APPENDIX: THEORY MODEL

To analyze the learning process in more detail, let education production be dependent on (potentially) multiple tasks or skillsets that must be honed.[1] That is, test score gains y is defined as the product of many tasks/skills. We also introduce the time subscript τ to denote some measure of cumulative trials or earned knowledge through time periods $t = 0,1,2, \dots, \tau - 1$. Test score evolves through time (y_t), but we are interested in the last (current) period τ output.

$$y_\tau = A \prod_{i=1}^{I} (B - (\theta_{i\tau} - x_{i\tau})^2)$$

with:

$$\theta_{it} = \gamma_i + \varepsilon_{it}$$

θ_i are measures of the school environment, potentially defined on I different dimensions, that is reflective of the student body (gender, ethnicity, academic ability, etc.) and the faculty (gender, ethnicity, age, experience, education level, etc.) that is centered around some value γ_i . The quadratic term captures the fact that choice of inputs x_i has a non-monotonic impact on test scores. The A and B terms are normalizing constants. In this framework, learning about or adjusting x_i is costless. The only negative consequence of choosing a non-optimal x_i is the decrease in test scores. Test scores would be maximized each period when ($\theta_{it} = x_{it}$). The principal must choose x_{it} before observing θ_{it}. We assume that each period, there is a stochastic shock, which makes solving for θ_i a non-trivial problem. Furthermore, she knows the distribution of γ_i is normal with variance $\sigma^2_{\gamma_i}$. However, she does not know the mean of the distribution. We assume ε_{it} has mean zero, is serially uncorrelated, and has variance σ^2_ε.[2]

Assuming a risk neutral principal, at period τ, she should set:

$$x_{i\tau} = E_\tau(\theta_{i\tau}) = E_\tau(\gamma_i) \quad \forall i$$

[1] This set up is most similar to Jovanovic and Nyarko (1995). We abstract away from a utility or objective function, except to assume that the principal wants to maximize test scores. This simplification is similar in spirit to the standard education production function.

[2] The mean zero, serially uncorrelated error term implies that test score gains $(y_t - y_{t-1})$ also does not suffer from serial correlation issues. The dynamics in the model is only through $E_t(\gamma_i)$, which evolves through time via learning.

Let $\delta_{i\tau} = E_\tau\big(\gamma_i - E_\tau(\gamma_i)\big)^2$ be the posterior variance over γ_i, with the information over the previous test score observations. Using Bayes' rule, $\delta_\tau = \frac{\sigma_\varepsilon^2 \sigma_\gamma^2}{\sigma_\varepsilon^2 + \sigma_\gamma^2 \cdot \tau}$.

Then, the expected test score at τ is:

$$E(y_\tau) = A(B - \delta_\tau - \sigma_\varepsilon^2)^I$$

Now, it is possible to define the average change in test score after one round of learning as:

$$\frac{d\,E(y_\tau)}{d\,\tau} = AI(B - \delta_\tau - \sigma_\varepsilon^2)^{I-1}\frac{\delta_\tau^2}{\sigma_\varepsilon^2} > 0$$

As expected, on average, school output increases with learning. The *rate* of increase of average test score is defined as:

$$\frac{d^2E(y_\tau)}{d\,\tau^2} = -AI(B - \delta_\tau - \sigma_\varepsilon^2)^{I-2}\frac{\delta_\tau^2}{\sigma_\varepsilon^4}[(I-1)\sigma_\varepsilon^2 + 2\delta_\tau(B - \delta_\tau - \sigma_\varepsilon^2)]$$

Proposition: When the principal learns (each time she fails to qualify for the bonus), the increase in school output in the next period depends on $\frac{d^2E(y_\tau)}{d\,\tau^2}$.

From the expression for the second derivative, $\frac{d^2E(y_\tau)}{d\,\tau^2} < 0$ if $I < \frac{2\delta_\tau(B-\delta_\tau-\sigma_\varepsilon^2)}{\sigma_\varepsilon^2}$. However, as δ_τ changes in size as τ increases, it is possible for some value of I, where it is positive for some values of τ, and negative for others.

For low values of I, the school output curve is concave. See Figure 1. This implies that principals with low number of failures can very quickly increase school output. Note that low value of I means that relative low number of skills must be honed to increase test scores, implying that increasing output is a relatively easy. Beyond some level of learning, the shape of the curve implies that decreasing marginal returns set in. In the data, if this model were true, we should see low experience principals generating large gains comparable to seasoned counterparts.

Alternatively, for high values of I, the school output curve is convex. See Figure 2. The large number of skills to be mastered implies that increasing school outputs is very complex, which shows why the increase scores is slow until the principal has experienced a high number of failures. The complexity implies that there are greater gains are left to be had, even for highly experienced principals. Empirical results should show, if this model were true, that high experience principals generate the largest gains.

Finally, in Figure 3, for mid-range values of I where small changes in τ can flip the inequality, the school output curve has a S-shape, with convex learning initially, followed by concave learning after an inflection point. This curve implies that increase in school output would be slow for low and high number of failures, but principals near the inflection point would be able to generate large gains.[3] Initially, the lack of knowledge on θ leads to many mistakes, with little improvement in output. However, principals gradually update their posterior on θ enough to make substantive improvements with enough chances to experiment through failures. In the long run, principals exhaust possible gains from learning. Additional changes to the production process would no longer yield discontinuous increases in school output.

While we will demonstrate out model of learning aligns with our empirical analysis, it is worth considering the merits of an alternative theory: leadership turnover. Sustained failures to marshal the school's resources to attain the yearly bonus may lead to a higher likelihood of school leadership turnover, whether from the loss of confidence from teachers, or from the desire of the principal to avoid potential professional stigma from failures. A replacement principal may bring fresh ideas and a new energy to the school, or perhaps merely the act of "resetting" galvanizes the teachers. In a similar fashion, if failures lead to higher teacher turnover, the arrival of new teachers may also alter the measured perception of the school environment. A strong assumption for this other theory is that although nothing substantive has necessarily changed, new arrivals are sometime enough to change not only perception, but also performance. We explicitly test for leadership and teacher turnover in the analysis section.

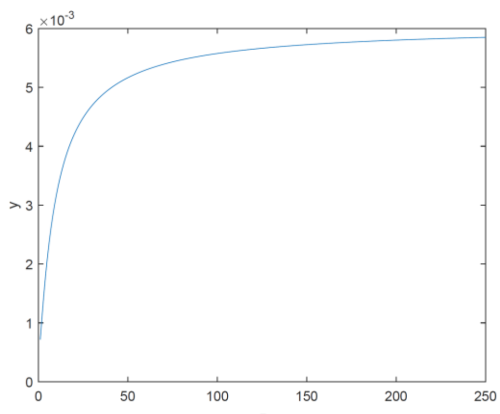

Figure 1: Concave school output with learning. $\sigma_\theta^2 = 0.15, \sigma_\varepsilon^2 = 0.5, I = 10, A = 1, B = 1.1$.

[3] Alternatively, if principals are sophisticated enough to know that their output curve is S-shaped in this manner, high (and perhaps low) experienced principals would choose not to exert effort to learn, as gains in test scores are expected to be low. Forward-looking low experience principals may be willing to experiment now for potential gains later in their careers, but this type of dynamic model is beyond the scope of this model.

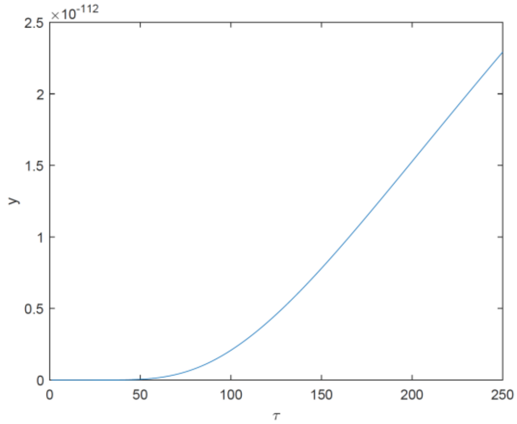

Figure 2: Convex school output with learning. $\sigma_\theta^2 = 0.15, \sigma_\varepsilon^2 = 0.5, I = 500, A = 1, B = 1.1.$

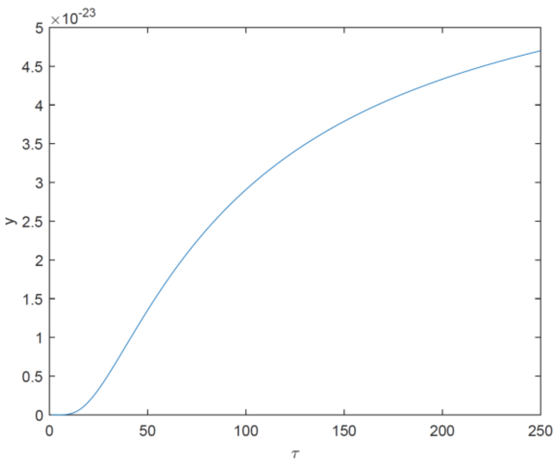

Figure 3: S-shaped school output with learning. $\sigma_\theta^2 = 0.15, \sigma_\varepsilon^2 = 0.5, I = 100, A = 1, B = 1.1.$