Same-Sex Teacher Effects in Education

A Meta-Analysis and a Multi-Country Study

Alexandra de Gendre | University of Melbourne
Jan Feld | Victoria University of Wellington
Nicolas Salamanca | University of Melbourne
Ulf Zoelitz | University of Zurich

ENS Lyon
17 September 2026

Same-Sex Teachers Matter

- Powerful narratives...
  - *"Girls need role models in STEM"* (UNICEF 2020, World Bank 2020, OECD 2012)
    - Lack of female role models often cited as key contributor to gender gap in STEM performance
    - More female STEM teachers could increase women's representation in STEM studies and jobs
  - *"Boys lack male role models in primary school"*
    - Many countries run campaigns to attract more men to become primary school teachers – hoping to stop boys' performance decline

Same-Sex Teachers Matter

- ... And policies implemented...
  - Switzerland: Men for Primary Schools Campaign featuring newspaper ads & taster days
  - Australia: Males in Primary Initiative including male only scholarships, mentorship programs, promotional campaigns
  - England: Marketing Campaigns to attract more men into teaching, highlighting positive impact and the rewarding nature of the profession
  - United States: Troops to Teachers Program addressing gender imbalance in primary schools
  - Norway: introduced and later abolished Male teacher quotas

- ... Yet **no systematic review of evidence**

Research Question

Is there a **same-sex teacher effect**

on **test scores** and **job preferences**?

What is a Same-sex Teacher Effect?

The test-score **premium** of having a **same-sex teacher** *on top of student and teacher gender effects.*

$$Outcome = \beta_0 + \beta_1 FemaleStudent + \beta_2 FemaleTeacher + \underbrace{\beta_3}_{\substack{\text{Same-sex} \\ \text{Teacher} \\ \text{Effect}}} FemaleStudent \times FemaleTeacher + u$$

- β_3 is the sum of 4 type of match effects:
  - γ_{FT}: female teachers are more effective at teaching female students (compared to male students)
  - γ_{MT}: male teachers are more effective at teaching male students (compared to female students)
  - δ_{FS}: girls are more studious when studying with a female teacher (compared to male teacher)
  - δ_{MS}: boys are more studious when working with a male teacher (compared to female teacher)
- β_3 can be identified with **difference-in-differences** (Muralidharan and Sheth, 2016)

This Paper

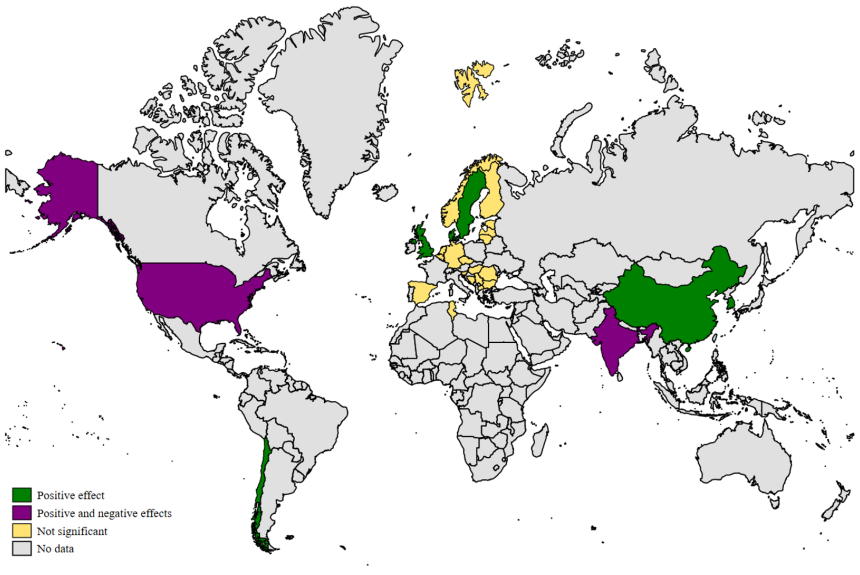
1. **Meta-analysis** of existing studies on same-sex teacher effects in primary & secondary education
   - 538 estimates from 24 studies
   - Average effect of **0.030 SD** on grades and test scores
   - After correcting for **publication bias**, estimates vary widely–**from -0.038 SD to 0.039 SD**
2. **Multi-country study** on same-sex teacher effects in primary & secondary education
   - Global evidence on 3 million students from 90 countries
   - Same-sex teacher effect on test scores **0.015 SD**
   - Same-sex teacher effect on job preferences **0.063 SD**
3. We study the **global distribution** and **generalizability** of same-sex teacher effects across subjects & grades.

Meta Analysis

Data Collection Methodology

- Pre-registered meta-analysis with Open Science Foundation (`https://osf.io/prereg/`)
- Inclusion criteria:
  1. empirical estimates of same-sex teacher effects
  2. on test scores or grades
  3. in primary or secondary education
- Systematic search: Google Scholar, Web of Science, pre-registered trials (at `https://socialscienceregistry.org`, `https://www.cos.io` and `https://www.researchregistry.com`)
- For all estimates in paper & appendices, we extract & standardize effect size + standard error

The Meta-Data Collected



Meta-Analysis Data & Empirics

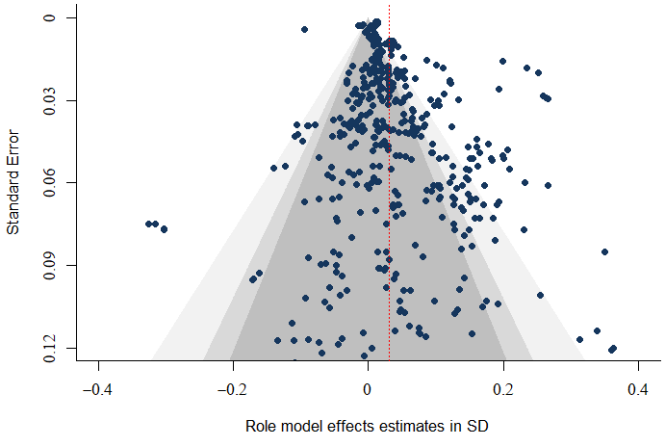
- 538 estimates from 24 studies
  - Median study: 10,916 observations from one country
  - Median ex-post MDE is 0.129 SD (huge)

- Estimate **3-level random effects model** (REML)
  - allows true same-sex teacher effects to differ by study + accounts the dependence of estimates within study
  - allows estimates to differ because of sampling error & differences in true effects between studies

- Use Knapp-Hartung adjustment to s.e. to account for uncertainty in between-study heterogeneity

Meta-Analysis Results

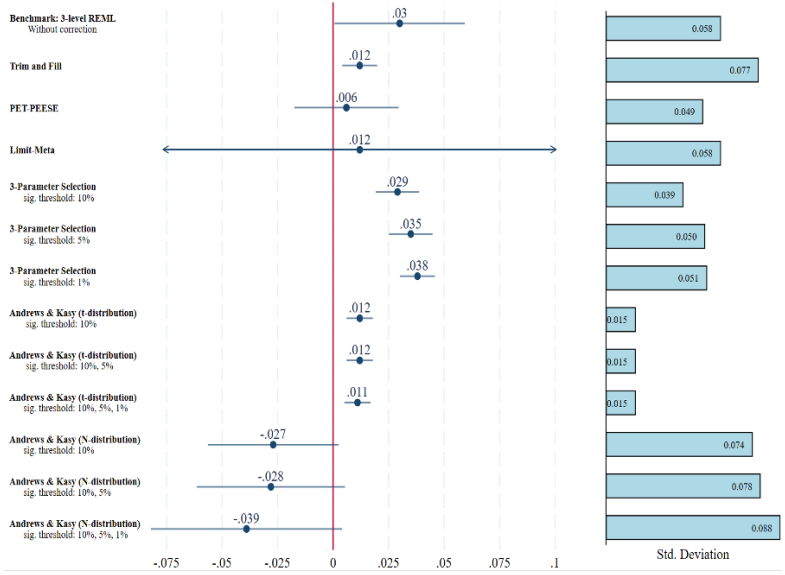
- Estimate of *true* underlying average effect: **0.030 SD**
- Estimates are all over the place! Estimate of std deviation of underlying effect: **0.058 SD**
- **Gain in statistical power**: ex-post MDE is 0.036 SD – 3.6 times smaller than 0.129 SD.
- **Underpowered estimates**: 79/538 estimates would have enough statistical power to detect the average effect of 0.030 SD!



Publication Bias?

- **Publication bias**: editors, referees, authors often prefer significant results
  - Can lead to *p*-hacking, selective reporting, file-drawer problem
- There are many **different methods** to correct for publication bias
  - Different assumptions, not one agreed-upon correction method
  - We apply **12 popular procedures** to correct for publication bias

Sensitivity to Publication Bias Corrections



Conclusions on Meta-Analysis

- **same-sex teacher effects are small**
- They disappear depending on how we correct for **publication bias**
- Our meta-regressions are **underpowered**
  - We can't say much about heterogeneous effects
  - We cannot identify settings where same-sex teacher effects matter more, because papers use different methodologies
- **We still do not know the true size of rule model effects.** They could be...
  - Zero?
  - Small and positive?

Multi-Country Analysis

Multi-Country Study Design

Our multi-country study answers

one research question

using **consistent methodology**

with **data** from **different settings**.

Benefits of Multi-Country Studies

- Increased **generalizability**
- More **statistical power** – the ability to detect small effects
- Potential to **explain heterogeneity** with institutional factors
- Greatly reduces risk of publication bias

Benefits of Multi-Country Studies

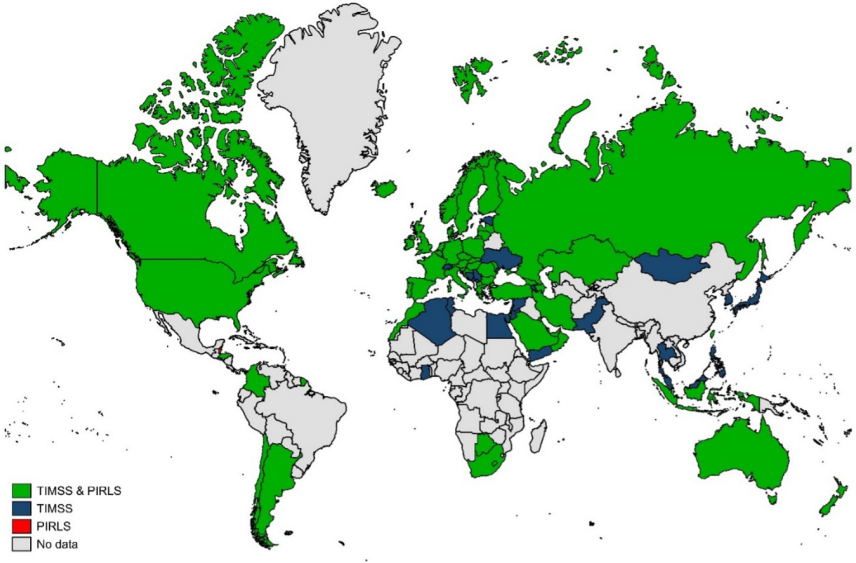
- Estimating **causal effects** in **multiple settings** is key for generalizing findings
- Using data from multiple settings is **not new**: increasing trend in social sciences
  - Anthropology: discussion of costs and benefits of multi-site studies discussed in 1995
  - Psychology: multi-lab replication studies
  - Developmental Psychology: "Many Babies" ongoing initiative
  - Neuroscience: discussions around use of secondary data because data collection is costly
- In economics
  - Altmejd et al (2021)
  - Bergbauer et al (2021)
  - Dhuey and Bedard (2006)
  - Woessmann and West (2005)

The Data

Data

- We combine **TIMSS** & **PIRLS** datasets from 1995 to 2019
  - 90 countries across 6 continents – of which 55 where same-sex teacher effects were never studies before
  - 3,047,752 students – over 90 times larger sample size of the median study
  - 231,942 teachers
  - 105,916 primary and secondary schools
- Large-scale international studies with standardized tests
  - **Math** and **Science** (TIMSS) and **Reading** (PIRLS)
  - Detailed Student, Parent and Teacher questionnaires
- Sampling: Grades 4 and 8
  - Random sample of 150 to 200 schools in each country
  - 1 to 3 classrooms within target grade in sampled school
  - all students in sampled classroom
  - min. 5,000 students per country-wave

Data



Empirical Strategy

Empirical Strategy

$$Y_{isj} = \beta_1 FemaleStud._i + \beta_2 FemaleTeach._j + \beta_3 FemaleStud._i \times FemaleTeach._j + \gamma' X_{isj} + u_{isj}$$

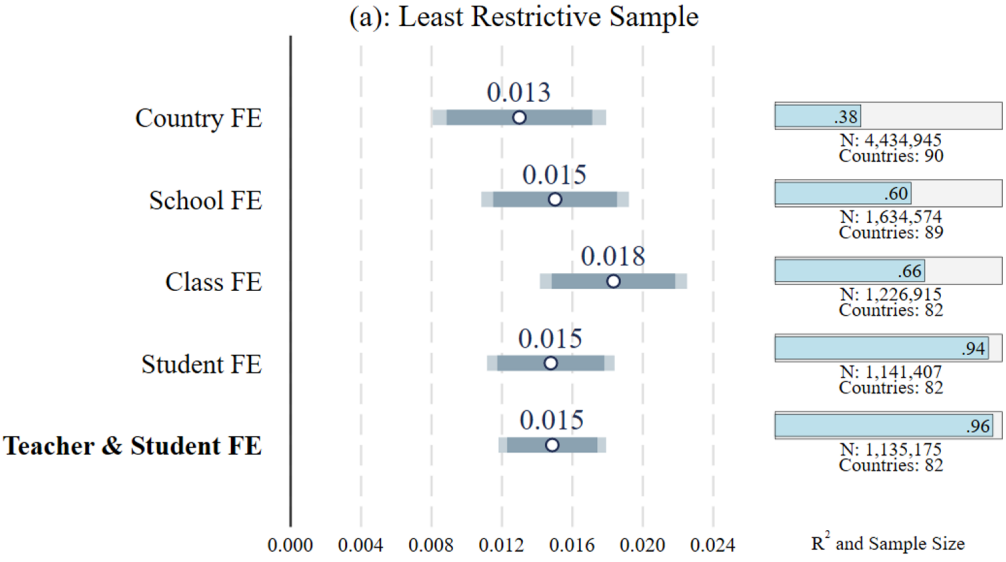
- β_3: same-sex teacher effect of interest
- Y_{isj}: **test scores, occupational preference**, subject enjoyment or subject confidence
- X_{isj} always includes:
  - subject fixed effects (e.g. science)
  - subject X female student interaction terms (e.g. sicence X female student)
  - subject X female teacher interaction terms (e.g. sicence X female teacher)
- We cluster **standard errors** at the **classroom level**.

Fixed Effects and Identifying Variation

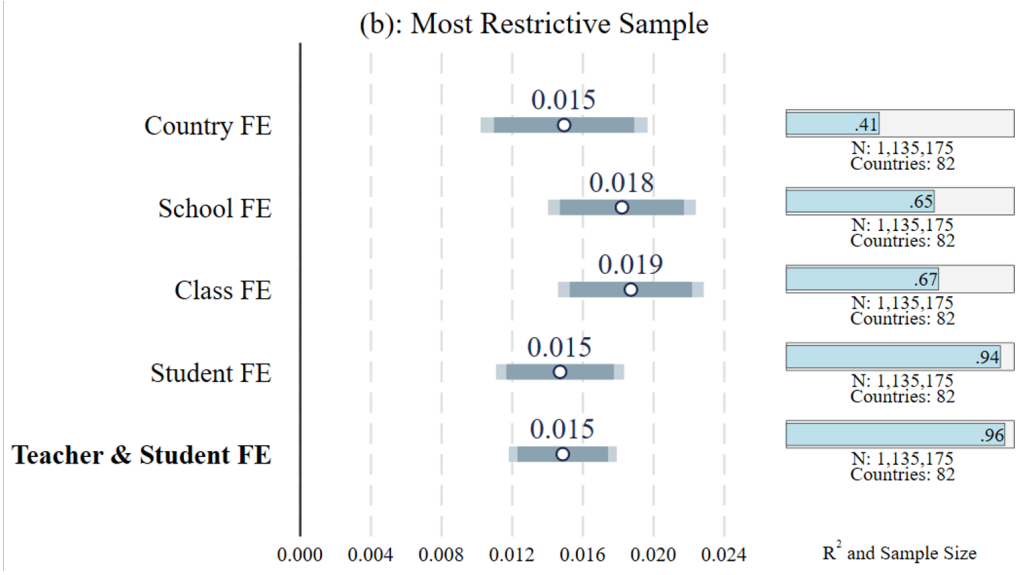
- **School FE**
  - Benchmark – common in literature (e.g. Dee, 2007; Lim & Meer, 2017, 2020...)
  - Addresses concerns of sorting of students/teachers into schools.
  - Remaining sorting concern:
    - if particularly high-ability students are assigned to same-sex teachers,
    - if high-quality teachers assigned to more same-sex students.
  - Balancing Table here
- Classroom FE
- Teacher FE
- Student FE
- **Student & Teacher** two-way **fixed effects** – rules out most concerns about omitted variable bias.
  - Our preferred specification – exploits within-student across-subject variation
  - Addresses concern that better teachers could be assigned to higher share of same-sex students.

Main Results

Main Results on Test Scores



Main Results on Test Scores

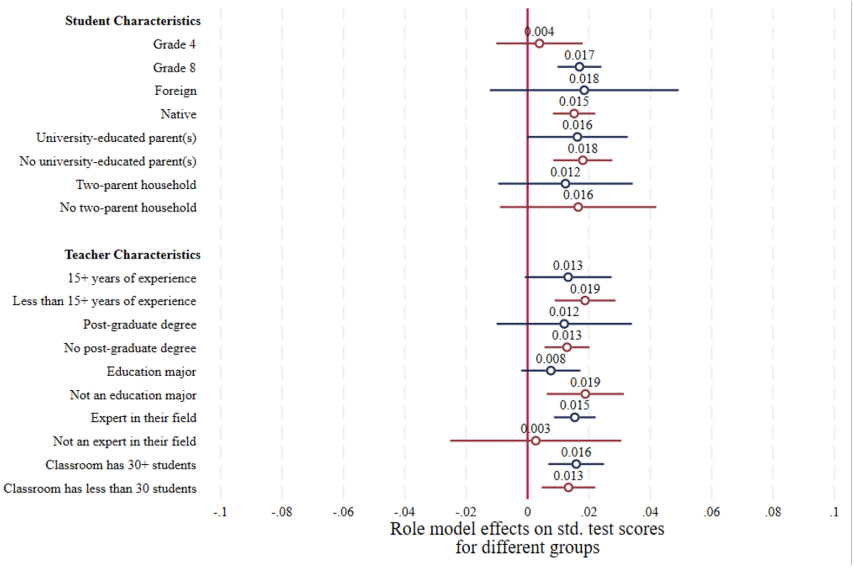


Discussion of Results

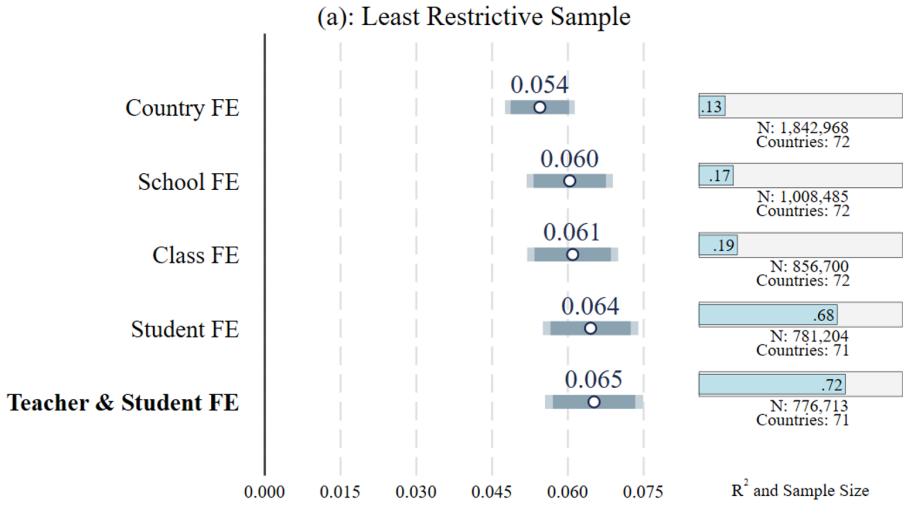
- **Effect is half the size of meta-estimate**. What drives this difference?
  - With methodology constant and no publication bias concerns, we get a sense of what is NOT driving same-sex teacher estimates.

- **Our same-sex teacher effect of 0.015 SD is small**:
  - 1.5 point increase on TIMSS/PIRLS tests
  - 40 times smaller than predicted effect on test score of having at least one uni-educated parent
  - 10 times smaller than the effect of a 1 SD increase in Teacher VA on math test scores (Chetty, Friedman and Rockoff, 2014)
  - 8 times smaller than effect of experienced v. rookie teacher on math test scores (Clotfelder, Ladd and Vigdor, 2006; Hanushek et al. 2005)

Heterogeneity of Effects on Test Scores

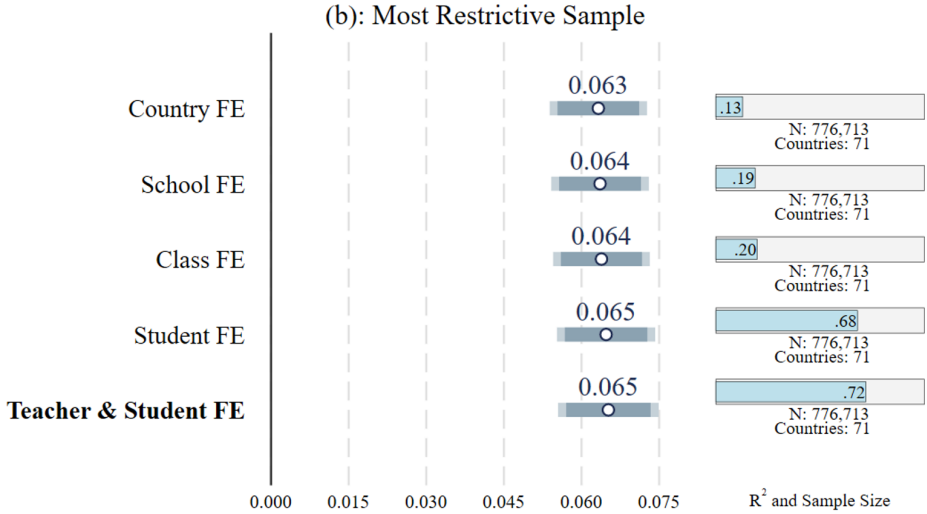


Main Results on Job Preferences



"I would like a **job** that involves using **mathematics/science**"

Main Results on Job Preferences



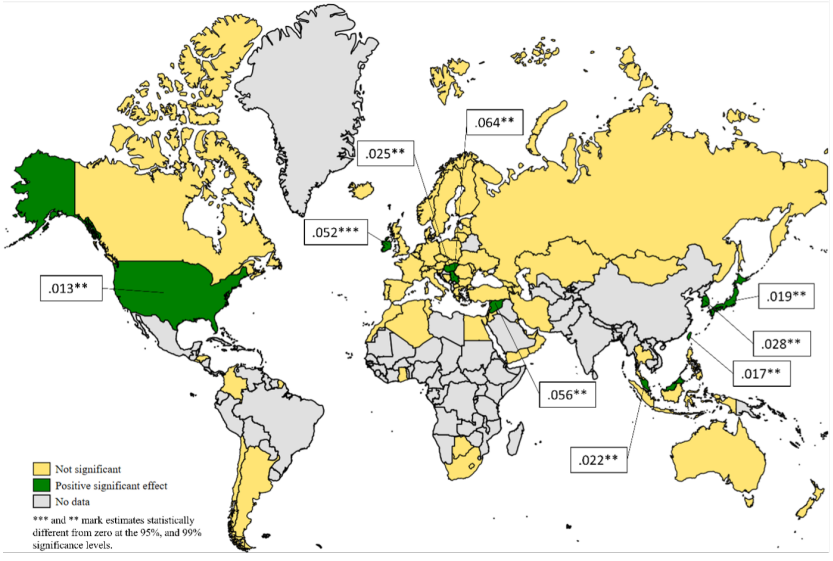
"I would like a **job** that involves using **mathematics/science**"

Global Heterogeneity

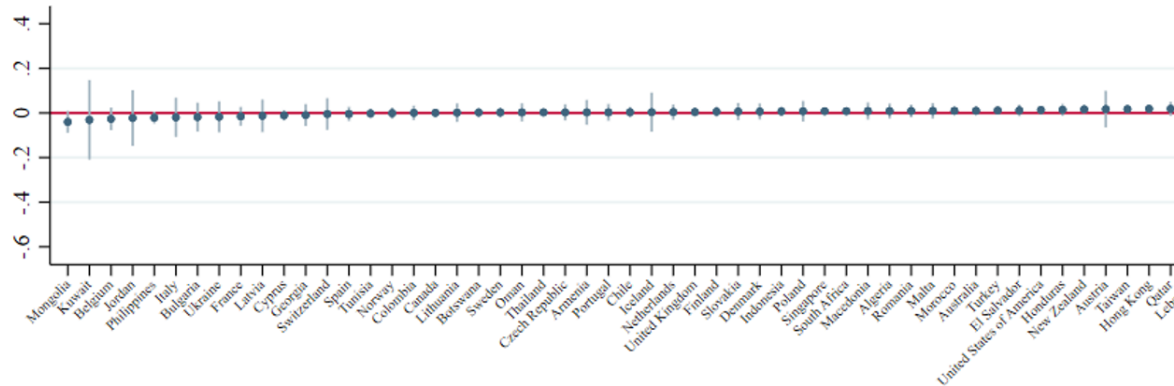
Global Heterogeneity in Same-Sex Teacher Effects

- How **generalizable** are same-sex teacher effects across contexts?
- We estimate country-specific same-sex teacher effects **holding methodology constant**
- No publication bias concerns because no incentives to ignore insignificant estimates

Effects on Test Scores by Country

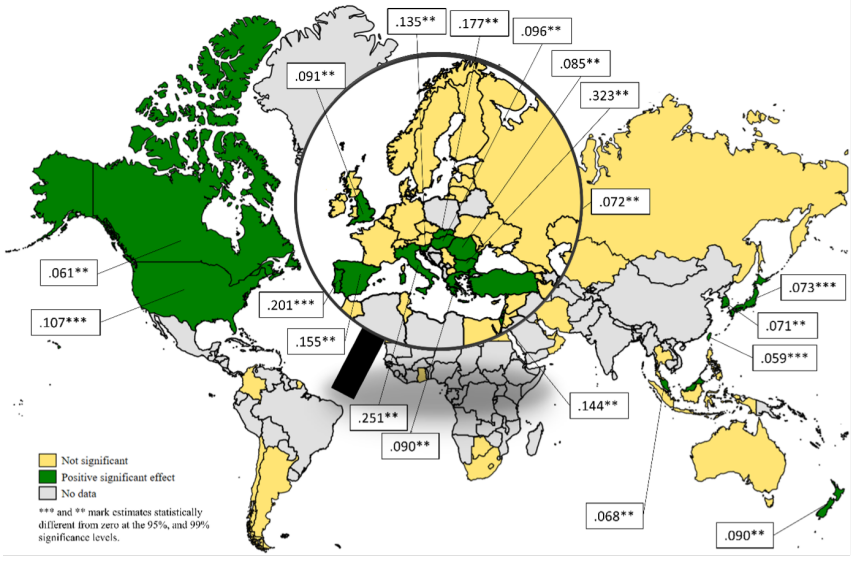


Effects on Test Scores by Country

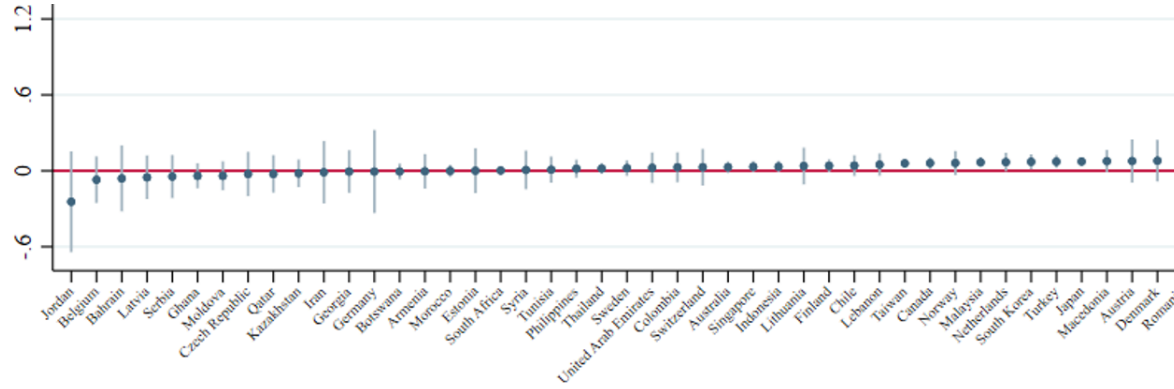


- Differences in effects sizes could be largely driven by **sampling error!**
- Estimated SD of underlying effect: **0.004**

Effects on Job Preferences by Country



Effects on Job Preferences by Country



- Here we do find meaningful heterogeneity
- Estimated SD of underlying effect: **0.028**

How Universal are same-sex teacher effects?

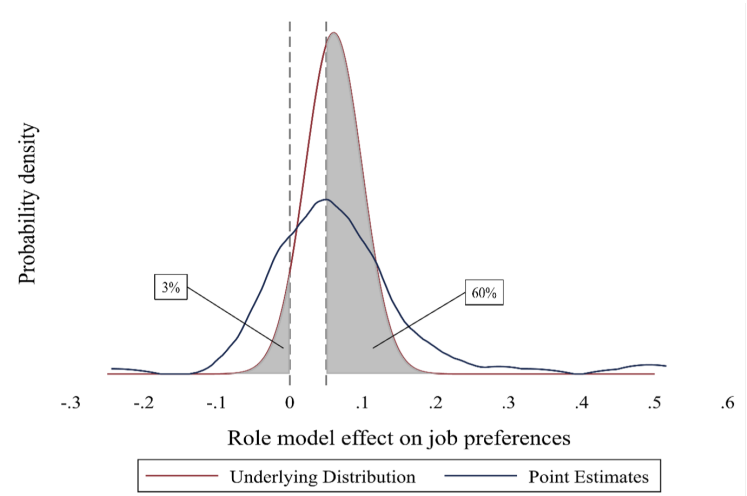
- Small mean effects may not be very informative for policy.
- Differences in country-level estimates can reflect **true heterogeneity** or **sampling error**
- We estimate the underlying **distribution** of effects clean of sampling error.
- Knowing the distribution informs us about the probability of finding meaningful **positive same-sex teacher effects**.

The Distribution of same-sex teacher effects

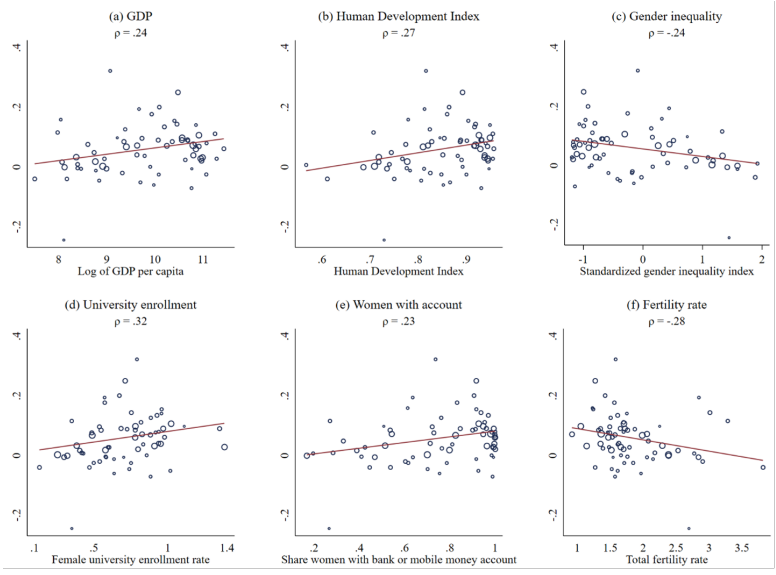
- In **primary education**, same-sex teacher effects vary substantially for different outcomes.
  - Positive effects in 46-94% of countries.
- In **secondary education** same-sex teacher effects are near **universally** positive.
  - Positive effects in more than 95% of the countries for **all outcomes**.
- For **test scores effects** are **universally tiny** – the estimated SD is 0.0002.
- What does the distribution look like for same-sex teacher effects on **job preferences**?

The Distribution of same-sex teacher effects on job preferences

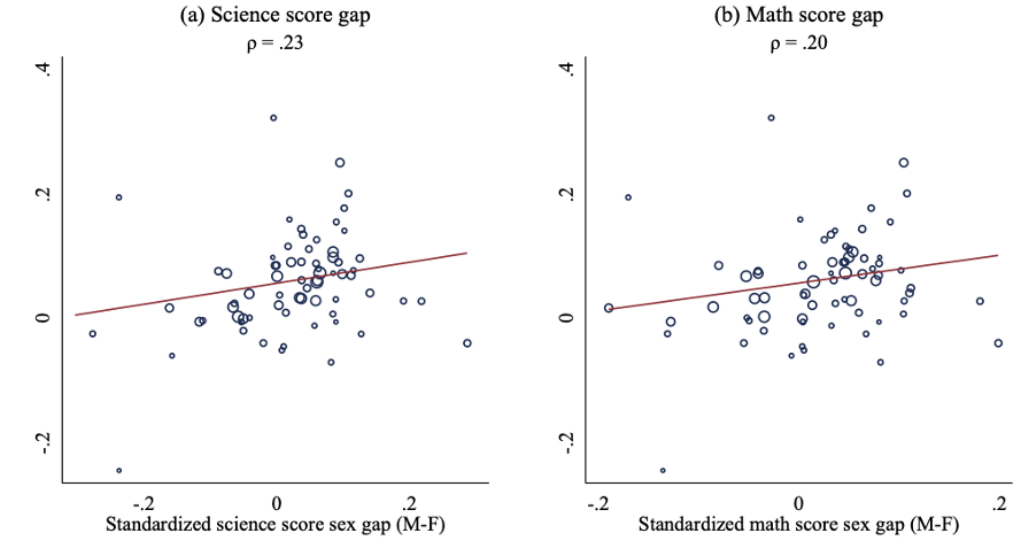
- Average effect: 0.06 SD
- Estimated SD of effect: 0.03 SD
- Prob. effect positive: 97%
- Prob. effect > 0.05 SD: 60%



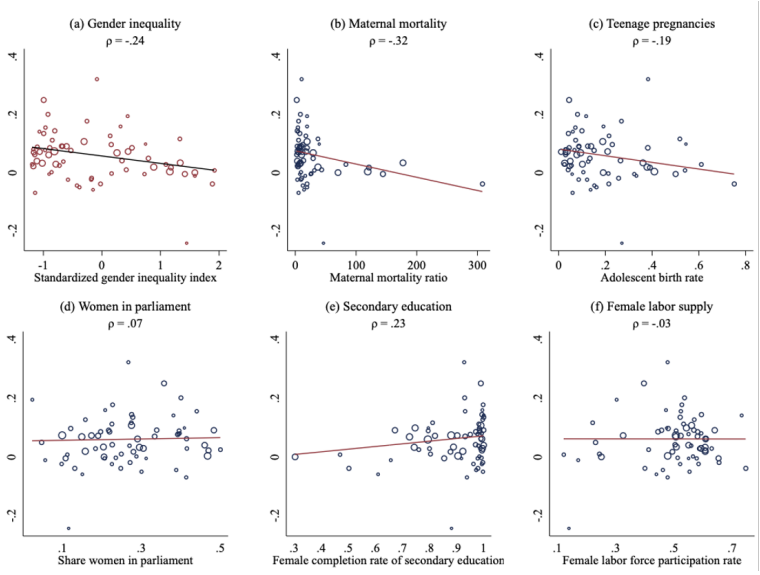
What Explains Global Heterogeneity on Job Preferences?



What Explains Global Heterogeneity on Job Preferences?



What Explains Global Heterogeneity on Job Preferences?



Some Speculation

- Same-sex teacher effects on job preferences are only visible in **high-income** and **gender-equal** countries.
- Same-sex teacher effects may be smaller when **job choices** are more **constrained** by necessity and tradition.
- **Fewer opportunities for STEM careers** when children expected to work in the family farm or business.

Conclusions

Conclusions

- **Meta-analysis** shows same-sex teacher effects for test scores are small.
  - Sensitive to publication bias correction.
  - Low power to analyze heterogeneity.
- **Multi-country analysis** reveals:
  - Tiny effects on test scores
  - Small, positive in primary education in 46% of countries.
  - Tiny but near-universally positive in secondary education.
  - Modest effects on job preferences, subject enjoyment and self-confidence
  - Concentrated in rich and gender-equal countries
- **Policy conclusions**:
  - Same-sex teachers will NOT close performance gaps.
  - But they may reduce occupational segregation in high-income countries.

Reflections

- For many research questions, **conventional studies are underpowered**!
- Lack of power and publication bias (p-hacking, selective reporting, file-drawer problem) mean that any one additional study is often not very helpful.
- Meta-analyses face an uphill battle because of publication bias
- Statistical corrections for publication bias are not reliable (depend on assumptions, lead to different results)

When to conduct a Multi-Country Study?

- For **mature strands of literature** that have **failed to reach a consensus**
  - When expect effects to be small or non-existent
  - Effects appear to be strongly **context-specific**
  - When you want to **settle a debate**

- Conventional studies will continue to make important contributions
  - They introduce novel research questions
  - Provide first-time evidence
  - Start new strands of literature

Multi-Setting Studies

Benefits of multi-setting studies:

- More statistical power – necessary when expect small effects
- No missing estimates
- Can detect & explain heterogeneity (holding methodology constant)
- Can truly explore universality of effects, which is essential to assess generalizability of findings

Thank you!

Alexandra de Gendre – a.degendre@unimelb.edu.au

Appendix

Balancing Table

Back

		Role model effect				
	Mean (1)	Coef. (2)	Std. err. (3)	R-Squared (4)	Countries (5)	N (6)
Student characteristics:						
Age (in years)	12.9	**0.0094**	(0.0019)	0.88	89	1,628,689
Foreign-born	0.09	-0.0008	(0.0008)	0.22	88	1,470,027
Parent(s) with uni. degree	0.38	-0.0009	(0.0015)	0.31	76	963,126
Two-parent household	0.66	-0.0018	(0.0023)	0.39	52	364,662
Teacher characteristics:						
40+ years old	0.76	-0.0017	(0.0021)	0.57	89	1,630,965
Experience (in years)	15.6	0.0297	(0.0283)	0.57	89	1,605,102
Has post-graduate degree	0.29	-0.0007	(0.0012)	0.64	89	1,571,995
Majored in education	0.63	**0.0035**	(0.0015)	0.60	86	1,301,828
Teaches field of expertise	0.86	0.0002	(0.0010)	0.67	82	1,273,757