Understanding Gender Discrimination by Managers

Christina Brown (University of Chicago)

July 25, 2023
NBER SI Gender

Support from DFID, JPAL, the Weiss Fund, CEGA, the Strandberg Fund, NAEd/Spencer Foundation, DERF and IRLE

Women in the Workforce

- Extensive evidence showing bias in the evaluation of women's on-the-job productivity (hiring, wages, promotions)

Ratio of female to male labor force participation (2019)

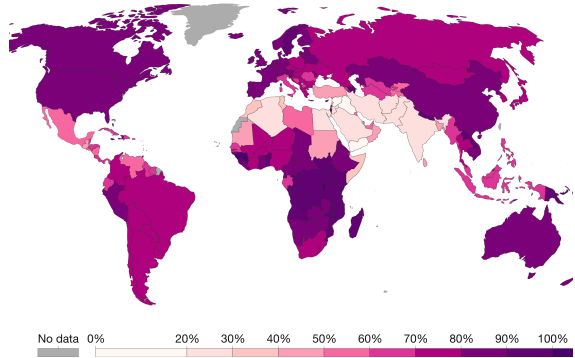


Figure: Our World in Data; Data: ILO

Women in the Workforce

- Extensive evidence showing bias in the evaluation of women's on-the-job productivity (hiring, wages, promotions)
- South Asia: FLFP is 1/4 of MLFP

Ratio of female to male labor force participation (2019)

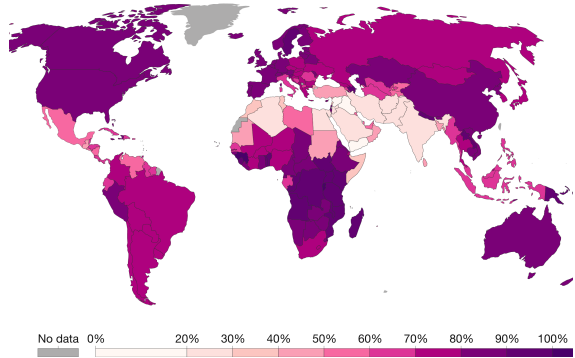


Figure: Our World in Data; Data: ILO

Women in the Workforce

- Extensive evidence showing bias in the evaluation of women's on-the-job productivity (hiring, wages, promotions)
- South Asia: FLFP is 1/4 of MLFP
- This paper →
  - How do employment policies affect bias?
  - What does that tell us about the underlying features of the bias?

Ratio of female to male labor force participation (2019)

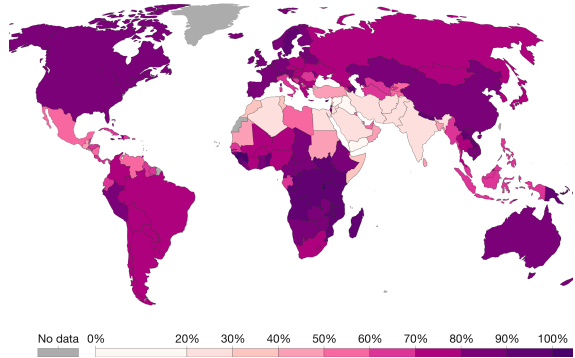


Figure: Our World in Data; Data: ILO

Overview - Design

(1) Research questions

a. **Policy:** How do personnel policies affect the extent of gender bias in performance evaluations?

b. **Theoretical:** What are the underlying mechanisms for gender bias in performance evaluations: taste-based, statistical, other?

Overview - Design

(1) Research questions

a. **Policy:** How do personnel policies affect the extent of gender bias in performance evaluations?
b. **Theoretical:** What are the underlying mechanisms for gender bias in performance evaluations: taste-based, statistical, other?

(2) Design: RCT with 5,000 teachers and 200 managers in Pakistan

a. Vary features of performance evaluation process
   - **Financial stakes:** Whether manager's evaluation determines employee's raise
   - **Observation:** Vary frequency of classroom observations done by managers

Overview - Design

(1) Research questions

a. **Policy:** How do personnel policies affect the extent of gender bias in performance evaluations?
b. **Theoretical:** What are the underlying mechanisms for gender bias in performance evaluations: taste-based, statistical, other?

(2) Design: RCT with 5,000 teachers and 200 managers in Pakistan

a. Vary features of performance evaluation process
   - **Financial stakes:** Whether manager's evaluation determines employee's raise
   - **Observation:** Vary frequency of classroom observations done by managers
b. Measure:
   - **Employee effort and productivity:** value-added, clock in times, videos of classes
   - **Performance evaluation:** Evaluation criteria and scores
   - **Manager beliefs:** Employee effort, gender bias, preferences
c. Follow up vignette survey to test mechanisms

Overview - Results

(1) Explicit Bias: Minimal explicit bias on ability as measured by manager and teacher surveys

Overview - Results

(1) Explicit Bias: Minimal explicit bias on ability as measured by manager and teacher surveys

(2) Evaluation Scores (controlling for productivity)

Overview - Results

(1) Explicit Bias: Minimal explicit bias on ability as measured by manager and teacher surveys

(2) Evaluation Scores (controlling for productivity)

- No financial stakes $\rightarrow$ *No gender bias*

Overview - Results

(1) Explicit Bias: Minimal explicit bias on ability as measured by manager and teacher surveys

(2) Evaluation Scores (controlling for productivity)

- No financial stakes $\rightarrow$ *No gender bias*
- Adding financial stakes $\rightarrow$ *Women receive 10% lower raise*

Overview - Results

(1) Explicit Bias: Minimal explicit bias on ability as measured by manager and teacher surveys

(2) Evaluation Scores (controlling for productivity)

- No financial stakes → *No gender bias*
- Adding financial stakes → *Women receive 10% lower raise*
- Increasing monitoring → *Reduces effect of financial stakes by 66%*

Overview - Results

(1) Explicit Bias: Minimal explicit bias on ability as measured by manager and teacher surveys

(2) Evaluation Scores (controlling for productivity)

- No financial stakes → *No gender bias*
- Adding financial stakes → *Women receive 10% lower raise*
- Increasing monitoring → *Reduces effect of financial stakes by 66%*

(3) Discrimination Framework

Overview - Results

(1) Explicit Bias: Minimal explicit bias on ability as measured by manager and teacher surveys

(2) Evaluation Scores (controlling for productivity)

- No financial stakes → *No gender bias*
- Adding financial stakes → *Women receive 10% lower raise*
- Increasing monitoring → *Reduces effect of financial stakes by 66%*

(3) Discrimination Framework

- No differences in mean, variance or noisiness of productivity

Overview - Results

(1) Explicit Bias: Minimal explicit bias on ability as measured by manager and teacher surveys

(2) Evaluation Scores (controlling for productivity)

- No financial stakes → *No gender bias*
- Adding financial stakes → *Women receive 10% lower raise*
- Increasing monitoring → *Reduces effect of financial stakes by 66%*

(3) Discrimination Framework

- No differences in mean, variance or noisiness of productivity
- Effort is imperfectly observable + differences in dis-utility of low raise by gender
  → *Disparate outcomes*

Overview - Results

(1) Explicit Bias: Minimal explicit bias on ability as measured by manager and teacher surveys

(2) Evaluation Scores (controlling for productivity)

- No financial stakes → *No gender bias*
- Adding financial stakes → *Women receive 10% lower raise*
- Increasing monitoring → *Reduces effect of financial stakes by 66%*

(3) Discrimination Framework

- No differences in mean, variance or noisiness of productivity
- Effort is imperfectly observable + differences in dis-utility of low raise by gender
  → *Disparate outcomes*

(4) Mechanisms (source of differential dis-utility):

Overview - Results

(1) Explicit Bias: Minimal explicit bias on ability as measured by manager and teacher surveys

(2) Evaluation Scores (controlling for productivity)

- No financial stakes → *No gender bias*
- Adding financial stakes → *Women receive 10% lower raise*
- Increasing monitoring → *Reduces effect of financial stakes by 66%*

(3) Discrimination Framework

- No differences in mean, variance or noisiness of productivity
- Effort is imperfectly observable + differences in dis-utility of low raise by gender
  → *Disparate outcomes*

(4) Mechanisms (source of differential dis-utility):

- Differences in perceived "deservedness" (household income) ✓

Overview - Results

(1) Explicit Bias: Minimal explicit bias on ability as measured by manager and teacher surveys

(2) Evaluation Scores (controlling for productivity)

- No financial stakes → *No gender bias*
- Adding financial stakes → *Women receive 10% lower raise*
- Increasing monitoring → *Reduces effect of financial stakes by 66%*

(3) Discrimination Framework

- No differences in mean, variance or noisiness of productivity
- Effort is imperfectly observable + differences in dis-utility of low raise by gender → *Disparate outcomes*

(4) Mechanisms (source of differential dis-utility):

- Differences in perceived "deservedness" (household income) ✓
- Differences in response (turnover, complaining) ✗

Contribution

- Interaction between bias and HR policies: Better information and less financial stakes decreases bias
  Biasi and Sarsons, 2022; Beg, Fitzpatrick and Lucas, 2021; Blau and Kahn, 2017
- Information and Discrimination: Better information about worker productivity lowers bias
  Laouénan and Rathelot, 2022; Bohren, Imas and Rosenberg, 2019; Sarsons, 2017; Bordalo et al, 2017
- Financial discrimination: Disparate employment outcomes without discrimination on productivity

Setting

Experiment

- Large private school network operating hundreds of schools across urban Pakistan
- Grades 4-13 in English, Urdu, math and science
- Managers are principal or vice principals

Setting

Experiment

- Large private school network operating hundreds of schools across urban Pakistan
- Grades 4-13 in English, Urdu, math and science
- Managers are principal or vice principals

Relevance

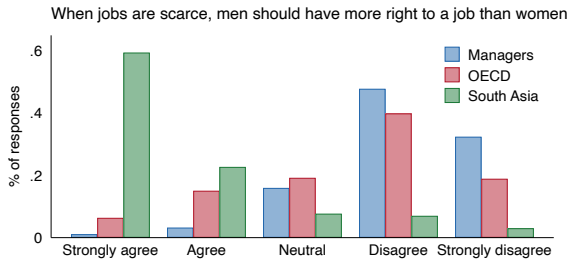
- **Useful for personnel econ:** Multiple, hard to measure outcomes
- **Very relevant for FLFP:** 51% of women in labor force with HS degree are teachers (8% of entire female labor force)

Conceptual Framework - Context

1). Taste-based

- Minimal stated/perceived gender bias on ability

Managers are very gender progressive



On average, teachers do not think there is bias in favor or against female teachers in evaluations.

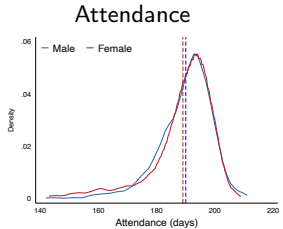
Conceptual Framework - Context

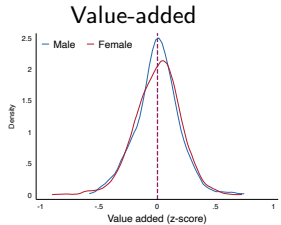
1). Taste-based

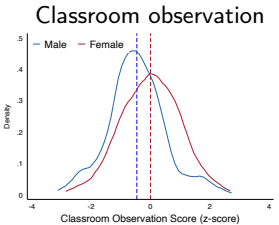
- Minimal stated/perceived gender bias on ability

2). Statistical

- Same productivity (mean & sd)







Conceptual Framework - Context

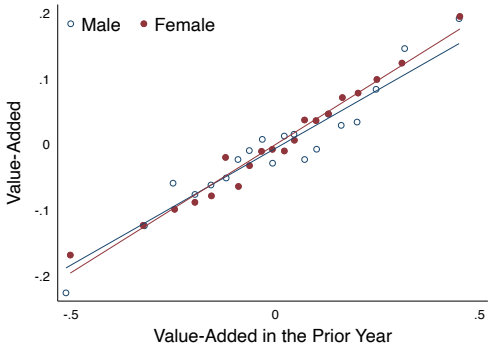
1). Taste-based

- Minimal stated/perceived gender bias on ability

2). Statistical

- Same productivity (mean & sd)
- Same production function noise

Correlation in VA over time by gender



Conceptual Framework - Context

1). Taste-based

- Minimal stated/perceived gender bias on ability

2). Statistical

- Same productivity (mean & sd)
- Same production function noise

3). This paper –
Disparate outcomes arise from:

- Noisy production function
- Dis-utility to manager of giving low wages varies by gender

Conceptual Framework - Set-up

The performance evaluation system takes place in three stages

1). Production: Employees work and produce output which is a noisy function of their effort

2). Evaluation: Manager provides evaluation score

Conceptual Framework - Set-up

The performance evaluation system takes place in three stages

1). Production: Employees work and produce output which is a noisy function of their effort

2). Evaluation: Manager provides evaluation score

3). Post-Evaluation:

a. Dis-utility from inaccurate scores

- e.g. system legitimacy, psychic cost of lying, punishment
- ***Decreasing in production function noisiness***

Conceptual Framework - Set-up

The performance evaluation system takes place in three stages

1). Production: Employees work and produce output which is a noisy function of their effort

2). Evaluation: Manager provides evaluation score

3). Post-Evaluation:

a. Dis-utility from inaccurate scores
   - e.g. system legitimacy, psychic cost of lying, punishment
   - ***Decreasing in production function noisiness***

b. Dis-utility from low wages
   - e.g. complaints, guilt, turnover
   - ***Increasing in financial stakes of evaluation***

Conceptual Framework - Set-up

The performance evaluation system takes place in three stages

1). Production: Employees work and produce output which is a noisy function of their effort

2). Evaluation: Manager provides evaluation score

3). Post-Evaluation:

a. Dis-utility from inaccurate scores
   - e.g. system legitimacy, psychic cost of lying, punishment
   - ***Decreasing in production function noisiness***

b. Dis-utility from low wages
   - e.g. complaints, guilt, turnover
   - ***Increasing in financial stakes of evaluation***
   - Allow this to vary by worker type

Conceptual Framework - Set-up

The performance evaluation system takes place in three stages

1). Production: Employees work and produce output which is a noisy function of their effort

2). Evaluation: Manager provides evaluation score

3). Post-Evaluation:

   a. Dis-utility from inaccurate scores
      - e.g. system legitimacy, psychic cost of lying, punishment
      - ***Decreasing in production function noisiness***

   b. Dis-utility from low wages
      - e.g. complaints, guilt, turnover
      - ***Increasing in financial stakes of evaluation***
      - Allow this to vary by worker type

Manager chooses evaluation score given post-evaluation costs and benefits.

Conceptual Framework - Set-up

The performance evaluation system takes place in three stages

1). Production: Employees work and produce output which is a noisy function of their effort

2). Evaluation: Manager provides evaluation score

3). Post-Evaluation:

a. Dis-utility from inaccurate scores
- e.g. system legitimacy, psychic cost of lying, punishment
- ***Decreasing in production function noisiness***

b. Dis-utility from low wages
- e.g. complaints, guilt, turnover
- ***Increasing in financial stakes of evaluation***
- Allow this to vary by worker type

Gender bias increasing in noise & financial stakes

Manager chooses evaluation score given post-evaluation costs and benefits.

Design - Treatments

- **All teachers:** Managers set performance evaluation criteria and rate teachers

Design - Treatments

- **All teachers:** Managers set performance evaluation criteria and rate teachers

- **Treatment 1:** Financial Stakes of Evaluation (randomized at school level)
  - *Control:* Employee's end of year raise is determined by:
    - *Flat Raise:* Employees receive a raise of 5% of their base salary
    - *Objective Raise:* Teachers receive a raise from 0-10% based on student test scores

Design - Treatments

- **All teachers:** Managers set performance evaluation criteria and rate teachers

- **Treatment 1:** Financial Stakes of Evaluation (randomized at school level)
  - *Control:* Employee's end of year raise is determined by:
    - *Flat Raise:* Employees receive a raise of 5% of their base salary
    - *Objective Raise:* Teachers receive a raise from 0-10% based on student test scores
  - *Treatment: Subjective Raise:* Teachers receive a raise from 0-10% based on their performance evaluation score

Design - Treatments

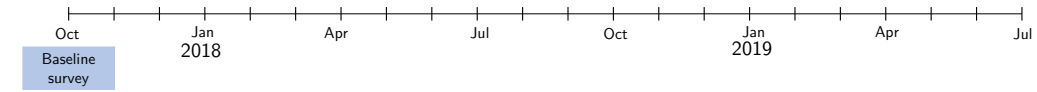
- **All teachers:** Managers set performance evaluation criteria and rate teachers

- **Treatment 1:** Financial Stakes of Evaluation (randomized at school level)
  - *Control:* Employee's end of year raise is determined by:
    - *Flat Raise:* Employees receive a raise of 5% of their base salary
    - *Objective Raise:* Teachers receive a raise from 0-10% based on student test scores
  - *Treatment: Subjective Raise:* Teachers receive a raise from 0-10% based on their performance evaluation score

- **Treatment 2:** Classroom Observations (randomized at teacher level)
  - Manager told to conduct monthly, unannounced 20 minute observations for 4 months before evaluation (relative to status quo)
    → *Treatment increases number of observations received by 50%*

Baseline balance: 2 of 27 coefficients are stat. sig. Table
Attrition: Administrative data available for all; 12% attrition for endline teacher survey

Design - Timeline



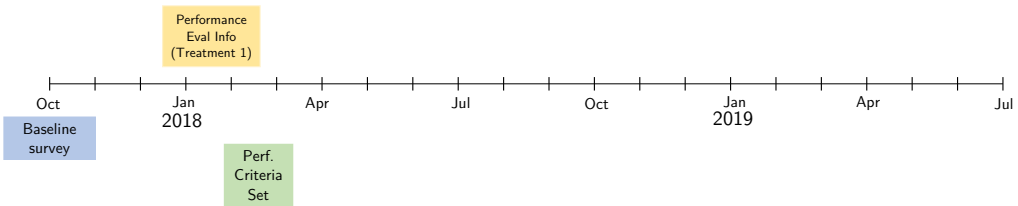
- Informed about incentive schemes
- Measure teacher characteristics

= Data collection activity
= Treatment implementation
= Both

Example criteria

Design - Timeline

- Research team meets in person with managers
- School system HR does in person presentation at each school
- Email information
- Displayed on teacher's dashboard

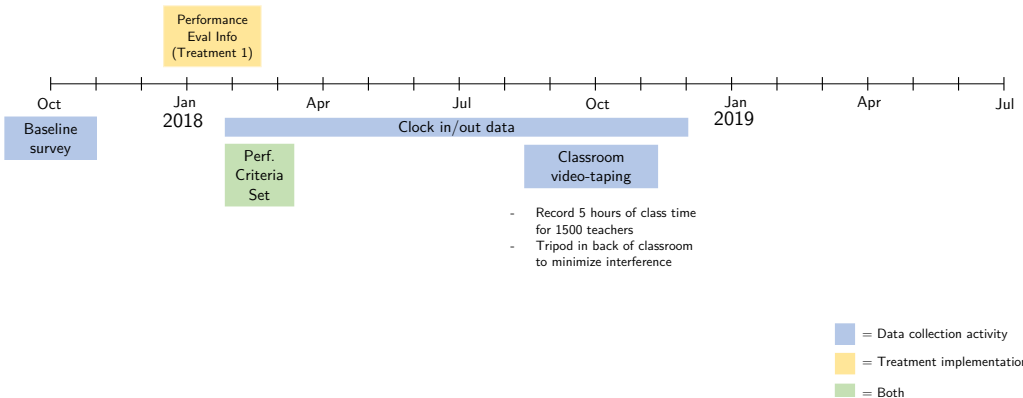


- Managers set criteria for ALL teachers (not new to experiment)

= Data collection activity
= Treatment implementation
= Both

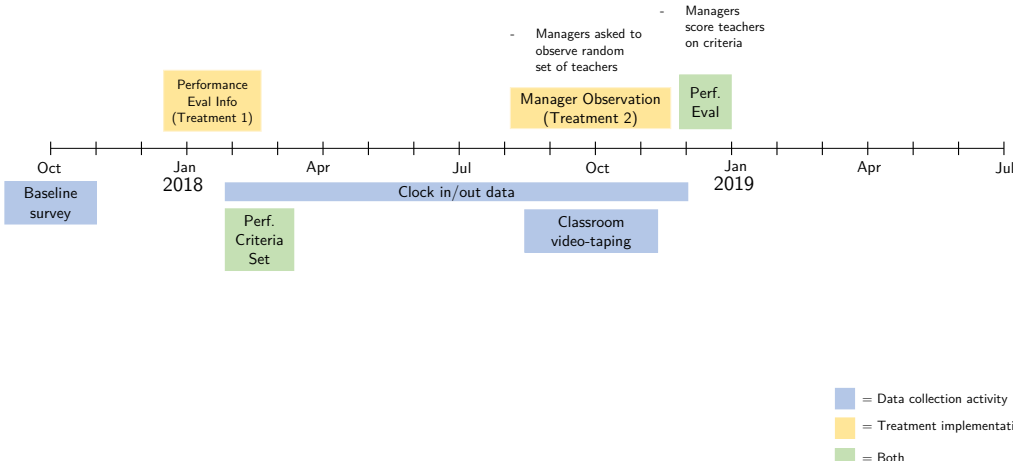
Example criteria

Design - Timeline



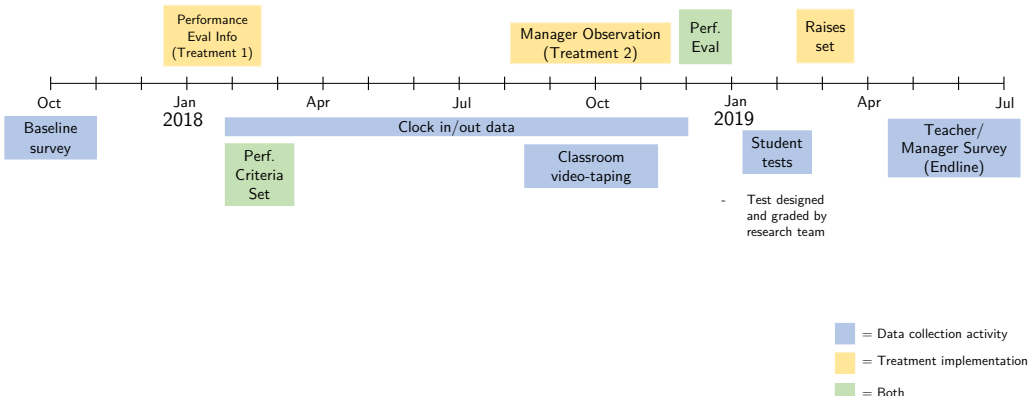
Example criteria

Design - Timeline



Example criteria

Design - Timeline



Example criteria

Design - Data

Type	N	Source	Outcomes
Teachers			
"Ground truth"	1,500	Class video	Rubric covering 20 aspects of pedagogy (Araujo et al, 2016)
	3,600	Admin data	Value-added (From 5 years of student test scores)
	9,100	Admin data	Daily clock in and out time

Design - Data

Type	N	Source	Outcomes
Teachers			
"Ground truth"	1,500	Class video	Rubric covering 20 aspects of pedagogy (Araujo et al, 2016)
	3,600	Admin data	Value-added (From 5 years of student test scores)
	9,100	Admin data	Daily clock in and out time
Managers			
Beliefs	189	Survey	Rate teachers on several criteria
	189	Admin data	Rate teachers on several criteria (after observation)
Preferences	189	Survey	Vignettes (rating hypothetical teachers)
	189	Survey	Rank importance of teacher behaviors
	189	Evaluation	Points allocated to criteria

Design - Data

Type	N	Source	Outcomes
Teachers			
"Ground truth"	1,500	Class video	Rubric covering 20 aspects of pedagogy (Araujo et al, 2016)
	3,600	Admin data	Value-added (From 5 years of student test scores)
	9,100	Admin data	Daily clock in and out time
Managers			
Beliefs	189	Survey	Rate teachers on several criteria
	189	Admin data	Rate teachers on several criteria (after observation)
Preferences	189	Survey	Vignettes (rating hypothetical teachers)
	189	Survey	Rank importance of teacher behaviors
	189	Evaluation	Points allocated to criteria
Evaluation	189	Evaluation	Total score and criteria-level score
Bias	189	Survey	World Values Survey questions
	189	Survey	Teacher's rating of manager's bias
	189	Survey	Varying gender of name in vignette

Teacher and Manager Sample

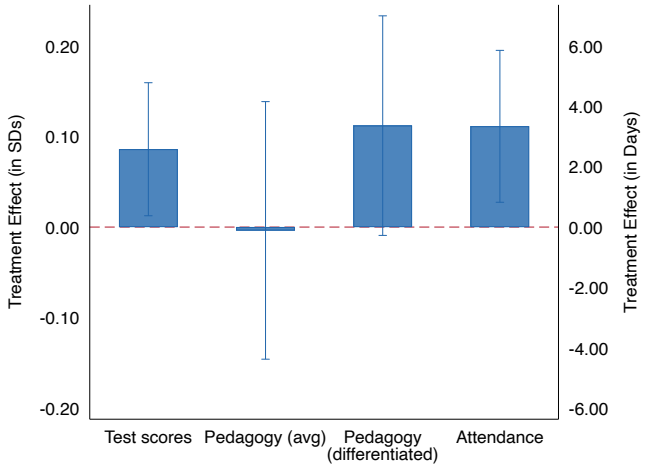
	Study Sample		US Sample	
	Mean	St. Dev.	Mean	St. Dev.
Panel A. Teacher Characteristics				
Age	35.0	8.9	41.8	7.5
Female	0.80	0.40	0.76	0.43
Years of experience	5.1	5.2	13.8	9.6
Has Post BA Education	0.68	0.47	0.54	0.50
Salary, USD	4,000	1,700	52,400	18,400
Panel B. Manager Characteristics				
Age	44.9	9.2	48.8	9.7
Female	0.61	0.49	0.53	0.50
Years of experience	9.6	7.9	13.0	7.5
Panel C. Manager Time Use				
Total hours worked	47.2	16.3	57.0	13.2
Hours spent on:				
- Administrative tasks	18.5	10.3	18.2	2.3
- Teacher management and teaching	17.5	8.2	15.1	2.0
- Student and parent interactions	6.3	4.4	20.2	2.7
- Other tasks	6.9	12.3	4.0	2.6

Col. 3 and 4 Source: School and Staff Survey (National Center for Education Statistics)

Treatment "First Stage"

Treatments effect teacher and manager behavior:

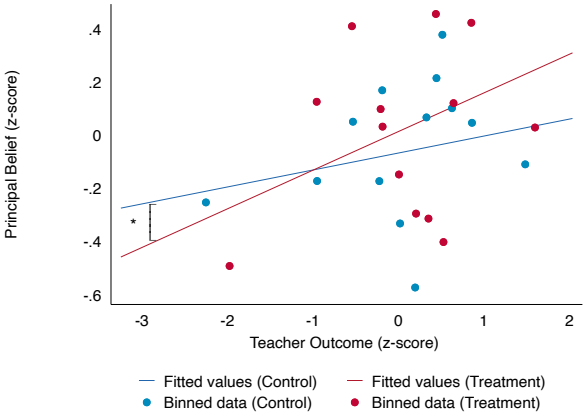
- Financial treatment: Teachers work harder and this effects student outcomes



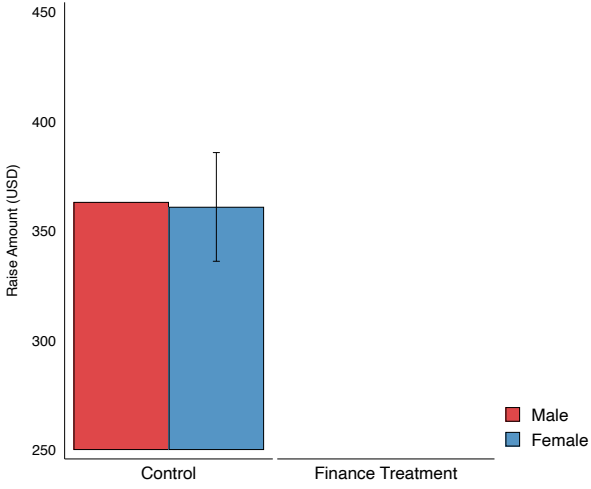
Treatment "First Stage"

Treatments effect teacher and manager behavior:

- Financial treatment: Teachers work harder and this effects student outcomes
- Observation treatment: Accuracy of managers' beliefs about teacher effort improves



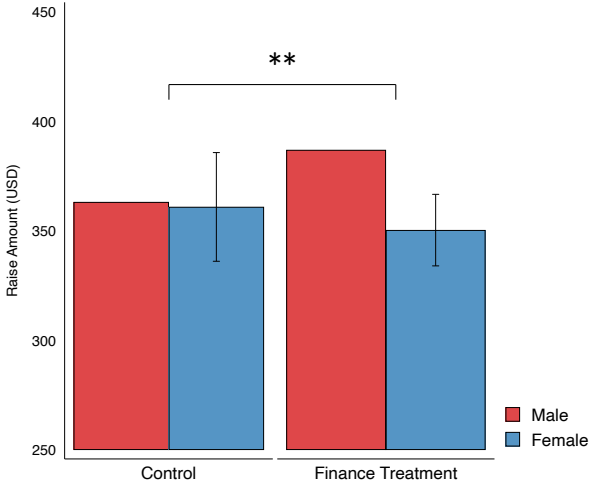
Effect of Financial Stakes on Bias



- No difference in evaluation score when no financial stakes
  Control for value-added, attendance, time use, video scores

Table

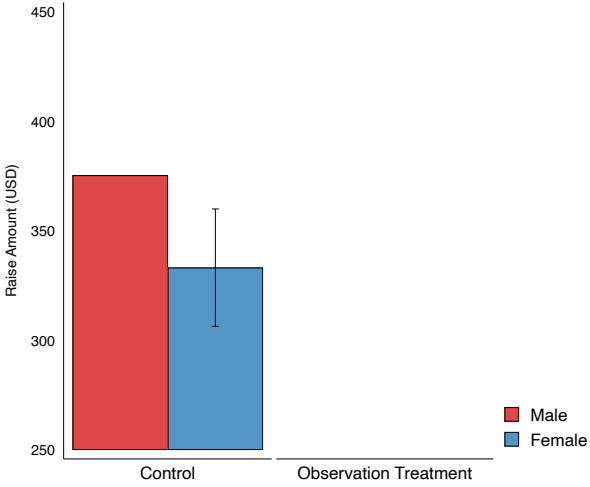
Effect of Financial Stakes on Bias



- No difference in evaluation score when no financial stakes
  Control for value-added, attendance, time use, video scores
- 10% lower raise for women when there are financial stakes of evaluation

Table

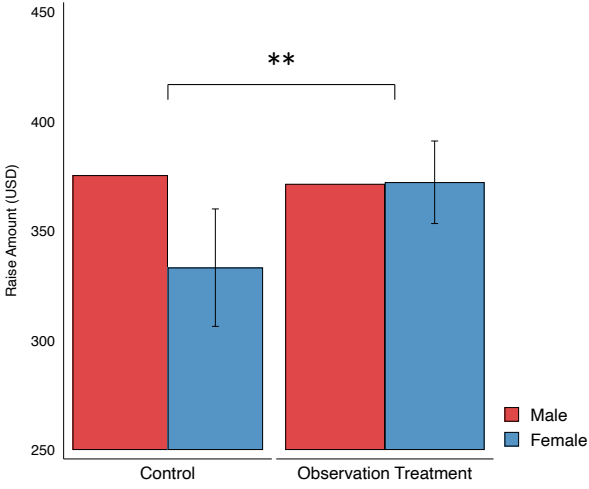
Effect of Observation on Bias



- On average, 12% lower evaluations for women (controlling for productivity)

Table

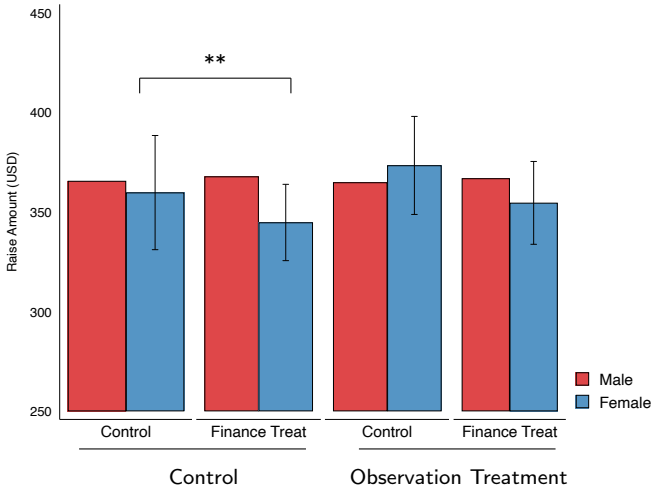
Effect of Observation on Bias



- On average, 12% lower evaluations for women (controlling for productivity)
- Gender gap disappears with better monitoring

Table

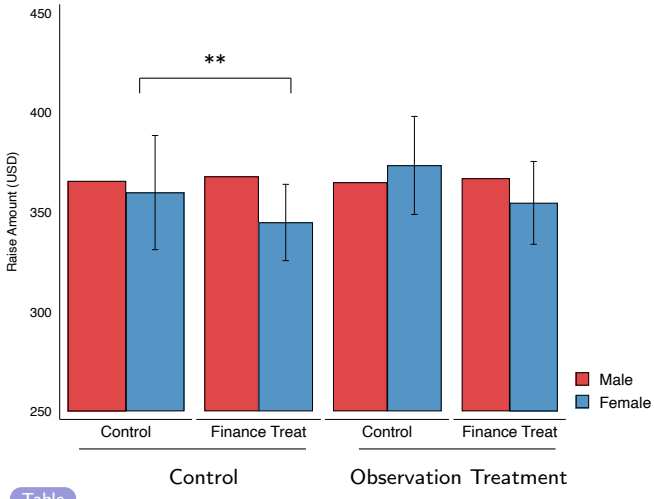
Effect of Both Treatments on Bias



Table

- More monitoring reduces the negative effect of financial stakes on gender bias by 2/3

Effect of Both Treatments on Bias



Table

- More monitoring reduces the negative effect of financial stakes on gender bias by 2/3
- No stat. sig. difference of treatments by manager gender, experience and baseline bias

Table

Mechanisms

- No detectable effect of financial or observation treatment on:
  - "Care" spent on evaluation scores (use of round numbers, variance across sub-criteria)
  - Evaluation criteria selected
  - No heterogeneous teacher effort response by gender (consistent with literature, Bandiera et al, 2021)

Mechanisms

Potential mechanisms from focus groups with teachers and managers

Mechanisms

Potential mechanisms from focus groups with teachers and managers

1). Differential **response to raise** by gender which creates dis-utility for the manager
E.g. More likely to complain, turnover, reduce public good provision, etc.

Mechanisms

Potential mechanisms from focus groups with teachers and managers

1). Differential **response to raise** by gender which creates dis-utility for the manager
E.g. More likely to complain, turnover, reduce public good provision, etc.

2). Differential beliefs about **"deservedness"** of financial rewards by gender
E.g. Differences in total household income, breadwinner norms, etc.

Mechanisms

Potential mechanisms from focus groups with teachers and managers

1). Differential **response to raise** by gender which creates dis-utility for the manager
E.g. More likely to complain, turnover, reduce public good provision, etc.

2). Differential beliefs about **"deservedness"** of financial rewards by gender
E.g. Differences in total household income, breadwinner norms, etc.

Test mechanisms: Vary teacher attributes and evaluation score

Mechanisms

Potential mechanisms from focus groups with teachers and managers

1). Differential **response to raise** by gender which creates dis-utility for the manager
E.g. More likely to complain, turnover, reduce public good provision, etc.

2). Differential beliefs about **"deservedness"** of financial rewards by gender
E.g. Differences in total household income, breadwinner norms, etc.

Test mechanisms: Vary teacher attributes and evaluation score
$\rightarrow$ Vignette survey experiment

Mechanisms: Vignette Survey

- Separate manager sample

Mechanisms: Vignette Survey

- Separate manager sample
- Vary teacher attributes

Name	Ahmad	Zainab	Iqbal
Teach	Class 5 Urdu	Class 1 Urdu	Class 1 Math
Test score growth	average	above average	below average
Classroom environment	organized and supportive of learning	disorganized and noisy	organized and supportive of learning
Days of leave	0 days, much less than average	7 days, about average	10 days, more than average
Classroom observation	You have observed the teacher frequently, so you are confident in your assessment of them.	You have not observed the teacher this year, so you are uncertain about their performance.	You have observed the teacher frequently, so you are confident in your assessment of them.
Plans for next year	Staying at your school	Transferring to another school	Transferring to another school
Years working with teacher	1	8	1
Spouse's job	Does not work	Doctor	Teacher

Mechanisms: Vignette Survey

- Separate manager sample
- Vary teacher attributes
- Vary evaluation features (across subject)

Financial stakes:

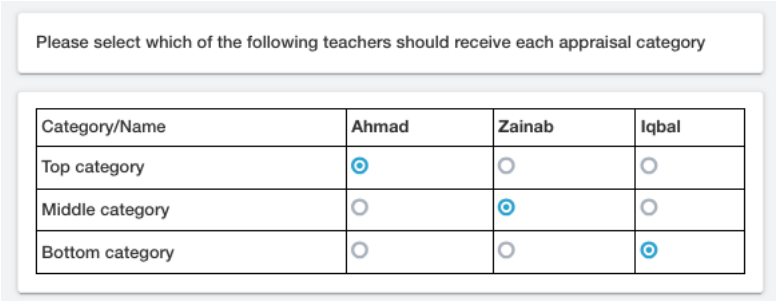
- Affects pay: "The score would affect the teacher's pay for the next year"
- Doesn't affect pay: "The score will **not** affect the teacher's salary or promotion opportunities"

Privacy of Decision

- Private: "The teacher will not learn who gave them this appraisal score so they will not know you made the decision"
- Public: "You would need to tell the teacher what appraisal score you gave them"

Mechanisms: Vignette Survey

- Separate manager sample
- Vary teacher attributes
- Vary evaluation features (across subject)
- Manager ranks teachers



Please select which of the following teachers should receive each appraisal category

Category/Name	Ahmad	Zainab	Iqbal
Top category	◉	○	○
Middle category	○	◉	○
Bottom category	○	○	◉

Mechanisms: Vignette Survey

- Separate manager sample
- Vary teacher attributes
- Vary evaluation features (across subject)
- Manager ranks teachers
- Managers predict response to hypothetical scores

- Provide attributes for a teacher and a hypothetical score
- Rate whether you agree or disagree with response
  - Teacher would:
    - Complain
    - Feel happy
    - Look for another job
    - Be less willing to help with extra tasks
    - Feel disappointed
    - Volunteer for extra duties
    - Suffer financial hardship
  - I (the manager) would:
    - Feel bad for the teacher
    - Feel good about the decision

Mechanisms: Vignette Survey

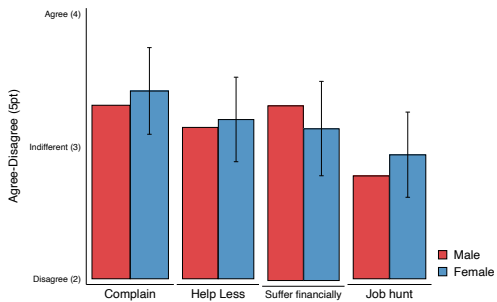
- Separate manager sample
- Vary teacher attributes
- Vary evaluation features (across subject)
- Manager ranks teachers
- Managers predict response to hypothetical scores
- Usual concerns

- Checks for inattention (response time, internal consistency)
- Consistent with World Values Survey bias responses
- Conservative test of mechanisms

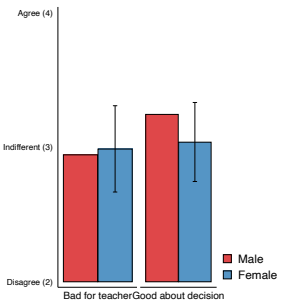
Mechanisms: Response to Raise

- Minimal differences in predicted response

Teacher would:



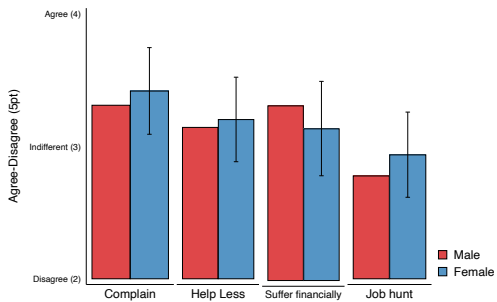
I would feel:



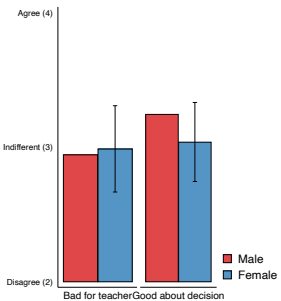
Mechanisms: Response to Raise

- Minimal differences in predicted response
- No differential turnover after low raise (in cross section)

Teacher would:

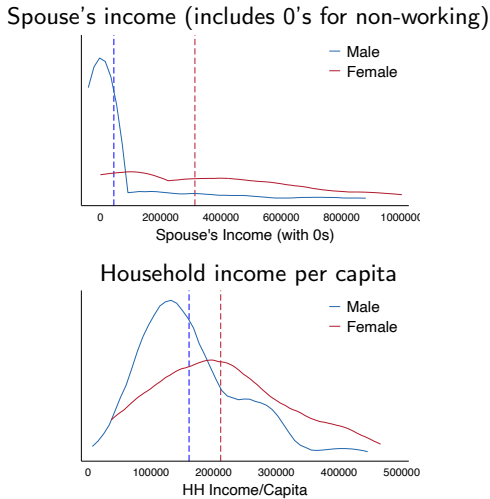


I would feel:



Mechanisms: Deservedness

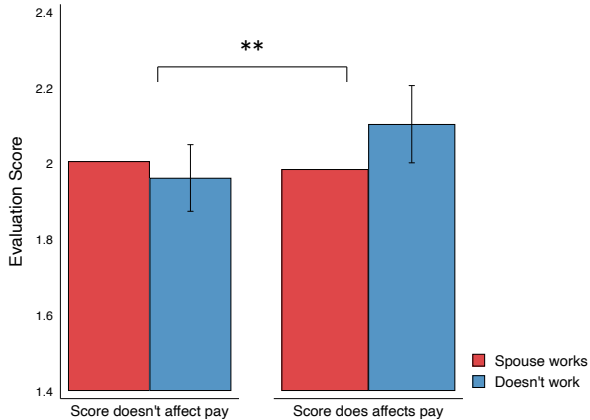
- Large differences in HH income by employee gender



Pakistan Social and Living Standard Measurement Survey

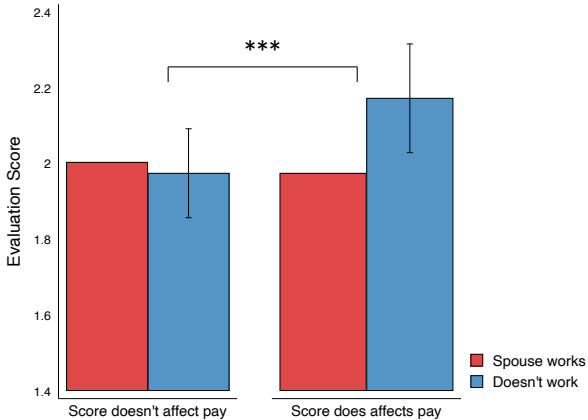
Mechanisms: Deservedness

- Large differences in HH income by employee gender
- Teachers rated higher when their spouse does not work *under financial stakes*



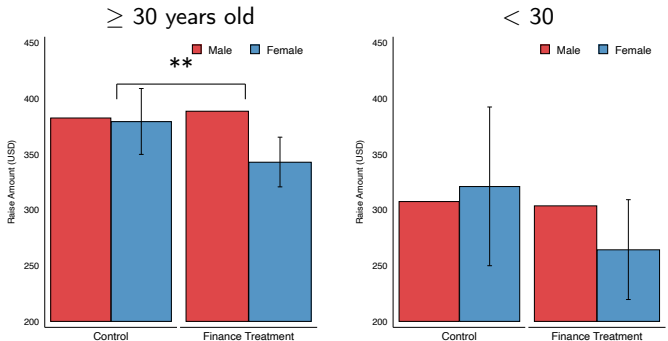
Mechanisms: Deservedness

- Large differences in HH income by employee gender
- Teachers rated higher when their spouse does not work *under financial stakes*
- Effects are larger when decision is public



Mechanisms: Deservedness

- Large differences in HH income by employee gender
- Teachers rated higher when their spouse does not work *under financial stakes*
- Effects are larger when decision is public
- Effects of financial treatment (RCT) are smaller for young teachers



Conclusion

- Women receive lower evaluation scores (controlling for productivity) only when the evaluation affects wages
- Gender bias decreases when managers have better information about employee effort

Conclusion

- Women receive lower evaluation scores (controlling for productivity) only when the evaluation affects wages
- Gender bias decreases when managers have better information about employee effort
- Suggests trade-off between manager's desire for accuracy and dis-utility from low wages

Conclusion

- Women receive lower evaluation scores (controlling for productivity) only when the evaluation affects wages
- Gender bias decreases when managers have better information about employee effort
- Suggests trade-off between manager's desire for accuracy and dis-utility from low wages
- Evidence that differential household income by gender contributes to effects

Conclusion

- Women receive lower evaluation scores (controlling for productivity) only when the evaluation affects wages
- Gender bias decreases when managers have better information about employee effort
- Suggests trade-off between manager's desire for accuracy and dis-utility from low wages
- Evidence that differential household income by gender contributes to effects
- Understand whether household income could be important omitted variable

Thank you!

Contact: christinabrown@uchicago.edu
christinalbrown.com

Research Team-Centre of Economic Research in Pakistan (CERP)


Haya Mubasher


Anam Tariq


Attefaq Ahmed


Mujahid Murtaza

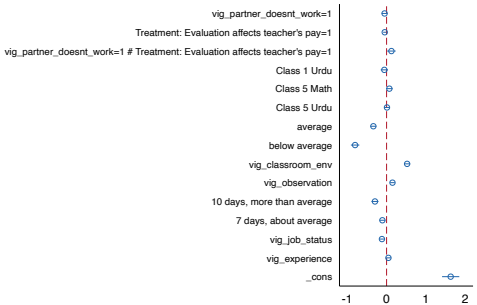

Maheen Rashid


Zahra Niazi


Zohaib Hassan

Appendix

Valuation of Teacher Attributes



Conceptual Framework

1. **Production:** Employee i produces output y_i, the sum of their true ability/effort, θ_i and noise, $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$.

$$y_i = \theta_i + \epsilon_i$$

Conceptual Framework

1. Production: Employee i produces output y_i, the sum of their true ability/effort, θ_i and noise, $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$.

$$y_i = \theta_i + \epsilon_i$$

2. Evaluation: Managers observe y_i but not θ_i. Evaluation score is a function of output and a discretionary component, d_i:

$$s_i = y_i + d_i = \theta + \epsilon_i + d_i$$

Conceptual Framework

1. **Production:** Employee i produces output y_i, the sum of their true ability/effort, θ_i and noise, $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$.

$$y_i = \theta_i + \epsilon_i$$

2. **Evaluation:** Managers observe y_i but not θ_i. Evaluation score is a function of output and a discretionary component, d_i:

$$s_i = y_i + d_i = \theta + \epsilon_i + d_i$$

3. **Post-Evaluation Manager Utility**

i. Dis-utility from inaccurate scores (system legitimacy, psychic cost of lying, punishment)

$$\begin{aligned} E[P_i] &= p(s_i - \theta_i)^2 \\ &= \frac{1}{\sigma_\epsilon^2}(s_i - \theta_i)^2 \end{aligned}$$

p selected to maximize benefit (punishment for inaccurate scores) - cost (unnecessarily punishing for noisy production functions)

Conceptual Framework

1. **Production:** Employee i produces output y_i, the sum of their true ability/effort, θ_i and noise, $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$.

$$y_i = \theta_i + \epsilon_i$$

2. **Evaluation:** Managers observe y_i but not θ_i. Evaluation score is a function of output and a discretionary component, d_i:

$$s_i = y_i + d_i = \theta + \epsilon_i + d_i$$

3. **Post-Evaluation Manager Utility**

i. Dis-utility from inaccurate scores (system legitimacy, psychic cost of lying, punishment)

$$\begin{aligned} E[P_i] &= p(s_i - \theta_i)^2 \\ &= \frac{1}{\sigma_\epsilon^2}(s_i - \theta_i)^2 \end{aligned}$$

p selected to maximize benefit (punishment for inaccurate scores) - cost (unnecessarily punishing for noisy production functions)

ii. Dis-utility of low eval scores (complaints from employees, guilt, turnover)

$$C_i = -c\rho_i s_i$$

- c is the unit-cost
- ρ_i is the dis-utility from a given employee, conditional on the score

Conceptual Framework

Managers select the discretionary component of the salary to minimize the dis-utility from inaccuracy and low scores they expect to face in the next period:

$$
\begin{aligned}
u(d_i) &= \min_{d_i} E[-c\rho_i s_i + p(s_i - \theta_i)^2] \qquad (1)\\
&= \min_{d_i} E[-c\rho_i(\theta_i + \epsilon_i + d_i) + \frac{1}{\sigma_\epsilon^2}(\epsilon_i + d_i)^2]\\
\frac{\partial u_i}{\partial d_i} &= E[-c\rho_i + 2\frac{1}{\sigma_\epsilon^2}(\epsilon_i + d_i)] = 0\\
d_i^* &= \frac{c\rho_i\sigma_\epsilon^2}{2} \qquad (2)
\end{aligned}
$$

Therefore an employee's evaluation score will be $s_i^* = y_i + \frac{c\rho_i\sigma_\epsilon^2}{2}$.

Conceptual Framework

Gender differences

- Same mean and variance: $\theta_i^f, \theta_i^m \sim \mathcal{N}(\mu, \sigma_\theta^2)$
- Same noisiness: $\epsilon_i^f, \epsilon_i^m \sim \mathcal{N}(0, \sigma_\epsilon^2)$
- Difference in dis-utility to manager from low evaluation (guilt, turnover, complaints): $\rho_m > \rho_f$

The difference in expected scores, conditional on ability, by gender then is:

$$\begin{aligned} \frac{\partial s_i^*}{\partial female}|_{\theta_i} &= (y_f - y_m)|_{\theta_i} + \frac{c\sigma_\epsilon^2}{2}(\rho_f - \rho_m) \\ &= \frac{c\sigma_\epsilon^2}{2}(\rho_f - \rho_m) < 0 \end{aligned} \tag{3}$$

Conceptual Framework

Effect of changes in:
- the magnitude of the inconvenience cost (c)
- the accuracy of information managers have (σ_ϵ^2)

i. On Evaluation scores:

$$\frac{\partial s_i^*}{\partial c} = \frac{\rho_i \sigma_\epsilon^2}{2} > 0 \qquad \frac{\partial s_i^*}{\partial \sigma_\epsilon^2} = \frac{c\rho_i}{2} > 0$$

ii. On Gender gap:

$$\textit{Prediction 1:} \frac{\partial^2 s_i^*}{\partial c \partial female}|_{\theta_i} = \frac{\sigma_\epsilon^2}{2}(\rho_f - \rho_m) < 0$$

$$\textit{Prediction 2:} \frac{\partial^2 s_i^*}{\partial \sigma_\epsilon^2 \partial female}|_{\theta_i} = \frac{c}{2}(\rho_f - \rho_m) < 0$$

$$\textit{Prediction 3:} \frac{\partial^3 s_i^*}{\partial c \partial \sigma_\epsilon^2 \partial female}|_{\theta_i} = \frac{1}{2}(\rho_f - \rho_m) < 0$$

Heterogeneous Effects of Financial Treatment

	Predicted Raise Amount (USD)							
	(1) Male	(2) Age	(3) Avg. Bias	(4) Math	(5) Jobs	(6) Family	(7) Teacher Age	(8) Young
Female	-31.77*	-88.23	-44.40	-55.51	-38.07	-18.46	-6.121	-32.78*
	(16.26)	(103.0)	(66.30)	(66.46)	(48.12)	(36.18)	(71.86)	(17.83)
Interaction	-84.65	-3.522*	-17.54	-15.80	-9.782	-4.119	8.982***	-137.7**
	(64.42)	(1.967)	(43.21)	(41.70)	(25.62)	(29.20)	(2.061)	(58.94)
Financial Treatment	22.93	-141.6	197.5	126.6	158.8	116.2	49.21	56.29
	(39.60)	(198.7)	(149.3)	(125.2)	(109.2)	(120.2)	(100.0)	(48.31)
Financial Treatment*Female	-61.62**	19.73	-201.7**	-131.6	-194.3**	-136.1*	-20.12	-97.11**
	(30.06)	(164.7)	(99.26)	(93.66)	(85.51)	(74.71)	(106.0)	(37.39)
Interaction*Financial Treatment	56.39	3.678	-76.17	-39.97	-67.75	-36.53	-0.252	-33.87
	(76.74)	(3.890)	(59.53)	(51.85)	(45.28)	(39.77)	(2.914)	(80.30)
Interaction*Female	47.03	1.355	8.217	14.63	6.094	-3.274	-0.639	24.16
	(49.76)	(2.082)	(27.58)	(29.44)	(20.89)	(13.30)	(1.941)	(51.83)
Interaction*Financial Treatment*Female	-1.904	-1.824	63.31	27.27	68.35*	30.47	-1.269	75.91
	(64.59)	(3.277)	(40.08)	(37.85)	(36.45)	(25.17)	(2.825)	(71.88)
Constant	415.7***	571.2***	444.9***	438.5***	425.6***	417.4***	66.57	430.8***
	(25.15)	(103.2)	(107.1)	(91.62)	(63.23)	(86.96)	(69.55)	(28.35)
Observations	3650	3650	3650	3650	3650	3650	3018	3018
Clusters	208	208	208	208	208	208	188	188
Dep. Var. Mean	368.4	368.4	368.4	368.4	368.4	368.4	368.4	368.4
Dep. Var. SD	176.3	176.3	176.3	176.3	176.3	176.3	176.3	176.3

Back

Heterogeneous Effects of Observation Treatment

	Predicted Raise Amount (USD)							
	(1) Male	(2) Age	(3) Avg. Bias	(4) Math	(5) Jobs	(6) Family	(7) Teacher Age	(8) Young
Female	-60.06***	-225.0	-91.40	-100.3*	-117.7**	-30.64	12.87	-71.06***
	(19.47)	(191.0)	(58.34)	(54.16)	(46.15)	(39.70)	(93.04)	(23.10)
Interaction	-99.09*	-5.532	-36.96	-32.40	-25.69	-9.613	8.936***	-148.9**
	(54.62)	(4.113)	(29.40)	(29.00)	(21.10)	(19.39)	(2.500)	(68.29)
Observation Treatment	-12.47	-212.9	22.77	4.496	-31.04	30.11	42.34	-22.15
	(27.35)	(194.5)	(87.14)	(65.33)	(60.49)	(73.69)	(140.8)	(35.58)
Observation Treatment*Female	18.25	266.0	-21.73	16.57	60.82	-53.11	-78.88	21.19
	(29.29)	(228.2)	(102.1)	(68.18)	(68.37)	(83.01)	(143.4)	(35.24)
Interaction*Observation Treatment	48.14	4.318	-14.03	-7.731	10.41	-13.37	-1.396	55.31
	(103.5)	(4.062)	(44.75)	(29.80)	(29.92)	(33.83)	(3.919)	(96.14)
Interaction*Female	80.48	3.609	16.45	21.53	31.23	-9.175	-1.872	52.71
	(52.47)	(3.952)	(25.28)	(23.73)	(19.48)	(16.29)	(2.496)	(63.01)
Interaction*Observation Treatment*Female	-40.40	-5.288	16.61	0.621	-22.24	25.05	2.360	-46.73
	(112.3)	(4.670)	(47.72)	(30.75)	(32.75)	(34.35)	(3.927)	(95.41)
Constant	418.2***	674.0***	494.1***	480.7***	464.0***	435.7***	77.17	444.4***
	(22.38)	(198.4)	(65.65)	(64.46)	(50.80)	(47.43)	(90.57)	(27.49)
Observations	2614	2614	2614	2614	2614	2614	2269	2269
Clusters	147	147	147	147	147	147	135	135
Dep. Var. Mean	368.4	368.4	368.4	368.4	368.4	368.4	368.4	368.4
Dep. Var. SD	176.3	176.3	176.3	176.3	176.3	176.3	176.3	176.3

Back

Heterogeneous Raise

	Predicted Raise Amount (USD)							
	(1) Male	(2) Age	(3) Avg. Bias	(4) Math	(5) Jobs	(6) Family	(7) Teacher Age	(8) Young
Female	16.41	-29.07	-2.460	-2.877	-13.83	26.73	-86.27	18.62
	(12.73)	(97.02)	(38.67)	(30.58)	(28.62)	(28.11)	(53.38)	(15.21)
Interaction	3.722	-0.198	-21.65	-18.71*	-13.96	-5.125	2.987*	-51.79
	(41.23)	(1.870)	(13.57)	(11.24)	(11.13)	(8.248)	(1.580)	(35.30)
Interaction*Female	1.292	0.958	7.797	7.767	15.55	-4.596	2.821**	-2.957
	(39.72)	(2.003)	(15.40)	(12.21)	(12.13)	(9.462)	(1.395)	(32.51)
Observations	2018	2018	2018	2018	2018	2018	1728	1728
Clusters	170	170	170	170	170	170	156	156
Dep. Var. Mean	368.4	368.4	368.4	368.4	368.4	368.4	368.4	368.4
Dep. Var. SD	176.3	176.3	176.3	176.3	176.3	176.3	176.3	176.3

Effect of Treatments on Evaluation

	Predicted Raise Amount (USD)			
	(1)	(2)	(3)	(4)
Female	-28.21***	-2.228	-41.90***	-18.43
	(6.600)	(12.59)	(13.15)	(18.71)
Financial Treatment		23.36		36.88
		(20.81)		(28.91)
Financial Treatment*Female		-34.33**		-51.01**
		(15.32)		(25.65)
Observation Treatment			-36.46**	-15.51
			(15.93)	(25.09)
Observation Treatment*Female			41.55**	31.46
			(17.07)	(27.42)
Financial Treatment*Observation Treatment				-46.46
				(31.71)
Financial Treatment*Observation Treatment*Female				37.52
				(34.62)
Observations	5051	4300	2626	2326
Clusters	.	263	.	158
Dep. Var. Mean	365.4	365.4	365.4	365.4
Dep. Var. SD	164.7	164.7	164.7	164.7

Back

Percentile Value Added

- Construction of the value added percentile:
  - Within each grade/year/subject bin, calculate each student's percentile rank.
  - For the following year's score, construct the student's percentile within the lagged percentile-grade-subject bin.
  - Compute the teacher's percentile in a given year by taking the average across all students
- Reasons for using percentile measure
  - Barlevy and Neal (2016) show results are similar to other value added models
  - Only relies on ordinal information allowing for new tests each year (less susceptible to manipulation)
  - Muralidharan/Walters and Lucas/Neal use same approach in India and Uganda, respectively

Back

Percentile Value Added

- Validating the Percentile Value Added
  - Year to year correlation
    - Standard models: 0.4
    - Our measure: 0.56
  - Increase in first 5 years of teaching
    - Standard models: 0.5
    - Our measure: 0.35
- Correlation with Other VA Models
  - Controlling for lagged score in the same subject: 0.44
  - CFR 2013: 0.25

Back

Balance in Baseline Covariates

Variable	(1) Control N/ [Clusters]	(1) Control Mean/ SE	(2) Objective Treatment N/ [Clusters]	(2) Objective Treatment Mean/ SE	(3) Subjective Treatment N/ [Clusters]	(3) Subjective Treatment Mean/ SE	T-test Difference (1)-(2)	T-test Difference (1)-(3)	T-test Difference (2)-(3)
Panel A: Teacher Characteristics									
Performance evaluation score	656 [40]	3.360 (0.030)	384 [32]	3.362 (0.039)	3566 [139]	3.338 (0.010)	-0.002	0.022	0.024
Salary (USD)	920 [40]	5417.984 (313.504)	535 [32]	5125.462 (295.013)	4928 [145]	5329.416 (124.042)	292.523	88.569	-203.954
Age	921 [40]	36.591 (0.738)	539 [32]	36.083 (0.846)	4926 [145]	36.630 (0.298)	0.507	-0.039	-0.546
Years of experience	918 [40]	5.505 (0.277)	534 [32]	5.487 (0.425)	4897 [145]	5.725 (0.156)	0.019	-0.220	-0.238
Panel B: Student Test Scores									
Math Test Z-Score	9959 [40]	0.071 (0.070)	5292 [33]	-0.146 (0.065)	51775 [137]	-0.014 (0.026)	0.217**	0.085	-0.132*
Urdu Test Z-Score	9702 [40]	0.041 (0.072)	5259 [33]	-0.048 (0.063)	50915 [138]	-0.002 (0.028)	0.089	0.043	-0.046
English Test Z-Score	9755 [40]	0.017 (0.056)	5289 [33]	-0.049 (0.050)	51356 [137]	0.002 (0.032)	0.067	0.016	-0.051
Social Studies Test Z-Score	9171 [40]	0.041 (0.046)	5030 [33]	-0.064 (0.056)	49411 [137]	0.007 (0.022)	0.105	0.033	-0.071
Science Test Z-Score	9636 [40]	-0.010 (0.041)	5065 [33]	-0.064 (0.042)	50268 [137]	0.001 (0.024)	0.055	-0.011	-0.066

Back

Endline Student Survey

Question	Category	Source
1. I enjoy my math/science/English/Urdu class	Love of learning	National Student Survey
2. When work is difficult, I either give up or study only the easy part (reversed)	Love of learning	Learning and Study Strategies Inventory
3. I get very easily distracted when I am studying or in class (reversed)	Love of learning	Learning and Study Strategies Inventory
4. I can spend hours on a single problem because I just can't rest without knowing the answer	Love of learning	Big Five (childrens)
5. I feel sorry for other kids who don't have toys and clothes	Ethical	Eisenberg's Child-Report Sympathy Scale
6. Seeing a child who is crying makes me feel like crying	Ethical	Bryant's Index of Empathy Measurement
7. It is ok if a student lies to get out a test they are worried about failing (reversed)	Ethical	

Back

Endline Student Survey

Question	Category	Source
8. The pressure to do well is very high, so it is ok to cheat sometimes (reversed)	Ethical	
9. I am interested in public affairs	Global	Afrobarometer/World Values Survey
10. This world is run by a few people in power, and there is not much that someone like me can do about it (reversed)	Global	Afrobarometer
11. People who are poor should work harder and not be given charity (reversed)	Global	Afrobarometer
12. It is important to protect the environment even if this means we cannot consume as much today	Global	Afrobarometer
13. People from other places can't really be trusted (reversed)	Global	Afrobarometer
14. I am comfortable asking my math/science/Urdu/English teacher for help or support	Inquisitive	Learning and Study Strategies Inventory
15. I enjoy learning about subjects that are unfamiliar to me.	Inquisitive	Litman and Spielberger, Epistemic Curiosity questionnaire
16. I would like to change to a different school	Dislike school	Learning and Study Strategies Inventory

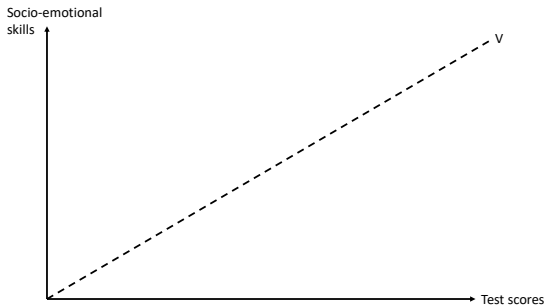
Back

What we know

1. What we know about the ability for contracts to screen types
- Lazear (other general ad sel lit)
2. Make clear tension between lit that suggests effects should be large vs. lit that predicts effects are zero and why this setting is different than Lazear 2000
- Mention barbara, jesse and owen
3. Performance Pay literature: lots of great stuff but missing sorting

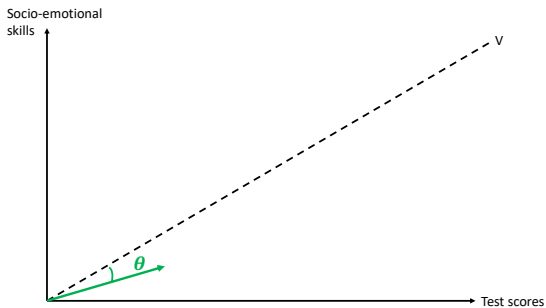
Distortion and Noise

For example, a school's value function, V may be that they value test scores and socio-emotional outcomes at a 2:1 ratio



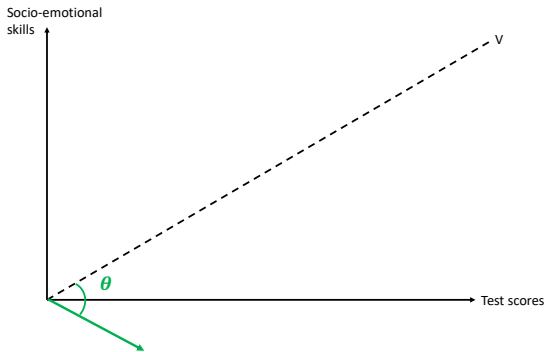
Distortion and Noise

Distortion is captures how aligned the incentive scheme is with the actions which produce V



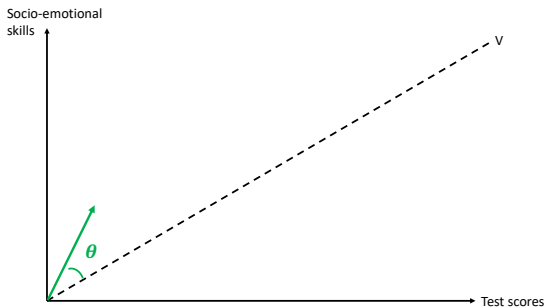
Distortion and Noise

Distortion is captures how aligned the incentive scheme is with the actions which produce V



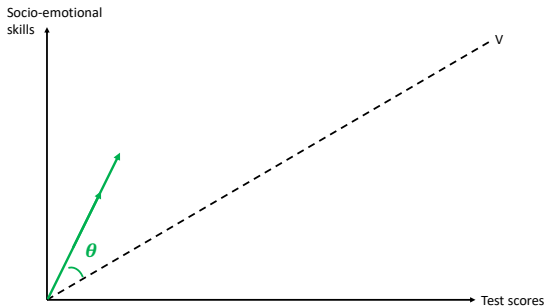
Distortion and Noise

Distortion is captures how aligned the incentive scheme is with the actions which produce V



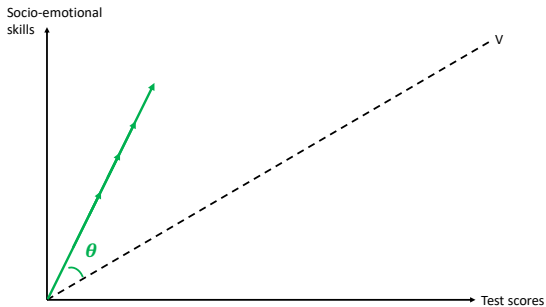
Distortion and Noise

Noise determines how high-powered the incentives are and hence, how large the effort response is



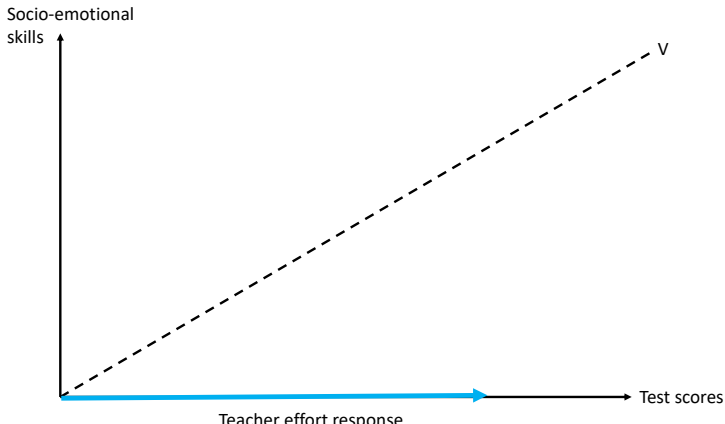
Distortion and Noise

Noise determines how high-powered the incentives are and hence, how large the effort response is



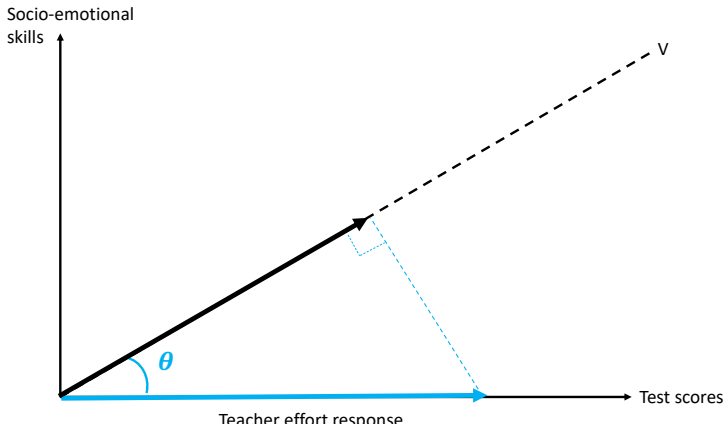
Distortion and Noise

For example, here is an incentive scheme which pays based on endline test scores

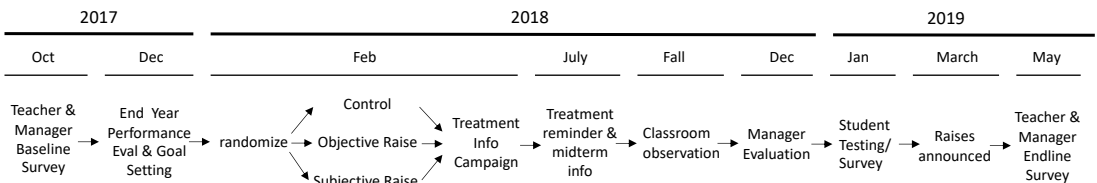


Distortion and Noise

For example, here is an incentive scheme which pays based on endline test scores



Experimental Design



Teacher Evaluation Vignettes

Example vignette:

"Haya is in the bottom 10% of teachers in terms of students' test score growth, in the middle 10% of teachers in terms of behavioral management, and is in the top 10% in terms of attendance and timeliness at work. If you had to given her a performance evaluation score, what score would you assign to her?"

Back

Student Outcomes - Test Scores

	Endline Test (z-score)				
	All	Remedial	External	Math/Science	English/Urdu
	(1)	(2)	(3)	(4)	(5)
Objective Treatment	0.0918*	0.189***	0.119**	0.104*	0.0917
	(0.0575)	(0.00518)	(0.0335)	(0.0668)	(0.166)
	[0.0730]	[0.0260]	[0.0200]	[0.194]	[0.144]
Subjective Treatment	0.0859**	0.142**	0.0855*	0.0884*	0.0986**
	(0.0220)	(0.0113)	(0.0601)	(0.0646)	(0.0267)
	[0.0130]	[0.0240]	[0.0170]	[0.121]	[0.0260]
F-test pval (subj=obj)	0.89	0.38	0.43	0.77	0.90
Randomiz infer pval (subj=obj)	0.884	0.453	0.388	0.819	0.873
Control Group Mean	-0.04	-0.09	-0.05	-0.04	-0.04
Clusters	234	204	225	223	225
Observations	141566	31944	100318	72714	68852

Back

Student Outcomes - Socio-emotional

	Endline Test (z-score)				
	All (1)	Remedial (2)	External (3)	Math/Science (4)	English/Urdu (5)
Objective Treatment	0.0918*	0.189***	0.119**	0.104*	0.0917
	(0.0575)	(0.00518)	(0.0335)	(0.0668)	(0.166)
	[0.0730]	[0.0260]	[0.0200]	[0.194]	[0.144]
Subjective Treatment	0.0859**	0.142**	0.0855*	0.0884*	0.0986**
	(0.0220)	(0.0113)	(0.0601)	(0.0646)	(0.0267)
	[0.0130]	[0.0240]	[0.0170]	[0.121]	[0.0260]
F-test pval (subj=obj)	0.89	0.38	0.43	0.77	0.90
Randomiz infer pval (subj=obj)	0.884	0.453	0.388	0.819	0.873
Control Group Mean	-0.04	-0.09	-0.05	-0.04	-0.04
Clusters	234	204	225	223	225
Observations	141566	31944	100318	72714	68852

Back

Teacher Effort - Teaching Practices

	Classroom Observation Rubric				Test Prep
	All	Class Climate	Differentiation	Student-Centered	Minutes
	(1)	(2)	(3)	(4)	(5)
Objective Treatment	-0.0713	-0.0791*	0.110*	-0.115**	0.577***
	(0.123)	(0.0788)	(0.0719)	(0.0346)	(0.00455)
	[0.171]	[0.101]	[0.149]	[0.0480]	[0.0120]
Subjective Treatment	-0.00206	-0.00704	0.105*	-0.0276	0.110
	(0.959)	(0.822)	(0.0699)	(0.521)	(0.255)
	[0.946]	[0.838]	[0.0690]	[0.559]	[0.649]
F-test pval (subj=obj)	0.10	0.10	0.93	0.09	0.02
Randomiz infer pval (subj=obj)	0.109	0.0830	0.940	0.0940	0.0140
Control Group Mean	4.67	5.64	2.65	4.93	0.14
Clusters	142	142	142	142	142
Observations	6827	6827	6827	6827	6827

Back

Design - Teacher Sample

	Study Sample		US Sample	
	Mean (1)	St. Dev. (2)	Mean (3)	St. Dev. (4)
Panel A. Teacher Characteristics				
Age	35.0	8.9	41.8	7.5
Female	0.80	0.40	0.76	0.43
Years of experience	5.1	5.2	13.8	9.6
Has Post BA Education	0.68	0.47	0.54	0.50
Salary, USD(PPP)	17,160	5,700	52,400	18,400
Panel B. Teacher Evaluation				
Number of observations per year	4.7	8.2	2.5	2.9
Use evaluation for compensation	-	-	0.12	0.32
Frequency of evaluation (months)	-	-	13.0	7.0
Performance metric used for evaluation:				
- Principal evaluation	-	-	0.90	0.30
- Test scores	-	-	0.35	0.48
- Peer evaluations	-	-	0.26	0.44
- Student ratings	-	-	0.05	0.22

Back

Design - Manager Sample

	Study Sample		US Sample	
	Mean (1)	St. Dev. (2)	Mean (3)	St. Dev. (4)
Panel A. Manager Characteristics				
Age	44.9	9.2	48.8	9.7
Female	0.61	0.49	0.53	0.50
Years of experience	9.6	7.9	13.0	7.5
Panel B. Manager Time Use				
Total hours worked	47.2	16.3	57.0	13.2
Hours spent on:				
- Administrative tasks	18.5	10.3	18.2	2.3
- Teacher management and teaching	17.5	8.2	15.1	2.0
- Student and parent interactions	6.3	4.4	20.2	2.7
- Other tasks	6.9	12.3	4.0	2.6
Panel C. Management Practice Rating				
Overall Management Score (out of 5)	4.27	0.43	2.76	0.43
People management (out of 5)	4.14	0.53	2.51	0.49
Operations (out of 5)	4.32	0.61	2.89	0.49
Performance monitoring (out of 5)	4.32	0.49	2.81	0.75

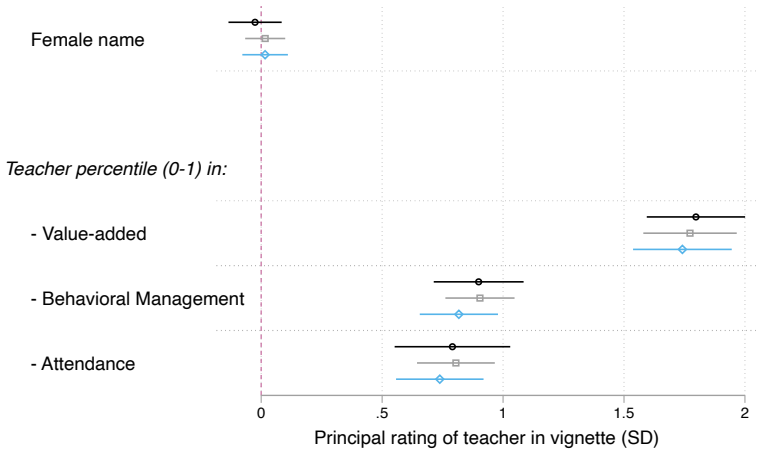
Back

What do principals value?

We give principal short vignettes describing an example teacher and ask them to give a hypothetical evaluation score. Vignette text

What do principals value?

We give principal short vignettes describing an example teacher and ask them to give a hypothetical evaluation score. Vignette text



Heterogeneous Effects by Manager Characteristics

Subjective incentives appear to be effective for all but the bottom quintile of managers

	Endline Test Scores			
	(1)	(2)	(3)	(4)
Subjective Treatment	-0.0156	0.169**	-0.0566	0.249***
	(0.197)	(0.0688)	(0.117)	(0.0775)
Interaction	0.00111	0.00827	0.0159	0.142*
	(0.00274)	(0.00503)	(0.0977)	(0.0763)
Interaction*Subjective Treatment	0.00205	-0.00883	0.148	-0.211**
	(0.00420)	(0.00648)	(0.127)	(0.0910)
Interaction	Age	Experience (years)	Female	Manager innacuracy (z-score)
Clusters	255	255	255	255
Observations	440595	440595	440595	440595

Principal-Agent Model with Moral Hazard

Using Baker (2002), V, outcome (student learning), is a function of their teacher's effort, $\vec{a}$, the return to those actions $\vec{f}$:

$$V(\vec{a}, e) = \vec{f} \cdot \vec{a} + e = f_1 a_1 + f_2 a_2 + ... + e \tag{4}$$

Principal-Agent Model with Moral Hazard

Using Baker (2002), V, outcome (student learning), is a function of their teacher's effort, $\vec{a}$, the return to those actions $\vec{f}$:

$$V(\vec{a}, e) = \vec{f} \cdot \vec{a} + e = f_1 a_1 + f_2 a_2 + ... + e \tag{4}$$

Teacher's pay under an incentive contract is a function of those actions, the piece rate for each action $\vec{g}$, and noise, ϕ:

$$P(\vec{a}, \phi) = \vec{g} \cdot \vec{a} + \phi = g_1 a_1 + g_2 a_2 + ... + \phi \tag{5}$$

Principal-Agent Model with Moral Hazard

Using Baker (2002), V, outcome (student learning), is a function of their teacher's effort, $\vec{a}$, the return to those actions $\vec{f}$:

$$V(\vec{a}, e) = \vec{f} \cdot \vec{a} + e = f_1 a_1 + f_2 a_2 + ... + e \tag{4}$$

Teacher's pay under an incentive contract is a function of those actions, the piece rate for each action $\vec{g}$, and noise, ϕ:

$$P(\vec{a}, \phi) = \vec{g} \cdot \vec{a} + \phi = g_1 a_1 + g_2 a_2 + ... + \phi \tag{5}$$

Assuming a quadratic cost of effort, then optimal effort will be $\vec{a^*} = \vec{g}$, and average student learning will be:

$$E[V^*(\vec{a^*}, e)] = \vec{f} \cdot \vec{g} = |f||g|cos\theta \tag{6}$$

Principal-Agent Model with Moral Hazard

Taking the variance of (2), we have $var(P) = |g|^2 var(\vec{a}) + \sigma_\phi^2$. Re-arranging, we can substitute this in for $|g|$ in to (3).

Principal-Agent Model with Moral Hazard

Taking the variance of (2), we have $var(P) = |g|^2 var(\vec{a}) + \sigma_\phi^2$. Re-arranging, we can substitute this in for $|g|$ in to (3).

Average student learning under a given incentive scheme is:

$$\begin{aligned} E[V^*(\vec{a}^*, e)] &= |f||g|cos\theta \\ &= |f| \frac{\sqrt{var(P) - \sigma_\phi^2}}{\sqrt{var(\vec{a})}} cos\theta \end{aligned}$$

Principal-Agent Model with Moral Hazard

Taking the variance of (2), we have $var(P) = |g|^2 var(\vec{a}) + \sigma_\phi^2$. Re-arranging, we can substitute this in for $|g|$ in to (3).

Average student learning under a given incentive scheme is:

$$\begin{aligned} E[V^*(\vec{a}^*, e)] &= |f||g|cos\theta \\ &= |f| \frac{\sqrt{var(P) - \sigma_\phi^2}}{\sqrt{var(\vec{a})}} cos\theta \end{aligned}$$

$|f|$, $var(P)$ and $var(\vec{a})$ are constant across the two incentive schemes we'll compare (feature of any within-firm tournament)

Principal-Agent Model with Moral Hazard

Taking the variance of (2), we have $var(P) = |g|^2 var(\vec{a}) + \sigma_\phi^2$. Re-arranging, we can substitute this in for $|g|$ in to (3).

Average student learning under a given incentive scheme is:

$$\begin{aligned} E[V^*(\vec{a}^*, e)] &= |f||g|cos\theta \\ &= |f| \frac{\sqrt{var(P) - \sigma_\phi^2}}{\sqrt{var(\vec{a})}} cos\theta \end{aligned}$$

$|f|$, $var(P)$ and $var(\vec{a})$ are constant across the two incentive schemes we'll compare (feature of any within-firm tournament)

Principal-Agent Model with Moral Hazard

Taking the variance of (2), we have $var(P) = |g|^2 var(\vec{a}) + \sigma_\phi^2$. Re-arranging, we can substitute this in for $|g|$ in to (3).

Average student learning under a given incentive scheme is:

$$E[V^*(\vec{a}^*, e)] = |f||g|cos\theta$$
$$= |f| \frac{\sqrt{var(P) - \sigma_\phi^2}}{\sqrt{var(\vec{a})}} cos\theta$$

$|f|$, $var(P)$ and $var(\vec{a})$ are constant across the two incentive schemes we'll compare (feature of any within-firm tournament)

Key predictions – Student learning is:

(b) increasing in alignment (lack of distortion), $cos(\theta)$

(a) decreasing in contract noise, σ_ϕ^2

Example evaluation criteria

Plan 1: Manager Appraisal of Effort

Effort Criteria	Objective Score
Assessment of student understanding (monitoring of student learning, effective and timely copy checking)	20
Differentiated lessons for varying learning needs	30
Effectively delivering accurate and relevant content (effective implementation of the curriculum)	30
Providing caring, supportive environment	20
Total	100

Back

Contribution

- **Financial Discrimination**
  Demonstrate gender discrimination
  cites

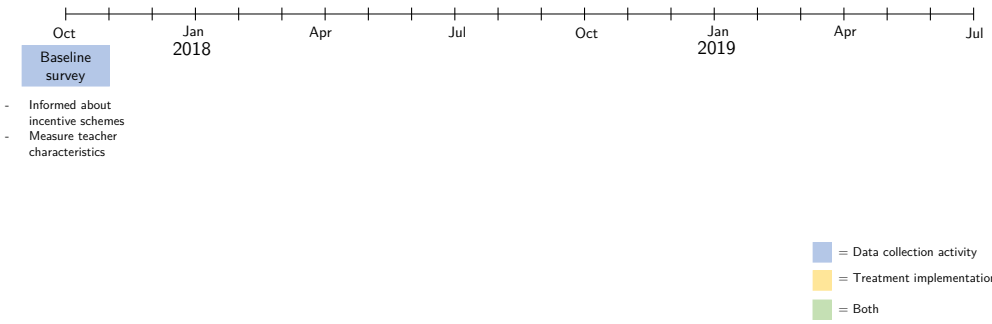
Contribution

- **Financial Discrimination**
  Demonstrate gender discrimination
  cites
- **Statistical discrimination**
  Text
  Cites

Contribution

- **Financial Discrimination**
  Demonstrate gender discrimination
  cites
- **Statistical discrimination**
  Text
  Cites
- **Disparate outcomes**
  Model with
  Baker, 2002; Prendergast, 1999; Prendergast and Topel, 1993; Prendergast, 2007

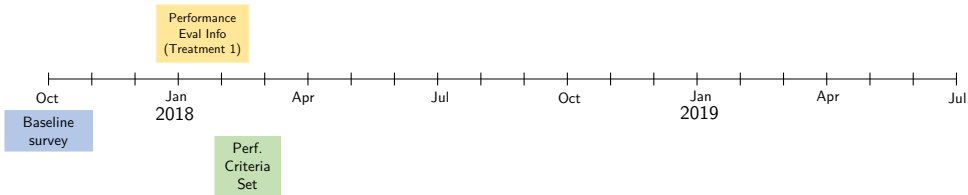
Design - Timeline



Example criteria

Design - Timeline

- Research team meets in person with managers
- School system HR does in person presentation at each school
- Email information
- Displayed on teacher's dashboard

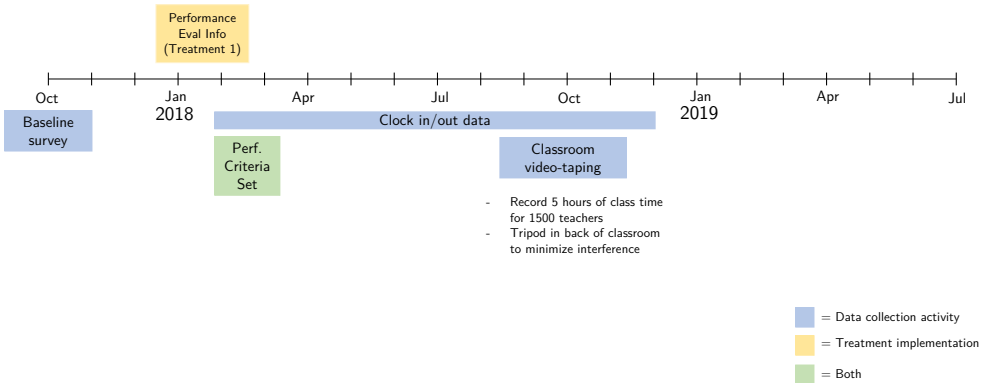


- Managers set criteria for ALL teachers (not new to experiment)

= Data collection activity
= Treatment implementation
= Both

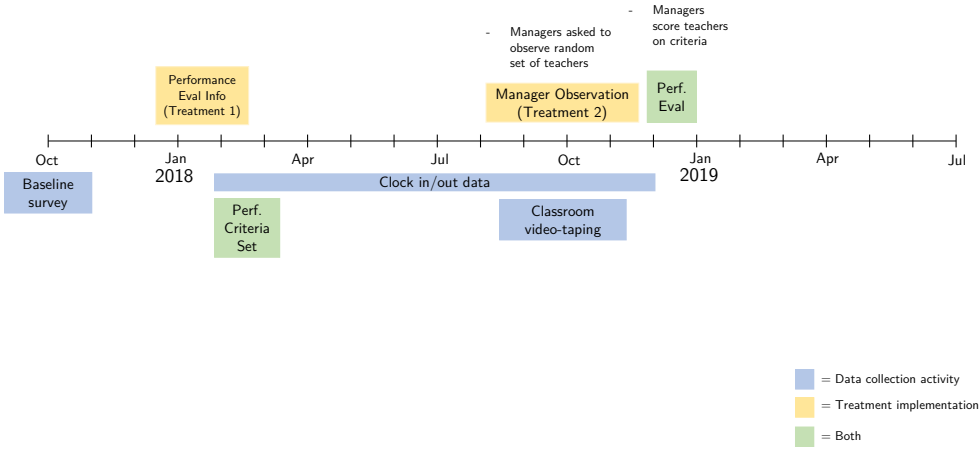
Example criteria

Design - Timeline



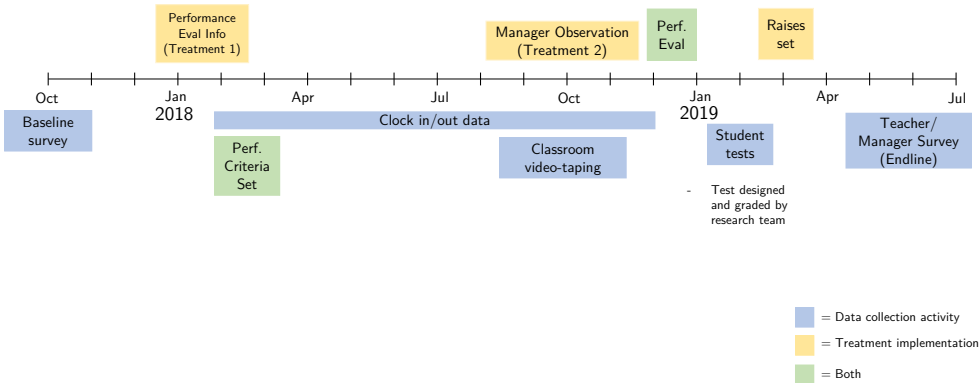
Example criteria

Design - Timeline



Example criteria

Design - Timeline



Example criteria

Teacher and Manager Sample

	Study Sample		US Sample	
	Mean	St. Dev.	Mean	St. Dev.
Panel A. Teacher Characteristics				
Age	35.0	8.9	41.8	7.5
Female	0.80	0.40	0.76	0.43
Years of experience	5.1	5.2	13.8	9.6
Has Post BA Education	0.68	0.47	0.54	0.50
Salary, USD	4,000	1,700	52,400	18,400
Panel B. Manager Characteristics				
Age	44.9	9.2	48.8	9.7
Female	0.61	0.49	0.53	0.50
Years of experience	9.6	7.9	13.0	7.5
Panel C. Manager Time Use				
Total hours worked	47.2	16.3	57.0	13.2
Hours spent on:				
- Administrative tasks	18.5	10.3	18.2	2.3
- Teacher management and teaching	17.5	8.2	15.1	2.0
- Student and parent interactions	6.3	4.4	20.2	2.7
- Other tasks	6.9	12.3	4.0	2.6

Col. 3 and 4 Source: School and Staff Survey (National Center for Education Statistics)