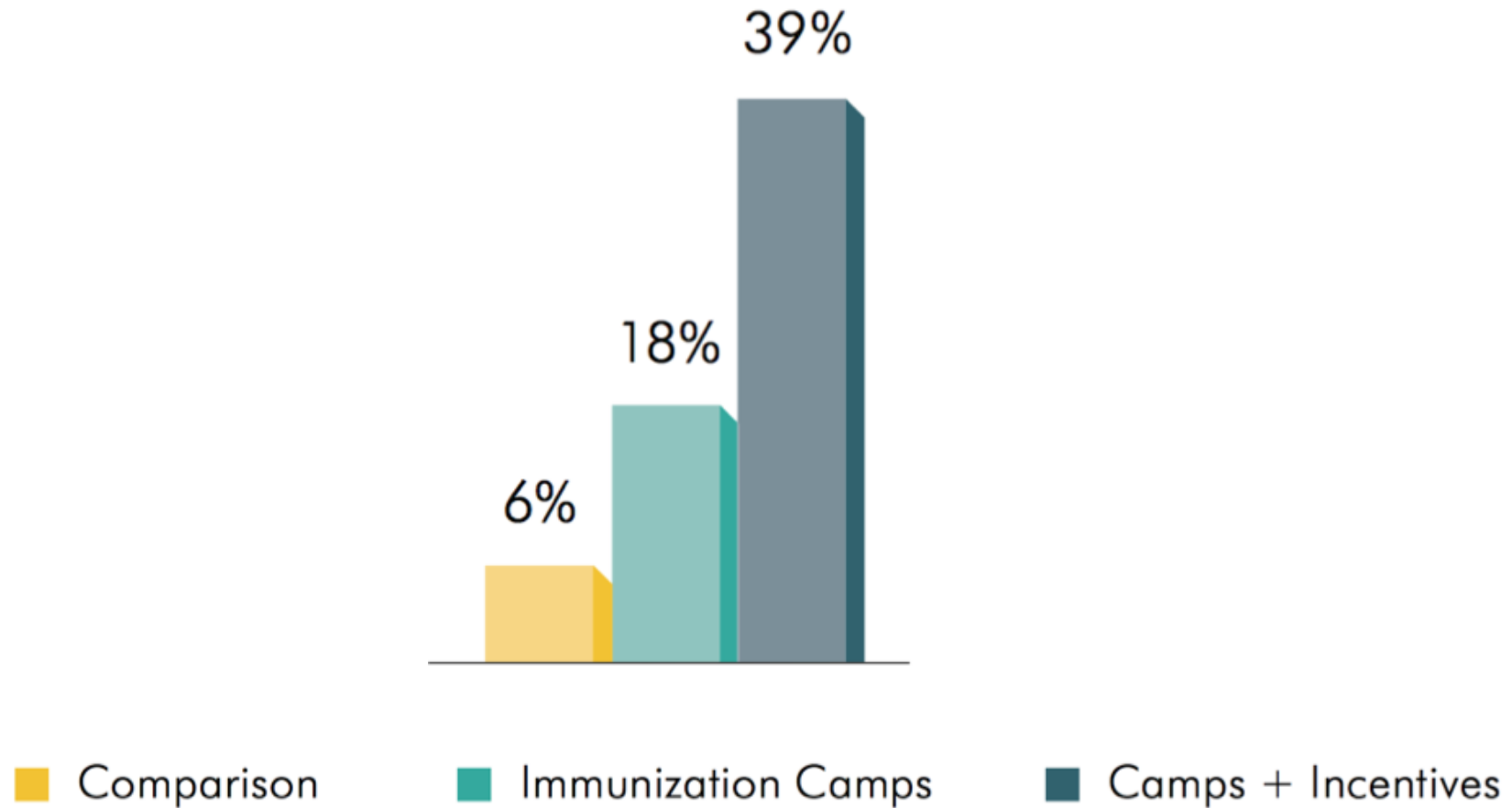
Viewing evidence in isolation

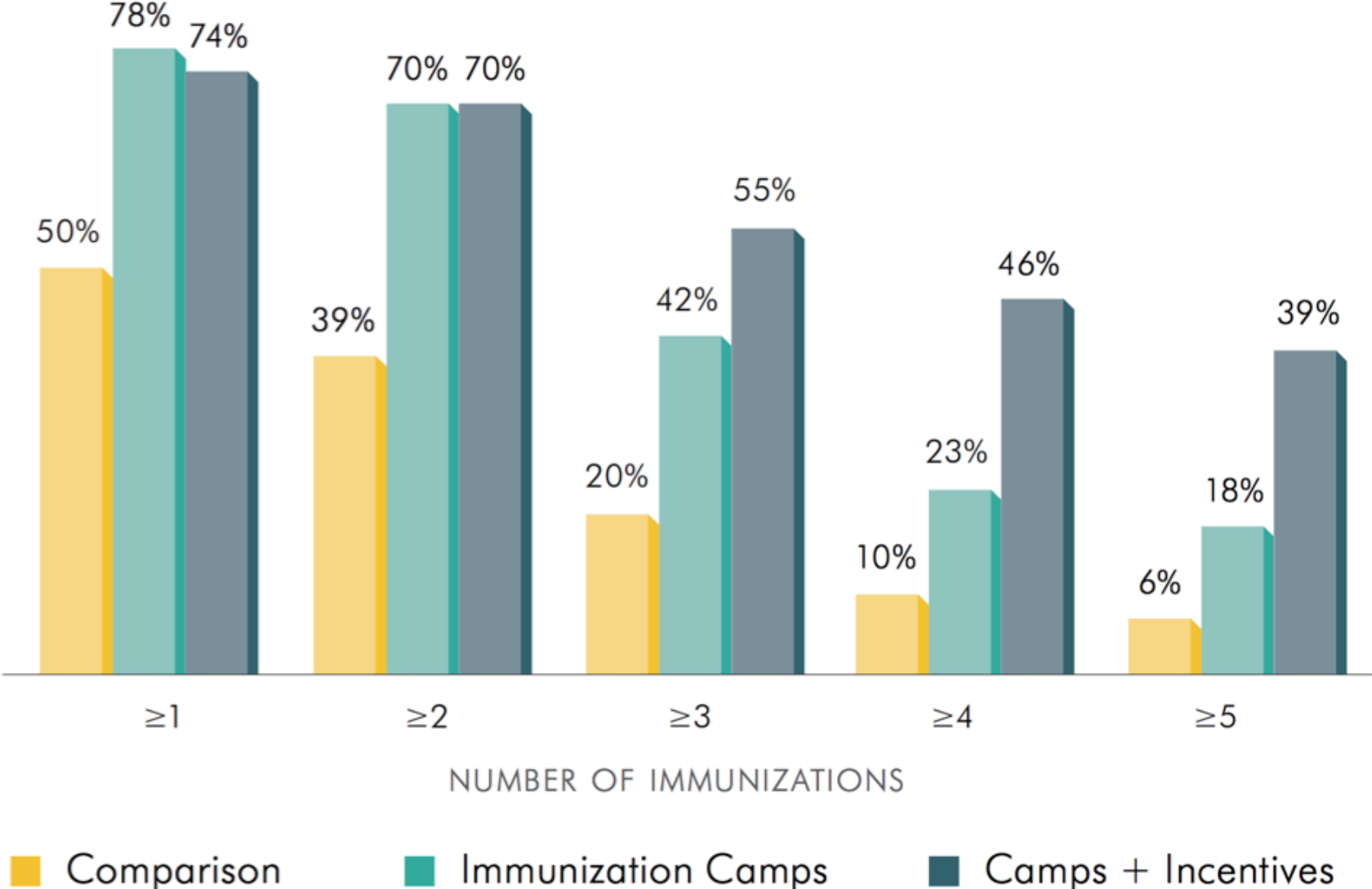If a government in West Africa wanted to improved immunization rate, should they consider noncash incentives?

- Only one RCT in South Asia; not Africa

- Program conducted by NGO not government

- Lentils not core part of local diet

INCENTIVES FOR IMMUNIZATION PROGRAM

?

COMPLETED IMMUNIZATION RATES RISE

Generalizability Framework

NUMBER OF IMMUNIZATIONS RECEIVED BY CHILDREN AGED 1-3 YEARS

Generalizability Framework

INCENTIVES FOR IMMUNIZATION PROGRAM

LOCAL CONDITIONS

1. Parents want to vaccinate
2. Parents can access clinic
3. Provider presence sufficient

GENERALIZED LESSONS ON BEHAVIOR

LOCAL IMPLEMENTATION

COMPLETED IMMUNIZATION RATES RISE
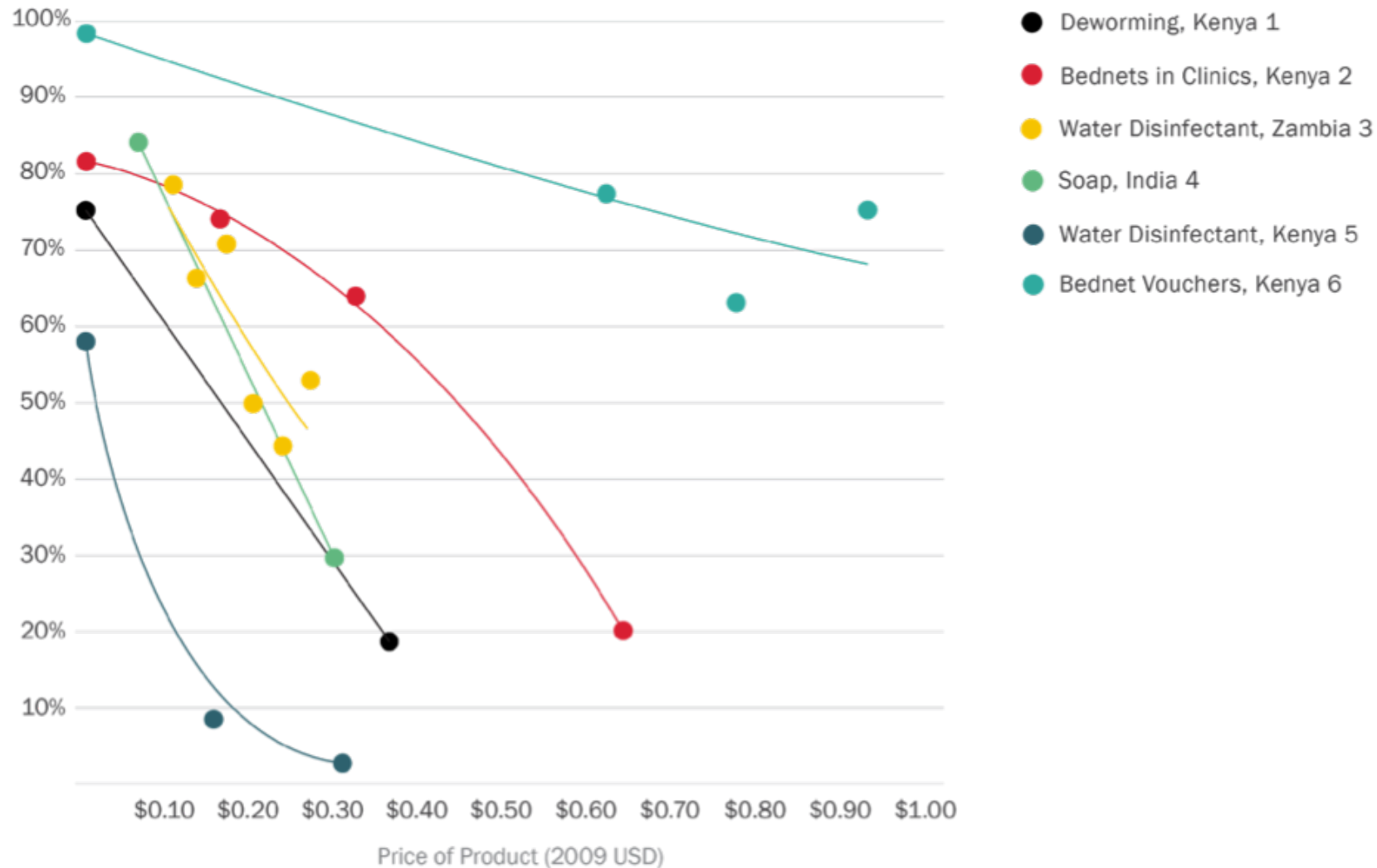
We all struggle with procrastination and prevention

- People procrastinate and find it hard to stick with behavior they believe is good for them and their children

- People are sensitive to the price of preventive health products: Small changes to the prices of preventive health products sharply reduce take-up

Small changes in price of preventive health products sharply reduce demand

Legend:
- Deworming, Kenya 1
- Bednets in Clinics, Kenya 2
- Water Disinfectant, Zambia 3
- Soap, India 4
- Water Disinfectant, Kenya 5
- Bednet Vouchers, Kenya 6

Y-axis: 100%, 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20%, 10%

X-axis: $0.10, $0.20, $0.30, $0.40, $0.50, $0.60, $0.70, $0.80, $0.90, $1.00

Price of Product (2009 USD)

24

What happens if we go beyond making a product free and adding a small incentive?

- Even very small incentives can influence non-trivial health decisions:
    - encouraging HIV testing (Thornton 2008, Malawi)
    - preventing child marriage (Buchmann et al. 2017, Bangladesh)
    - increasing take-up of flu vaccinations (Alsan et al. 2017, United States)
    - combating diabetes (Aggarwal et al. 2020, India)

Generalizability Framework

Left diagram:

INCENTIVES FOR IMMUNIZATION PROGRAM

?

COMPLETED IMMUNIZATION RATES RISE

Right diagram:

INCENTIVES FOR IMMUNIZATION PROGRAM

LOCAL CONDITIONS
1. Parents want to vaccinate
2. Parents can access clinic
3. Provider presence sufficient

GENERALIZED LESSONS ON BEHAVIOR
1. Parents procrastinate or fail to persist
2. Parents are highly sensitive to price of preventative health

LOCAL IMPLEMENTATION
1. Incentives delivered to clinics
2. Incentives delivered to parents

COMPLETED IMMUNIZATION RATES RISE

Should immunization incentives be replicated and scaled up in either country?

Immunization rates by antigen		
	Country 1	Country 2
DPT1	84%	47%
DPT#	74%	41%
Measles	67%	41%
Fully immunized	49%	38%

Breakout Discussion II
10 minutes

Which country is a good potential scale up location?

Immunization rates by antigen		
	Country 1	Country 2
DPT1	84%	47%
DPT#	74%	41%
Measles	67%	41%
Fully immunized	49%	38%

Country 1, Country 2, Both, or Neither?

Considerations in favor of Country 1

- Immunization rates are high for the first shot, but drop off considerably thereafter
- This suggests that supply is not a barrier – most families can access clinics!
- Instead, "demand-side factors" such as lack of follow-through appear to be at play (possibly due to procrastination)
  - Here, incentives continue with the full immunization schedule might make a difference.

Considerations in favor of (and against) Country 2

- Only 47% of children get their first vaccination
  - This suggests that *access* to clinics might be the primary barrier, rather than incentives to continue through to full immunization
- Nevertheless, immunization rates drop over the immunization schedule, suggesting some room for incentives to encourage follow-through

What local implementation issues would you consider?

End of Breakout II

Applying the Generalizability Framework

Three examples

1. Scaling immunization incentives
2. Teaching at the right level
3. Promoting refugee acceptance through sport

COVID-19 response in education

The School Year Really Ended in March

Abrupt closings have stalled the learning of millions of students. U.S. education needs a rescue, an economist says, and it won't be cheap.

Lily Padula

By Susan Dynarski

https://www.nytimes.com/2020/05/07/business/school-education-online-money.html

Saga Innovations

- Chicago Public Schools
- Evaluate Saga Education's model of individualized math tutoring on academic outcomes for 9th and 10th grade male students
  - **One-hour daily tutoring session as part of their regular class schedule**.
  - **Tutors met with two students at a time**
  - **Reviewed foundational skills—targeting instruction—and worked on current topics from students' regular math classes**.
- Results:
  - Students in Saga learned an **extra one to two years' worth of math** beyond what their peers learned in an academic year.
  - Tutoring raised participants' average national percentile rank on 9th and 10th grade math exams by more than 20 percent.
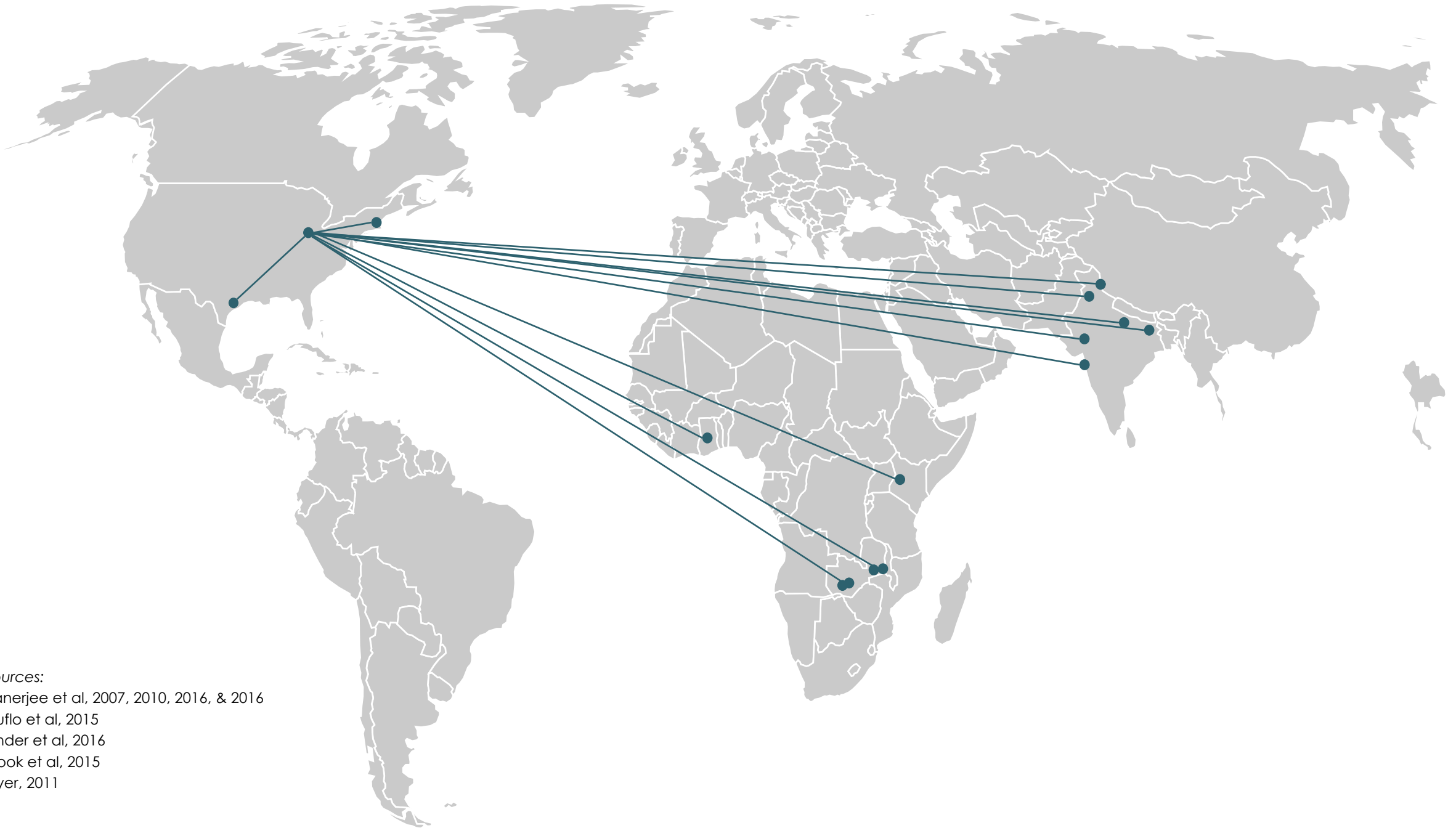
Sources:
Banerjee et al, 2007, 2010, 2016, & 2016
Duflo et al, 2015
Ander et al, 2016
Cook et al, 2015
Fryer, 2011

Teaching at the right level

Saga Education tutoring session
www.povertyactionlab.org/case-study/
individualized-tutoring-improve-learning

TaRL activities in a classroom in Gujarat, India
www.povertyactionlab.org/case-study/
teaching-right-level-improve-learning

Teaching at the right level

The approach works by
- dividing students into groups based on learning needs rather than age or grade;
- dedicating time to basic skills rather than focusing solely on the curriculum; and
- regularly assessing student performance, rather than relying only on end-of-year examinations.

https://www.teachingattherightlevel.org/

Targeted instruction increases learning

Series of studies shows targeted instruction can work in a variety of contexts:

1. Balsakhi Assistant Programme in India (Duflo et al 2007)
2. Read India Programme (Banerjee et al 2007)
3. Computer Assisted Learning (Duflo et al 2007)
4. India Reading Camps (Banerjee et al 2010)
5. Extra Teacher Programme in Kenya (Duflo et al 2011)
6. Haryana Learning Enhancement Programme (Berry et al 2013)
7. TCAI Programme in Ghana (Duflo and Kiessel 2012)
8. Match Education and Youth Guidance in Chicago (Cook et al 2014)
9. Match Education in Boston (Cook et al 2015)
10. Saga Innovations in Chicago (Davis et al 2017)

For more, see: J-PAL Evidence Review. 2019. "Will Technology Transform Education for the Better?"

TaRL · teaching at the right level

TEACHING AT THE RIGHT LEVEL

Zambia

This case study shares J-PAL Africa and Pratham's experience of working with the Ministry of General Education in Zambia to build the Catch Up programme, where global evidence, local adaptation, and iterative testing were used to improve learning outcomes.

TARGETED INSTRUCTION / TUTORING PROGRAM

LOCAL CONDITIONS

1. Children attend school, but literacy and numeracy rates are low
2. Teachers face incentives to teach grade-level material, not catch-up material

GENERALIZED LESSONS ON BEHAVIOR

1. Catch-up program instruction is at the student's level
2. Students learn when material is at their level

LOCAL IMPLEMENTATION

1. Teachers/volunteers trained in catch-up program
2. Time is devoted to catch-up program
3. Students attend catch-up classes targeted to their learning level

LITERACY AND NUMERACY RATES RISE

Applying the Generalizability Framework

Three examples
1. Scaling immunization incentives
2. Teaching at the Right Level
3. Promoting refugee acceptance through sport

Inter-religious Soccer Leagues in Iraq

- In Qaraqosh and Erbil, Iraq around 100 male soccer teams existed
- Most teams were composed of the same religion–either Christian or Muslim
- Researchers recruited around 51 Christian teams: half received 3 additional Muslim players, while the other half received 3 additional Christian players
- Teams performed in a tournament, with the top 3 earning prizes
- Results:
  - Contact with Muslim teammates increased Christians' acceptability of interacting with Muslims
  - Two-thirds of mixed teams had integrated Muslims as core team members six months after the intervention ended

Mousa, 2018

Scenario

- You are an IRC program officer in Lebanon working to combat social and economic discrimination against Syrian refugees

- Previous studies have shown that inter-religious football leagues promoted trust and cooperation across religious lines

- How might you decide whether to adopt this approach to improve relations between Syrian refugees and host communities in Lebanon?

Generalizability Framework in four steps

1. What is the theory behind the program?

2. Do the local conditions hold for the theory to apply?

3. How strong is the evidence for the required general behavioral change?

4. What is the evidence that the implementation process can be carried out well?

Contact Hypothesis: Theory

- Positive social contact across group lines towards a common goal can reduce prejudice, and increase cooperation

- Additional conditions: members should have equal status within the intervention (i.e. no hierarchies), and integration should be endorsed by communal authorities

Levy Paluck et al., 2018*

Do the local conditions hold for theory to apply?

- Is there ongoing conflict or competition between refugees and hosts that would likely outweigh positive effects of contact?

- Are there opportunities for cooperation outside of football?

  - E.g. opportunities for commerce or trade, or residential integration

- Is the broader community receptive to closer integration?

  - Or will beneficiaries face social sanction for reaching across group lines?

What is the existing evidence for general behavioral change?

- In post-conflict or conflict-affected settings:

  – Evidence is limited, but generally positive

    - Mousa (2020) finds positive effects for inter-religious football teams in Iraq

    - Scacco & Warren (2018) find positive effects for inter-religious vocational training in Nigeria

- Outside of conflict settings:

  – Considerable evidence that contact reduces prejudice

  – A 'meta-analysis' of 27 studies supports this hypothesis (Paluck, Green, & Green, 2017)

Can the intervention be carried out well?

- Does the implementing partner have the capacity to organize football leagues in host communities?

- Does the necessary infrastructure exist? (E.g. football fields)

- Is there demand for football leagues among refugees and host communities? Is there willingness to join integrated teams?

- Will community leaders support or endorse the program?

Breakout Discussion III
12 minutes

Should the IRC implement mixed-group football leagues to improve relations between refugees and host communities in Lebanon?

- Do the local conditions hold for Contact Theory to apply?

- How strong is existing evidence that inter-group sport leads to lower prejudice?

- Can the IRC implement the program well?

End of Breakout III

Conclusion

Does evidence from RCTs replicate to new context?
Too big a question, need to break it down:

- What is the theory of change behind the RCT?

- Do the local conditions hold for that theory to apply?

- How strong is the evidence for the general behavioral change?

- What is the evidence that the implementation process can be carried out well?

Conclusion

- If we have enough evidence to act, do we have enough evidence to stop evaluating impact?
    - We often need to act even when evidence is thin
    - Always monitor

- Often big overlap between when have enough evidence to launch big new initiative and when still worth evaluating

    – Questions may remain about best way to implement

- Trade off between evidence in new areas, versus more on improving evidence on refining a program

Further reading and resources

- Kremer and Glennerster, 2012, Chapter in Handbook of Health Economics

- Bates and Glennerster, 2017, "The Generalizability Framework," Stanford Social Innovation Review

- J-PAL Evidence to Policy page: http://www.povertyactionlab.org/evidence-to-policy/