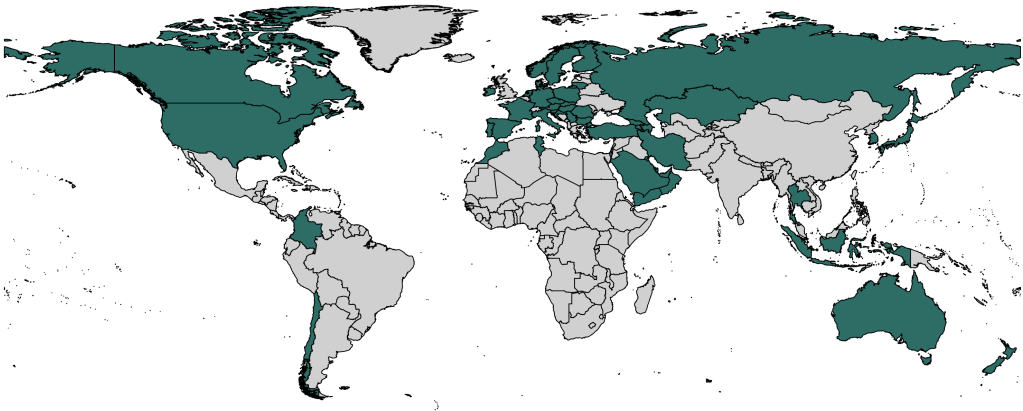
A Rosetta Stone for Human Capital

Dev Patel Justin Sandefur

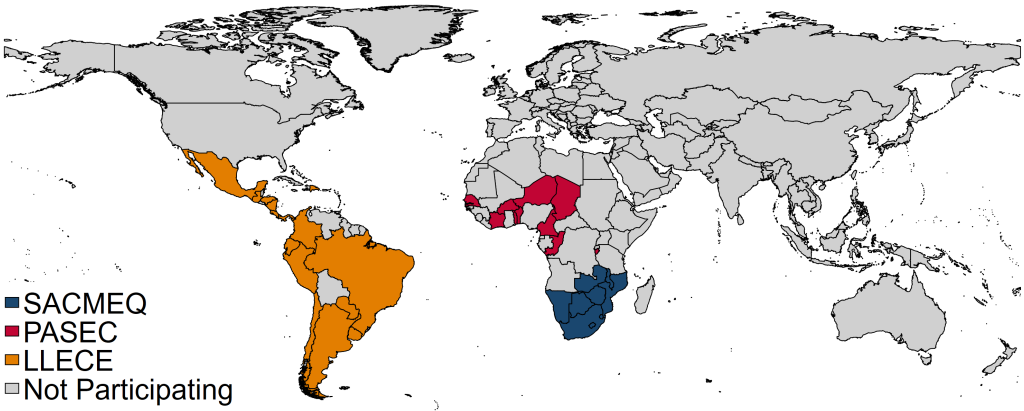
Who can read better, a 4^{th}-Grader in Niger or Peru?

Coverage of International Assessments



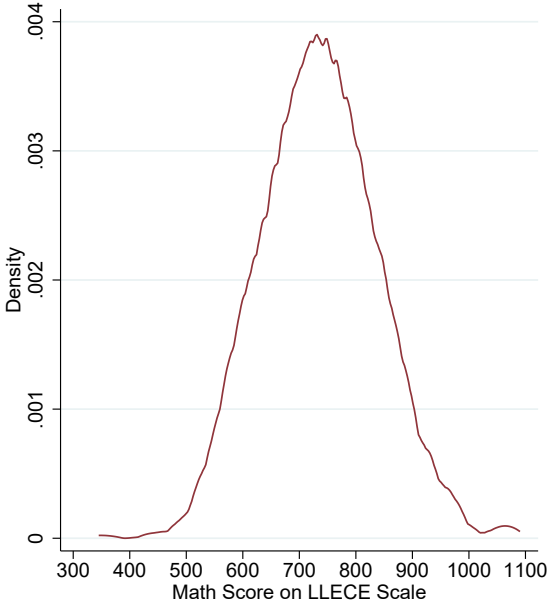
Note: Data from World Bank's Education Statistics using TIMSS 2011 and PIRLS 2011, PASEC 2014, LLECE 2013, and SACMEQ 2007.

Coverage of International Assessments

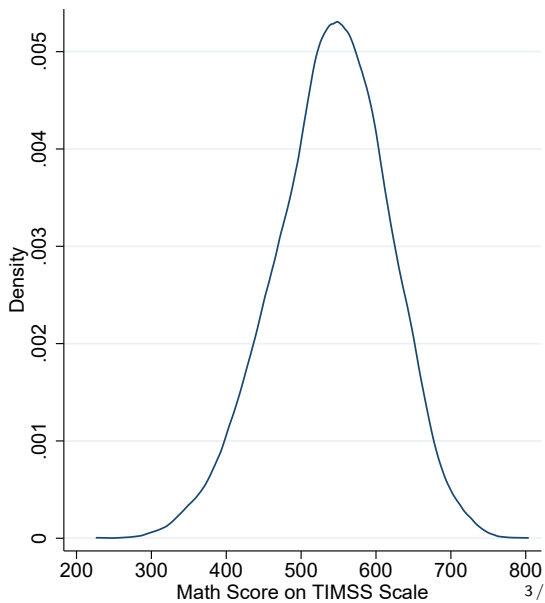


Note: Data from World Bank's Education Statistics using TIMSS 2011 and PIRLS 2011, PASEC 2014, LLECE 2013, and SACMEQ 2007.

Brazil (LLECE Scale)



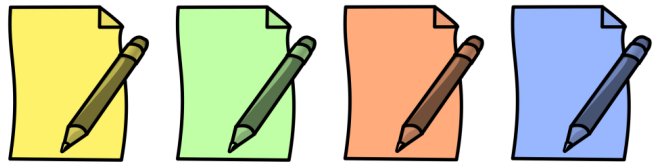
United States (TIMSS Scale)



Part #1: Method to “Translate” Human Capital Across Scales

Part #1: Method to "Translate" Human Capital Across Scales

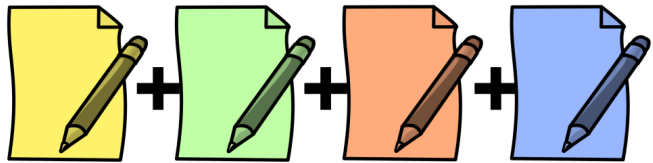
Methodological Overview



Part #1: Method to "Translate" Human Capital Across Scales

Methodological Overview

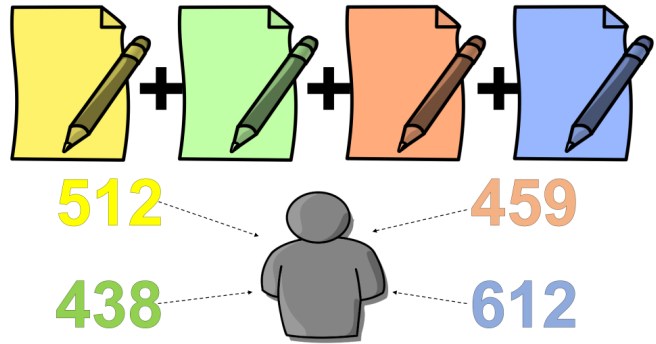
- Build hybrid test combining questions from source exams



Part #1: Method to "Translate" Human Capital Across Scales

Methodological Overview

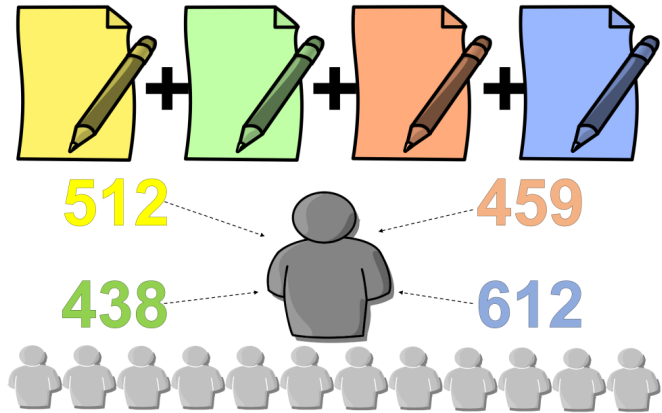
- Build hybrid test combining questions from source exams
- Use item response theory to grade performance on each original scale



Part #1: Method to "Translate" Human Capital Across Scales

Methodological Overview

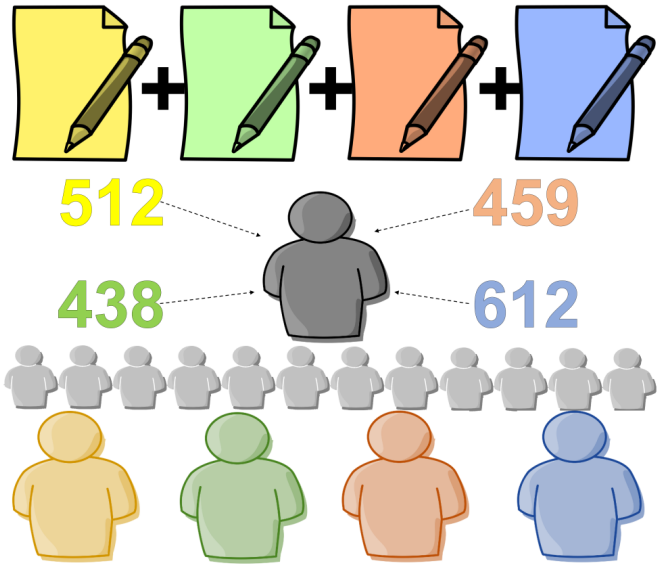
- Build hybrid test combining questions from source exams
- Use item response theory to grade performance on each original scale
- Repeat for many students, producing "translations" across the ability distribution



Part #1: Method to "Translate" Human Capital Across Scales

Methodological Overview

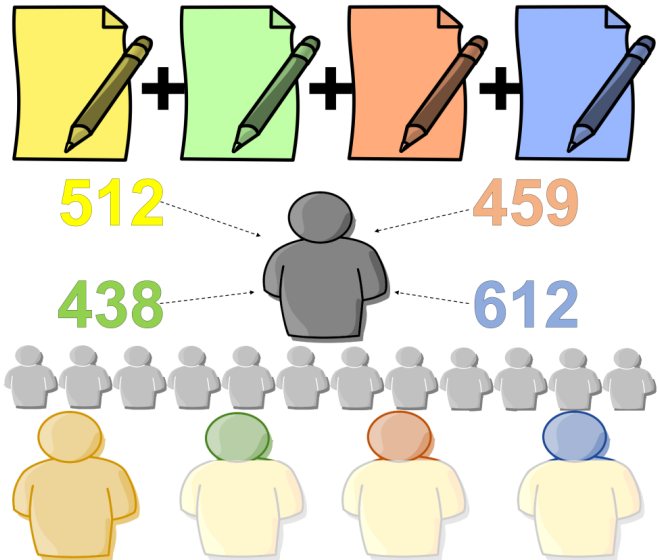
- Build hybrid test combining questions from source exams
- Use item response theory to grade performance on each original scale
- Repeat for many students, producing "translations" across the ability distribution
- Apply "Rosetta Stone" out of sample to translate any score



Part #1: Method to "Translate" Human Capital Across Scales

Methodological Overview

- Build hybrid test combining questions from source exams
- Use item response theory to grade performance on each original scale
- Repeat for many students, producing "translations" across the ability distribution
- Apply "Rosetta Stone" out of sample to translate any score



Part #2: Applications to the Global Learning Distribution

Link four largest primary-age standardized tests using data from India and U.S.
Translate math and reading scores of 600,000 pupils across 80 countries

Part #2: Applications to the Global Learning Distribution

Link four largest primary-age standardized tests using data from India and U.S.
Translate math and reading scores of 600,000 pupils across 80 countries

Preview of results:

1. **Where children live matters more than their income**
   - Students with the same household resources living in different countries can have very different educational outcomes

Part #2: Applications to the Global Learning Distribution

Link four largest primary-age standardized tests using data from India and U.S.
Translate math and reading scores of 600,000 pupils across 80 countries

Preview of results:

1. Where children live matters more than their income

2. **Income-test score gradient is steeper where income inequality is higher**
   - E.g., Score-income coefficient in Finland (-0.0017) vs. Brazil (.41), echoing "Great Gatsby Curve" (Krueger, 2012; Corak, 2013)

Part #2: Applications to the Global Learning Distribution

Link four largest primary-age standardized tests using data from India and U.S.
Translate math and reading scores of 600,000 pupils across 80 countries

Preview of results:

1. Where children live matters more than their income

2. Income-test score gradient is steeper where income inequality is higher

3. **Girls read better than boys on average at all household incomes; outperform in math only below $3,000 (PPP) per capita**

Part #2: Applications to the Global Learning Distribution

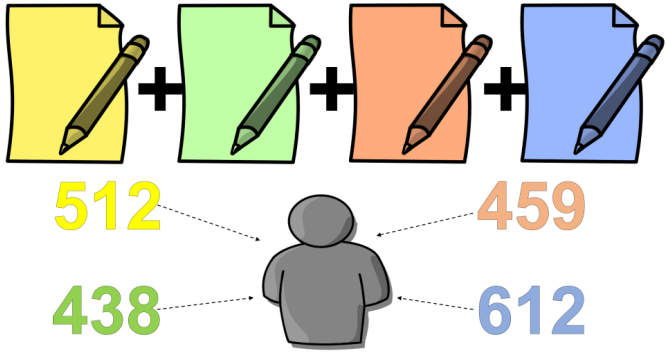
Link four largest primary-age standardized tests using data from India and U.S.
Translate math and reading scores of 600,000 pupils across 80 countries

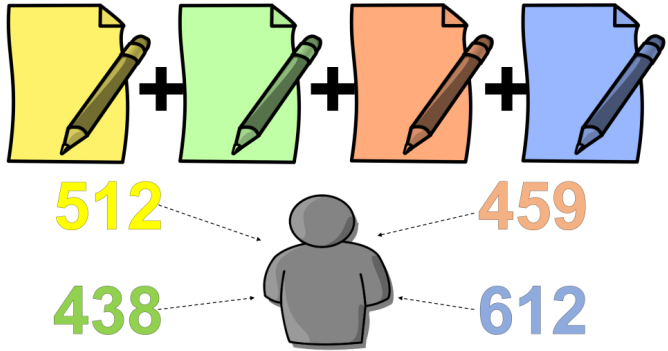
Preview of results:

1. Where children live matters more than their income
2. Income-test score gradient is steeper where income inequality is higher
3. Girls read better than boys on average at all household incomes; outperform in math only below $3,000 (PPP) per capita
4. **Score gap between public and private schools increases with income inequality**

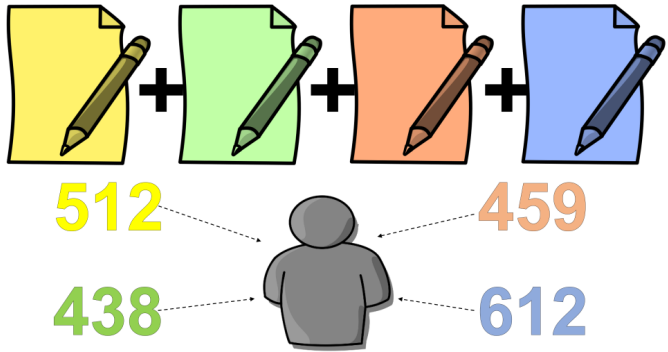
Today's Talk

1. **"Rosetta Stone" Methodology**
   - Overview of methodology
   - Implementation using fieldwork from India and U.S.
2. **Application to Global Learning Distribution**

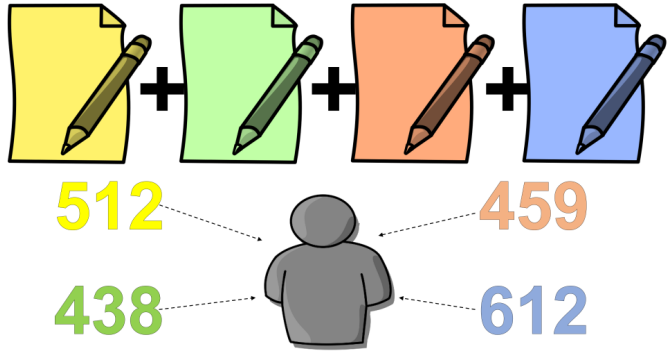

512
459
438
612



1. Most standardized tests are long $\Rightarrow$ impractical for a student to sit multiple in a day

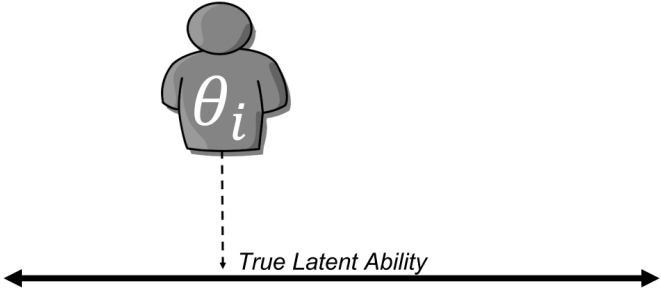


1. Most standardized tests are long $\Rightarrow$ impractical for a student to sit multiple in a day
2. Testing firms only publicly release a small number of questions

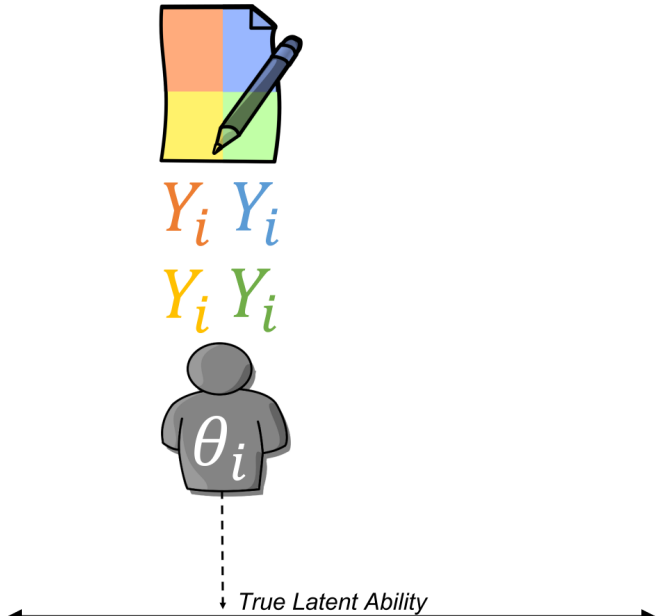


1. Most standardized tests are long $\Rightarrow$ impractical for a student to sit multiple in a day
2. Testing firms only publicly release a small number of questions

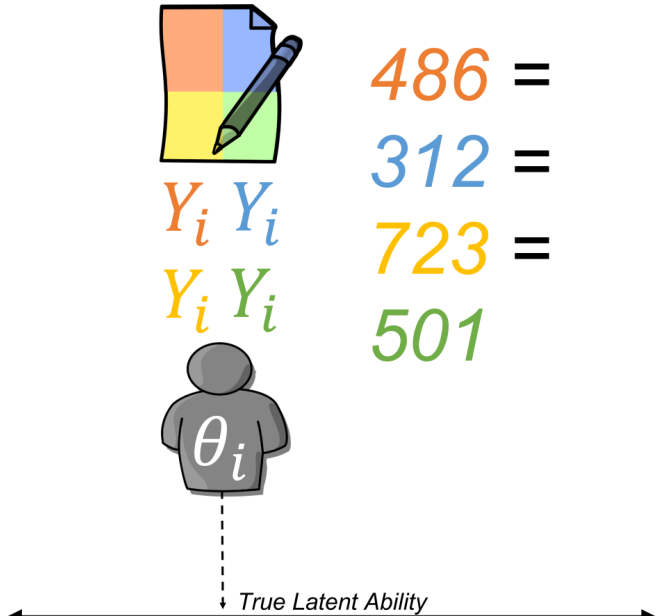
$\Rightarrow$ Build a hybrid exam consisting of some questions from each source test



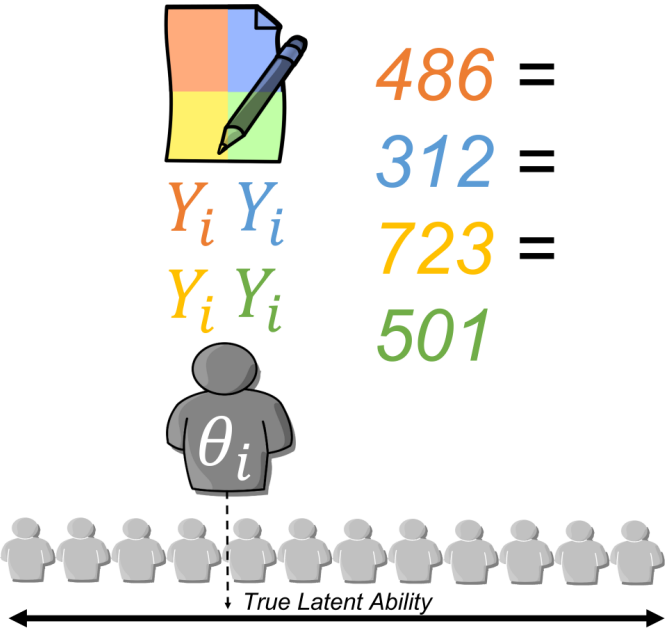
- Administer hybrid exam pooling questions from target tests



- Administer hybrid exam pooling questions from target tests
- For student i, grade performance on each source scale using anchored IRT parameters



- Administer hybrid exam pooling questions from target tests
- For student i, grade performance on each source scale using anchored IRT parameters
- Because it is the same student on the same day, scores *within* pupil are comparable



- Administer hybrid exam pooling questions from target tests
- For student i, grade performance on each source scale using anchored IRT parameters
- Because it is the same student on the same day, scores *within* pupil are comparable
- Repeat for lots of students to estimate "translations" across skill distribution

Application: Assessing the Global Learning Distribution

- Apply methodology to link four of the world's largest standardized tests spanning 80 countries:
  - TIMSS and PIRLS (OECD)
  - LLECE (Latin America)
  - PASEC (West Africa)

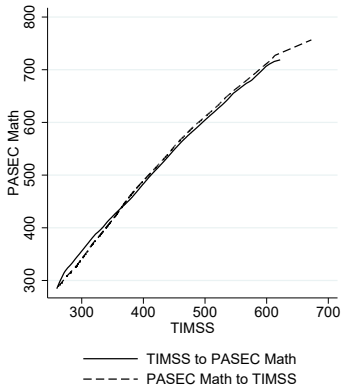
- Administer hybrid exam to two samples of elementary school students:
  - Bihar, India: 2,300 students in 4^{th}-8^{th} grade Details
  - Florida, United States: 2,200 students in 6^{th}-8^{th} grade Details

  Provides coverage across skill distribution

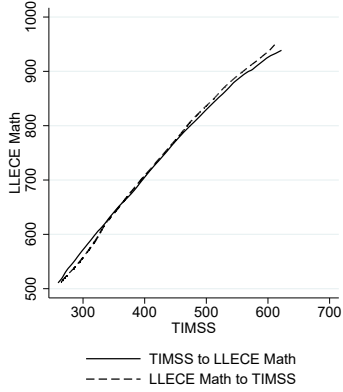
- Exams cover both math and reading
  - *Focus on math today for sake of time*

Test Score Conversion Functions

OECD ⇔ West Africa

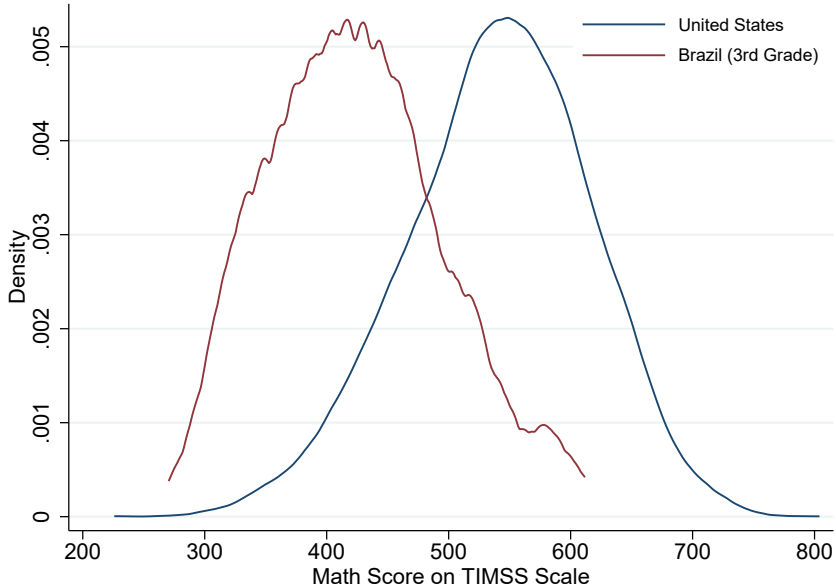


OECD ⇔ Latin America



T-test results

Placing Brazil and U.S. on the Same Scale

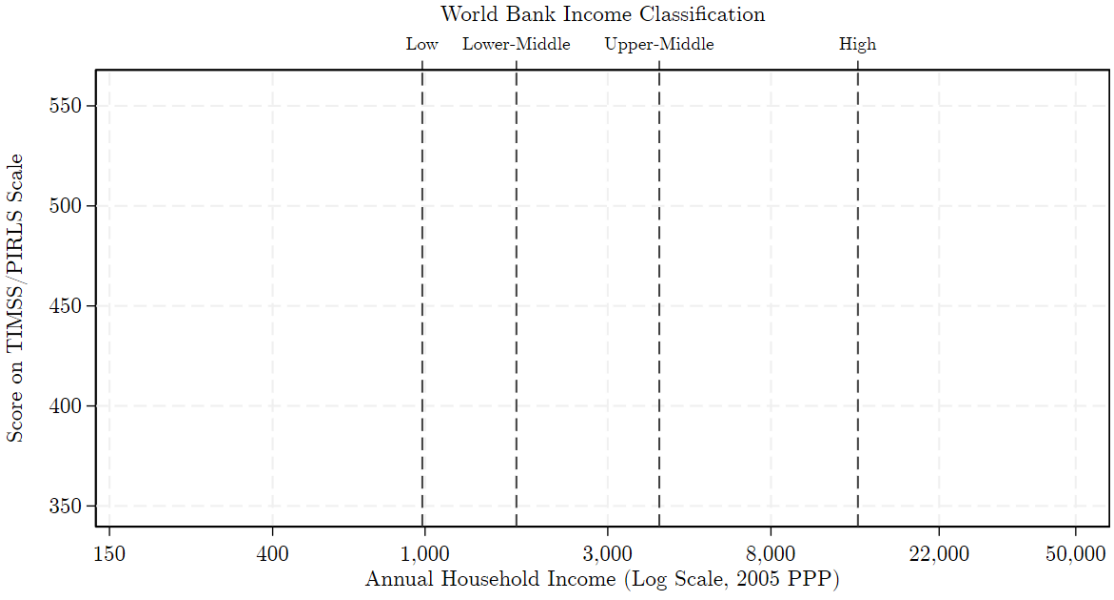


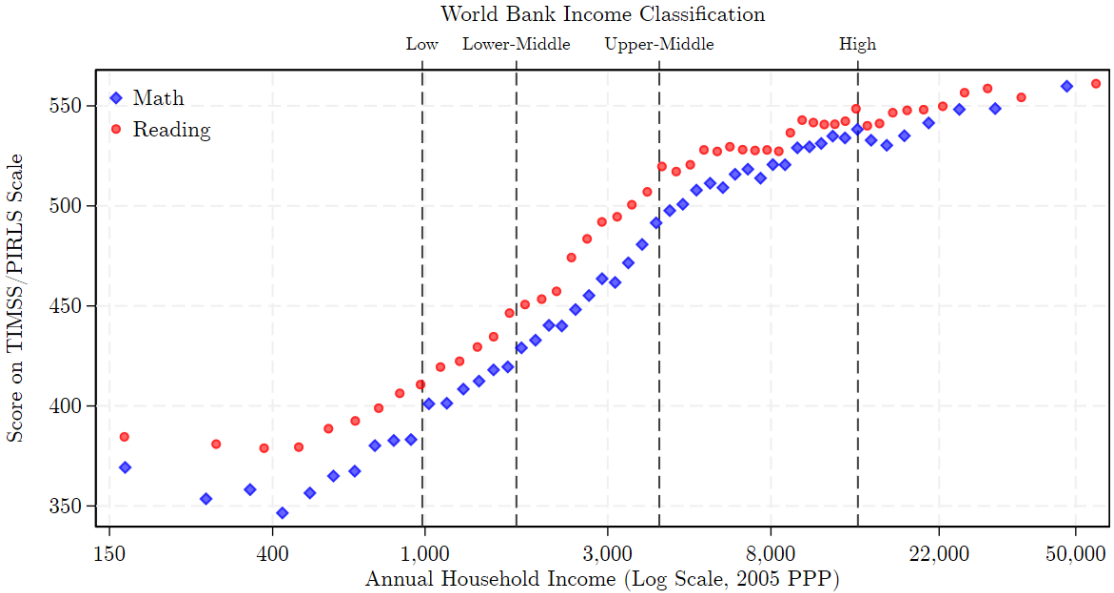
Today's Talk

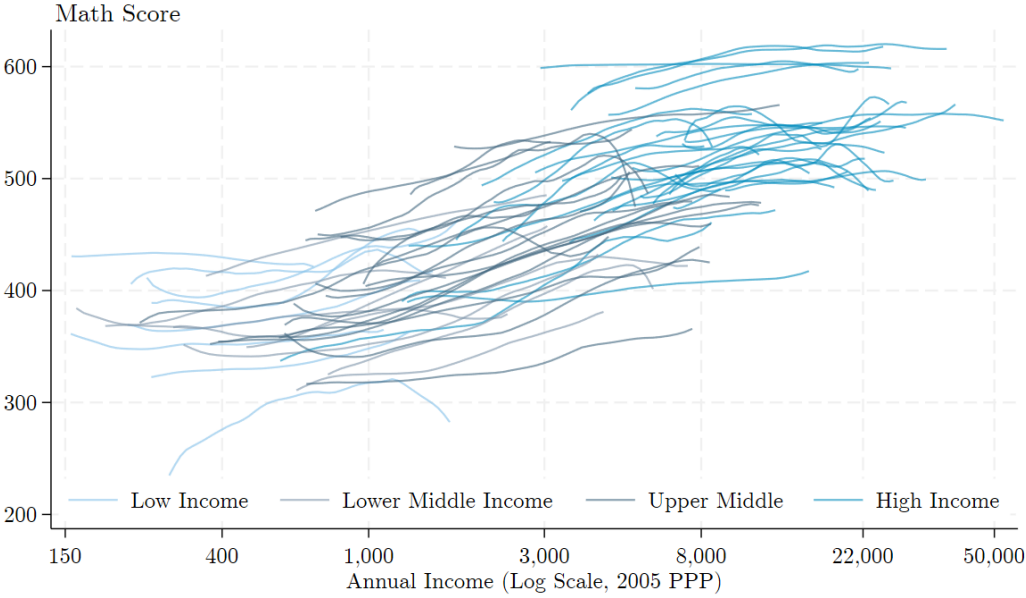
1. "Rosetta Stone" Methodology
2. **Application to Global Learning Distribution**
   - Apply these translations to microdata from 628,601 students from 80 countries who participated in the source tests
   - Results that follow use TIMSS/PIRLS scales ($100 = 1\sigma$)
   - Four facts about the learning around the world

Four Facts About the Global Learning Distribution

1. **Where children live matters more than their income**
   - Students with the same household resources living in different countries can have very different educational outcomes


World Bank Income Classification
Low
Lower-Middle
Upper-Middle
High
550
500
450
400
350
Score on TIMSS / PIRLS Scale
150
400
1,000
3,000
8,000
22,000
50,000
Annual Household Income (Log Scale, 2005 PPP)

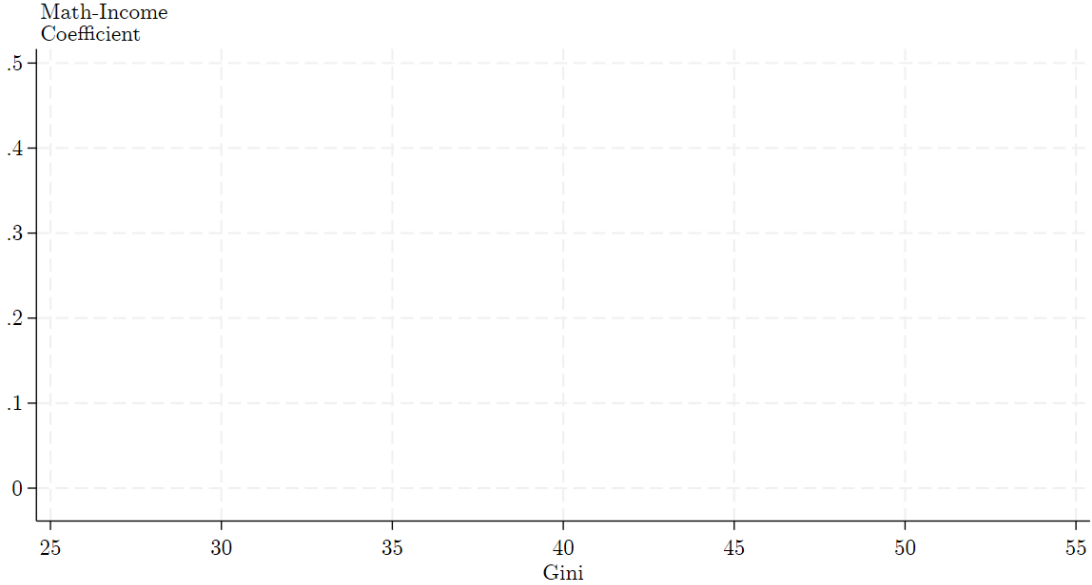

World Bank Income Classification
Low
Lower-Middle
Upper-Middle
High
Math
Reading
Score on TIMSS / PIRLS Scale
550
500
450
400
350
150
400
1,000
3,000
8,000
22,000
50,000
Annual Household Income (Log Scale, 2005 PPP)

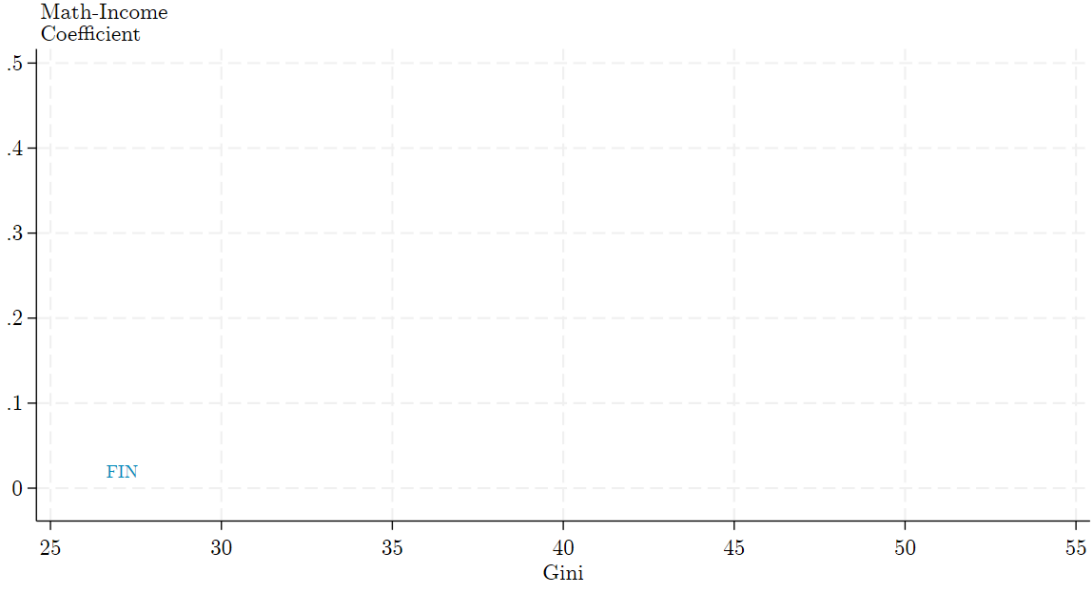

Math Score
600
500
400
300
200
Low Income
Lower Middle Income
Upper Middle
High Income
150
400
1,000
3,000
8,000
22,000
50,000
Annual Income (Log Scale, 2005 PPP)

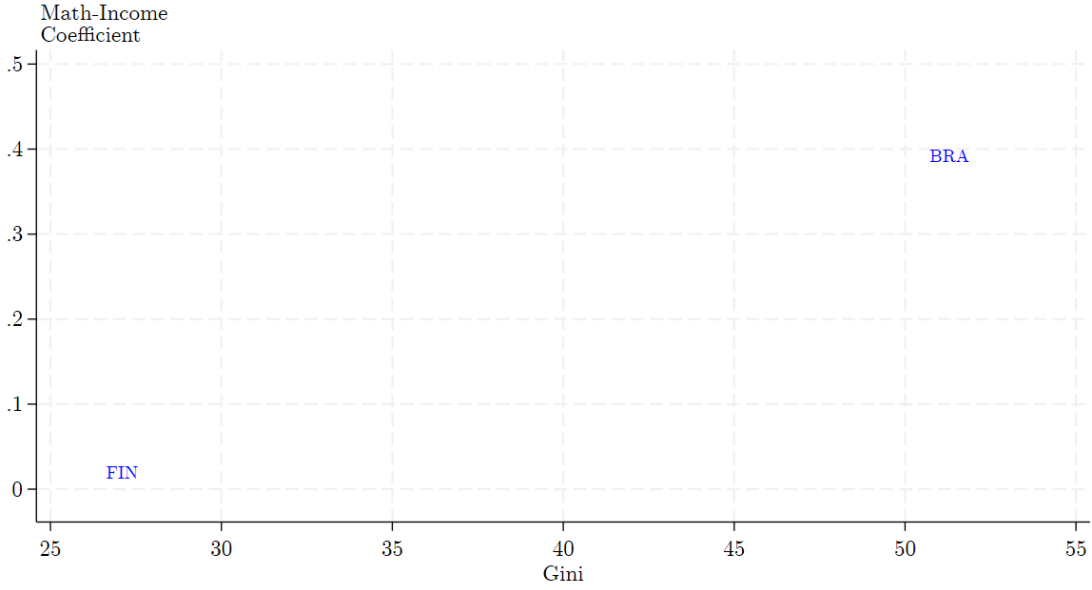
Four Facts About the Global Learning Distribution

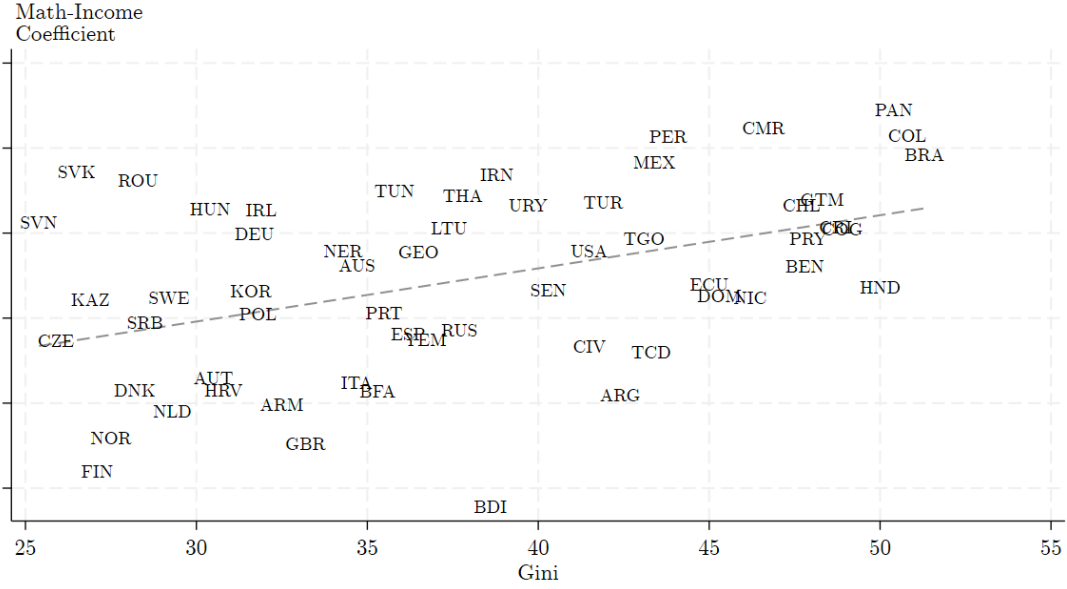
1. Where children live matters more than their income Shapley Decomposition

2. **Income-test score gradient is steeper where income inequality is higher**
   - E.g., Score-income coefficient in Finland (-0.0017) vs. Brazil (.41), echoing "Great Gatsby Curve" (Krueger, 2012; Corak, 2013)


Math-Income
Coefficient
.5
.4
.3
.2
.1
0
25
30
35
40
45
50
55
Gini

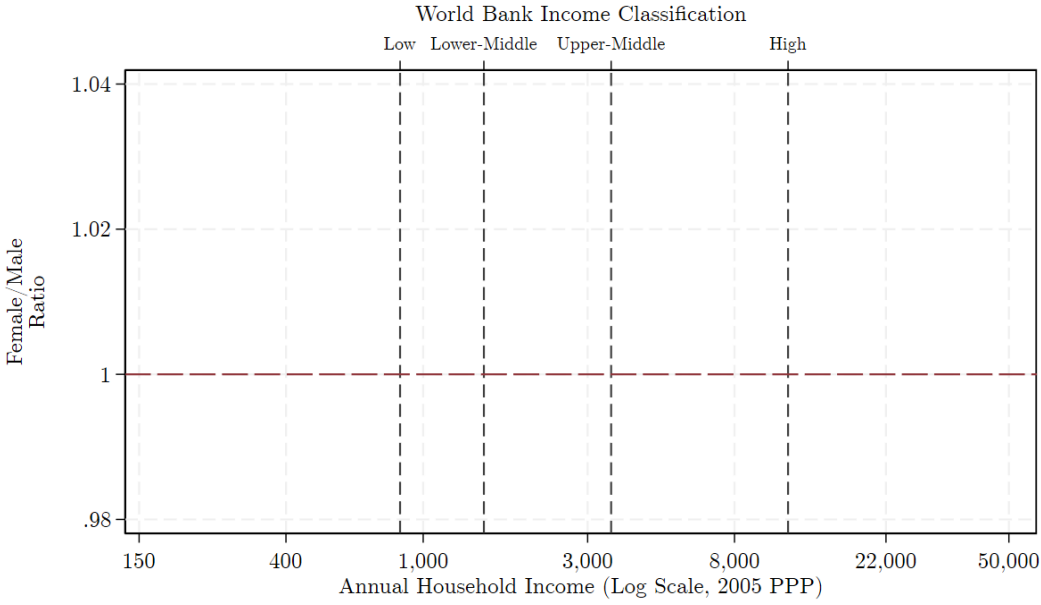

Math-Income
Coefficient
.5
.4
.3
.2
.1
0
FIN
25
30
35
40
45
50
55
Gini

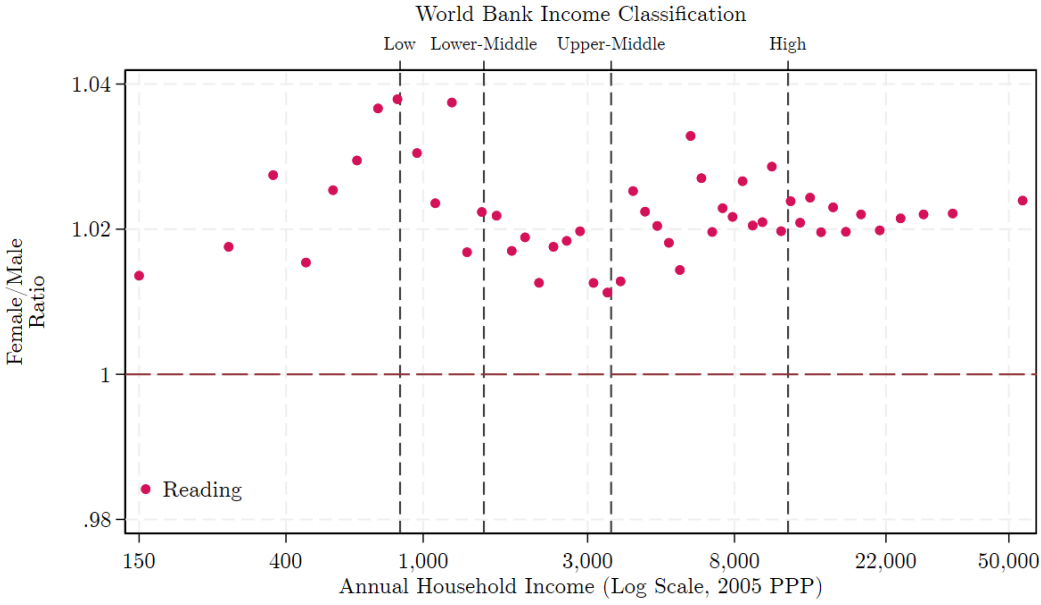

Math-Income
Coefficient
.5
.4
.3
.2
.1
0
BRA
FIN
25
30
35
40
45
50
55
Gini

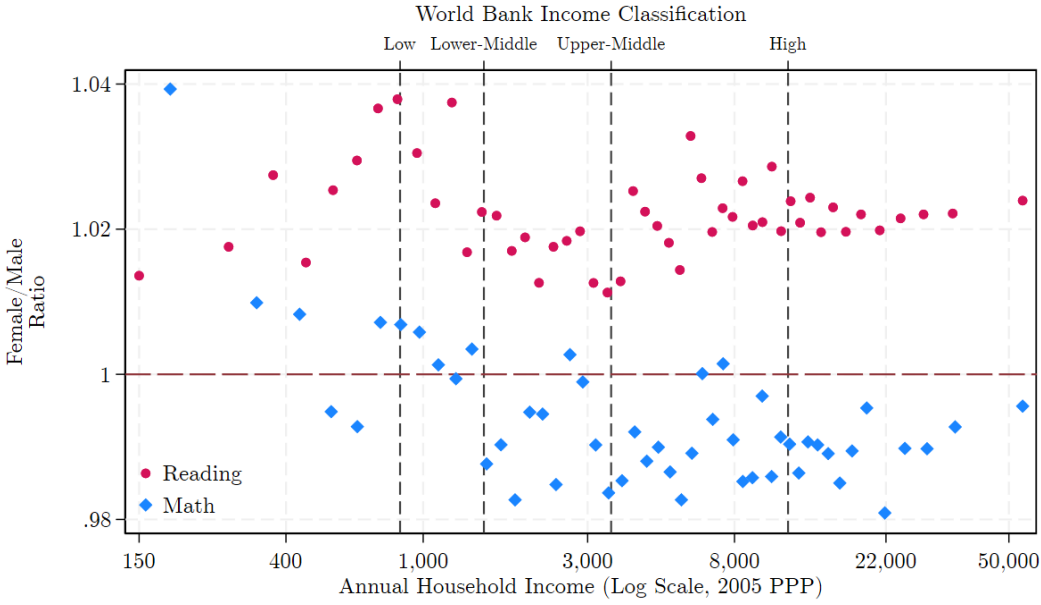

Math-Income
Coefficient
.5
.4
.3
.2
.1
0
25
30
35
40
45
50
55
Gini
PAN
PER
CMR
COL
BRA
MEX
SVK
ROU
IRN
TUN
THA
URY
TUR
GTM
CHL
SVN
HUN
IRL
DEU
LTU
TGO
PRY
CRI
NER
GEO
AUS
USA
BEN
KAZ
SWE
KOR
SEN
ECU
DOM
NIC
HND
SRB
POL
PRT
CZE
ESP
YEM
RUS
CIV
TCD
AUT
DNK
HRV
ITA
BFA
ARG
NLD
ARM
NOR
GBR
FIN
BDI

Four Facts About the Global Learning Distribution

1. Where children live matters more than their income
2. Income-test score gradient is steeper where income inequality is higher
3. **Girls read better than boys on average at all household incomes; outperform in math only below $3,000 (PPP) per capita**

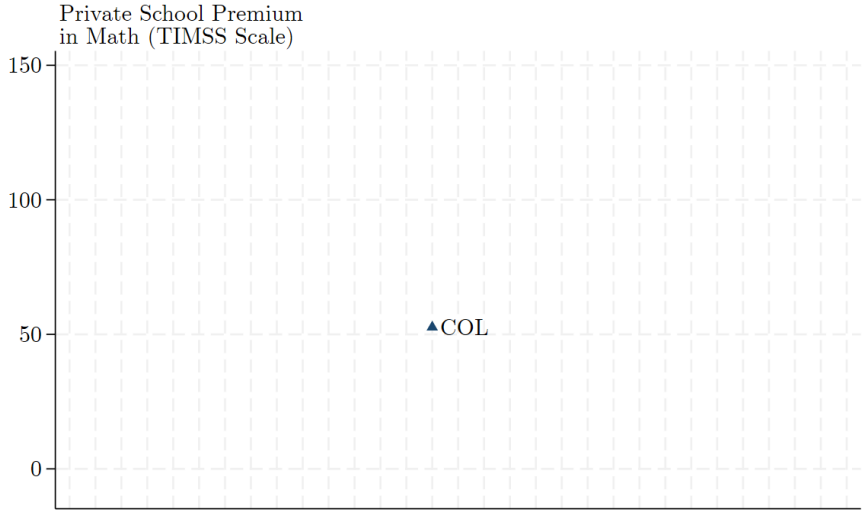

World Bank Income Classification
Low
Lower-Middle
Upper-Middle
High
1.04
1.02
1
.98
Female/Male Ratio
150
400
1,000
3,000
8,000
22,000
50,000
Annual Household Income (Log Scale, 2005 PPP)

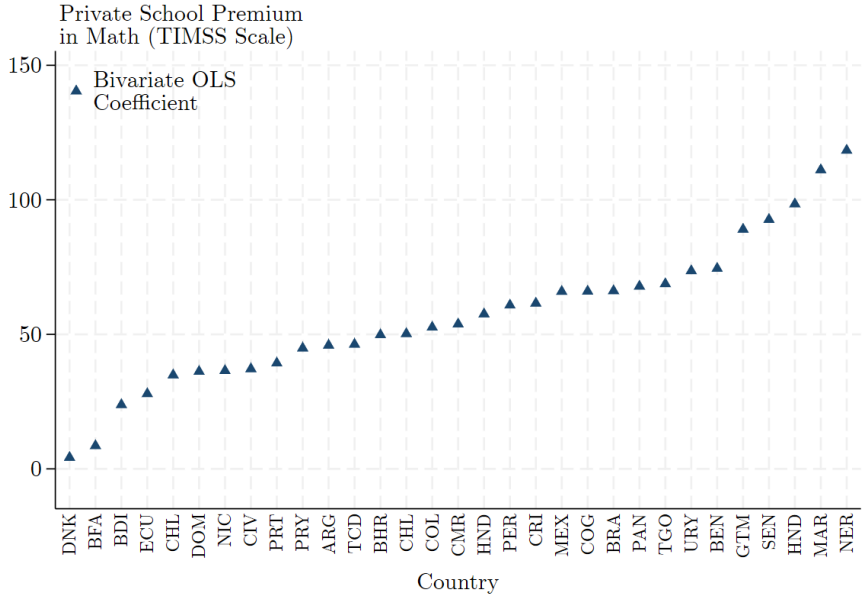

World Bank Income Classification
Low
Lower-Middle
Upper-Middle
High
1.04
1.02
1
.98
Female/Male Ratio
Reading
150
400
1,000
3,000
8,000
22,000
50,000
Annual Household Income (Log Scale, 2005 PPP)

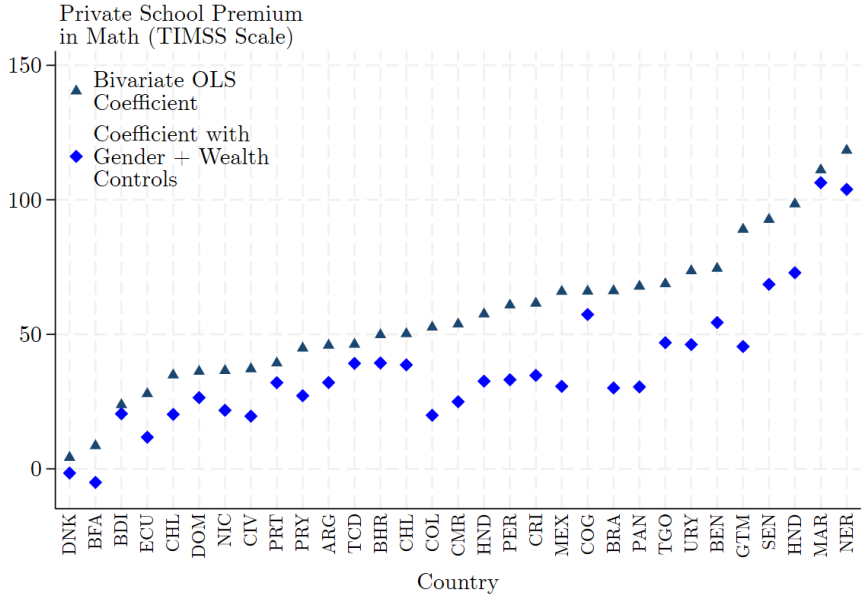

World Bank Income Classification
Low
Lower-Middle
Upper-Middle
High
1.04
1.02
1
.98
Female/Male Ratio
Reading
Math
150
400
1,000
3,000
8,000
22,000
50,000
Annual Household Income (Log Scale, 2005 PPP)

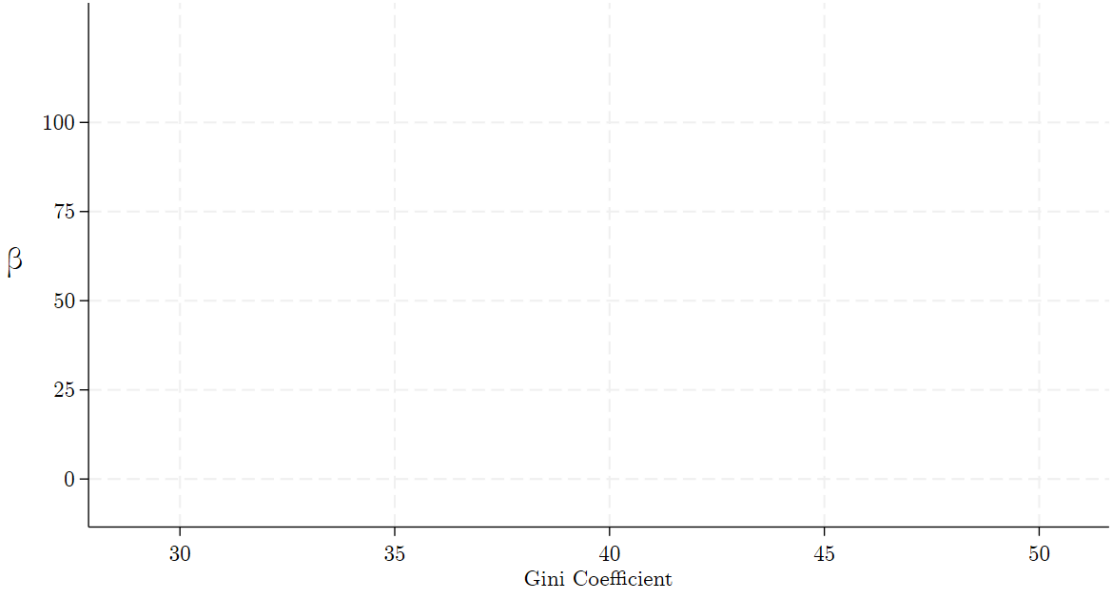
Four Facts About the Global Learning Distribution

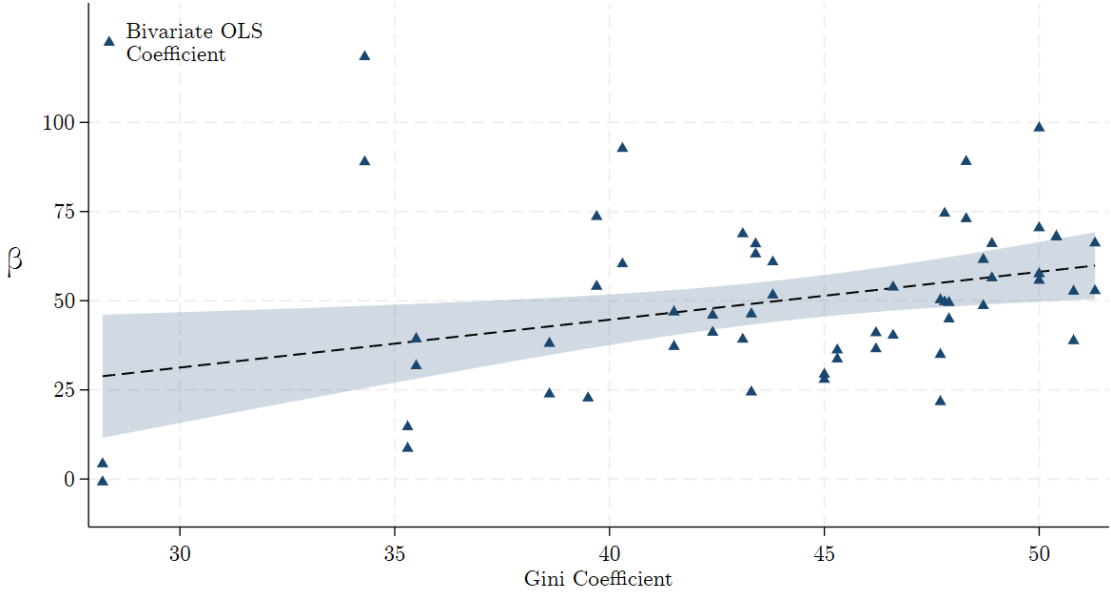
1. Where children live matters more than their income
2. Income-test score gradient is steeper where income inequality is higher
3. Girls read better than boys on average at all household incomes; outperform in math only below $3,000 (PPP) per capita
4. **Score gap between public and private schools increases with income inequality**

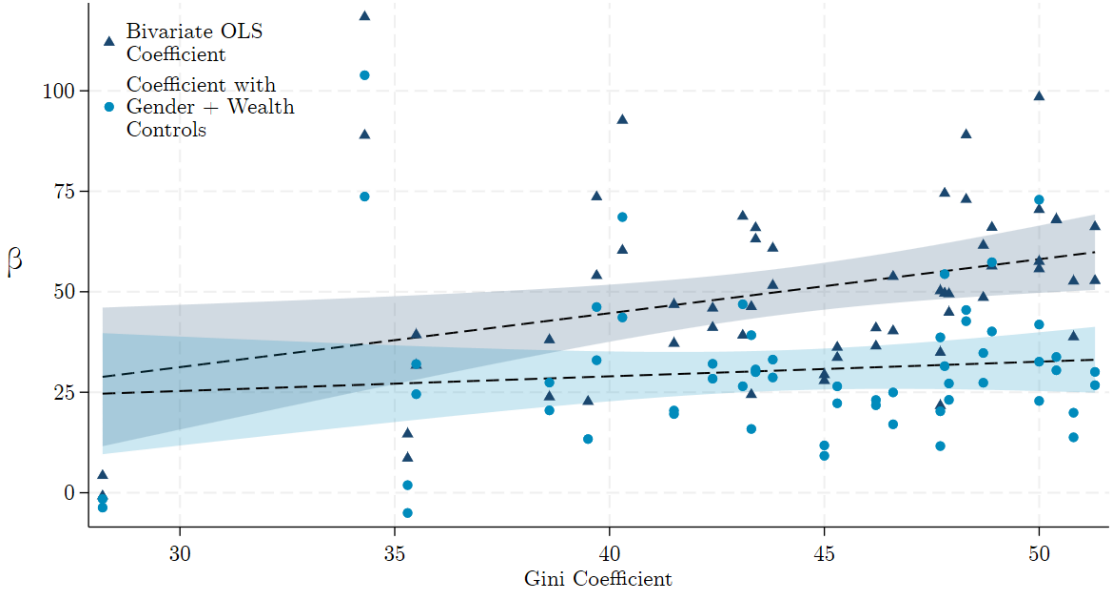

Private School Premium
in Math (TIMSS Scale)
150
100
50
0
COL


Private School Premium
in Math (TIMSS Scale)
Bivariate OLS
Coefficient
150
100
50
0
DNK
BFA
BDI
ECU
CHL
DOM
NIC
CIV
PRT
PRY
ARG
TCD
BHR
CHL
COL
CMR
HND
PER
CRI
MEX
COG
BRA
PAN
TGO
URY
BEN
GTM
SEN
HND
MAR
NER
Country


Private School Premium
in Math (TIMSS Scale)
150
100
50
0
Bivariate OLS
Coefficient
Coefficient with
Gender + Wealth
Controls
DNK
BFA
BDI
ECU
CHL
DOM
NIC
CIV
PRT
PRY
ARG
TCD
BHR
CHL
COL
CMR
HND
PER
CRI
MEX
COG
BRA
PAN
TGO
URY
BEN
GTM
SEN
HND
MAR
NER
Country


100
75
β
50
25
0
30
35
40
45
50
Gini Coefficient


Bivariate OLS
Coefficient
β
100
75
50
25
0
30
35
40
45
50
Gini Coefficient


Bivariate OLS Coefficient
Coefficient with Gender + Wealth Controls
β
100
75
50
25
0
30
35
40
45
50
Gini Coefficient

Conclusion

- Propose method to compare measures of ability elicited on different scales
- Apply to link microdata on primary student math and reading skills across 80 countries
- Large gaps exist between rich and poor countries *above and beyond* household income
- "Rosetta Stone" approach exhibits network externality

Thank you!

devpatel@fas.harvard.edu

Generalized Partial Credit Model

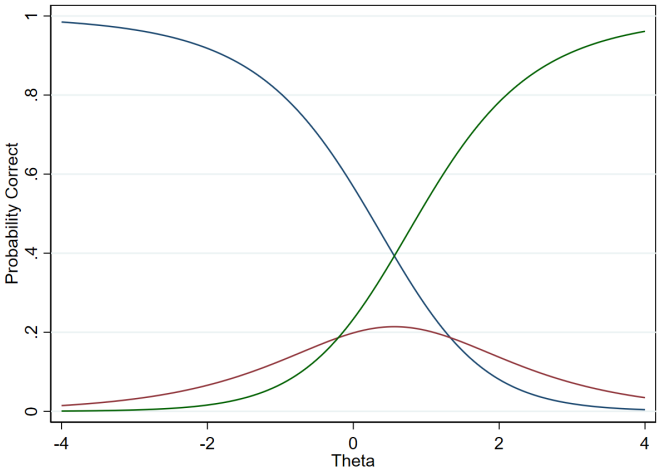
Back

For non-dichotomous items (partial credit), adopt general partial credit model:

$$P(x_{ij} = l|\theta_i, a_j, b_j, d_{j,1}, ..., d_{j,m_j-1}) =$$

$$\frac{\exp\left(\sum_{v=0}^{l} a_j\left(\theta_i - b_j + d_{j,v}\right)\right)}{\sum_{g=0}^{m_i-1}\exp\left(\sum_{v=0}^{g} a_j\left(\theta_i - b_j + d_{j,v}\right)\right)}$$

- Blue curve: 0 points
- Red curve: 1 point
- Green curve: 2 points



Data Collection in Bihar, India

Back

- Roughly 57 items per test
- Sample of 2,314 students in 51 schools across six districts in Bihar, India
- Grades 4-8, with a median age of 11
- 52.1 percent female

Data Collection in Florida, USA

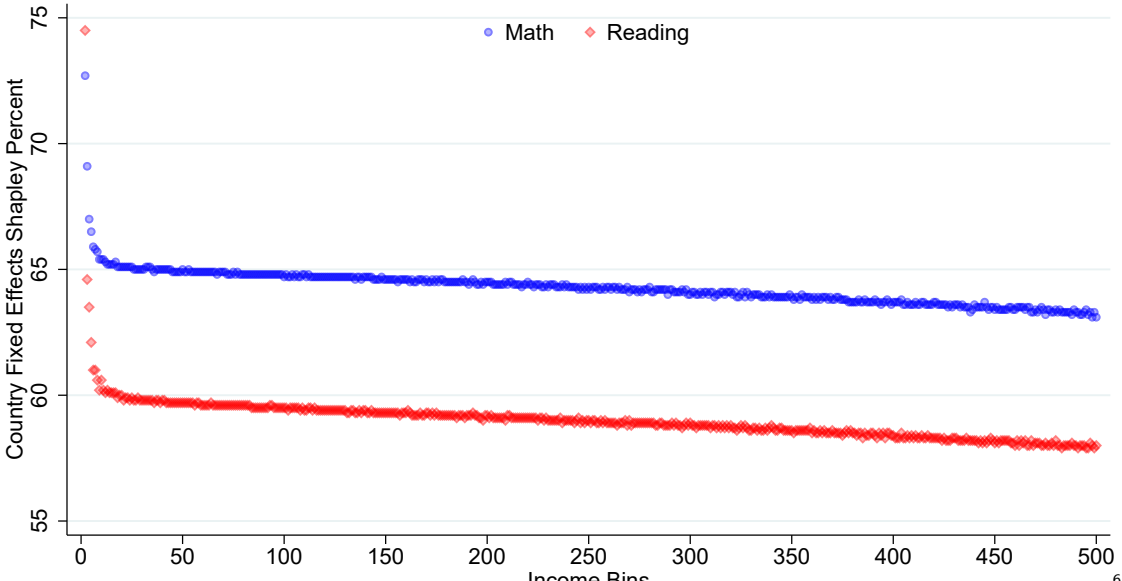
Back

- Partner with Character Labs Research Network
- Conducted online via a Qualtrics survey (most students took test at home due to COVID)
- Students randomized into either math or English
- Within subject, students randomized to questions, stratified by source test
- Much shorter than Bihar: 25 minutes

Symmetry Tests

Back

Variable	(1) Original N	(1) Original Mean/SE	(2) Converted Twice N	(2) Converted Twice Mean/SE	T-test P-value (1)-(2)
PASEC Math	2314	468.488 (1.955)	2313	468.559 (1.809)	0.979
PASEC Reading	2314	480.378 (2.084)	2314	480.791 (1.913)	0.884
LLECE Math	2314	688.897 (2.008)	2313	688.901 (1.886)	0.999
LLECE Reading	2314	652.112 (1.774)	2314	652.290 (1.719)	0.943


Math
Reading
Country Fixed Effects Shapley Percent
75
70
65
60
55
0
50
100
150
200
250
300
350
400
450
500
Income Bins

Relative Role of Country vs. Household Income

Back

Shapley decomposition of R^2

$$R_j^2 = \sum_{T \subseteq Z \ \{x_j\}} \frac{k!\,(p-k-1)!}{p!} \left[R^2(T \cup \{x_j\}) - R^2(T)\right]$$

T is model with k regressors except x_j, Z is fully saturated model