

The Potential Output Gains from Using Optimal Teacher Incentives:

An Illustrative Calibration of a Hidden Action Model

Nirav Mehta

University of Western Ontario *

April 23, 2018

Abstract

This paper examines the potential output gains from the implementation of optimal teacher incentive pay schemes, by calibrating the Hölmstrom and Milgrom (1987) hidden action model using data from Muralidharan and Sundararaman (2011), a teacher incentive pay experiment implemented in Andhra Pradesh, India. Findings suggest that the introduction of optimal individual incentive-pay schemes could result in very large increases in output, about six times the size of the (significant) results obtained in the experiment.

Keywords: hidden action, empirical contracts, teacher incentive pay

JEL: I2, J3, J4

*nirav.mehta@uwo.ca I thank Rui Castro and Rachel Margolis for useful discussions pertaining to this paper, and the SSHRC Insight Development Grant program and Jacobs Foundation for funding. I also thank Enrique Martin Luccioni for research assistance.

1 Introduction

Evidence that teacher quality is an important determinant of human capital that is hard to measure (Goldhaber and Brewer (1997); Rivkin et al. (2005); Hanushek (2011)) has generated substantial policy interest in output-based teacher incentive schemes and motivated a research agenda using randomized controlled trials to estimate whether such schemes affect teacher inputs. In theory, many characteristics of an incentive scheme’s design, together with the context in which the scheme is implemented, could be important determinants of its efficacy. Broadly, teacher incentive schemes may differ by their *structure*—e.g., whether they provide teachers with individual bonuses that are linear in only their own output, or assign bonuses via tournaments at the district level— and their *strength*—e.g., the “slope” of the individual bonus in output, or the size of the prize given to the winner of the tournament.

Most research studying the design of output-based teacher incentives has focused on varying scheme structure. This research has been carried out across a wide variety of different contexts. Given this heterogeneity, it is perhaps unsurprising that the effects of output-based teacher incentives have been widely mixed. For example, even among randomized controlled trials (one of the many research designs used in this literature), there is no consensus on the broad question of whether incentivizing teachers based on output measures improves student achievement:¹ Muralidharan and Sundararaman (2011) find significant positive effects of individual and group-based linear incentive schemes implemented in Andhra Pradesh, India; Springer et al. (2010) find no significant effect on student achievement of POINT, an individual, threshold-based incentive implemented in Nashville, Tennessee; and Fryer Jr. (2013) finds significant *negative* effects of group-based (and typically threshold-based) incentive schemes chosen by public schools in New York City.²

Perhaps due to the wide variety of possible structures, incentive strengths, and contexts, little attention has been paid to the important question of what are the *potential gains* from

¹As would be expected, there is a similar lack of consensus among research using non-experimental data to study different types of teacher incentive schemes. For example, Dee and Wyckoff (2015) use a regression-discontinuity design to estimate the effect of an individual, threshold-based scheme in Washington, D.C., and find significant, positive effects of both dismissal threats and performance-based bonuses on student achievement. Sojourner et al. (2014) use a difference-in-differences approach to estimate the effect of Q-Comp, a reform in Minnesota in which districts were allowed to choose from a set of possible schemes, all of which included fairly weak group-based threshold-based teacher incentives; while it seems that Q-Comp marginally improved students’ reading, there was not a large significant effect on their mathematics achievement.

²Springer et al. (2010) and Muralidharan and Sundararaman (2011) study teacher incentive schemes that were based on absolute (as opposed to relative) performance; Fryer Jr. (2013) considers a scheme based on relative performance. For another example of a scheme based on relative performance, see Imberman and Lovenheim (2015), which studies a tournament scheme in Houston, Texas, in which groups of teachers competed for performance bonuses.

their implementation, which can only be computed using the *optimal incentive scheme* for a given environment, could be. Despite the aforementioned availability of high-quality experimental findings and a well-developed theoretical literature, the gains from using *optimal* incentives are unknown for two reasons. First, characterizing optimal contracts is technically very demanding in most economic environments. Second, quantifying the effects of implementing such a contract would also require knowledge of the parameters of the model describing the underlying economic environment.

If output is a noisy measure of teacher effort, then output-based incentives could be suboptimally strong if, as incentive pay opponents argue, they expose teachers to too much risk. Alternatively, they could be too weak if they do not appreciably change teacher effort inputs, making it hard to discern significant effects. Theoretical work such as Barlevy and Neal (2012), which takes the first step, by developing a multitask model of teacher effort provision, cannot quantify the gains from optimal contracts without taking the second step. At the same time, cleanly identified and precisely estimated causal effects from RCTs cannot speak to the gains from optimal incentives without the additional structure provided by theoretical work. Moreover, since RCT implementation is expensive, experiments typically study only a small number of different levels of incentive strength (i.e., treatment groups). Searching for optimal incentive strength via pure experimentation would be prohibitively expensive, motivating the use of additional structure to maximally leverage findings from RCTs for use in education policy.

This paper takes a step towards filling this gap by using the framework of Hölmstrom and Milgrom (1987), a hidden action, or “moral hazard” model of effort choice, to interpret findings from Muralidharan and Sundararaman (2011), an experimental study of teacher incentive pay implemented in Andhra Pradesh, a state in India. In the model, teachers choose an unobserved effort level, which determines their quality. The main advantage of this model is its closed-form solution of the optimal contract: the optimal incentive scheme is linear in output, which depends on teacher effort and a shock. A larger error variance or higher teacher risk aversion would reduce optimal incentive strength, or slope of remuneration in output. Hölmstrom and Milgrom (1987) is equivalent to the widely used CARA-Normal model, making it a particularly salient example environment.

Muralidharan and Sundararaman (2011) is particularly good for calibrating model parameters because the experiment introduced an output-based incentive scheme that, like the optimal contract in Hölmstrom and Milgrom (1987), is linear in the output of individual teachers. Additionally, the authors estimate a significant effect of individual incentive-pay schemes. Because this paper’s goal is to assess potential gains from optimal contracts, it is most natural to focus on a well-designed incentive pay experiment reporting a statistically

significant effect.

2 Methods

2.1 Background on Muralidharan and Sundararaman (2011)

Muralidharan and Sundararaman (2011) conducted a large-scale experimental evaluation of group- and individual-based performance-based bonus schemes for primary school teachers in rural Andhra Pradesh, which was the fifth most populous state in India. Muralidharan and Sundararaman (2011) write that the quality of education in rural Andhra Pradesh was similar to that elsewhere in rural India, where “nearly 60 percent of children aged 6–14 . . . could not read at the second-grade level, though over 95 percent of them were enrolled in school,” (page 45). In addition to this context of fairly low achievement, prior research had already shown that potentially important inputs were also fairly low in Andhra Pradesh. For example, Kremer et al. (2005) documented a 25% absence rate among teachers in the state. Taken together, these facts point to large potential gains from increasing teachers’ provision of costly effort inputs.

The experiment, which spanned the 2005–06 and 2006–07 school years, randomly assigned 100 schools from a representative sample of government-run schools in rural Andhra Pradesh to either treatment arm, setting aside another 100 government-run schools to serve as a control group.³ Although Muralidharan and Sundararaman (2011) find similar positive and significant average treatment effects for the group- and individual-based incentives for their main output measure (the average of achievement score increases in math and language) in the first year, I focus on the individual-based incentive scheme because it is most closely linked with the theoretical model. The individual bonus scheme assigned a teacher a bonus linear in the amount a teacher’s student test score growth exceeded a minimal threshold: in particular, the teacher was paid 500 rupees per percent increase in mean test scores, for test score gains above 5%.⁴ The primary schools studied in Muralidharan and Sundararaman (2011) were fairly small, with most schools having either two or three teachers and a typical class size of around 40 students. The average teacher in the sample was 37 years old, the standard deviation of teacher age was 8.8 years, and 57% of teachers were male; these statistics are germane for the later sensitivity analysis, in which I consider the robustness of

³Muralidharan and Sundararaman (2011) also featured a treatment arm that provided resources to schools; please see Muralidharan and Sundararaman (2011) for further detail.

⁴This threshold was chosen to minimize concerns that threshold-induced nonlinearities would substantially affect teacher behavior; see footnote 16 on page 51 of Muralidharan and Sundararaman (2011). At a conversion of 45 rupees per dollar, 500 rupees corresponds to \$11.11.

potential gains.⁵

2.2 Model

This section presents the workhorse CARA-Normal model of moral hazard, as developed in Bolton and Dewatripont (2005). Although this model assumes a linear contract, which need not be optimal, the solution is the same as that in Hölmstrom and Milgrom (1987), which studies a static one-period model split into a number of sub-periods, where in each sub-period an agent (i.e., teacher) controls the probability of success for a binomial random variable. In particular, Hölmstrom and Milgrom (1987) show that the optimal contract features an end-of-period payment that is a linear function of aggregated signals. The interpretation for our education context would be that, in each infinitesimal unit of time, the teacher could exert more or less effort to increase the probability a student obtains a sub-period-specific “bit” of human capital measured by an end-of-year exam.

The administrator has utility $q - w$, where q is output and w is the wage paid to the teacher. The teacher has constant absolute risk aversion (CARA) utility $-e^{-\xi(w-\psi(a))}$, where ξ is their coefficient of absolute risk-aversion and the cost of exerting effort a is $\psi(a) = \gamma a^2/2$. The teacher requires an expected utility of \underline{u} to participate. Output from teacher i depends on their effort according to $q_i = a_i + \eta_i$, where the IID ex-post shock $\eta_i \sim N(0, \sigma_\eta^2)$ renders output a noisy measure of teacher effort.

Hölmstrom and Milgrom (1987) show that it is optimal for the administrator to pay the teacher using the linear contract $w = \beta_0 + \beta_1 q$, where β_1 is the share of output paid to the teacher. Therefore, the administrator solves

$$\begin{aligned} \max_{\beta_0, \beta_1} \quad & \mathbb{E}_\eta [a + \eta - w(a, \eta)] & (1) \\ \text{s.t.} \quad & w(a, \eta) = \beta_0 + \beta_1(a + \eta) \\ & \mathbb{E}_\eta [-e^{-\xi(w(a, \eta) - \psi(a))}] \geq \underline{u} & (\text{IR}) \\ & a \in \arg \max \mathbb{E}_\eta [-e^{-\xi(w(a, \eta) - \psi(a))}] & (\text{IC}) \end{aligned}$$

The teacher problem yields a unique optimal effort level $a^* = \beta_1/\gamma$ by differentiating (IC) with respect to effort, and the optimal linear contract sets $\beta_1^* = 1/(1 + \xi\gamma\sigma_\eta^2)$. Therefore, expected output is $\mathbb{E}[q^*] = \mathbb{E}_\eta [a^* + \eta] = a^* = 1/(\gamma(1 + \xi\gamma\sigma_\eta^2))$. Intuitively, as the signal quality worsens (i.e., σ_η^2 increases) the contract becomes lower powered (i.e., β_1^* decreases), resulting in lower effort a^* and expected output $\mathbb{E}[q^*]$.

⁵These statistics were computed using the supplementary data associated with Muralidharan and Sundararaman (2011), which are available at <https://doi.org/10.1086/659655>.

If noise increased, the resulting optimal contract would partially protect a risk-averse teacher by making incentives weaker in output, by reducing the slope of the linear contract β_1^* . The more risk-averse the teacher, the more protected they would be from fluctuations in η .

2.3 Calibration

I calibrate the model parameters $(\gamma, \xi, \sigma_\eta^2)$ using a “sophisticated” back-of-the-envelope method, which is “sophisticated” because I calibrate using equilibrium implications of the hidden action model. As I show below, values for ξ and σ_η^2 can be obtained either directly from external sources or by transforming external data. However, to calibrate the effort cost parameter γ , we need to know how much teachers respond to incentive pay. Note that the “causal” or composite effect of teacher incentive pay reported in the experimental results could, in theory, also include changes in student and/or family inputs. However, assigning the total effect to changes in teacher effort is consistent with the theoretical model used to interpret these results. Note that effort and output are compared to their baseline levels, i.e., that obtained absent output-based incentives.

As with any mapping between theory and data, assumptions have to be made. The benefit of using Hölmstrom and Milgrom (1987) to interpret Muralidharan and Sundararaman (2011) is that the linear scheme employed in the latter affords a clean mapping between their findings and the hidden action model. The same is true of their experimental research design, which obviates having to account for mean differences in output between treatment and control groups being based on selection on hidden types, allowing the calibration to proceed for a representative (average) teacher.⁶ I convert currency into U.S. dollars for convenience. While this might raise concerns about external validity, CARA utility implies that risk aversion is independent of wealth, meaning the large wealth differences between teachers in India and the U.S. would only affect the intercept, not optimal teacher effort and output.

There were on average 3.14 teachers and 37.5 pupils per teacher in the incentive schools. Student achievement increased by an average of 0.1415 sd, per year.⁷ Students’ annual wages increased by an average of 2,812 rupees per student⁸; the average cost of the incentive

⁶The linearity of the administrator’s objective implies that she can solve a separate problem for each teacher.

⁷This was obtained from the top row of Table 8 of Muralidharan and Sundararaman (2011), by dividing the composite two-year gain by two.

⁸See footnote 34 on page 72 of Muralidharan and Sundararaman (2011); note, however, that a 10.04% increase was applied instead of 7.7%, as the former was based on individual scheme gains and the latter was based on pooled individual and group incentive scheme gains.

scheme was 20,000 rupees.⁹ With a conversion rate of 45 rupees per dollar, this corresponds to \$2,343.60 ($=\62.50×37.5) in total output produced by the average teacher and \$141.54 ($=\$444.44/3.14$) paid to the average teacher. Then, the slope of the contract is the per-teacher income increase (\$141.54) divided by the increase in output (\$2,343.60), or 0.0604; i.e., teachers were paid a piece rate of 6.04% of output.

We can exploit the teacher’s optimal choice of action, which solves (IC) in (1) but does not rely on optimality of the slope β_1 , to map (β_1, a) to the effort cost γ . The value of γ which rationalizes this increase is then $\gamma = \beta_1/a = 0.0604/2,343.60 = 2.577 \times 10^{-5}$. Teacher risk aversion matters for how incentives are structured (Nadler and Wiswall (2011)). I set the CARA parameter to $\xi = 6.7 \times 10^{-3}$, the mean estimated coefficient of absolute risk aversion from the benchmark model of Cohen and Einav (2007), Table 5.

Assuming mean test scores¹⁰ \bar{y} are converted to output via $q = \beta_q \bar{y}$, the conversion factor β_q can be calibrated by noting that the scheme increased mean test scores by 0.1415 sd and output per teacher by \$2,343.60, resulting in a conversion factor $\beta_q = \$16,562.54$ ($=\$2,343.60/0.1415$). Student j ’s test score depends on their teacher i ’s effort and a student-specific shock distributed IID according to $\epsilon_{ji} \sim N(0, \sigma_\epsilon^2)$, which captures idiosyncratic factors affecting student achievement on the administered test instrument. The variance of the mean test score then can be computed by dividing the variance of test score error σ_ϵ^2 by the average number of students per teacher in the data, i.e., $\sigma_{\bar{y}}^2 = 0.953/(37.5)$.¹¹ To obtain the variance of output σ_η^2 we then square the test-score-to-income parameter and multiply by the variance of mean test score, i.e., $\sigma_\eta^2 = 6,971,331\$^2 (= \$16,562.54^2 \times 0.953/(37.5))$.¹²

3 Results

Using the calibrated parameter values, we can solve for the optimal slope of $\beta_1^* = 0.454$, which is over six times steeper than in the experiment. This results in an optimal effort level/output gain of $a^* = \$17,608.83$, which corresponds to an average increase in student achievement of 1.063 sd. Accordingly, these increases are also more than six times larger than the estimated increases stemming from the much weaker incentives provided under the experiment.

⁹The incentive scheme cost an average of 10,000 rupees for each of two years (Muralidharan and Sundaraman, 2011, pp. 70-71).

¹⁰Note that everywhere, I refer to test score gains.

¹¹Schochet and Chiang (2012) compile estimates of the variances from a large number of studies in their study of error rates in value-added models, providing a good source for typical values for σ_ϵ^2 of $\sigma_\epsilon^2 = 0.953$. Results available upon request.

¹²This is because the variance of q , i.e., σ_η^2 , is $\beta_q^2 \sigma_{\bar{y}}^2$.

Sensitivity Analysis Figure 1 presents contour maps of model outcomes for a grid of points covering a wide range of alternative values of the logarithms of σ_η^2 and ξ , ranging from one half to ten times the calibrated value of each parameter.¹³ Note that, because γ was recovered using the teacher’s effort action choice and can be recovered by using the slope of incentives in the experiment and increase in output, it does not depend on (σ_η^2, ξ) . Figure 1a is a contour map of the optimal output share, or β_1^* . Figure 1b is a contour map of optimal output, i.e., $E[q^*]$. In both figures, the value corresponding to the calibrated values of σ_η^2 and ξ is indicated by a red dot. We can see that as teachers become more risk averse (increasing ξ) or output becomes noisier (increasing σ_η^2), both incentive strength (Figure 1a) and output gains decrease (Figure 1b). For example, fixing σ_η^2 at its calibrated value, the increase in output ranges from about 1.4 sd in student achievement to around 0.3 sd when teachers are ten times more risk averse than their calibrated value of $\xi = 6.7 \times 10^{-3}$. This latter figure is only about twice the estimated effect of the incentive scheme, but still of considerable magnitude when compared with the effects of other educational interventions, while not being implausibly large.¹⁴ Put another way, teachers would have to be extremely risk averse and/or output would have to be far noisier than is typically the case for optimal incentives be even close to as flat as those in Muralidharan and Sundararaman (2011).

We can use additional results from Cohen and Einav (2007) to make the sensitivity analysis with respect to the coefficient of absolute risk aversion a bit more concrete. For example, taking the benchmark model estimates from Cohen and Einav (2007), Table 4, the mean of which I use as the calibrated coefficient of absolute risk aversion, the coefficient on *Female* in predicting the log of the coefficient of absolute risk aversion is 0.2049. That is, ceteris paribus, females have a 20% higher coefficient of absolute risk aversion than males, which means that increasing the calibrated CARA parameter ξ by 20% would provide a very conservative upper bound for how switching from male to female would affect ξ . The resulting value, $\xi = 8.1 \times 10^{-3}$, is well within the range of values explored in the sensitivity analysis. Similarly, Cohen and Einav (2007) reports that the coefficient of absolute risk aversion is U-shaped in age, achieving a minimum at 48 years. Using the estimated coefficients from Table 4 on *Age* and *Age squared*, the log coefficient of absolute risk aversion would be 0.51 higher for a 20-year old than it would be for a 48-year old, implying a very conservative upper bound for the coefficient of absolute risk aversion for a 20-year old—about two standard

¹³Table 2 in Babcock et al. (1993) shows that a higher-end estimate of ξ is about 0.35, well above the range considered in the parameter grid here. The lower bound for the range of the CARA parameter is very close to the mean and median values of the CARA parameter when Cohen and Einav (2007) assume CARA utility (in the third panel of their Table 5); therefore, using a larger (i.e., more risk-averse) value for ξ as the baseline results in weaker optimal incentives. In this particular sense, we may view the results for the calibrated values as conservative, in terms of the potential output gains.

¹⁴Cohen (1988) classifies gains of 0.80 sd and higher as “large”.

deviations younger than the average teacher in the sample—that is 51% higher than the calibrated value, i.e., $\xi = 1.0 \times 10^{-2}$. As was the case for females, this value is also well within the range of values explored in the sensitivity analysis.

Finally, Figure 1c presents a contour map of the expected share of teacher income comprised by variable compensation, i.e., $E[\beta_1^* q^*] / E[\beta_0^* + \beta_1^* q^*]$. As with the slope and output, this share declines as the output shock variance and degree of risk aversion increase.¹⁵ The optimal expected share of income that is variable pay under the calibrated parameter values would be around 10%.

4 Discussion and Conclusion

This paper produces the first quantitative assessment of the potential gains to implementing optimal teacher incentive pay. The findings point to large potential gains to implementing optimal contracts, which are six times steeper than those in the experiment, which were already significant and positive. This suggests that the estimated null effect found in some implemented studies of incentive pay could potentially be attributed to weaker-than-optimal incentive strength.

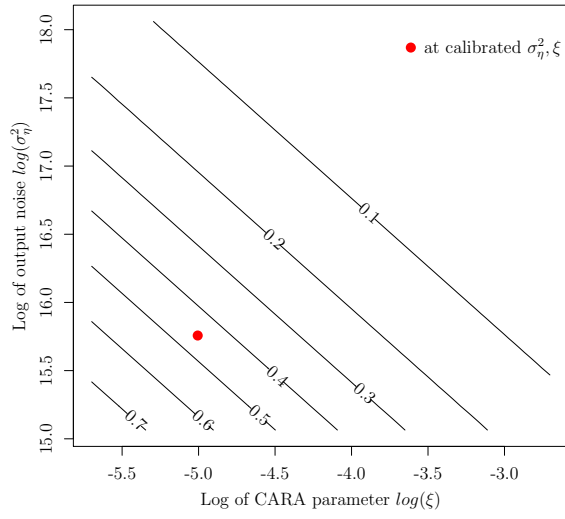
The simplicity of this paper’s approach allows me to study an environment for which the optimal contract has already been characterized and then use a well-implemented empirical study to recover the relevant parameters. It provides an example of the potential gains to adopting optimal contracts in educational production. That being said, caveats regarding to the interpretation of this paper’s results are in order.

To start, there are two reasons, stemming from this paper’s strategy of combining information from Indian and other (e.g., the US and Israeli) contexts to calibrate the model, why it would be prudent to view the calibrated six-fold increase in output resulting from the introduction of optimal incentives as an upper bound. First, if teachers are substantially more risk-averse than other agents (e.g., Dohmen and Falk (2010) document that teachers are more risk-averse than other workers), optimal incentives (and the associated potential output gains) would naturally be weaker (and smaller). In a similar vein, although teachers are assumed to have CARA preferences, it is worth considering how a deviation from constant absolute risk aversion could affect this paper’s findings. The typical deviation from constant absolute risk aversion is decreasing absolute risk aversion (Friend and Blume (1975)).¹⁶ Therefore, as a rough approximation, suppose that we maintained CARA pref-

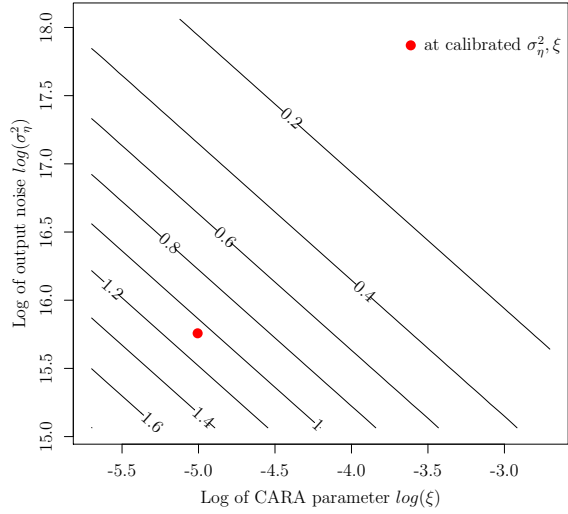
¹⁵This was computed using a certainty equivalent value of \$70,000 (Himes (2015)), along with a binding IR constraint in (1).

¹⁶For example, macroeconomists often assume that preferences exhibit constant relative risk aversion (i.e., CRRA utility), which implies decreasing absolute risk aversion.

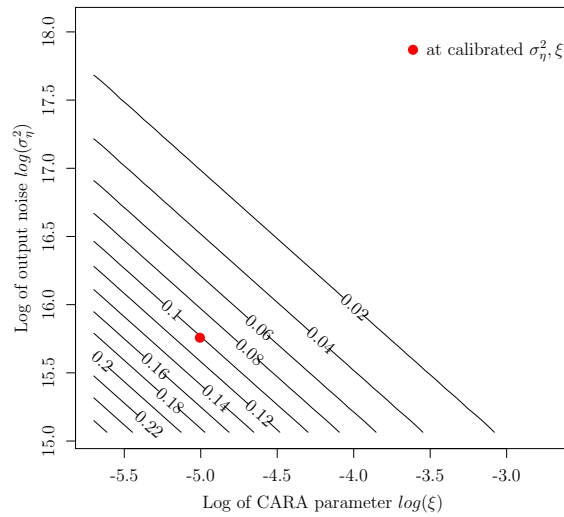
Figure 1: Optimal output share and ratio of output for (σ_η^2, ξ) -grid



(a) Output share,
 β_1^*



(b) Output (sd), $E[q^*]$



(c) Variable share of income,
 $E[\beta_1^* q^*] / E[\beta_0^* + \beta_1^* q^*]$

erences, but adjusted ξ to reflect the potentially large differences in income between, say, India and the US. In this case, it might seem reasonable to think that the calibrated value of ξ (which was based on data from Israel, which is much more similar to the US, in terms of income¹⁷) would be too low to properly account for the potentially greater risk aversion of Indian teachers. Thus, here too, the optimal incentive strength, and potential output gains, would be smaller than those using the calibrated value of ξ . This highlights the importance of the sensitivity analysis, which shows that there would be substantial potential output gains even when ξ was ten times larger than its calibrated value.

Second, output in this paper is measured in terms of value added, and output gains (i.e., increases in value added) are determined by γ , the effort cost parameter. The large treatment effects found in Muralidharan and Sundararaman (2011) imply a small calibrated value of γ . Such large effects may seem quite reasonable in a context where, e.g., in the control group over 20% of teachers were absent and less than half were found to be actively teaching when audited by enumerators (Muralidharan and Sundararaman (2011), Table 9); indeed, the low baseline achievement level of rural Andhra Pradesh was what motivated the experimental intervention. The convexity of the effort cost function implies that in a context such as the US, which perhaps could be viewed as having higher baseline student achievement and teacher effort (e.g., lower absence rates), would imply a higher value for γ . This would reduce optimal incentive strength and the associated potential output gains. Quantifying γ for the US, or other, contexts would thus be a useful line of future research.

A final caveat is that the linear, individual teacher incentive scheme is only one of the many potential structures of incentive schemes, as the Hölmstrom and Milgrom (1987) environment is only one of many potentially relevant economic environments; different underlying environments would likely imply different optimal incentive schemes. For example, in this paper’s model, the lack of teacher effort is the channel through which poor student performance can be explained. While there is some external support to the hypothesis that teacher effort matters, e.g., Jones (2013), another potentially important design element is whether (and how strongly) students should be incentivized. For example, Behrman et al. (2015) conducted a randomized controlled trial in Mexican high schools that featured three treatment arms, with threshold-based incentives for (i) only students, (ii) only teachers, and (iii) both students and teachers; only those with student incentives significantly improved achievement. Calculating the potential output gains from moving to optimal contracts in other environments constitutes an important avenue for future research.

¹⁷Real per-capita GDP in 2005, in 2011 dollars, was \$3,179 in India, \$26,761 in Israel, and \$49,762 in the US. Source: World Bank national accounts data, and OECD National Accounts data files, <https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD>.

References

- Babcock, B. A., E. K. Choi and E. Feinerman, “Risk and Probability Premiums for CARA Utility Functions,” *Journal of Agricultural and Resource Economics*, pp. 17–24, 1993.
- Barlevy, G. and D. Neal, “Pay for Percentile,” *American Economic Review*, 102(5):1805–31, 2012.
- Behrman, J. R., S. W. Parker, P. E. Todd and K. I. Wolpin, “Aligning Learning Incentives of Students and Teachers: Results from a Social Experiment in Mexican High Schools,” *Journal of Political Economy*, 123(2):325–364, 2015.
- Bolton, P. and M. Dewatripont, *Contract Theory*, MIT Press, 2005.
- Cohen, A. and L. Einav, “Estimating Risk Preferences From Deductible Choice,” *American Economic Review*, pp. 745–788, 2007.
- Cohen, J., *Statistical Power Analysis for the Behavioral Sciences*, vol. 2, Hillsdale, NJ: Lawrence Earlbaum Associates, 1988.
- Dee, T. S. and J. Wyckoff, “Incentives, Selection, and Teacher Performance: Evidence from IMPACT,” *Journal of Policy Analysis and Management*, 34(2):267–297, 2015.
- Dohmen, T. and A. Falk, “You Get What You Pay For: Incentives and Selection in the Education System,” *Economic Journal*, 120(546):F256–F271, 2010.
- Friend, I. and M. E. Blume, “The Demand for Risky Assets,” *American Economic Review*, 65(5):900–922, 1975.
- Fryer Jr., R. G., “Teacher Incentives and Student Achievement: Evidence from New York City Public Schools,” *Journal of Labor Economics*, 31(2):373–407, 2013.
- Goldhaber, D. D. and D. J. Brewer, “Why Don’t Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables On Educational Productivity.” *Journal of Human Resources*, 32(3), 1997.
- Hanushek, E. A., “The Economic Value of Higher Teacher Quality,” *Economics of Education Review*, 30(3):466–479, 2011.
- Himes, T., “LAUSD Educators Typically Earned \$75,504 Last Year,” *Los Angeles Daily News*, 2015.

- Hölmstrom, B. and P. Milgrom, "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, pp. 303–328, 1987.
- Imberman, S. A. and M. F. Lovenheim, "Incentive Strength and Teacher Productivity: Evidence from a Group-Based Teacher Incentive Pay System," *Review of Economics and Statistics*, 97(2):364–386, 2015.
- Jones, M. D., "Teacher Behavior Under Performance Pay Incentives," *Economics of Education Review*, 37:148–164, 2013.
- Kremer, M., N. Chaudhury, F. H. Rogers, K. Muralidharan and J. Hammer, "Teacher Absence in India: A Snapshot," *Journal of the European Economic Association*, 3(2-3):658–667, 2005.
- Muralidharan, K. and V. Sundararaman, "Teacher Performance Pay: Experimental Evidence from India," *Journal of Political Economy*, 119(1):39–77, 2011.
- Nadler, C. and M. Wiswall, "Risk Aversion and Support for Merit Pay: Theory and Evidence From Minnesota's Q Comp Program," *Education Finance and Policy*, 6(1):75–104, 2011.
- Rivkin, S., E. Hanushek and J. Kain, "Teachers, Schools, and Academic Achievement," *Econometrica*, 73(2):417–458, 2005, ISSN 1468-0262.
- Schochet, P. and H. Chiang, "What Are Error Rates for Classifying Teacher and School Performance Using Value-Added Models?" *Journal of Educational and Behavioral Statistics*, 2012.
- Sojourner, A. J., E. Mykerezzi et al., "Teacher Pay Reform and Productivity: Panel Data Evidence from Adoptions of Q-Comp in Minnesota," *Journal of Human Resources*, 49(4):945–981, 2014.
- Springer, M. G., L. Hamilton, D. F. McCaffrey, D. Ballou, V.-N. Le, M. Pepper, J. Lockwood and B. M. Stecher, "Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching." *National Center on Performance Incentives*, 2010.