Education RCTs

Mauricio Romero (heavily inspired by Abhijeet Singh notes)

Education RCTs

- Introduction
- Five stylized facts on education in developing countries
- Measuring learning outcomes
- Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)
- A quick review of classic papers
- Final remarks

Education RCTs

Introduction

Five stylized facts on education in developing countries

Measuring learning outcomes

Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)

A quick review of classic papers

Final remarks

Education in developing countries

- Education is central to **policy agendas** globally, including in developing countries
  - Prominent in the vision of national governments
  - Also prominent in international policy discussions: e.g. MDGs, SDGs, the WDR 2018 on Education

Education in developing countries

- Education is central to **policy agendas** globally, including in developing countries
  - Prominent in the vision of national governments
  - Also prominent in international policy discussions: e.g. MDGs, SDGs, the WDR 2018 on Education
- Education is also central in the **research agendas** of development economists

Education in developing countries

- Education is central to **policy agendas** globally, including in developing countries
  - Prominent in the vision of national governments
  - Also prominent in international policy discussions: e.g. MDGs, SDGs, the WDR 2018 on Education
- Education is also central in the **research agendas** of development economists
- This lecture is a (selective) introduction to the field:
  - Why are economists interested in this area?
  - What are currently prominent (classes of) questions?
  - What are some examples of the best experimental work in the area?

Economics of Education reflects many strands of econ research

- Important for **individual welfare**:
  - Expanding "capabilities" (Sen, 1998)
  - Private returns on e.g. labour market
  - Relevant for studying inequality, anti-poverty policy

Economics of Education reflects many strands of econ research

- Important for **individual welfare**:
  - Expanding "capabilities" (Sen, 1998)
  - Private returns on e.g. labour market
  - Relevant for studying inequality, anti-poverty policy
- Potentially imp for **economic growth**
  - see e.g. macro HK and growth decomposition literature

Economics of Education reflects many strands of econ research

- Important for **individual welfare**:
  - Expanding "capabilities" (Sen, 1998)
  - Private returns on e.g. labour market
  - Relevant for studying inequality, anti-poverty policy
- Potentially imp for **economic growth**
  - see e.g. macro HK and growth decomposition literature
- Important **investments** for households and individuals
  - links to e.g. intra-household resource allocation

Economics of Education reflects many strands of econ research

- Important for **individual welfare**:
  - Expanding "capabilities" (Sen, 1998)
  - Private returns on e.g. labour market
  - Relevant for studying inequality, anti-poverty policy
- Potentially imp for **economic growth**
  - see e.g. macro HK and growth decomposition literature
- Important **investments** for households and individuals
  - links to e.g. intra-household resource allocation
- Core govt function, substantial share of **public expenditure**
  - interesting issues of state capacity and public finance
- Schools are **major employers**
  - issues of personnel economics (contracts, incentives, teams)

Economics of Education reflects many strands of econ research

- Important for **individual welfare**:
  - Expanding "capabilities" (Sen, 1998)
  - Private returns on e.g. labour market
  - Relevant for studying inequality, anti-poverty policy
- Potentially imp for **economic growth**
  - see e.g. macro HK and growth decomposition literature
- Important **investments** for households and individuals
  - links to e.g. intra-household resource allocation
- Core govt function, substantial share of **public expenditure**
  - interesting issues of state capacity and public finance
- Schools are **major employers**
  - issues of personnel economics (contracts, incentives, teams)
- Substantial **private sector**, esp in developing countries
  - major issues of IO such as competition and choice

Education RCTs

- Introduction
- Five stylized facts on education in developing countries
- Measuring learning outcomes
- Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)
- A quick review of classic papers
- Final remarks

Introduction

Five stylized facts on education in developing countries

Measuring learning outcomes

Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)

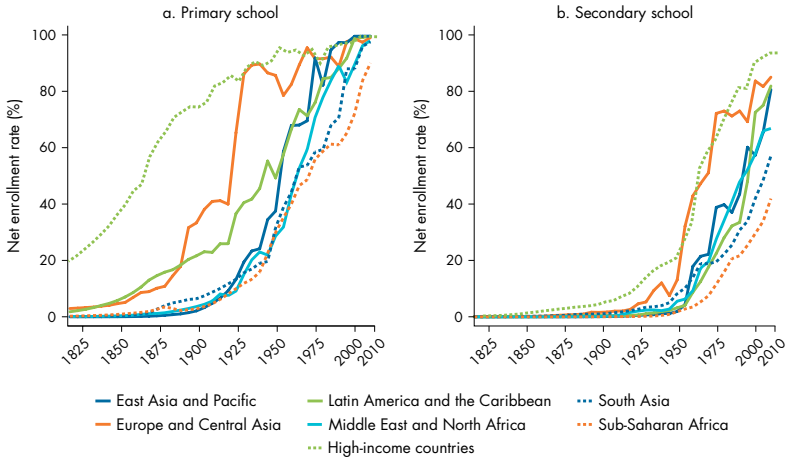
A quick review of classic papers

Final remarks

I. Enrollment is near universal, years of schooling rising rapidly

Figure 2.1 School enrollments have shot up in developing countries

Net enrollment rates, by country group (1820–2010)

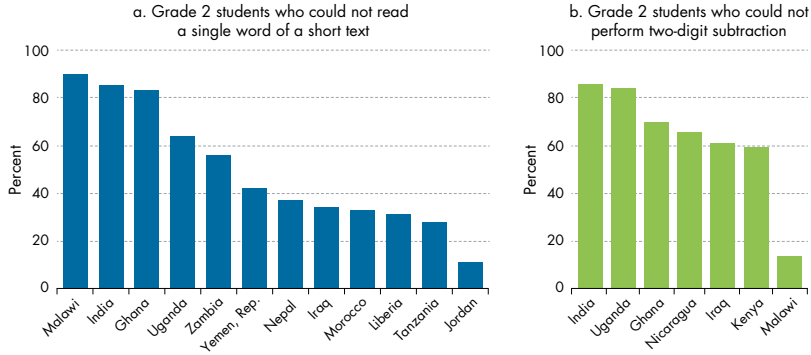


Source: WDR 2018 team, using data from Lee and Lee (2016). Data at http://bit.do/WDR2018-Fig_2-1.

II. Learning levels are very poor

Figure O.1 Shortfalls in learning start early

Percentage of grade 2 students who could not perform simple reading or math tasks, selected countries



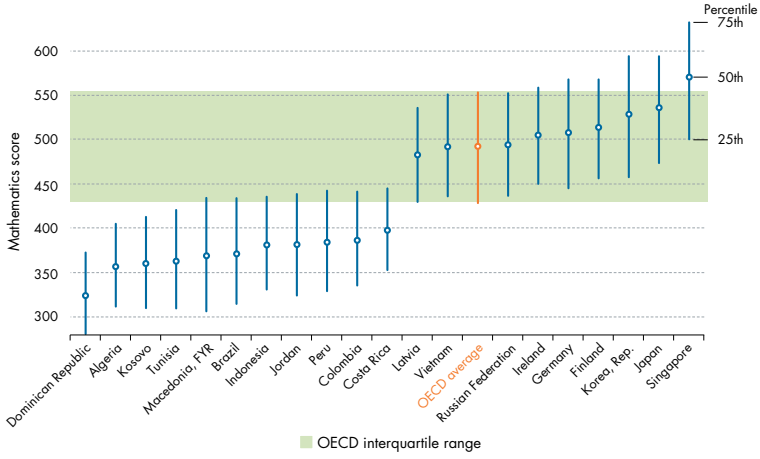
Sources: WDR 2018 team, using reading and mathematics data for Kenya and Uganda from Uwezo, Annual Assessment Reports, 2015 (http://www.uwezo.net/); reading and mathematics data for rural India from ASER Centre (2017); reading data for all other countries from U.S. Agency for International Development (USAID), Early Grade Reading Barometer, 2017, accessed May 30, 2017 (http://www.earlygradereadingbarometer.org/); and mathematics data for all other countries from USAID/RTI Early Grade Mathematics Assessment intervention reports, 2012–15 (https://shared.rti.org/sub-topic/early-grade-math-assessment-egma). Data at http://bit.do/WDR2018-Fig_O-1.

Note: These data typically pertain to selected regions in the countries and are not necessarily nationally representative. Data for India pertain to rural areas.

II. Learning levels are very poor - II

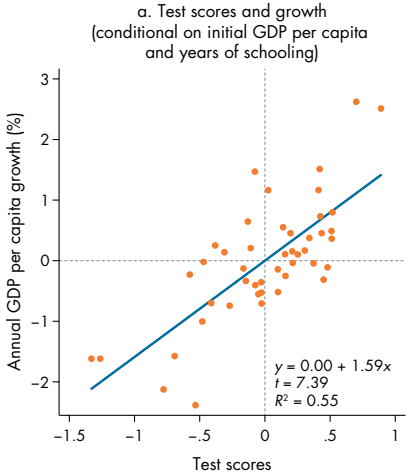
Figure O.2 **In several countries, the 75th percentile of PISA test takers performs below the 25th percentile of the OECD average**

Performance of 25th, 50th, and 75th percentiles in 2015 PISA mathematics assessment, selected countries



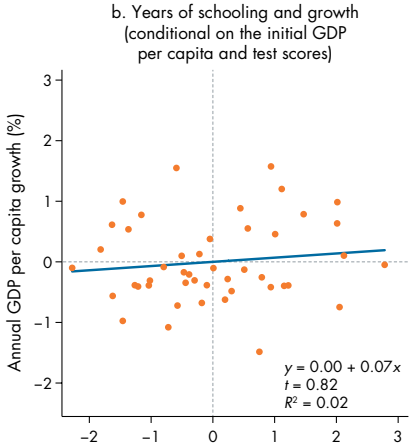
Source: WDR 2018 team, using data from Programme for International Student Assessment (PISA) 2015 (OECD 2016). Data at http://bit.do/WDR2018-Fig_O-2.

IIb. Quality not quantity matters for growth



Source: WDR 2018

IIb. Quality not quantity matters for growth

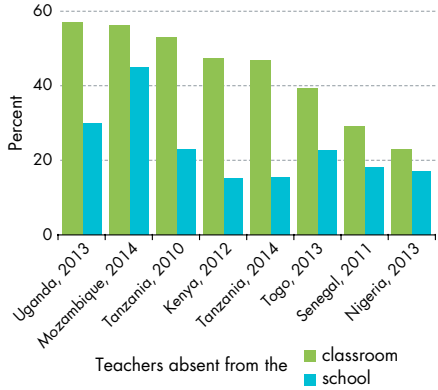


Source: WDR 2018

III. Teachers are often absent from schools and classrooms

Figure O.9 **In Africa, teachers are often absent from school or from classrooms while at school**

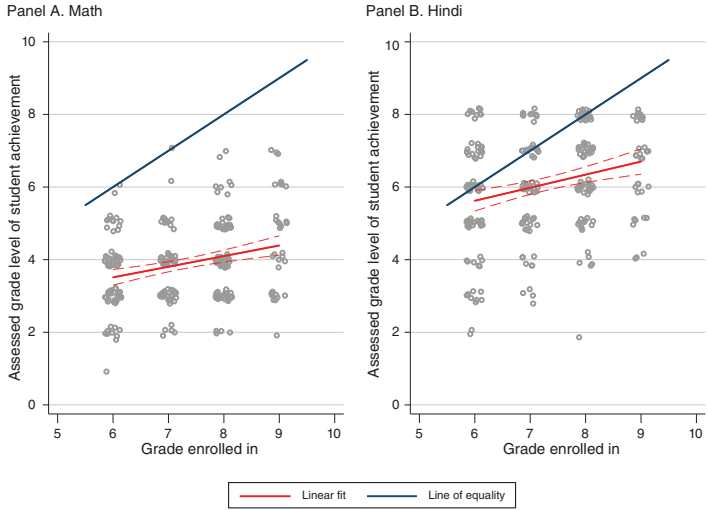
Percentage of teachers absent from school and from class on the day of an unannounced visit, participating countries



Source: Bold and others (2017). Data at http://bit.do/WDR2018-Fig_O-9.

Note: "Absent from the classroom" combines absences from school with absences from class among teachers who are at school. Data are from the

IV. Curricula and academic preparation are misaligned

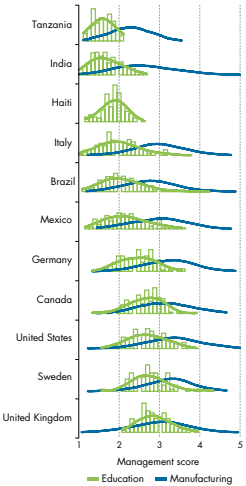


Muralidharan, Singh, and Ganimian (2019)

V. Management quality in schools is very poor

Figure O.10 Management capacity is low in schools in low- and middle-income countries

Distribution of management scores by sector, participating countries



Sources: Bloom and others (2014, 2015); Lemos and Scur (2016), with updates.
Data at http://bit.do/WDR2018-Fig_O-10.

Education RCTs

- Introduction
- Five stylized facts on education in developing countries
- Measuring learning outcomes
- Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)
- A quick review of classic papers
- Final remarks

Education RCTs

Introduction

Five stylized facts on education in developing countries

Measuring learning outcomes

Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)

A quick review of classic papers

Final remarks

Introduction

- Student learning is a core outcome for all education research
  - understanding mechanisms and processes is important but ultimate goal is improving student outcomes (e.g., earnings and test scores)

Introduction

- Student learning is a core outcome for all education research
  - understanding mechanisms and processes is important but ultimate goal is improving student outcomes (e.g., earnings and test scores)
- Measuring student achievement is central to education RCTs designs

Introduction

- Student learning is a core outcome for all education research
  - understanding mechanisms and processes is important but ultimate goal is improving student outcomes (e.g., earnings and test scores)
- Measuring student achievement is central to education RCTs designs
- What we will cove:
  - Objectives in test design
  - How we intend to score these tests
  - Implications of the above for test design and administration
  - Analysis of test scores
  - Practical issues in test implementation

Education RCTs
- Introduction
- Five stylized facts on education in developing countries
- Measuring learning outcomes
  - Objectives of test design
  - Implications for test design
  - Item Response theory
  - How should we make sense of test score impacts?
  - Analysis
- Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)
- A quick review of classic papers
  - The role of information
  - Vouchers
  - Public-Private Partnerships
  - Market-level dynamics
  - Remedial instruction with low-cost volunteers
  - Tracking
  - Personalized computer-aided instruction
- Final remarks

What does a good test look like?

Content Validity

A test is useful only if it is measuring the right things:

- The test content is **appropriate to the context**
  - Major need for piloting, adaptation of instruments

What does a good test look like?

Content Validity

A test is useful only if it is measuring the right things:

- The test content is **appropriate to the context**
  - Major need for piloting, adaptation of instruments
- The test measures **what we think it measures**
  - We want to measure learning, not test-taking skills or speed

What does a good test look like?

Content Validity

A test is useful only if it is measuring the right things:

- The test content is **appropriate to the context**
  - Major need for piloting, adaptation of instruments
- The test measures **what we think it measures**
  - We want to measure learning, not test-taking skills or speed
- The test is focused on dimensions that we think the intervention might improve
  - Requires thinking carefully about what kind of test domains we want to focus on
  - Also requires thinking about how the assessment might be 'gamed'

What does a good test look like?

Distribution and Discrimination

- The test should give us a **continuous well-distributed measure** of student achievement
  - No ceiling or floor effects
  - The test should not be “too easy”, “too hard” or “too short”
    - This Goldilocks zone can often be very hard to achieve!

What does a good test look like?

Distribution and Discrimination

- The test should give us a **continuous well-distributed measure** of student achievement
  - No ceiling or floor effects
  - The test should not be "too easy", "too hard" or "too short"
    - This Goldilocks zone can often be very hard to achieve!
- The test should be **discriminating** — i.e., informative at all levels of ability
  - Should be able to distinguish differences in absolute achievement around 10th percentile as well as around median ability

What does a good test look like?

Distribution and Discrimination

- The test should give us a **continuous well-distributed measure** of student achievement
  - No ceiling or floor effects
  - The test should not be "too easy", "too hard" or "too short"
    - This Goldilocks zone can often be very hard to achieve!
- The test should be **discriminating** — i.e., informative at all levels of ability
  - Should be able to distinguish differences in absolute achievement around 10th percentile as well as around median ability
  - This is often hard to do:
    - PISA, TIMSS etc. not informative at very low achievement levels
    - ASER not informative at high achievement levels

What does a good test look like?

Comparability and benchmarking

- Two different assessments, even measuring the same construct, are not comparable **unless they are designed to be**
  - An SD of achievement is **not the same thing** across contexts, test design and scoring methods!

What does a good test look like?

Comparability and benchmarking

- Two different assessments, even measuring the same construct, are not comparable **unless they are designed to be**
  - An SD of achievement is **not the same thing** across contexts, test design and scoring methods!
- We want to ideally target three levels of comparability

What does a good test look like?

Comparability and benchmarking

- Two different assessments, even measuring the same construct, are not comparable **unless they are designed to be**
  - An SD of achievement is **not the same thing** across contexts, test design and scoring methods!
- We want to ideally target three levels of comparability
  - **Dynamic comparability:** Learning is a cumulative process. Our test measure should be comparable over different rounds to allow us to study dynamics effectively

What does a good test look like?

Comparability and benchmarking

- Two different assessments, even measuring the same construct, are not comparable **unless they are designed to be**
  - An SD of achievement is **not the same thing** across contexts, test design and scoring methods!
- We want to ideally target three levels of comparability
  - **Dynamic comparability:** Learning is a cumulative process. Our test measure should be comparable over different rounds to allow us to study dynamics effectively
  - **Cross-sectional comparability:** Ideally, should be comparable to other studies (including our own!)

What does a good test look like?

Comparability and benchmarking

- Two different assessments, even measuring the same construct, are not comparable **unless they are designed to be**
  - An SD of achievement is **not the same thing** across contexts, test design and scoring methods!
- We want to ideally target three levels of comparability
  - **Dynamic comparability:** Learning is a cumulative process. Our test measure should be comparable over different rounds to allow us to study dynamics effectively
  - **Cross-sectional comparability:** Ideally, should be comparable to other studies (including our own!)
  - **Benchmarking:**Ideally, should be comparable to external metrics, such as TIMSS and PISA, so that we can benchmark our samples against global distributions.

Education RCTs
- Introduction
- Five stylized facts on education in developing countries
- Measuring learning outcomes
  - Objectives of test design
  - Implications for test design
  - Item Response theory
  - How should we make sense of test score impacts?
  - Analysis
- Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)
- A quick review of classic papers
  - The role of information
  - Vouchers
  - Public-Private Partnerships
  - Market-level dynamics
  - Remedial instruction with low-cost volunteers
  - Tracking
  - Personalized computer-aided instruction
- Final remarks

Practical implications for test design

Choosing items to administer

- Tests should contain items targeting a wide distribution of achievement
  - Grade-appropriate tests particularly inappropriate for our purposes!

Practical implications for test design

Choosing items to administer

- Tests should contain items targeting a wide distribution of achievement
  - Grade-appropriate tests particularly inappropriate for our purposes!
- Each item should map into a concrete skill that we want to test
  - Requires an item map at the very beginning!

Practical implications for test design

Choosing items to administer

- Tests should contain items targeting a wide distribution of achievement
  - Grade-appropriate tests particularly inappropriate for our purposes!
- Each item should map into a concrete skill that we want to test
  - Requires an item map at the very beginning!
- Subset of items should be repeated across rounds for comparability
  - Possible through separate out-of-sample linking but requires additional data collection

Practical implications for test design

Choosing items to administer

- Tests should contain items targeting a wide distribution of achievement
  - Grade-appropriate tests particularly inappropriate for our purposes!
- Each item should map into a concrete skill that we want to test
  - Requires an item map at the very beginning!
- Subset of items should be repeated across rounds for comparability
  - Possible through separate out-of-sample linking but requires additional data collection
- Subset of items should be drawn from other assessments
  - To allow for comparability across tests (although this could fail)

Practical implications for test design

Choosing how to administer the test

- Tests may be administered in a variety of formats:
  - Individually-administered by the interviewer
  - Group-oral administration
  - Written administration

Practical implications for test design

Choosing how to administer the test

- Tests may be administered in a variety of formats:
  - Individually-administered by the interviewer
  - Group-oral administration
  - Written administration
- There are clear advantages and disadvantages to all of these:
  - Individual oral much better for assessing children at young ages but very burdensome in the field
  - Group oral attempts to replicate above at scale but classroom management is very difficult, answers less precise
  - Written tests are ideal for later grades but with a strong possibility of floor effects in primary grades

Practical implications for test design

Choosing how to administer the test

- Tests may be administered in a variety of formats:
  - Individually-administered by the interviewer
  - Group-oral administration
  - Written administration
- There are clear advantages and disadvantages to all of these:
  - Individual oral much better for assessing children at young ages but very burdensome in the field
  - Group oral attempts to replicate above at scale but classroom management is very difficult, answers less precise
  - Written tests are ideal for later grades but with a strong possibility of floor effects in primary grades
- Balancing across these is strongly influenced by fieldwork logistics
  - But results of an inappropriate choice will plague for a long time...

Constructing aggregate test scores

- Tests administer a sequence of single items. Aggregating these into a test metric involves important choices.
- There are four common ways of seeing test scores reported:

Constructing aggregate test scores

- Tests administer a sequence of single items. Aggregating these into a test metric involves important choices.
- There are four common ways of seeing test scores reported:
  - Binary/Categorical against a benchmark:
    - Pass/fail in official exams; ASER/EGRA type categories
  - Percentage correct
  - Internally normalized standard deviations:
    - Typically within-grade and within-test booklet
  - Item Response Theory (IRT) linked scale scores with common normalizations across overlapping assessments
    - Probably the most desirable but with much more prepwork needed before and much analysis after!

Education RCTs
- Introduction
- Five stylized facts on education in developing countries
- Measuring learning outcomes
  - Objectives of test design
  - Implications for test design
  - Item Response theory
  - How should we make sense of test score impacts?
  - Analysis
- Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)
- A quick review of classic papers
  - The role of information
  - Vouchers
  - Public-Private Partnerships
  - Market-level dynamics
  - Remedial instruction with low-cost volunteers
  - Tracking
  - Personalized computer-aided instruction
- Final remarks

Item Response Theory — A basic introduction

- Item Response Theory (IRT) is a statistical tool for designing, validating and analyzing tests
  - Decades long history in education and psychometrics – GRE, GMAT, SAT, NAEP, TIMSS

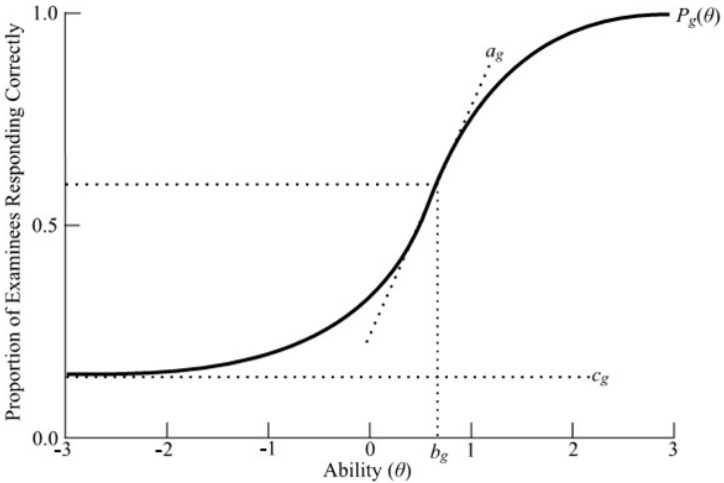
Item Response Theory — A basic introduction

- Item Response Theory (IRT) is a statistical tool for designing, validating and analyzing tests
  - Decades long history in education and psychometrics – GRE, GMAT, SAT, NAEP, TIMSS
- The basic idea: The focus of IRT is at the item level.
  - Models the probability that an individual with given ability will get an item right
  - The overall ability estimate (test score) generated by analyzing an individual's response to different items each defined by their own characteristics

Item Response Theory — A basic introduction

- Item Response Theory (IRT) is a statistical tool for designing, validating and analyzing tests
  - Decades long history in education and psychometrics – GRE, GMAT, SAT, NAEP, TIMSS
- The basic idea: The focus of IRT is at the item level.
  - Models the probability that an individual with given ability will get an item right
  - The overall ability estimate (test score) generated by analyzing an individual's response to different items each defined by their own characteristics
- Many advantages (see e.g. Das and Zajonc, 2010):
  - Most importantly (for me) the ability to link
  - But also much better diagnostics for cross-cultural comparisons
  - Less arbitrary than summing up correct responses

Item Characteristic Curve



3 Parameter Logistic (3PL) Model

Item Response Function:

$$P_g(\theta_i) = c_g + \frac{1 - c_g}{1 + exp(-1.7.a_g.(\theta_i - b_g))}$$

- c_g is the pseudo-guessing parameter - with multiple choice questions, even the lowest ability can get some answers right. Set to zero for non-MCQ to get 2PL model
- b_g is the difficulty parameter - the level at which the probability of getting item right is 0.5 in 2 PL
- a_g is the discrimination parameter - slope of the ICC at b – how quickly the likelihood of success changes with respect to ability

Item Response Theory — Core Assumptions

1. Unidimensionality - A single latent individual-specific trait determines performance on the test

2. No Differential Item Functioning: Implicit in ICC, item characteristics are person-invariant

    2.1 particularly important in cross-cultural settings

3. (Conditional) local independence:

    3.1 Item responses are independent across individuals (no cheating!)

    3.2 Conditional on ability, item responses are locally independent across questions for the same individual

Under these assumptions, can recover estimates of ability and item characteristics given matrix of responses by individuals

Item Response Theory — How does linking work?

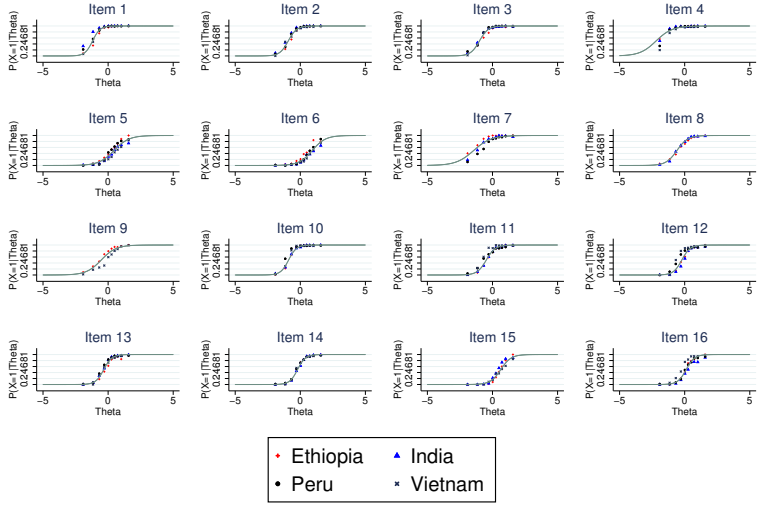
- Item characteristics are fixed and can be used to link across samples
  - Common items serve as ‘anchors’ which bring two assessments on a common scale
  - only a subset of items need to be common

Item Response Theory — How does linking work?

- Item characteristics are fixed and can be used to link across samples
  - Common items serve as ‘anchors’ which bring two assessments on a common scale
  - only a subset of items need to be common
- Without sufficient common items:
  - Still can do IRT but it’s like having temperature data in Celsius and Fahrenheit across rounds (and not knowing what the transformation is!)
  - You cannot make statements about whether students know more or less than they knew before!

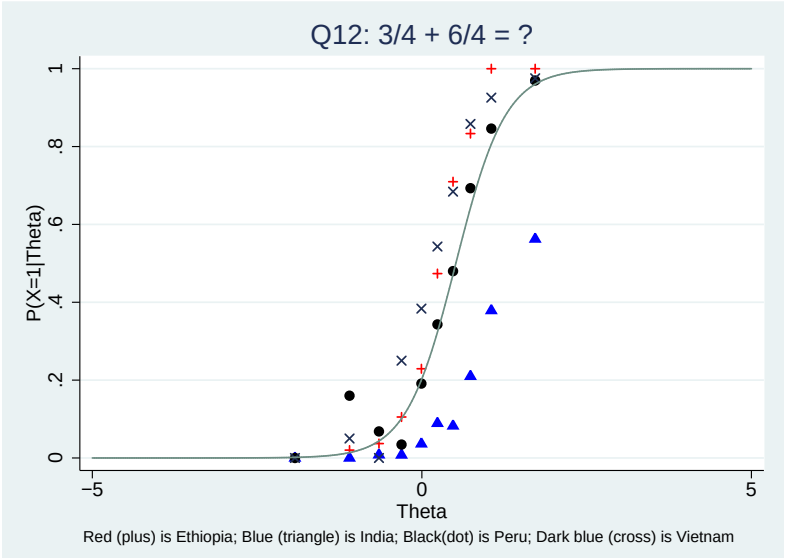
When linking works (mostly) well

No Differential Item Functioning (DIF)



When linking works less well

Clear DIF

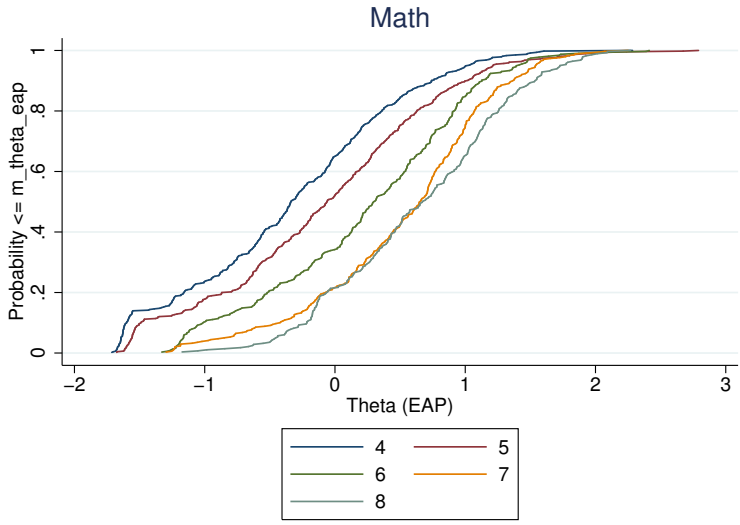


Steps for estimating IRT scores

- I use OpenIRT suite of commands in Stata
  - Better than the native IRT commands in Stata 14
- Step 1: Create a pooled dataset across all samples to be linked
- Step 2: Run the OpenIRT commands
  - 3 PL models for MCQ, 2 PL for open-ended responses
- Step 3: Generate Item Characteristic Curves
  - Do items fit reasonably well?
- Step 4: Generate DIF graphs
  - Do items perform similarly across linked samples?
- Step 5: If DIF is found, split items in assessment
  - Rerun Step 2-Step 5
  - Repeat until satisfactory diagnostics (or give up!)

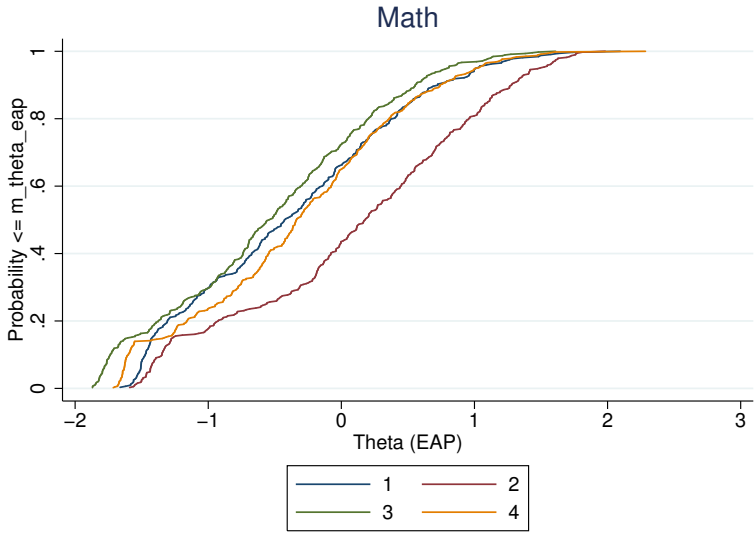
Distributions of student achievement

When test design and linking provides reasonable results...



Distributions of student achievement

When test design and linking works less well...



Education RCTs
- Introduction
- Five stylized facts on education in developing countries
- Measuring learning outcomes
  - Objectives of test design
  - Implications for test design
  - Item Response theory
  - How should we make sense of test score impacts?
  - Analysis
- Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)
- A quick review of classic papers
  - The role of information
  - Vouchers
  - Public-Private Partnerships
  - Market-level dynamics
  - Remedial instruction with low-cost volunteers
  - Tracking
  - Personalized computer-aided instruction
- Final remarks

How should we make sense of test score impacts?

- By itself, very little!
  - For a normal distribution, gives you a move from the median to the 66th percentile
  - But we don't even know for most studies whether tests are near-normal
- And it depends on the test and the sample
- This is a rant for another time:
  - Go Abhijeet Singh's blog post from a few years ago on Development Impact blog: `https://blogs.worldbank.org/impactevaluations/how-standard-standard-deviation-cautionary-note-using-sds-compare-across-impact-evaluations`

Education RCTs
- Introduction
- Five stylized facts on education in developing countries
- Measuring learning outcomes
  - Objectives of test design
  - Implications for test design
  - Item Response theory
  - How should we make sense of test score impacts?
  - Analysis
- Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)
- A quick review of classic papers
  - The role of information
  - Vouchers
  - Public-Private Partnerships
  - Market-level dynamics
  - Remedial instruction with low-cost volunteers
  - Tracking
  - Personalized computer-aided instruction
- Final remarks

Due diligence

- Probably one of the more important things to do: do the test results seem to make sense?
  - Are they well-distributed?
  - Do they increase over time?
  - Do they display sensible inter-period correlations?
  - Do they display sensible inter-subject correlations?
  - Do they display sensible correlations with wealth and parental education?
- This is akin to a "sniff test" for your test metrics: if it smells fishy, it deserves digging deeper!

Ordinality of test scores

- The fundamental problem is that ability has no natural metric
  - Test scores are only a proxy and inherently ordinal
  - They present only a rank-ordering of individuals and so any rank-preserving transformation is a valid measure
- Typically ignored in most applied education work but potentially very serious consequences when looking at inter-group differences and trends
  - See e.g. Bond and Lang (2013), Nielsen (2014a, b), Neal (2006)
- What can you do?
  - Look at the full distribution (CDFs, kernel densities) in addition to mean
  - Look non-parametrically at learning dynamics
  - Not guaranteed to give an answer you like, but very powerful when it does

Education RCTs

- Introduction
- Five stylized facts on education in developing countries
- Measuring learning outcomes
- Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)
- A quick review of classic papers
- Final remarks

Introduction

Five stylized facts on education in developing countries

Measuring learning outcomes

Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)

A quick review of classic papers

Final remarks

A quick review

- Spot checks on whether the school is opened (!)
- Spot checks on teacher attendance (paper rosters tend to be wrong)
- Spot checks on student attendance (paper rosters tend to be wrong)
- Household investment in education (financial and time)
- Classroom observations: Measure time on-task (within the class) and pedagogical practices
  - Stalling classroom observation
  - TEACH from the world bank

Education RCTs

- Introduction
- Five stylized facts on education in developing countries
- Measuring learning outcomes
- Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)
- A quick review of classic papers
- Final remarks

Education RCTs

Introduction

Five stylized facts on education in developing countries

Measuring learning outcomes

Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)

A quick review of classic papers

Final remarks

Education RCTs
Introduction
Five stylized facts on education in developing countries
Measuring learning outcomes
Objectives of test design
Implications for test design
Item Response theory
How should we make sense of test score impacts?
Analysis
Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)
A quick review of classic papers
The role of information
Vouchers
Public-Private Partnerships
Market-level dynamics
Remedial instruction with low-cost volunteers
Tracking
Personalized computer-aided instruction
Final remarks

The role of information in household decision making — Jensen (2010, QJE)

- Starts from a premise that there are high returns to schooling in developing countries
  - Setting: Dominican Republic, 2001
  - Secondary school completers earn 41% more than primary school completers
  - Implied returns 8% per year, similar to Duflo (2001)

The role of information in household decision making — Jensen (2010, QJE)

- Starts from a premise that there are high returns to schooling in developing countries
  - Setting: Dominican Republic, 2001
  - Secondary school completers earn 41% more than primary school completers
  - Implied returns 8% per year, similar to Duflo (2001)
- Makes the crucial point that, for HH investment decisions, what matters are **perceived** returns
  - And it is not clear that households in dev countries have accurate information
- In that case, providing correct information may lead to substantial changes in edu investments
  - Attractive for policy since this is easily scaleable, low marginal cost

The experiment — Providing information to middle-schoolers

- Targets male students in Grade 8, the last year of compulsory schooling
  - Representative sample from 30 largest towns and cities
- Elicits perceptions of returns to education:
  - For the individuals themselves when 30-40, if they completed current school year/secondary/university
  - For adult men between 30-40, if they completed primary/secondary/university
- **Intervention:**At the end of the survey, each respondent in randomly selected **schools** was given info about the mean differences in the earning levels of adult men with primary/secondary/university
- Collects data for the next 4 years, till 2005, to actually see if students completed secondary schooling

Do perceived returns predict schooling?

	Panel A. Round 1 implied perceived returns (control group only)					
	(1)	(2)	(3)	(4)	(5)	(6)
	Returned next year	Returned next year	Finished school	Finished school	Years of schooling	Years of schooling
Implied perceived returns	0.11***	0.083**	0.14***	0.092**	0.53***	0.37**
	(0.030)	(0.034)	(0.036)	(0.038)	(0.13)	(0.14)
Log (inc. per capita)		0.090		0.25***		0.76***
		(0.062)		(0.063)		(0.24)
School performance		0.015		0.015		0.093**
		(0.014)		(0.011)		(0.045)
Father finished secondary		0.036		−0.014		0.045
		(0.041)		(0.044)		(0.16)
Age		−0.017		0.006		−0.045
		(0.024)		(0.025)		(0.093)
R^2	.008	.016	.017	.048	.016	.042
Observations	1,003	1,003	1,003	1,003	918	918

Effect on perceived returns and schooling: Full Sample

	Full sample			
	(1) Returned next year	(2) Finished school	(3) Years of schooling	(4) Perceived returns
Treatment	0.041*	0.023	0.20**	367***
	(0.023)	(0.020)	(0.082)	(28)
Log	0.095**	0.23***	0.79***	29.0
(inc. per capita)	(0.040)	(0.044)	(0.16)	(47)
School	0.011	0.019**	0.086**	0.74
performance	(0.010)	(0.009)	(0.034)	(14)
Father	0.074**	0.050*	0.26**	−24
finished sec.	(0.030)	(0.030)	(0.12)	(32)
Age	−0.010	0.004	−0.006	−42*
	(0.016)	(0.015)	(0.059)	(21)
R^2	.016	.040	.049	.090
Observations	2,241	2,205	2,074	1,859

Some further considerations

- The "returns" communicated to students are differences in mean earnings, not causal estimates
- Also, the average return might not be informative of what I should expect the return to look like for me
  - "essential heterogeneity": Could be systematic, e.g. by race or location, which means info might be systematically misleading
  - Jensen has very thoughtful responses (see footnotes 22, 23)
  - A great example of interpretational issues that crop up in any realistic policy experiment
    - (And the trade-offs needed between simplicity of implementation and an "optimal" design)
  - But clearly, heterogeneous returns could be super-imp: e.g. Munshi & Rosenzweig (2006, AER), Jensen (2012, QJE)
- Dizon-Ross (2019) follows in the spirit of Jensen (2001) and looks at the investments in individual children in levels and appropriate investments

Education RCTs
Introduction
Five stylized facts on education in developing countries
Measuring learning outcomes
Objectives of test design
Implications for test design
Item Response theory
How should we make sense of test score impacts?
Analysis
Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)
A quick review of classic papers
The role of information
Vouchers
Public-Private Partnerships
Market-level dynamics
Remedial instruction with low-cost volunteers
Tracking
Personalized computer-aided instruction
Final remarks

Comparing government and private schools — Muralidharan and Sundaraman (2015, QJE)

School and Teacher Characteristics

	(1) Private schools	(2) Public schools	(3) Difference
Panel A: School characteristics			
Total enrollment	296.21	74.04	222.17***
Total working days	229.81	218.66	11.15***
Pupil-teacher ratio	17.62	25.28	−7.67***
Drinking water available	0.99	0.92	0.07***
Functional toilets	0.86	0.68	0.18***
Separate functional toilets for girls	0.77	0.40	0.37***
Functional electricity	0.88	0.61	0.28***
Functional computers	0.52	0.05	0.48***
Functional library	0.80	0.97	−0.18***
Functional radio	0.13	0.81	−0.68***
Observations	289	346	
Panel B: Teacher characteristics			
Male	0.24	0.46	−0.21***
Age	27.58	40.00	−12.42***
Years of teaching	5.14	14.96	−9.82***
Completed at least college or masters	0.69	0.88	−0.19***
Teacher training completed	0.34	0.99	−0.65***
Come from the same village	0.44	0.13	0.32***
Current gross salary per month (Rs)	2,606.66	14,285.94	−11,679.27***
Observations	2,000	1,358	
Panel C: School expenditures			
Annual cost per child (Rs/child)	1,848.88	8,390.00	−6,542***
Observations	311	335	

Comparing government and private schools

Teacher and School Effort

	(1) Private schools	(2) Public schools	(3) Difference
Panel A: Measures of classroom activity			
Class is engaged in active teaching	0.51	0.34	0.17***
A teacher is present in class	0.97	0.92	0.048***
Teacher is effective in teaching and maintaining discipline	0.50	0.36	0.14***
Teacher has complete control over class	0.69	0.41	0.28***
Teachers teaching mutliple classes at the same time	0.24	0.79	−0.55***
Observations	2,738	2,784	
Panel B: Measures of teacher activity			
Teacher is absent	0.09	0.24	−0.15***
Teacher is actively teaching	0.50	0.35	0.15***
Teacher is in school and not teaching	0.01	0.03	−0.02***
Observations	6,577	5,552	
Panel C: Measures of school hygiene			
Flies heavily present on premises of the school	0.14	0.19	−0.05**
Stagnant water present on premises of the school	0.18	0.28	−0.10***
Garbage dumped on premises of the school	0.33	0.44	−0.11***
Observations	426	614	

Experimental design

Panel A: Treatment Villages

Group 1T	Group 2T	Group 3T	Group 4T
Non-Applicants in Public Schools	Applicants in Public Schools NOT awarded a Voucher	Applicants in Public Schools AWARDED a Voucher	Non-voucher Students in Private Schools

Panel B: Control Villages

Group 1C	Group 2C	Group 3C	Group 4C
Non-Applicants in Public Schools	Applicants in Public Schools NOT awarded a Voucher	Does not exist	Non-voucher Students in Private Schools

Impact after 2-4 years

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Year 2 assessments				Year 4 assessments						
	Telugu score	Math score	English score	Combined across tests	Telugu score	Math score	English score	EVS score	Combined across tests excluding Hindi	Hindi score	Combined across tests
Panel A: Impact of winning a voucher (intention to treat effects)											
Offered voucher	−0.079	−0.053	0.185**	0.016	−0.017	−0.031	0.116*	0.083	0.036	0.545***	0.133***
	(0.055)	(0.065)	(0.079)	(0.061)	(0.051)	(0.052)	(0.070)	(0.060)	(0.048)	(0.068)	(0.045)
Total observations	4,620	4,620	4,525	13,765	4,385	4,385	4,217	4,243	17,230	1,696	18,926
Treatment observations	1,778	1,778	1,738	5,294	1,674	1,675	1,607	1,628	6,584	867	7,451
Control observations	2,842	2,842	2,787	8,471	2,711	2,710	2,610	2,615	10,646	829	11,475
Panel B: Average treatment on the treated (ATT) effect of attending a private school (scaling up intention to treat effect by inverse of voucher take-up rate)											
Voucher recipient in private school	−0.156	−0.104	0.364**	0.032	−0.033	−0.061	0.229*	0.164	0.071	1.074***	0.262***
	(0.108)	(0.128)	(0.156)	(0.120)	(0.100)	(0.102)	(0.138)	(0.118)	(0.095)	(0.134)	(0.089)
Total observations	4,620	4,620	4,525	13,765	4,385	4,385	4,217	4,243	17,230	1,696	18,926
Voucher recipients	997	997	982	5,294	945	946	911	920	6,584	510	7,451
Nonrecipients	3,623	3,623	3,543	8,471	3,440	3,439	3,306	3,323	10,646	1,186	11,475

Putting results in context — School time-tables

TABLE VII

School Time Use: Instructional Time by Subject (Minutes per Week)

	(1) Private schools	(2) Public schools	(3) Difference
Telugu	307.72	511.52	−203.81***
	(6.36)	(3.60)	(6.99)
Math	339.75	500.69	−160.94***
	(7.50)	(3.36)	(8.63)
English	322.68	235.52	87.17***
	(7.96)	(5.39)	(9.69)
Social studies	239.21	173.24	65.96***
	(6.29)	(6.89)	(9.84)
Science	205.52	104.58	100.94***
	(9.09)	(5.78)	(9.44)
Hindi	215.78	0.01	215.77***
	(6.08)	(0.89)	(6.41)
Moral science	16.85	20.11	−3.26
	(4.82)	(3.20)	(5.56)
Computer use	46.7	0.51	46.19***
	(6.50)	(1.02)	(6.80)
Other	311.66	250.29	61.37***
	(14.55)	(6.70)	(16.20)
Total instructional time	2,005.87	1,796.47	209.4***
	(13.73)	(6.86)	(14.46)
Break	461	473.18	−12.18
	(9.14)	(3.05)	(10.58)
Total school time	2,466.87	2,269.65	197.22***
	(17.46)	(8.25)	(19.79)
Observations	325	200	

Summarizing results

- Private schools have little evidence of doing better in Math or Telugu
  - Do better consistently in English and Hindi
  - Hindi scores are explained by the longer instructional time
  - Overall, no sign that private schooling alone will make a big dent in the learning crisis

Summarizing results

- Private schools have little evidence of doing better in Math or Telugu
  - Do better consistently in English and Hindi
  - Hindi scores are explained by the longer instructional time
  - Overall, no sign that private schooling alone will make a big dent in the learning crisis
- Private schools are more productive though
  - same achievement delivered in math and Telugu but with lower instructional time
  - Delivered at a fraction of per-pupil spending in govt schools
  - Rao (2015), shows important effects on social outcomes

Summarizing results

- Private schools have little evidence of doing better in Math or Telugu
  - Do better consistently in English and Hindi
  - Hindi scores are explained by the longer instructional time
  - Overall, no sign that private schooling alone will make a big dent in the learning crisis
- Private schools are more productive though
  - same achievement delivered in math and Telugu but with lower instructional time
  - Delivered at a fraction of per-pupil spending in govt schools
  - Rao (2015), shows important effects on social outcomes
- The big open question: Can pvt schools deliver much higher gains at same cost?

Education RCTs
Introduction
Five stylized facts on education in developing countries
Measuring learning outcomes
Objectives of test design
Implications for test design
Item Response theory
How should we make sense of test score impacts?
Analysis
Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)
A quick review of classic papers
The role of information
Vouchers
Public-Private Partnerships
Market-level dynamics
Remedial instruction with low-cost volunteers
Tracking
Personalized computer-aided instruction
Final remarks

The logic of PPPs - Romero, Sandefur, Sandholtz (AER, 2020)

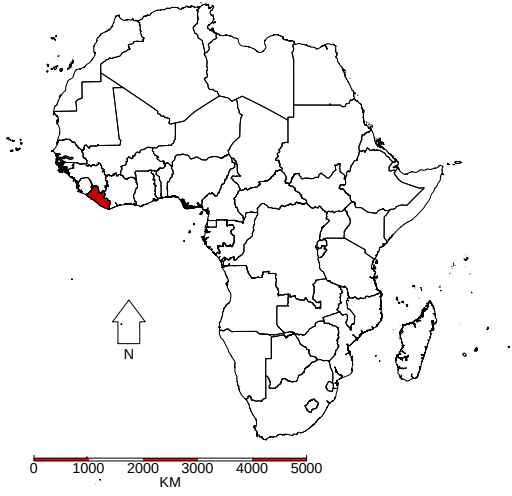
- Overcome **efficiency-equity trade-off**:
  - Efficiency: Private schools are on average better managed than public schools
  - Equity: Fee-charging private schools may increase inequality and sorting

The logic of PPPs - Romero, Sandefur, Sandholtz (AER, 2020)

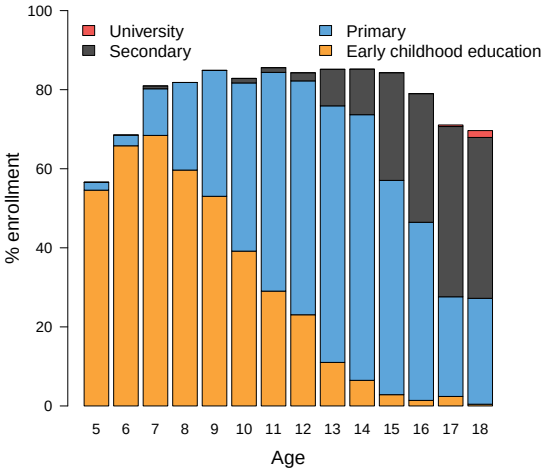
- Overcome **efficiency-equity trade-off**:
  - Efficiency: Private schools are on average better managed than public schools
  - Equity: Fee-charging private schools may increase inequality and sorting
- Overcome **financing constraints**:
  - Governments enter PPPs in large-part to raise capital
  - NB: impacts necessarily include resource and efficiency effects

The logic of PPPs - Romero, Sandefur, Sandholtz (AER, 2020)

- Overcome **efficiency-equity trade-off**:
  - Efficiency: Private schools are on average better managed than public schools
  - Equity: Fee-charging private schools may increase inequality and sorting
- Overcome **financing constraints**:
  - Governments enter PPPs in large-part to raise capital
  - NB: impacts necessarily include resource and efficiency effects
- Contractors have incentives to cut quality on non-contracted/non-monitored processes/outcomes

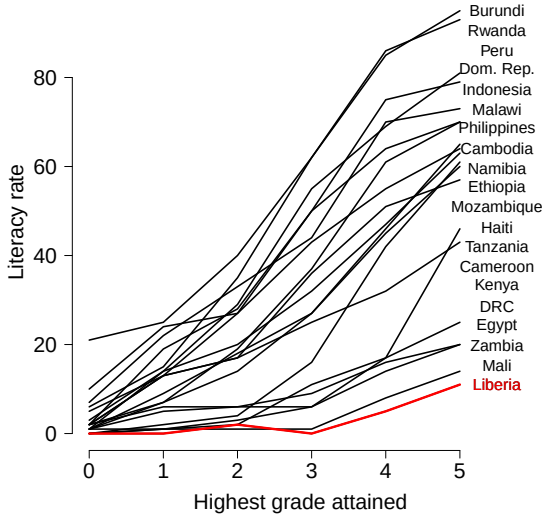

N
0
1000
2000
3000
4000
5000
KM

Low enrollment and backlog of overage children



Note: Based on 2014 Household Income and Expenditures Survey.

Schooling $\neq$ learning



Source: Oye, Pritchett, and Sandefur (2016)

FT Magazine + Add to myFT

Liberia is outsourcing education. Can it work?

FT Magazine + Add to myFT

Liberia is outsourcing education. Can it work?

The Economist Subscribe Welcome

Ashes to classes

Liberia's bold experiment in school reform

A war-scorched state where almost nothing works tries charter schools

Print edition | Middle East and Africa ›
Feb 23rd 2017 | MONROVIA

TIONS HOME SEARCH

FT Magazine + Add to myFT

Liberia is outsourcing education. Can it work?

The Economist Subscribe Welcome

The New York Times

The Opinion Pages

hool

Liberia, Desperate to Educate, Turns to Charter Schools

arter schools

Tina Rosenberg
FIXES JUNE 14, 2016

Print edition | Middle East and Africa ›
Feb 23rd 2017 | MONROVIA

What are "Partnership Schools for Liberia"?

- 93 schools
- free
- non-selective
- staffed by teachers on government payroll
- and managed by 8 private contractors
- with a $50 per pupil subsidy (+ fundraising)

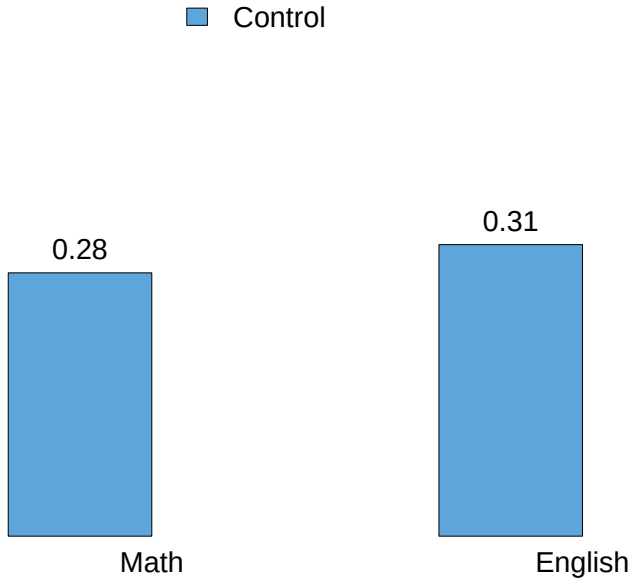
8 Private providers

- 5 are nonprofit
- 3 are local
- 6 contracted through competitive tender

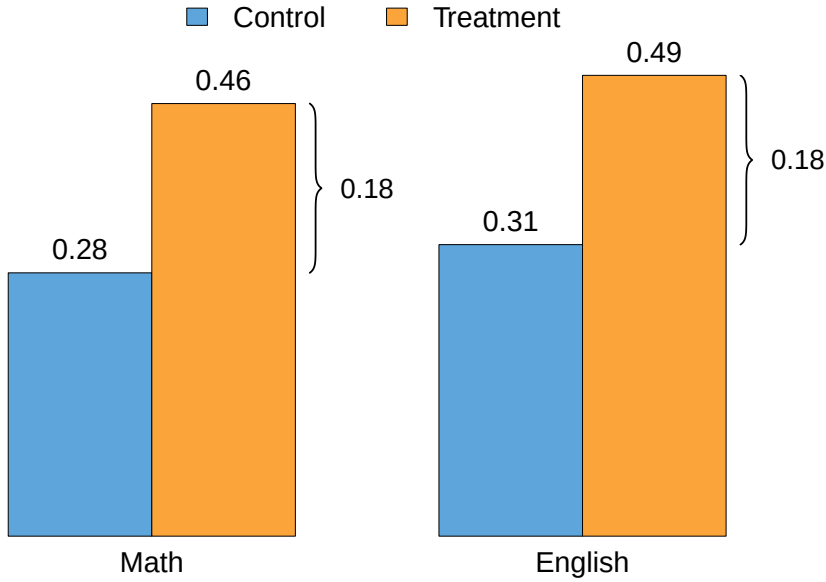
Test scores increased by .18σ

	Second wave (9-10 months after treatment)		
	ITT		ToT
	(1)	(2)	(3)
English	0.17***	0.18***	0.21***
	(0.04)	(0.03)	(0.04)
Math	0.19***	0.18***	0.22***
	(0.04)	(0.03)	(0.04)
Abstract	0.05	0.05	0.06
	(0.04)	(0.04)	(0.05)
Composite	0.19***	0.18***	0.22***
	(0.04)	(0.03)	(0.04)
Controls	No	Yes	Yes
Observations	3,492	3,492	3,492

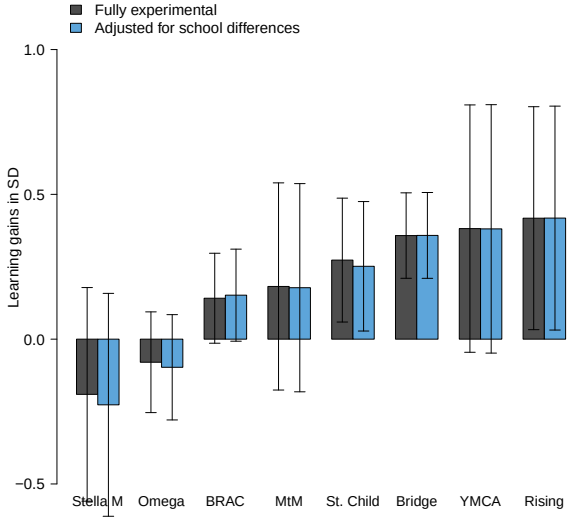
"Business as usual" learning is .3σ per academic year



Treatment is roughly $\sim$0.62 extra years of schooling

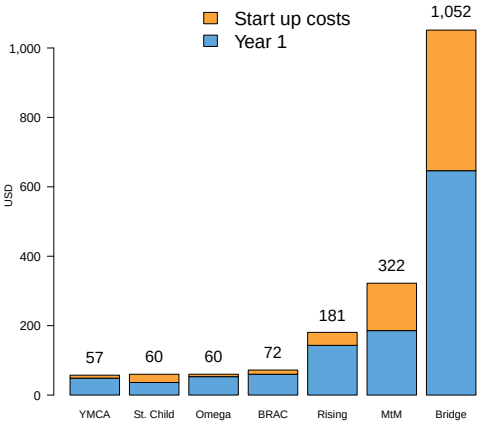


Learning outcomes by provider

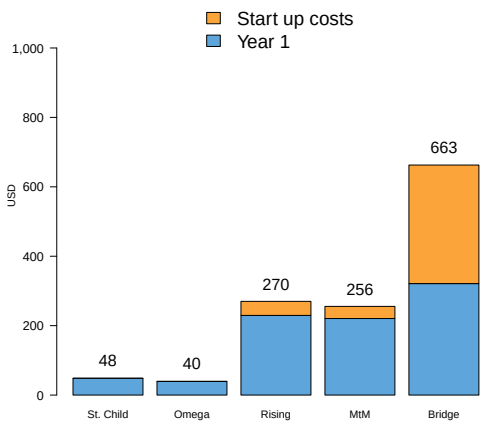


Cost per pupil varies across providers

Ex ante budget per pupil



Ex post cost per pupil



Education RCTs
- Introduction
- Five stylized facts on education in developing countries
- Measuring learning outcomes
  - Objectives of test design
  - Implications for test design
  - Item Response theory
  - How should we make sense of test score impacts?
  - Analysis
- Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)
- A quick review of classic papers
  - The role of information
  - Vouchers
  - Public-Private Partnerships
  - Market-level dynamics
  - Remedial instruction with low-cost volunteers
  - Tracking
  - Personalized computer-aided instruction
- Final remarks

Studying educational markets — Andrabi, Das and Khwaja (2017, AER)

- The typical household in many developing countries faces a choice between many providers of government and private schools
  - These schools differ on various characteristics, inputs, and prices charged, which are set endogenously
  - Unlike OECD economies, degree of effective regulation on the private sector is relatively low
  - Household demand responds to external information, income etc.

Studying educational markets — Andrabi, Das and Khwaja (2017, AER)

- The typical household in many developing countries faces a choice between many providers of government and private schools
  - These schools differ on various characteristics, inputs, and prices charged, which are set endogenously
  - Unlike OECD economies, degree of effective regulation on the private sector is relatively low
  - Household demand responds to external information, income etc.
- An important q is how **markets** respond to information
  - Educational markets are typically islands, especially in rural areas
  - Offers interesting possibilities for research more generally
- This is the main focus of Andrabi et al. (2017)
  - Also, a good intro to the LEAPS study in Pakistan
  - Major research undertaking w/ non-experimental and experimental work
  - Also major inputs to policy, advances in measurement

Setting: rural Punjab (Pakistan)

- 112 villages in 3 districts of Punjab province in Pakistan
  - Each village an effectively closed market
  - On average, 7.3 schools: 4.4 (sex-segregated) public schools, 2.9 co-ed private schools
- Annual surveys in these villages from 2004
  - Testing of students in all schools
  - Teacher and HM interviews
  - Parent interviews, hh surveys
- Wide variation in test scores within village, across schools, in fees
- Strong indications that the market is reasonably competitive

The intervention

- Tested all children in Grade 3 in all schools in the sample
- Experimentally allocated one-half of villages (within-district stratification) to receive report cards on **child and school performance**
  - Reported raw scores for English, math, Urdu for the child on first page with quintile rank
  - Reported scores for all the schools in the village, with quintile rank, and num of children
  - Report cards distributed to schools and parents at a school meeting

The intervention

- Tested all children in Grade 3 in all schools in the sample
- Experimentally allocated one-half of villages (within-district stratification) to receive report cards on **child and school performance**
  - Reported raw scores for English, math, Urdu for the child on first page with quintile rank
  - Reported scores for all the schools in the village, with quintile rank, and num of children
  - Report cards distributed to schools and parents at a school meeting
- The focus of the paper is on market-level impacts
  - Fees, test scores, enrollment and switching
  - Heterogeneity in effects across schools/children by baseline characteristics

Fees and Test score impacts

TABLE 3—FEE AND TEST SCORES: IMPACT ON MARKET OUTCOMES

	Village average fees (Year 2)			Village average test scores		
	School report		Household report			
	Basic (1)	Weighted by children (2)	Basic (3)	Year 2 (4)	Year 3 (5)	Year 2 (same kids) (6)
Panel A. No controls						
Report card	−288.4	−334.1	−193.9	0.128	0.140	0.129
	(92.58)	(107.9)	(99.97)	(0.0624)	(0.0584)	(0.0599)
Observations	104	104	83	112	112	112
R^2	0.336	0.473	0.259	0.328	0.292	0.399

Enrollment and switching

TABLE 4—ENROLLMENT AND SWITCHING: IMPACT ON MARKET OUTCOMES

	Village enrollment (Year 2)			Village average test scores: same kids, no switchers (Year 2)
	Primary enrollment rate	Switching rate (tested cohort only)	Dropout rate (tested cohort only)	
	(1)	(2)	(3)	(4)
Panel A. No controls				
Report card	0.0390	0.009	0.009	0.129
	(0.0263)	(0.007)	(0.006)	(0.0608)
Observations	112	112	112	112
R^2	0.473	0.0561	0.377	0.397

- Not presenting the results on heterogeneity here, but definitely worth taking a look

Education RCTs
Introduction
Five stylized facts on education in developing countries
Measuring learning outcomes
Objectives of test design
Implications for test design
Item Response theory
How should we make sense of test score impacts?
Analysis
Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)
A quick review of classic papers
The role of information
Vouchers
Public-Private Partnerships
Market-level dynamics
Remedial instruction with low-cost volunteers
Tracking
Personalized computer-aided instruction
Final remarks

Remedial Education with low-cost volunteers

Banerjee et al. (2007, QJE); Design

- Low-cost volunteers used for instruction in groups of 15-20 for 2 hours per day
- RCT in 2 cities, randomized at school*grade level

	Year 1 (2001-2002)		Year 2 (2002-2003)		Year 3 (2003-2004)	
	Grade 3	Grade 4	Grade 3	Grade 4	Grade 3	Grade 4
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: Vadodara						
A.1 Balsakhi						
Group A	Balsakhi	No balsakhi	No Balsakhi	Balsakhi	No Balsakhi	No Balsakhi
(5,264 students in 49 schools in year 1; 6,071 students in 61 schools in year 2)						
Group B	No balsakhi	Balsakhi	Balsakhi	No Balsakhi	No Balsakhi	No Balsakhi
(4934 students in 49 schools in year 1; 6,344 students in 61 schools in year 2)						
A.2 Computer Assisted Learning (CAL)						
Group A1B1	No CAL	No CAL	No CAL	CAL	No CAL	No CAL
(2,850 students in 55 schools in year 2; 2,814 students in 55 schools in year 3)						
Group A2B2	No CAL	No CAL	No CAL	No Cal	No CAL	CAL
(3,095 students in 56 schools in year 2; 3,131 students in 56 schools in year 3)						
PANEL B: Mumbai						
Balsakhi						
Group C	Balsakhi	No Balsakhi	No Balsakhi	Balsakhi	No Balsakhi	No Balsakhi
(2,592 students in 32 schools in year 1; 5,755 students in 38 schools in year 2)						
Group D	No Balsakhi	No Balsakhi	Balsakhi	No Blasakhi	No Balsakhi	No Balsakhi
(2,182 students in 35 schools year 1; 4,990 students in 39 schools in year 2)						

Remedial Education with low-cost volunteers

Banerjee et al. (2007, QJE); Results

TABLE III

ESTIMATES OF THE IMPACT OF THE BALSAKHI PROGRAM, BY CITY AND SAMPLE

	Number of observations	Dependent variable: test score improvement (posttest − pretest)		
		Math	Language	Total
	(1)	(2)	(3)	(4)
A: Pooling grades and locations				
Mumbai and Vadodara together year 1	12,855	0.182	0.076	0.138
		(0.046)	(0.056)	(0.047)
Mumbai and Vadodara together year 2	21,936	0.353	0.187	0.284
		(0.069)	(0.050)	(0.060)
B: Pooling both grades				
Vadodara year 1	8,426	0.189	0.109	0.161
		(0.057)	(0.057)	(0.057)
Vadodara year 2	11,950	0.371	0.246	0.331
		(0.073)	(0.061)	(0.070)
Mumbai year 1 (grade 3 only)	4,429	0.161	0.086	0.127
		(0.075)	(0.066)	(0.067)
Mumbai year 2	9,986	0.324	0.069	0.188
		(0.145)	(0.081)	(0.112)

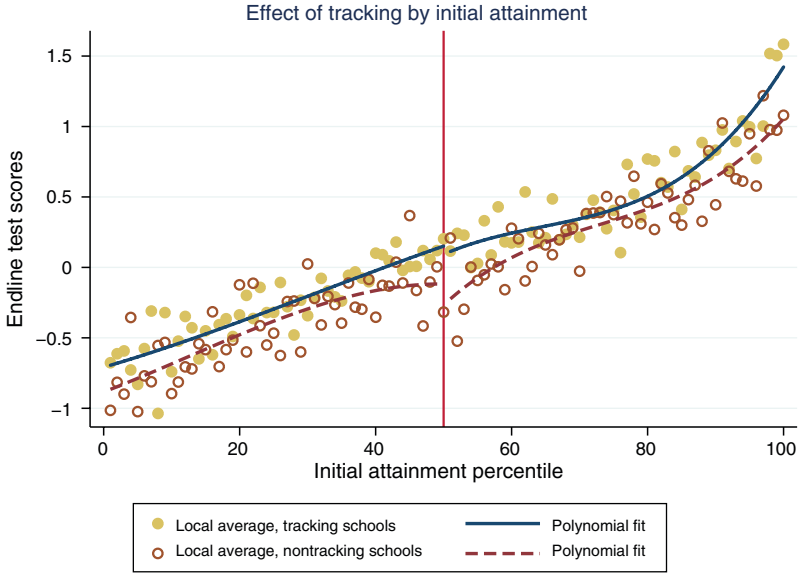
Education RCTs
- Introduction
- Five stylized facts on education in developing countries
- Measuring learning outcomes
  - Objectives of test design
  - Implications for test design
  - Item Response theory
  - How should we make sense of test score impacts?
  - Analysis
- Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)
- A quick review of classic papers
  - The role of information
  - Vouchers
  - Public-Private Partnerships
  - Market-level dynamics
  - Remedial instruction with low-cost volunteers
  - Tracking
  - Personalized computer-aided instruction
- Final remarks

Tracking by ability levels

Duflo, Dupas and Kremer (2011, AER)

- One solution to within-class heterogeneity is to track by achievement level
  - e.g. high-performing set and low-performing set
  - can allow for optimization of instruction to level of preparation
  - commonly used in developed countries, controversial because of labelling effects
  - but also because having high-achieving peers might be good, perhaps esp for low-achieving students
- This is typically less common in developing countries:
  - needs more resources, esp teachers and classrooms
  - in some settings like India, primary schools also lack the scale to do this
- DDK study this question in a very nice experiment in Kenya
  - randomization across schools into tracked and non-tracked Grade 1 classes
  - contract teacher assigned randomly to one of two sections

Raises achievement levels for students of all abilities



Raises achievement levels for students of all abilities

TABLE 2—OVERALL EFFECT OF TRACKING

	Total score				Math score		Literacy score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Short-run effects (after 18 months in program)								
(1) Tracking school	0.139 (0.078)*	0.176 (0.077)**	0.192 (0.093)**	0.182 (0.093)*	0.139 (0.073)*	0.156 (0.083)*	0.198 (0.108)*	0.166 (0.098)*
(2) In bottom half of initial distribution × tracking school			−0.036 (0.07)		0.04 (0.07)		−0.091 (0.08)	
(3) In bottom quarter × tracking school				−0.045 (0.08)		0.012 (0.09)		−0.083 (0.08)
(4) In second-to-bottom quarter × tracking school				−0.013 (0.07)		0.026 (0.08)		−0.042 (0.07)
(5) In top quarter × tracking school				0.027 (0.08)		−0.026 (0.07)		0.065 (0.08)
(6) Assigned to contract teacher		0.181 (0.038)***	0.18 (0.038)***	0.18 (0.038)***	0.16 (0.038)***	0.161 (0.037)***	0.16 (0.038)***	0.16 (0.038)***
Individual controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,795	5,279	5,279	5,279	5,280	5,280	5,280	5,280
Total effects on bottom half and bottom quarter								
Coeff (Row 1) + Coeff (Row 2)			0.156		0.179		0.107	
Coeff (Row 1) + Coeff (Row 3)				0.137		0.168		0.083
F-test: total effect = 0			4.40	2.843	5.97	3.949	2.37	1.411
p-value (total effect for bottom = 0)			0.038	0.095	0.016	0.049	0.127	0.237
p-value (effect for top quarter = effect for bottom quarter)				0.507		0.701		0.209

Effect on teacher effort

TABLE 6—TEACHER EFFORT AND STUDENT PRESENCE

	All teachers		Government teachers		ETP teachers		Students
	Teacher found in school on random school day (1)	Teacher found in class teaching (unconditional on presence) (2)	Teacher found in school on random school day (3)	Teacher found in class teaching (unconditional on presence) (4)	Teacher found in school on random school day (5)	Teacher found in class teaching (unconditional on presence) (6)	Student found in school on random school day (7)
Tracking school	0.041	0.096	0.054	0.112	−0.009	0.007	−0.015
	(0.021)**	(0.038)**	(0.025)**	(0.044)**	(0.034)	(0.045)	(0.014)
Bottom half × tracking school	−0.049	−0.062	−0.073	−0.076	0.036	−0.004	0.003
	(0.029)*	(0.040)	(0.034)**	(0.053)	(0.046)	(0.057)	(0.007)
Years of experience teaching	0.000	−0.005	0.002	0.002	−0.002	−0.008	
	(0.001)	(0.001)***	(0.001)*	(0.001)	(0.003)	(0.008)	
Female	−0.023	0.012	−0.004	0.101	−0.034	−0.061	−0.005
	(0.018)	(0.026)	(0.020)	(0.031)***	(0.032)	(0.043)	(0.004)
Assigned to contract teacher							0.011
							(0.005)**
Assigned to contract teacher × tracking school							0.004
							(0.008)
Observations	2,098	2,098	1,633	1,633	465	465	44,059
Mean in non-tracking schools	0.837	0.510	0.825	0.450	0.888	0.748	0.865
F (test of joint significance)	2.718	9.408	2.079	5.470	2.426	3.674	5.465
p-value	0.011	0.000	0.050	0.000	0.023	0.001	0.000

There's also a lot on both teacher incentives on peer effects in the paper which is worth looking at.

Education RCTs
- Introduction
- Five stylized facts on education in developing countries
- Measuring learning outcomes
  - Objectives of test design
  - Implications for test design
  - Item Response theory
  - How should we make sense of test score impacts?
  - Analysis
- Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)
- A quick review of classic papers
  - The role of information
  - Vouchers
  - Public-Private Partnerships
  - Market-level dynamics
  - Remedial instruction with low-cost volunteers
  - Tracking
  - Personalized computer-aided instruction
- Final remarks

Using technology to "Teach at the Right Level"

Muralidharan, Singh and Ganimian (2019, AER)

- One option that excites policy-makers is education technology

Using technology to "Teach at the Right Level"

Muralidharan, Singh and Ganimian (2019, AER)

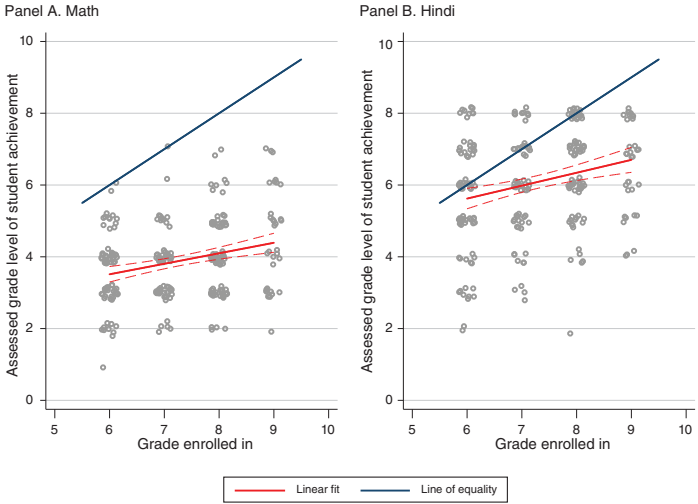
- One option that excites policy-makers is education technology
- In 2015, we evaluated a blended learning program (*Mindspark*)
  - Developed by a leading Indian education firm over a decade
    - Over 45,000 question Item Bank, used by over 400,000 students, administering over a million questions daily
  - **Individual, dynamically updated, assessment and content**
  - Instruction is targeted at children's actual level of achievement, **not the curriculum-mandated level**

Using technology to "Teach at the Right Level"

Muralidharan, Singh and Ganimian (2019, AER)

- One option that excites policy-makers is education technology
- In 2015, we evaluated a blended learning program (*Mindspark*)
  - Developed by a leading Indian education firm over a decade
    - Over 45,000 question Item Bank, used by over 400,000 students, administering over a million questions daily
  - **Individual, dynamically updated, assessment and content**
  - Instruction is targeted at children's actual level of achievement, **not the curriculum-mandated level**
- We evaluate the after-school model (Mindspark centers), which provide supplementary after-school instruction to students six days/week
  - 45 mins individual study using CAL software (Mindspark); 45 mins small group teaching (12-15 students)
  - 619 students, individual level randomization, 4.5 months treatment, treated students received a complete fee waiver
  - all students from government secondary schools in Delhi

Low and dispersed achievement, mismatch with curriculum



Source: Muralidharan, Singh and Ganimian (2019)

Main effects (ITT)

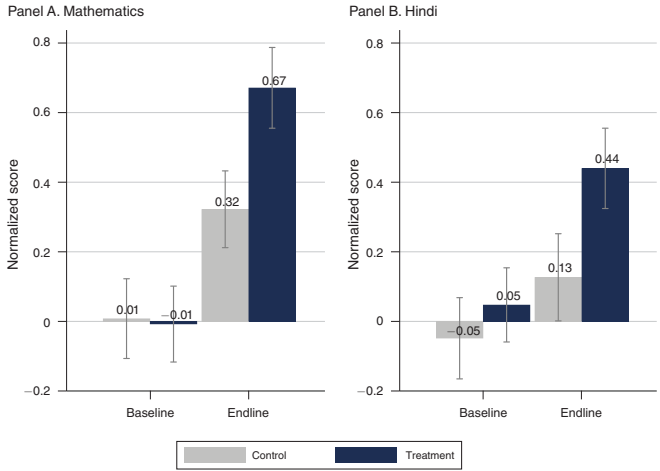


Figure 2. Mean Difference in Test Scores between Lottery Winners and Losers

Source: Muralidharan, Singh and Ganimian (2019)

Effects across the achievement distribution

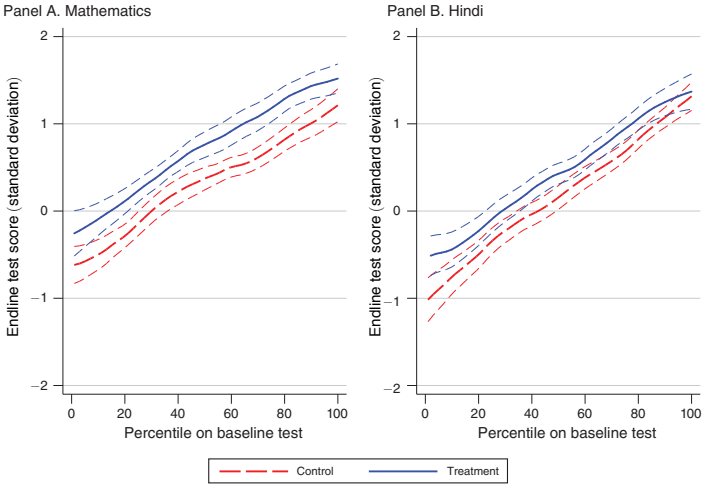


FIGURE 3. NONPARAMETRIC INVESTIGATION OF TREATMENT EFFECTS BY BASELINE PERCENTILES

Source: Muralidharan, Singh and Ganimian (2019)

Effect across terciles

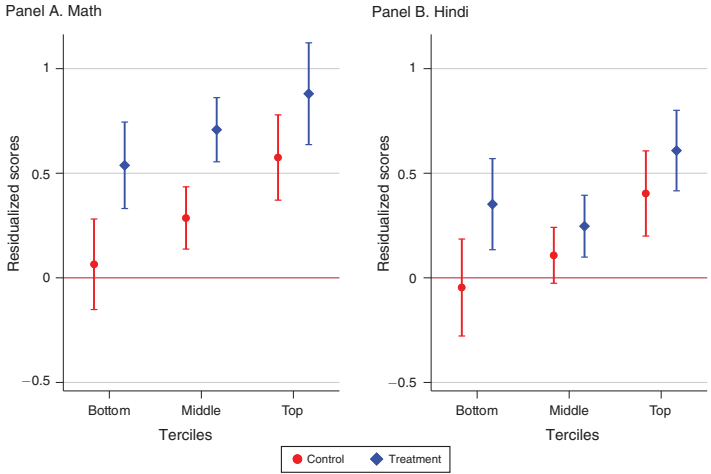


FIGURE 4. GROWTH IN ACHIEVEMENT IN TREATMENT AND CONTROL GROUPS

Source: Muralidharan, Singh and Ganimian (2019)

Education RCTs

- Introduction
- Five stylized facts on education in developing countries
- Measuring learning outcomes
- Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)
- A quick review of classic papers
- Final remarks

Education RCTs

Introduction

Five stylized facts on education in developing countries

Measuring learning outcomes

Other outcomes of interest (e.g., teacher's time-on-task, classroom observations)

A quick review of classic papers

Final remarks

Further reading

- The economics of education literature is sprawling
  - what we've covered is selective, even within applied micro dev
- Some themes (out of many) that are worth seeing:
  - Access to schooling: CCTs, free schooling, bicycles, scholarships etc.
  - ECE; production functions for human capital
  - School accountability, governance, political economy
  - Incentives and contracts in schooling
  - School inputs, school and teacher VA
  - Macro HK and growth literature