Generalizing and Applying Evidence

Hannah Reuter, J-PAL North America
North Carolina OSBM 2021 Performance Management Academy
Ground rules

• We will be using a tool called Menti today. It may be helpful to have your cell phone handy.

• Jot down any questions you might have as they come up. We’ll periodically pause for Q+A.
J-PAL’s mission is to reduce poverty by ensuring that policy is informed by scientific evidence.
Our evaluations
Learning Objectives

• Introduce and apply a generalizability framework, which provides a systematic way to assess how evidence applies or does not apply to a new context

• Review concrete examples of whether to scale up evidence-based interventions in new contexts
Menti open response

Go to www.menti.com
Enter code: 97 40 22 8
The Challenge

• In context of limited resources, how do we know which programs provide the most value?

• To date we have seen a number of rigorous impact evaluations across many policy areas, but finding evidence that is perfectly relevant to your context is probably unlikely
“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
A randomized evaluation of a text message parental notification program in a West Virginia school district reduced course failures by nearly 30% (Bergman, Chan, 2019)

If you’re a North Carolina education leader, how do you know if this program would also work in your context?
Example: Parental notification text messages

- Study by Peter Bergman and Eric Chan (published in the *Journal of Human Resources*)
- Location: Kanawha County, West Virginia
- 1,137 parents of students in 22 middle and high schools
- The researchers sent parents information about their child’s class absences, missed assignments, and grades via automated text message to assess the impact of providing information to parents on student achievement
  - Provided *information* to correct parents’ beliefs about their child’s academic behaviors and performance
  - Empowered parents to *take action* to monitor and improve their child’s academic behaviors and performance
Parent alert: Jaden has 5 missing assignments in science class. For more information log in online.
Results from Study

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course Failures</td>
<td>.71 classes</td>
<td>.97 classes</td>
</tr>
<tr>
<td>Courses Attended</td>
<td>312 classes</td>
<td>278 classes</td>
</tr>
</tbody>
</table>

Impacts were larger for students with below average GPA

Bergman & Chan 2019
If a school district in North Carolina wanted to reduce course failures, should they consider text message reminders?

• Only one RCT in West Virginia; not North Carolina

How can we determine if this program will work somewhere else?
Should school districts in North Carolina replicate the text message parental notification program?

Share some reasons why you said yes or no. Please unmute yourself to share.
Four misguided questions

• What counts as a “similar enough” new setting?
  – Can a study inform policy only in the location in which it was undertaken?

• Should we use only whatever evidence we have from our specific location?

• Should a new local randomized evaluation always precede scale up?

• Must an identical program or policy be replicated a specific number of times before it is scaled up?
Are the locations identical?

Is there a similar problem?

Why did a solution work?
Key Principles of The Generalizability Framework

• Instead of focusing on place and time, focus on people
  – Key conditions and general lessons about behavior

• Evidence from single study just one part of the puzzle
  – We weigh the evidence based on quality and adjust priors

• Combine theory, descriptive evidence, and results of rigorous impact evaluations to answer:
  – Whether results from one context likely to replicate in another
  – When we need more evaluation and when we don’t

Key Principles of The Generalizability Framework

• Instead of focusing on place and time, focus on people
  – Key conditions and general lessons about behavior

• Evidence from single study just one part of the puzzle
  – We weigh the evidence based on quality and adjust priors

• Combine theory, descriptive evidence, and results of rigorous impact evaluations to answer:
  – Whether results from one context likely to replicate in another
  – When we need more evaluation and when we don’t

Key Principles of The Generalizability Framework

• Instead of focusing on place and time, focus on people
  – Key conditions and general lessons about behavior
• Evidence from single study just one part of the puzzle
  – We weigh the evidence based on quality and adjust priors

• Combine theory, descriptive evidence, and results of rigorous impact evaluations to answer:
  – Whether results from one context likely to replicate in another
  – When we need more evaluation and when we don’t

Key Principles of The Generalizability Framework

• Instead of focusing on place and time, focus on people
  – Key conditions and general lessons about behavior

• Evidence from single study just one part of the puzzle
  – We weigh the evidence based on quality and adjust priors

• Combine theory, descriptive evidence, and results of rigorous impact evaluations to answer:
  – Whether results from one context likely to replicate in another
  – When we need more evaluation and when we don’t

The “how” of applying the Generalizability Framework

Start with a **theory of change** for the previous program ...
The “how” of applying the Generalizability Framework … then walk through a series of guided questions
Applying the Generalizability Puzzle Framework

Examples

1. Texting parents about student performance
2. Tutoring/teaching at the right level
Case Study 1: Text Message Notifications

- West Virginia program to reduce course failures, tested with RCT
  - Bergman, Chan 2019

- Provided **information** to correct parents’ beliefs about their child’s academic behaviors and performance

- Empowered parents to **take action** to improve their child’s academic behaviors and performance
Case Study 1: Text Message Notifications

How often do you talk to your child about schoolwork?

Most parents report speaking to their children every day about their homework.

How many assignments has your child missed?

Over 50 percent of parents believed their children had missed no assignments.
In actuality, only 20 percent of respondents’ children had missed no assignments even though over 50 percent of parents thought their children have never missed an assignment.
Case Study 1: Text Message Notifications

• **50 percent** of parents heard from the school less than once every three months

• **48 percent** of parents believed their child does not disclose enough information about their academic progress for them to be easily involved in the child’s education → negatively correlated with student’s GPA

• Parents with older or lower-performing children were more likely to perceive that their child is not telling them enough about their schoolwork.
Imagine you are considering replicating or adapting this program

What does this descriptive data suggest about the underlying challenge?
• Most parents report speaking to their children every day about their homework
• 20 percent of respondents’ children had never missed an assignment even though over 50 percent of parents thought their children have never missed an assignment.
• 50 percent of parents heard from the school less than once every three months
• 48 percent of parents believed their child does not disclose enough information about their academic progress.

Share your thoughts by unmuting yourselves.
Generalizability Framework

Step 1: Write down the theory of change for the original program(s) and articulate the key assumptions for why it worked.

- **Needs**: Middle and high school students fail courses
- **Inputs**: Text messages provide parents accurate information
- **Outputs**: Parents receive info on student achievement and attendance
- **Intermediate Outcomes**: Parents are more involved in education at home
- **Outcomes**: Students complete assignments and stay in class
- **Outcomes**: Achievement and attendance improve
Generalizability Framework

Step 2: Find descriptive data to better understand if the underlying issue from the original context is also at play in your community.

1. Parents have inaccurate beliefs about their children’s performance
2. Parents will take action when equipped with accurate information

1. Students are missing classes
2. Schools are not communicating information to parents with enough frequency

1. School information systems can collect accurate data
2. Information can be delivered to parents

1. Course failures reduced
Generalizability Framework

**Step 3:** Assess the strength of the evidence for the mechanisms that made program effective in the original context(s), and whether the key assumptions are likely to hold in your context.

1. **Parents have inaccurate beliefs about their children’s performance**
2. **Parents will take action when equipped with accurate information**

1. **Students are missing classes**
2. **Schools are not communicating information to parents with enough frequency**

1. **School information systems can collect accurate data**
2. **Information can be delivered to parents**

1. **Course failures reduced**
Providing information to correct misperceptions/increase salience can change human behavior

- Criminal justice - Providing information on the consequences of not showing up for your court date reduced failure to appear rates in New York. ([Cooke et al 2018](https://www.cooke.edu))

- Health - Giving students information about the distribution of HIV infection rates by age and gender groups reduced risky sexual behavior in Kenya. ([Dupas 2011](https://www.dupas.com))

- Education - Providing information about the financial returns to education led to improved learning outcomes in Chile. ([Neilson et al 2019](https://www.neilson.edu))
Imagine you can build in a notification system in a school district. Which school district might be a good fit?

A. District 1
B. District 2
C. Neither
D. Both

Menti poll:
Generalizability Framework

Step 4: Assess whether you or another organization can successfully implement the intervention with fidelity to the original model.

- Parents have inaccurate beliefs about their children’s performance
- Parents will take action when equipped with accurate information

1. Students are missing classes
2. Schools are not communicating information to parents with enough frequency

1. School information systems can collect accurate data
2. Information can be delivered to parents

Local Conditions

Generalized Lessons on Behavior

Local Implementation

Course Failures Reduced
Local Information to Guide Implementation

• Does your district collect regular and accurate data on missed courses, missed assignments, and student performance?

• Does your district have a way of integrating this data with an automated information system?

• Do parents in your district have regular access to text messages? Are the majority of the parents in your district able to read in English?
1. Parents have inaccurate beliefs about their children’s performance
2. Parents will take action when equipped with accurate information

1. Students are missing classes
2. Schools are not communicating information to parents with enough frequency

1. School information systems can collect accurate data
2. Information can be delivered to parents

COURSE FAILURES REDUCED

LOCAL CONDITIONS

GENERALIZED LESSONS ON BEHAVIOR

LOCAL IMPLEMENTATION

TEXT MESSAGE NOTIFICATIONS

TEXT MESSAGE NOTIFICATIONS

COURSE FAILURES REDUCED
Applying the Generalizability Puzzle Framework

Examples
1. Texting parents about student performance
2. Tutoring/teaching at the right level
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.

By Susan Dynarski

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

If \(3x - 10 = 24\), then \(x = ?\)

For all \(a\) and \(b\),

\[6a^2b^3 - 3a^2b\] is equivalent to

which of the expressions?

When your math literacy is like this:

\[8 + 14 - 7\]

\[7 \times 4\]
Example 2: Teaching at the right level/Tutoring

**Challenge:** Low reading and math performance among primary school students in rural India; Large classrooms → teachers are incentivized to teach at grade level.

**What has worked?**
Reorienting teaching to the level of the student consistently improves learning outcomes.

Banerjee et al. 2007; Banerjee et al. 2010; Banerjee et al. 2016; Banerjee et al. 2017
Example 2: Saga Education/Tutoring

**Challenge:** Low math performance among high school students in Chicago; Large classrooms → teachers are incentivized to teach at grade level.

**What has worked?**
Small group tutoring at the student's level consistently improves learning outcomes.

Cook et al. 2015
J-PAL affiliates and co-authors partnered with Chicago Public Schools to study the impact of Saga Education’s model of individualized math tutoring on academic outcomes for 9th and 10th grade male students

- Saga assigned students to a **one-hour tutoring session every day as part of their regular class schedule**.
- **Tutors met with two students at a time** and divided instructional time evenly between **reviewing foundational skills—targeting instruction—and working 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.

Cook et al. 2015
New Context:
Now let’s say you lead a school district in North Carolina and you want to improve student outcomes. Should you consider a similar program?

At first glance:
• The evaluations were conducted in India and Chicago; but will it work here?
• Intensive interventions may be costly
Teaching at the right level/Saga Tutoring

Step 1: Theory of Change

**Step 1:** Write down the theory of change for the original program(s) and articulate the key assumptions for why it worked.

- **Needs:** Children have fallen years behind grade level. Teachers are incentivized to teach at grade level.

- **Inputs:**
  - TaRL: Teachers are trained to assess children and regroup them by learning level.
  - SAGA: Tutors are trained to teach students at their level.

- **Output:** Children receive smaller group instruction by learning level by teachers (TaRL) or tutors (SAGA).

- **Intermediate Outcomes:** Children learn more quickly when material is at their level.

- **Impact:** Scores improve.
Teaching at the right level/Tutoring
Step 2: Local conditions

New Context: North Carolina school district

**Step 2:** Find descriptive data to better understand if the underlying issue from the original contexts is also at play in your community.

<table>
<thead>
<tr>
<th></th>
<th><strong>India</strong> (original context)</th>
<th><strong>Chicago</strong> (original context)</th>
<th><strong>North Carolina school district</strong> (new context)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Academic performance</strong></td>
<td>Grade 5 students are years behind in reading and math.</td>
<td>Grade 9 students are years behind in math.</td>
<td></td>
</tr>
<tr>
<td><strong>Variation in classrooms</strong></td>
<td>Learning levels within each classroom are varied, and students have little recourse to learn basic skills if they have not mastered them in the foundational years.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Teaching at the right level/Tutoring
Step 3: General Lessons from Existing Evidence

**Step 3:** Assess the strength of the evidence for the mechanisms that made the program effective in the original context(s), and whether the key assumptions are likely to hold in your context.

*Six randomized evaluations* in India find that programs that reorient classroom instruction to teach at the level of the student are *consistently effective*, raising test scores by between 0.07 and 0.70 standard deviations.
Tailored instruction: consistently positive impacts across contexts

Preterm TaRL programs

<table>
<thead>
<tr>
<th>Program</th>
<th>Standard deviation improvement in test score (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TaRL Balakrish Year 1 (India)</td>
<td>0.14</td>
</tr>
<tr>
<td>TaRL Balakrish Year 2 (India)</td>
<td>0.28</td>
</tr>
<tr>
<td>Computer-assisted learning (India)</td>
<td>0.47</td>
</tr>
<tr>
<td>One-month summer camps (India)</td>
<td>0.08</td>
</tr>
<tr>
<td>TaRL in-school (India)</td>
<td>0.12</td>
</tr>
<tr>
<td>TaRL in-school (India)</td>
<td>0.15</td>
</tr>
<tr>
<td>10-day TaRL camps (India)</td>
<td>0.70</td>
</tr>
<tr>
<td>20-day TaRL camps (India)</td>
<td>0.62</td>
</tr>
<tr>
<td>Mindspark (India)</td>
<td>0.29</td>
</tr>
<tr>
<td>Tracking (Kenya)</td>
<td>0.18</td>
</tr>
<tr>
<td>Match tutoring (Chicago)</td>
<td>0.65</td>
</tr>
<tr>
<td>STRIPES (India)</td>
<td>0.75</td>
</tr>
<tr>
<td>In-school with assistants (Ghana)</td>
<td>0.08</td>
</tr>
<tr>
<td>In-school with teachers (Ghana)</td>
<td>0.15</td>
</tr>
<tr>
<td>Tutoring (Clifie)</td>
<td>0.08</td>
</tr>
</tbody>
</table>
Targeted instruction increases learning

Number of Tutoring Programs by Effect Size

- Negative (<-0.14): 6
- Null (-0.15 - 0): 11
- Small (0.1-0.14): 19
- Medium (0.15-0.29): 17
- Large (>0.3): 76

For more, see: “The Transformative Potential of Tutoring for Pre K-12 Learning Outcomes”
Teaching at the right level
Step 4: Local Implementation

**Step 4:** Assess whether you or another organization can successfully implement the intervention with fidelity to the original model.

Adapt intervention to North Carolinian school district context:

- Can you align tutoring with relevant school materials/curriculum?
- Can your tutors assess where students are at?
- Can you train tutors from the local community and provide them with ongoing support?
1. Students learn when material is at their level

LOCAL CONDITIONS

1. Literacy and numeracy rates are below grade level
2. Teachers face incentives to teach grade-level material, not catch-up material

GENERALIZED LESSONS ON BEHAVIOR

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

LOCAL IMPLEMENTATION

LEARNING OUTCOMES IMPROVE
Findings from J-PAL’s tutoring research meta-analysis have informed state efforts to address learning loss caused by the Covid-19 pandemic. This includes California lawmakers’ decision to pass an education package allocating $4.6 billion to student recovery, tutoring, and integrated pupil supports, including $460 million to hire professional tutors. The research has also informed a Colorado bill advocating for the establishment of statewide high-impact tutoring programs.
Putting it all together: Decision tree

Finally, once you have completed the steps of the generalizability framework, you can use the following decision tree to help you interpret your analysis.

**Local Conditions**
- Does the problem that the original intervention solved exist in your community?
  - Yes: Go to **General Lessons from Existing Evidence**
  - No: Continue to **Local Implementation**

**General Lessons from Existing Evidence**
- Is the underlying mechanism of change valid in your context?
  - Yes: Go to **Local Implementation**
  - No: Consider **Capacity Building May Be Necessary** or **Evaluation Encouraged**

**Local Implementation**
- Who would implement the program and do they have the capacity?
  - Yes: Go to **Intervention Adaptations**
  - No: Consider **Limited Capacity** or **Slightly Modify Intervention**

**Capacity Building May Be Necessary**
- This program may be a good fit for your context, but some capacity-building work and/or partnership development is needed to successfully implement the program. We suggest brainstorming potential local implementation partners and looking for open source resources to develop an implementation plan.

**Evaluation Encouraged**
- This program may address the issue you hope to solve, but it does not fully meet all the criteria to be reasonably confident. Consider implementing with a robust evaluation to ensure intended impact.

**Good Match**
- Potentially replicate without evaluation.

**Local Conditions**
- Does the problem that the original intervention solved exist in your community?
  - Yes: Go to **General Lessons from Existing Evidence**
  - No: Continue to **Local Implementation**

**General Lessons from Existing Evidence**
- Is the underlying mechanism of change valid in your context?
  - Yes: Go to **Local Implementation**
  - No: Consider **Capacity Building May Be Necessary** or **Evaluation Encouraged**

**Local Implementation**
- Who would implement the program and do they have the capacity?
  - Yes: Go to **Intervention Adaptations**
  - No: Consider **Limited Capacity** or **Slightly Modify Intervention**

**Capacity Building May Be Necessary**
- This program may be a good fit for your context, but some capacity-building work and/or partnership development is needed to successfully implement the program. We suggest brainstorming potential local implementation partners and looking for open source resources to develop an implementation plan.

**Evaluation Encouraged**
- This program may address the issue you hope to solve, but it does not fully meet all the criteria to be reasonably confident. Consider implementing with a robust evaluation to ensure intended impact.

**Good Match**
- Potentially replicate without evaluation.
Concluding Remarks
Conclusion

Does evidence from RCTs replicate to new contexts? Too big a question. Break it down:

– What is the theory of change behind the intervention in the original 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? (Always monitor)
  - We often need to act even when evidence is thin
- Often big overlap between when have enough evidence to launch a new initiative and when it is still worth evaluating
  - Questions may remain about best way to implement
- Tradeoff between evidence in new areas, versus more on improving evidence on refining a program
Final Takeaway

Are the locations identical?
Final Takeaway

Are the locations identical?

Is there a similar problem?

Why did a solution work?
Over 400 million people reached by scaling up programs found to be effective by J-PAL RCTs

Evidence to Policy
Evidence to Policy

Evidence from randomized evaluations is changing how we understand and address problems related to poverty. Policymakers, practitioners, and funders worldwide are increasingly applying this learning to social policies and programs.

Over 400 million people have been reached by programs that were scaled up after being evaluated by J-PAL affiliated researchers. Many more have benefitted from the several broader ways evidence can inform policy, outlined below.

Pathways to Policy Change

Below, you will find six pathways through which evidence can have an impact on policy and case studies that illustrate partnerships leading to policy impact.

- **Shifting global thinking**
  Knowledge generated by randomized evaluations has fundamentally shaped our understanding of many social policies.
  
  Example case studies:
  - Free bednets to fight malaria
  - More...

- **Institutionalizing evidence use**
  Many organizations, including governments and large NGOs, have institutionalized processes for rigorously evaluating innovations and incorporating evidence into decision-making.
  
  Example case studies:
  - A government innovation lab to improve education
  - More...

- **Applying research insights**
  Lessons from randomized evaluations have informed the design of programs.

- **Adapting and scaling a program**
  Programs originally evaluated in one context have been adapted and scaled in others.
Thank you!
Further reading and resources

  https://ssir.org/articles/entry/the_generalizability_puzzle
- Kremer and Glennerster, 2012, Chapter in Handbook of Health Economics
- J-PAL Evidence to Policy page
  http://www.povertyactionlab.org/evidence-to-policy/
- J-PAL Self-Guided Case Study on Applying the Generalizability Framework to Complex Health Care