Subjective versus Objective Performance Pay and Teacher Productivity

Christina Brown (UC Berkeley) and
Tahir Andrabi (Lahore University of Management Sciences)

April 26, 2020
CIES

Support from DFID, JPAL, the Weiss Fund, CEGA, the Strandberg Fund, NAEd/Spencer Foundation, and IRLE
AEA Registry-0003835

Incentivizing Teachers

- How should schools incentivize teachers (when effort is non-verifiable/non-contractable)?

Incentivizing Teachers

- How should schools incentivize teachers (when effort is non-verifiable/non-contractable)?
- Second best: Incentivize on outcome(s)

Incentivizing Teachers

- How should schools incentivize teachers (when effort is non-verifiable/non-contractable)?
- Second best: Incentivize on outcome(s)
- Introduces two new problems

Incentivizing Teachers

- How should schools incentivize teachers (when effort is non-verifiable/non-contractable)?
- Second best: Incentivize on outcome(s)
- Introduces two new problems
  - **Distortion**: Unable to incentivize all outcomes

    Multi-tasking problem

Incentivizing Teachers

- How should schools incentivize teachers (when effort is non-verifiable/non-contractable)?
- Second best: Incentivize on outcome(s)
- Introduces two new problems
  - **Distortion**: Unable to incentivize all outcomes

    Multi-tasking problem
  - **Noise**: Imperfect mapping from teacher effort to outcome

    Introduces risk for teacher; reduces power of incentive (for large class of incentives)

Incentivizing Teachers

- How should schools incentivize teachers (when effort is non-verifiable/non-contractable)?
- Second best: Incentivize on outcome(s)
- Introduces two new problems
  - **Distortion**: Unable to incentivize all outcomes
    Multi-tasking problem
  - **Noise**: Imperfect mapping from teacher effort to outcome
    Introduces risk for teacher; reduces power of incentive (for large class of incentives)
- What do organizations actually do? $\rightarrow$ Use manager's knowledge to address distortion and noise issue

Subjective Performance Incentives

- 85% of employees in the US have their raise, promotion and/or termination based on managerial discretion

  Engellandt and Riphahn, 2011

Subjective Performance Incentives

- 85% of employees in the US have their raise, promotion and/or termination based on managerial discretion

  Engellandt and Riphahn, 2011

- Why could subjective incentives correct noise and distortion? Lazear and Oyer, 2009

Subjective Performance Incentives

- 85% of employees in the US have their raise, promotion and/or termination based on managerial discretion

  Engellandt and Riphahn, 2011

- Why could subjective incentives correct noise and distortion? Lazear and Oyer, 2009
  - Managers can account for negative shocks, prioritize multiple outcomes

Subjective Performance Incentives

- 85% of employees in the US have their raise, promotion and/or termination based on managerial discretion

  Engellandt and Riphahn, 2011

- Why could subjective incentives correct noise and distortion? Lazear and Oyer, 2009
  - Managers can account for negative shocks, prioritize multiple outcomes
  - But it could introduce new problems: bias, incorrect priorities, etc

Subjective Performance Incentives

- 85% of employees in the US have their raise, promotion and/or termination based on managerial discretion

  Engellandt and Riphahn, 2011

- Why could subjective incentives correct noise and distortion? Lazear and Oyer, 2009
  - Managers can account for negative shocks, prioritize multiple outcomes
  - But it could introduce new problems: bias, incorrect priorities, etc

- Limited evidence: Correlational studies and RCTs of bundled objective/subjective incentive schemes

  Oyer and Schaefer, 2011; Khan et al, 2016; Fryer, 2013; Engellandt and Riphahn, 2011; Kahn and Sherer, 1990

Overview

(1) What we want to learn:

   a. What is the effect of subjective versus objective incentives on student outcomes?
   b. Can subjective incentives reduce noise and distortion?
   c. When does it fail?

Overview

(1) What we want to learn:

a. What is the effect of subjective versus objective incentives on student outcomes?
b. Can subjective incentives reduce noise and distortion?
c. When does it fail?

(2) What we do:

a. Randomize teachers to:
   - Subjective incentives (manager discretionary)
   - Objective incentives (value-added)
   - Flat pay
b. Measure every aspect along the causal chain
   - Beliefs about incentive scheme noise and distortion
   - Effort across different types of actions
   - Outcomes: Student test scores and socio-emotional outcomes

Overview

(3) What we find:

a. Subjective performance incentives are equally effective at increasing test scores as objective incentives, without any negative effects on student socio-emotional outcomes
b. Mechanisms - Subjective is:
   - Less noisy: Produces a larger overall effort response
   - (Less) distorted: Prioritizes both testing and non-testing student outcomes
c. Subjective incentives dominate objective for all but the bottom quintile of managers

Talk Structure

- Experimental Setting and Design
- Reduced Form Results: Effect of Incentives on Student Outcomes and Teacher Effort
- Mechanism Results: Noise and Distortion

Design - Context

- Grades 4-13 in English, Urdu, math and science
- Large private school network operating hundreds of schools across urban Pakistan
- 51% of secondary students in South Asia attend private school
- Annual tuition is $900

Design - Treatments

Randomize contracts at school level:

- Control (46 schools): Flat raise: All teachers receive a raise of 5%
- Treatment (212 schools): Performance Raise: Teachers receive a raise from 0-10% based on within school ranking:
  - Objective (34 schools): Percentile Valued-Added (Barlevy and Neal, 2012)
  - Subjective (178 schools): Principal Rating of Teacher

Implemented from Oct 2017-May 2019 Timeline
0 of 10 variables are stat. sig. at baseline Balance Table

Design - Data

Subject	N	Source	Items
Principal	300	Administrative	Demographics, employment history, time use
	300	Endline	World Management Survey
Teacher	6,000	Administrative	Demographics, employment history, student link, performance evaluation
	1,500	Classroom video	CLASS rubric (Araujo et al, 2016), test preparation
	5,000	Endline	Time use, belief about contract noisiness/return to actions
Student	60,500	Administrative	Demographics, academic history
	46,600	Endline	Standardized exam in English, Urdu, Math and Science and Socio-emotional skills survey

Design - Data

Subject	N	Source	Items
Principal	300	Administrative	Demographics, employment history, time use
	300	Endline	World Management Survey
Teacher	6,000	Administrative	Demographics, employment history, student link, performance evaluation
	1,500	Classroom video	CLASS rubric (Araujo et al, 2016), test preparation
	5,000	Endline	Time use, belief about contract noisiness/return to actions
Student	60,500	Administrative	Demographics, academic history
	46,600	Endline	Standardized exam in English, Urdu, Math and Science and Socio-emotional skills survey

Design - Data

Subject	N	Source	Items
Principal	300	Administrative	Demographics, employment history, time use
	300	Endline	World Management Survey
Teacher	6,000	Administrative	Demographics, employment history, student link, performance evaluation
	1,500	Classroom video	CLASS rubric (Araujo et al, 2016), test preparation
	5,000	Endline	Time use, belief about contract noisiness/return to actions
Student	60,500	Administrative	Demographics, academic history
	46,600	Endline	Standardized exam in English, Urdu, Math and Science and Socio-emotional skills survey

Student Outcomes - Test Scores

Conduct endline test with students in grades 4-12 in four subjects

	Endline Test (z-score)				
	All (1)	Remedial (2)	External (3)	Math/Science (4)	English/Urdu (5)
Objective Treatment	0.0918* (0.0575) [0.0730]	0.189*** (0.00518) [0.0260]	0.119** (0.0335) [0.0200]	0.104* (0.0668) [0.194]	0.0917 (0.166) [0.144]
Subjective Treatment	0.0859** (0.0220) [0.0130]	0.142** (0.0113) [0.0240]	0.0855* (0.0601) [0.0170]	0.0884* (0.0646) [0.121]	0.0986** (0.0267) [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

Clustered standard errors * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Subjective and objective incentives increase student test scores by 0.09sd

Student Outcomes - Socio-emotional

Conduct endline student survey to measure socio-emotional skills Survey Items

	Socio-Emotional Indices (z-score)					
	All (1)	Love of learning (2)	Ethical (3)	Global (4)	Inquisitive (5)	Dislike school (6)
Objective Treatment	-0.0262	-0.0854	-0.0137	0.0278	0.00293	0.0860*
	(0.423)	(0.133)	(0.760)	(0.582)	(0.955)	(0.0719)
	[0.515]	[0.123]	[0.830]	[0.635]	[0.957]	[0.135]
Subjective Treatment	0.0171	0.000933	0.0115	0.0474	-0.0217	-0.0314
	(0.363)	(0.976)	(0.668)	(0.192)	(0.552)	(0.395)
	[0.576]	[0.985]	[0.792]	[0.225]	[0.649]	[0.513]
F-test pval (subj=obj)	0.16	0.09	0.55	0.65	0.59	0.00
Randomiz infer pval (subj=obj)	0.146	0.0420	0.626	0.682	0.614	0.00400
Control Group Mean	-0.00	-0.00	-0.00	0.00	-0.01	0.38
Clusters	126	126	126	125	126	124
Observations	15418	15401	14904	14168	14909	11505

Clustered standard errors * p<0.10, ** p<0.05, *** p<0.01

Objective incentives decrease socio-emotion skills, but subjective incentives have no effect on them

Teaching Effort

Conduct classroom observations for 1,500 teachers during intervention year to measure teacher pedagogy

	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

Clustered standard errors * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Objective incentives decrease classroom pedagogy quality

Teaching Effort

Measure teacher clock in and out time for all teachers using biometric data

	Days present at school		Hours worked per day	
	(1)	(2)	(3)	(4)
Objective Treatment	2.426	1.554	0.262	0.293
	(0.570)	(0.339)	(0.195)	(0.282)
	[0.618]	[0.392]	[0.318]	[0.319]
Subjective Treatment	5.927*	3.340***	0.0348	-0.0432
	(0.0719)	(0.00947)	(0.840)	(0.832)
	[0.0960]	[0.0100]	[0.855]	[0.823]
Sample	All	Restricted	All	Restricted
F-test pval (subj=obj)	0.30	0.15	0.13	0.12
Randomiz infer pval (subj=obj)	0.371	0.202	0.295	0.164
Control Group Mean	144.79	182.72	7.90	7.92
Clusters	295	277	295	277
Observations	6394	4363	6394	4363

Clustered standard errors * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Subject incentives increase attendance by 4%

Mechanisms

How can we understand the difference between the effects of subjective versus objective performance pay?

Mechanisms

How can we understand the difference between the effects of subjective versus objective performance pay?

Features that are the same:

- Within school tournament
- 0-10% raise
- Timing of roll out

Mechanisms

How can we understand the difference between the effects of subjective versus objective performance pay?

Features that are the same:

- Within school tournament
- 0-10% raise
- Timing of roll out
- Endline survey - No reported difference in:
    - when teachers said they understood what was expected
    - understanding of main features of contract
    - how often they thought about incentive
    - system unfairly favors certain types of teachers (age, gender, etc)

Mechanisms: Noise and Distortion

Teacher's beliefs about incentive → Teacher effort → Student outcomes

Teacher's beliefs about incentive

- Subjective incentive point categories
- Teachers' belief about:
  - Noise of incentive scheme
  - Accuracy of principal evaluation
  - Actions that are rewarded

Teacher effort

- Classroom observation
  - Research team
  - Managers
- Reported time use
- *Attendance/clock in and out*

Student outcomes

- Test scores (Math, Science, English and Urdu)
- Student survey measuring socio-emotional outcomes

Mechanism 1: Noise

Under subjective incentives, teachers are:

- Less likely to say "their raise is out of their control"
- More likely to say "those who work harder ear more"
- More likely to say "I feel motivated"

Mechanism 2: Distortion

Teachers also believe different actions are most important under subjective versus objective incentives. Under subjective incentives, they are:

- More likely to say helping with administrative and afterschool duties is important
- Less likely to say that doing test preparation is important

Heterogeneity by Principal Quality

Overall subjective incentives appear to dominate objective incentives. But does the effectiveness vary across managers?

No effect of subjective performance pay on test scores:

- For principals who teachers believe accept bribes (10% of managers)
- For principals who are in bottom quintile of perceived rating accuracy

Conclusion

- Subjective and objective incentives increase test scores
- Objective incentives decrease socio-emotional outcomes and teaching quality
- Subjective incentives appear less noisy and distorted
- Not all principals are able to implement subjective incentives well

Thank you!

Contact: christinabrown@berkeley.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

Performance Pay Effect Decomposition

$$\begin{aligned}
\Delta T &= E[\theta + \beta | b \geq 0] - E[\theta | b < 0] && (1)\\
&= E[\theta | b \geq 0] - E[\theta | b < 0] + E[\beta | b \geq 0] && (2)\\
&= E[\theta | b \geq 0] - E[\theta | b < 0] + (E[\beta] - XX && (3)\\
&= E[\theta | b \geq 0] - E[\theta | b < 0] + E[\beta] + (E[\beta] - E[\beta | b < 0])\frac{1-p}{p} && (4)
\end{aligned}$$

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) Flat Mean/SE	(2) Objective Mean/SE	(3) Subjective Mean/SE	T-test P-value (1)-(2)	T-test P-value (1)-(3)
Age	37.929	37.259	37.770	0.448	0.855
	(0.682)	(0.564)	(0.544)		
First year teacher	0.193	0.228	0.178	0.351	0.683
	(0.031)	(0.020)	(0.019)		
Years of experience	4.851	4.748	5.147	0.824	0.515
	(0.339)	(0.323)	(0.302)		
Female	0.755	0.785	0.746	0.566	0.880
	(0.038)	(0.034)	(0.048)		
N	1108	711	847		
Clusters	42	42	43		

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

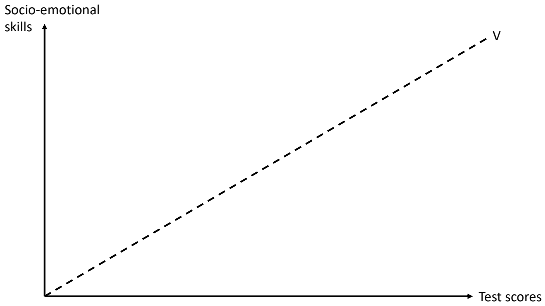
Back

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

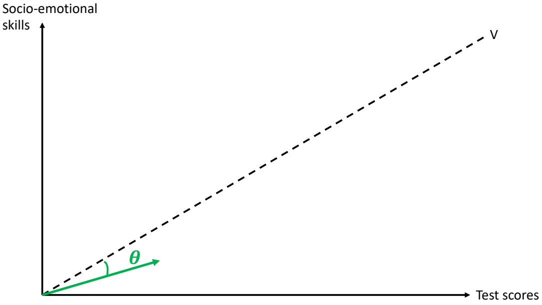
Distortion and Noise

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



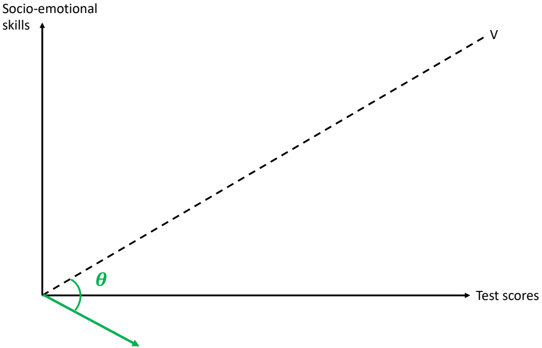
Distortion and Noise

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



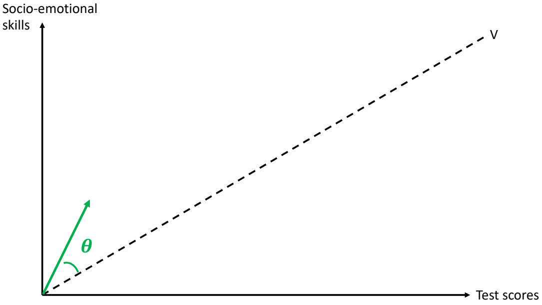
Distortion and Noise

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



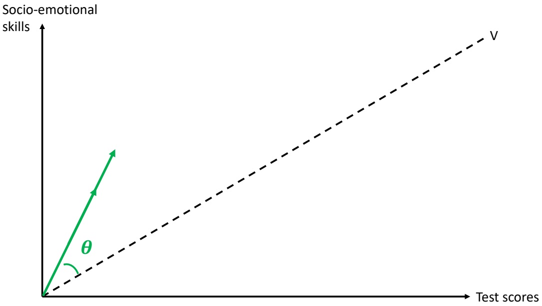
Distortion and Noise

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



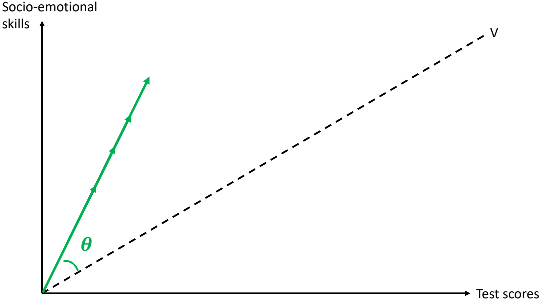
Distortion and Noise

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



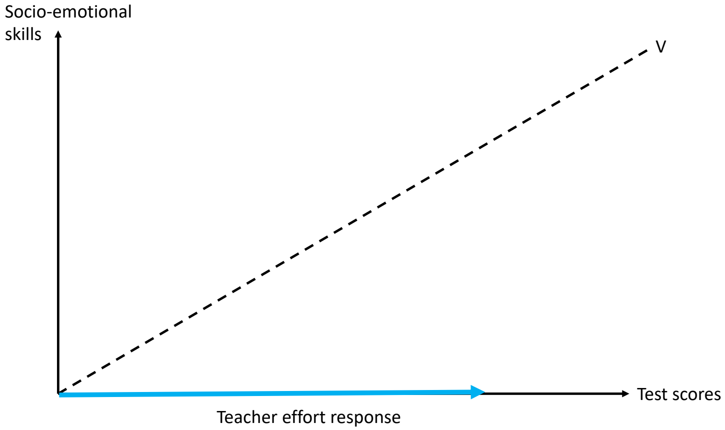
Distortion and Noise

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



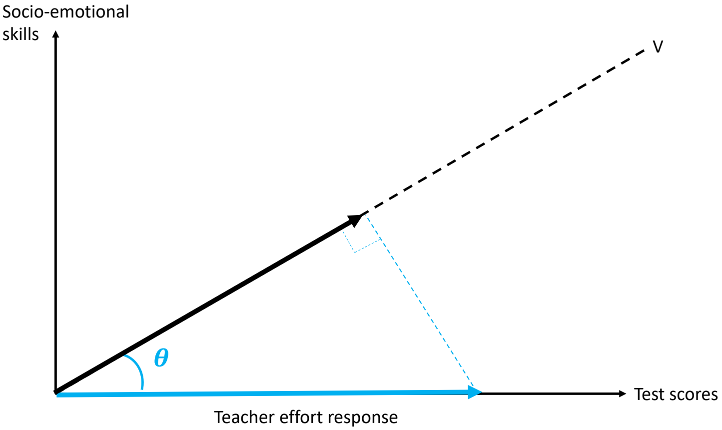
Distortion and Noise

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