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? → 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
(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

0 of 10 variables are stat. sig. at baseline
## Design - Data

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

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

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

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

<table>
<thead>
<tr>
<th></th>
<th>All (1)</th>
<th>Remedial (2)</th>
<th>External (3)</th>
<th>Math/Science (4)</th>
<th>English/Urdu (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objective Treatment</td>
<td>0.0918*</td>
<td>0.189***</td>
<td>0.119**</td>
<td>0.104*</td>
<td>0.0917</td>
</tr>
<tr>
<td></td>
<td>(0.0575)</td>
<td>(0.00518)</td>
<td>(0.0335)</td>
<td>(0.0668)</td>
<td>(0.166)</td>
</tr>
<tr>
<td></td>
<td>[0.0730]</td>
<td>[0.0260]</td>
<td>[0.0200]</td>
<td>[0.194]</td>
<td>[0.144]</td>
</tr>
<tr>
<td>Subjective Treatment</td>
<td>0.0859**</td>
<td>0.142**</td>
<td>0.0855*</td>
<td>0.0884*</td>
<td>0.0986**</td>
</tr>
<tr>
<td></td>
<td>(0.0220)</td>
<td>(0.0113)</td>
<td>(0.0601)</td>
<td>(0.0646)</td>
<td>(0.0267)</td>
</tr>
<tr>
<td></td>
<td>[0.0130]</td>
<td>[0.0240]</td>
<td>[0.0170]</td>
<td>[0.121]</td>
<td>[0.0260]</td>
</tr>
<tr>
<td>F-test pval (subj=obj)</td>
<td>0.89</td>
<td>0.38</td>
<td>0.43</td>
<td>0.77</td>
<td>0.90</td>
</tr>
<tr>
<td>Randomiz infer pval (subj=obj)</td>
<td>0.884</td>
<td>0.453</td>
<td>0.388</td>
<td>0.819</td>
<td>0.873</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Survey Items</th>
<th>Socio-Emotional Indices (z-score)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All (1)</td>
</tr>
<tr>
<td>Objective Treatment</td>
<td>-0.0262</td>
</tr>
<tr>
<td></td>
<td>(0.423)</td>
</tr>
<tr>
<td></td>
<td>[0.515]</td>
</tr>
<tr>
<td>Subjective Treatment</td>
<td>0.0171</td>
</tr>
<tr>
<td></td>
<td>(0.363)</td>
</tr>
<tr>
<td></td>
<td>[0.576]</td>
</tr>
<tr>
<td>F-test pval (subj=objc)</td>
<td>0.16</td>
</tr>
<tr>
<td>Randomiz infer pval (subj=objc)</td>
<td>0.146</td>
</tr>
<tr>
<td>Control Group Mean</td>
<td>-0.00</td>
</tr>
<tr>
<td>Clusters</td>
<td>126</td>
</tr>
<tr>
<td>Observations</td>
<td>15418</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Objective Treatment</th>
<th>Classroom Observation Rubric</th>
<th>Test Prep</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All (1)</td>
<td>(5)</td>
</tr>
<tr>
<td></td>
<td>Class Climate (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Differentiation (3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Student-Centered (4)</td>
<td></td>
</tr>
<tr>
<td>Objective Treatment</td>
<td>-0.0713</td>
<td>0.577***</td>
</tr>
<tr>
<td></td>
<td>(0.123)</td>
<td>(0.00455)</td>
</tr>
<tr>
<td></td>
<td>[0.171]</td>
<td>[0.0120]</td>
</tr>
<tr>
<td>Subjective Treatment</td>
<td>-0.00206</td>
<td>0.110</td>
</tr>
<tr>
<td></td>
<td>(0.959)</td>
<td>(0.255)</td>
</tr>
<tr>
<td></td>
<td>[0.946]</td>
<td>[0.649]</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Days present at school</th>
<th>Hours worked per day</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Objective Treatment</td>
<td>2.426</td>
<td>1.554</td>
</tr>
<tr>
<td></td>
<td>(0.570)</td>
<td>(0.339)</td>
</tr>
<tr>
<td></td>
<td>[0.618]</td>
<td>[0.392]</td>
</tr>
<tr>
<td>Subjective Treatment</td>
<td>5.927*</td>
<td>3.340***</td>
</tr>
<tr>
<td></td>
<td>(0.0719)</td>
<td>(0.00947)</td>
</tr>
<tr>
<td></td>
<td>[0.0960]</td>
<td>[0.0100]</td>
</tr>
</tbody>
</table>

Sample

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Restricted</th>
<th>All</th>
<th>Restricted</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-test pval (subj=obj)</td>
<td>0.30</td>
<td>0.15</td>
<td>0.13</td>
<td>0.12</td>
</tr>
<tr>
<td>Randomiz infer pval (subj=obj)</td>
<td>0.371</td>
<td>0.202</td>
<td>0.295</td>
<td>0.164</td>
</tr>
</tbody>
</table>

Control Group Mean

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Clusters</td>
<td>144.79</td>
<td>182.72</td>
<td>7.90</td>
<td>7.92</td>
</tr>
<tr>
<td>Observations</td>
<td>6394</td>
<td>4363</td>
<td>6394</td>
<td>4363</td>
</tr>
</tbody>
</table>

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

Teacher effort

Student outcomes
  - Classroom observation
    - Research team
    - Managers
    - Reported time use
    - Attendance/clock in and out
  - 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 earn 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

\[ \Delta T = E[\theta + \beta|b \geq 0] - E[\theta|b < 0] \]
\[ = E[\theta|b \geq 0] - E[\theta|b < 0] + E[\beta|b \geq 0] \]
\[ = E[\theta|b \geq 0] - E[\theta|b < 0] + (E[\beta] - XX) \]
\[ = E[\theta|b \geq 0] - E[\theta|b < 0] + E[\beta] + (E[\beta] - E[\beta|b < 0]) \frac{1-p}{p} \]
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
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
## Balance in Baseline Covariates

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

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

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

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 $V$. 

![Diagram showing the relationship between socio-emotional skills and test scores]
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.
Appendix

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.
Appendix

Experimental Design

2017

Oct

Teacher & Manager Baseline Survey

Dec

End Year Performance Eval & Goal Setting

2018

Feb

Control randomize Objective Raise

Objecive Raise

Treatment Info Campaign

2019

Jan

Student Testing/Survey

March

Raises announced

May

Teacher & Manager Endline Survey

2018

July

Treatment reminder & midterm info

Fall

Classroom observation

Dec

Manager Evaluation

2019

Jan

Student Testing/Survey

March

Raises announced

May

Teacher & Manager Endline Survey