RISE Working Paper 14-017

"Estimating a Coordination Game in the Classroom"

by

Petra E. Todd, Kenneth I. Wolpin

Estimating a Coordination Game in the Classroom[*]

Petra E. Todd[†]   Kenneth I. Wolpin[‡]

April, 2012

Abstract

This paper develops and estimates a strategic model of the joint effort decisions of students and teachers in a classroom setting to better understand the reasons for the low mathematics performance of students on curriculum-based examinations administered in Mexican high schools. The model allows for student heterogeneity in preferences for knowledge and in initial mathematics preparation, and for teacher heterogeneity in preferences for student knowledge and in instructional ability. Survey data of students and teachers, collected as part of a randomized controlled experiment (the ALI project), include multiple measures of student and teacher effort, student and teacher preferences, student initial knowledge and teacher ability, all of which are treated as latent variables with an underlying factor structure. A simulation-based maximum likelihood estimation procedure is used to recover the parameters of the knowledge production function and the parameters pertaining to the latent variables and measurement structure. Estimation results, based on a sample of 10th grade students, indicate that the most significant factor accounting for low mathematics performance is the lack of sufficient prior preparation and not the lack of student or teacher effort. Our results suggest a mismatch between the content of the curriculum and student prior preparation.

1   Introduction

There is concern in many countries that students are underachieving. This is particularly true in Mexico, where, based on both international and national measures, the performance of students is poor. For example, Mexico ranked last among the 34 OECD countries in the 2009 PISA (Program for International Student Assessment) examination in mathematics.[1] Similarly, only 9.2 percent of ninth grade students and 15.6 percent of 12th grade students scored at the proficient level or above on the 2008 ninth and twelfth grade national mathematics examination (ENLACE).[2]

There are many potential explanations for substandard student performance related to student effort, teacher effort, teacher preparation (subject matter knowledge, teaching methods), school-level physical resources (libraries, textbooks, computers), and the overall learning environment within the school (teacher morale, administrative leadership). Their relative importance has been debated, with initiatives focused on particular potential causes, e.g., redesigning teacher training programs, redesigning curricula, providing computers. An alternative approach to improving performance is to provide performance-based monetary rewards for students, teachers or both. This approach recognizes that the reasons for substandard performance may differ across schools in ways that are not transparent. Providing monetary incentives allows educational institutions (and each student and teacher) to implement policies most suitable to their own circumstance.[3] In that vein, the Mexican Ministry of Education conducted a pilot program called Aligning Learning Incentives (ALI), beginning in 2008 and extending over a three year period, in which 88 Federal high schools took part in a randomized experiment to determine the impact of student and teacher incentive payments on mathematics performance.

Briefly, the ALI experiment included three treatments randomly assigned to 20 schools each, consisting of bonus payments determined by performance on end-of-year curricula-based examinations in mathematics paid either to students alone (T1), to teachers alone (T2) or to both students

---

[1]PISA assessments, begun under the auspices of the OECD in 2000, are administered in reading, mathematics and science to 15 year olds. In 2009, 65 nations and territories participated.

[2]The percentage of students scoring at the proficient level (or above) on the ninth grade test increased to 15.8 in 2011 and 20.6 in 2012 and on the 12th grade test to 24.7 and 30.8. The high school graduation rate in Mexico is about 35 percent.

[3]Hanushek (1994) discusses the value of performance incentives as a decentralized mechanism for improving school efficiency.

and teachers (T3).[4] Behrman et. al.'s (2011) analysis of the experimental data found that incentive payments to teachers were mostly ineffective in increasing test score performance (T2) except when combined with student incentive payments (T3), and that student incentives alone were effective (T1), but were even more effective when combined with teacher incentives (T3).[5] For example, for the 10th grade in year 3, the effect for T1 is between .3-.4 of a test score standard deviation and for T3 between .6-.8 of a standard deviation.[6] Measured in standard deviation units, these treatment effects are large relative to the range of estimates reported in the experimental incentives literature. However, as measured by raw scores, the performance of the treatment groups is less striking. Students in T3 (T1) answer only 45 (42) percent of the questions correctly as compared to 38 percent for the controls.[7]

The first order question raised by these results is why these Mexican high school students, even with large monetary incentives, master less than 50 percent of the curriculum. The goal of this paper is to provide a framework within which to assess quantitatively the reasons for this level of performance. For that purpose, we develop and estimate a model of a classroom in which there are multiple students and a single teacher making effort decisions that affect student performance. Students begin the academic year with an initial knowledge level, which, in combination with their effort and their teacher's effort during the year as well as their teacher's instructional ability, produces end-of-year knowledge. We show that this specification of the cognitive achievement production function can be derived from a cumulative specification (see Todd and Wolpin (2003)) in which current knowledge depends on all past inputs and on initial ability. Students have preferences over end-of-year knowledge, while teachers care about the sum of the end-of-year knowledge levels of their students. The technology is such that student and teacher effort are productive only above some minimum effort level, where there are both fixed and variable costs of supplying effort above

---

[4]The high schools in the experiment are administered by the federal government and account for about 25 percent of all high school students. The schools selected for the experiment are more rural than federal high schools in general. Most of the non-federal high schools are administered by the separate states.

[5]Incentive payments to students were as large as 1,500 U.S. dollars and a teacher could earn an additional month of salary or more.

[6]Unlike all other grade-year combinations, there is a small effect of teacher incentives (T2) of about .1 sd.

[7]These results are based on a slightly different sample than in Behrman et. al., namely those students who took the ALI exam and, in addition, completed a student survey. The estimated treatment effects are based on the same method adopted for correcting for cheating and are of similar magnitude as in that paper.

minimum levels.

It is assumed that teacher effort is a pure public input and that student and teacher effort are complementary in producing student knowledge. Thus, all students in the class benefit from an increase in any one student's effort through the induced increase in teacher effort. Student initial knowledge and teacher instructional ability augment the marginal products of student and teacher effort. Student and teacher effort are assumed to be chosen within a Nash game. As in coordination games more generally, there are potentially multiple equilibria.[8] If there is no fixed cost of supplying effort, then there can be two equilibria, one in which all students and the teacher supply above-minimum effort and one in which they all supply minimum effort. With student fixed costs, however, there are up to 2^N equilibria, where N is the class size. This makes it computationally infeasible to determine the full set of equilibria, which requires checking whether each potential equilibrium is defection-proof. However, we show that the number of potential equilibria can be greatly reduced under an assumption that the ratio of the fixed-to-variable cost does not vary among students within a class.[9] In that case, students can be ordered in terms of their propensity to choose minimum effort and there are at most $N + 1$ equilibria that need to be checked, with different equilibria corresponding to different numbers of students supplying minimum effort. We describe later in the paper an algorithm for determining the full set of equilibria.

There are only a few previous studies that develop explicit models of teacher or student effort choices and, to our knowledge, none that implement a model of both student and teacher effort choices. Duflo, Dupas and Kremer (2008) develop a model in which teachers choose effort levels and a target level at which to orient their instruction, taking into account their students' previous performance levels. They test the model's implications using data from a tracking experiment in Kenya that randomly assigned some schools to a treatment where classroom assignment depended on prior performance. They find that teacher effort, measured by teacher attendance, is higher under the tracking regime and that both high and low ability students benefit from tracking in terms of performance. In another study, Duflo, Hanna and Ryan (2012) develop a dynamic model of teacher attendance in India to study compensation schemes and implications for student performance using data from a randomized experiment.[10] They find that financial incentives increase

---

[8] See Vives (2005) for a discussion of games with strategic complementarities.

[9] The ratio of fixed-to-variable cost may vary over classes within the same school and over schools.

[10] This paper builds on a growing literature that combines structural estimation with randomized control experi-

teacher attendance and improve student test scores. These models do not, however, incorporate student effort. Kremer, Miguel and Thorton (2009) reference, but do not explicitly develop, a model of strategic effort complementarities to interpret the results of a randomized merit scholarship program in which Kenyan girls received school fees and a grant depending academic exam performance. Although the incentives were provided only to high performing girls, they find that girls with low pre-test scores, who were unlikely to win the tournament, also showed improvement in performance as did boys. Also, teachers in the schools assigned to the scholarship program had higher attendance. They note that these results are consistent with there being positive classroom externalities to study effort and potentially a strategic complementary between student and teacher effort. The model developed and estimated in this paper can help explain the pattern of experimental results they find.[11]

This paper also contributes to the empirical literature on the estimation of models with strategic complementarities. Examples include models of the adoption by banks of the automated clearing house (ACH) system in Ackerberg and Gowrisankaran (2006), the timing of desertions during the Civil War in De Paula (2009) and the timing of radio commercials in Sweeting (2009). In contrast to these applications, in the model we estimate the objects of choice (student and teacher effort levels) are continuous rather than discrete and we assume complete information.

The data we use to estimate the model come from surveys of control-school students and teachers combined with test score data on curricula-based mathematics examinations administered under the ALI project. We develop a simulated maximum likelihood estimation procedure that uses multiple measures of effort and multiple measures and exogenous determinants of the model primitives, that is, student initial knowledge, teacher instructional ability and student and teacher preferences for knowledge.[12] A probabilistic equilibrium selection rule that depends on equilibrium

---

ments (see Todd and Wolpin (2006) and the citations within).

[11] A related literature develops models in which peer group norms influence individuals' educational investment choices. (e.g. Fryer, Austen-Smith, 2005, Brock and Durlauf, 2001). Lazear (2001) considers a model of educational production in which one disruptive student imposes negative spillovers on other students in the class and he uses the model to study implications for optimal class size. In our model, though, spillover effects on peers arise only indirectly through teacher effort choices.

[12] This econometric framework has antecedents in the MIMIC (multiple-indicator multiple-cause) framework (see, for example, Joreskog and Goldberger (1975)). For recent applications and extensions, see Cunha and Heckman (2008) and Cunha, Heckman and Schennach (2010).

characteristics is posited and estimated, jointly with the other parameters of the model.[13] In addition, the estimation incorporates unobserved heterogeneity in productivity across schools.

The model is estimated for 10th grade students in the control group in the third (and last) year of the ALI project, for which extensive data on measures of the model primitives are available.[14] The ALI test covers the 10th grade curriculum which includes algebra, geometry and trigonometry.[15] One finding from the estimation is that a student's end-of-year knowledge, as measured by a standardized score, is strongly affected by a student's place in the initial knowledge distribution. Production function estimates for end-of-year knowledge show that, holding student and teacher effort and teacher ability at their mean values, a student with initial knowledge two standard deviations above the mean will have end-of-year knowledge that is 1.6 standard deviations above the mean.[16] A second finding is that the marginal productivity of both student and teacher effort is low. A student whose effort is two standard deviations above the mean (equivalent to an increase of 1.7 hours per week spent studying math), holding initial knowledge, teacher effort and ability at their mean values, will have a level of end-of-year knowledge that is only .10 standard deviations above the mean.[17] Similarly, a teacher whose effort is two standard deviations above the mean (equivalent to an increase of 1.4 hours per week spent in class preparation), holding student initial knowledge, student effort and teacher ability at their mean values, will induce a student's end-of-year knowledge to be only .05 standard deviations above the mean.[18]

Although relative performance is strongly affected by initial knowledge, absolute performance as measured by raw scores is less affected.[19] The percentage of questions answered correctly given

---

[13]For other applications of this approach to equilibrium selection rules, see Ackerberg and Gowrisankaran (2006), Bjorn and Vuong (1984), Jia (2008), Bajari, Hong and Ryan (2010) and Card and Giuliano (forthcoming). See DePaula (2012) for a discussion and survey.

[14]Estimation on the T1 and T3 treatment groups is not feasible because the non-linearity of the student incentive schedule does not allow for a reduction in the size of the potential set of equilinbria.

[15]The 11th grade curriculum includes analytical geometry and differential calculus. The 12th grade curriculum includes probability and statistics and, in the third year of the program, integral calculus.

[16]The standard deviation of initial knowledge is estimated to be equivalent to 44.9 standardized points on the 9th grade mathematics ENLACE, which has a mean of 500 and a standard deviation of 100.

[17]A two standard deviation increase in student effort is equivalent to spending an extra 1.7 hours per week studying math outside of class, double the mean level of (above-minimum) effort.

[18]A two standard deviation increase in teacher effort is equivalent to spending an extra 1.4 hours per week in class preparation, approximately a 35 percent increase over the mean level of (above-minimum) effort.

[19]There were 79 questions on the test.

initial knowledge equal to two standard deviations above the mean (holding student and teacher effort and teacher ability at their mean values) is only 10 percentage points higher than the percentage correct given initial knowledge equal to the mean. This result is not altered when we account for effort changes induced by greater initial knowledge; a counterfactual experiment in which all students have an initial knowledge level at least two standard deviations above the mean would lead to an average raw score of only 47.6 percent. It appears that, given the technology, the curriculum is simply too difficult even for students with initial knowledge at the top of the distribution.[20]

The paper proceeds as follows. Section two presents the model, section three describes the estimation procedure, section four discusses the data and section five presents the main empirical results. The last section presents the conclusions and discusses the relevance of the control-group estimation for explaining the ALI treatment results.[21]

2 The Model

This section presents a model of the production of student knowledge within a classroom setting. End-of-year student knowledge depends on the student's initial level of knowledge and effort and on the teacher's instructional ability and effort. Student and teacher effort levels are assumed to be the outcome of a Nash game. The model primitives are teacher ability, teacher preferences, initial levels of student knowledge, student preferences, and fixed and variable costs of effort for students and teachers.

---

[20]Pritchett and Beaty (2012) argue that the slow pace of learning evident in a number of developing countries is the result of reliance on overly ambitious curricula.

[21]It is not possible to use the control-group estimates to forecast directly the effect of the ALI treatments because the model does not identify the value that students and teachers attach to money. It is conceptually possible to use the treatment groups, specifically T1 to identify the value of money for students and T2 for teachers, and then forecast the effect of T3 in which both students and teachers receive performance bonuses. It is not, however, feasible ro implement this strategy for computational reasons. In the case of T2, the teacher reaction function is non-linear and, in addition to that, in the case of T1 the invariance property that reduces the potential number of equilibria does not hold.

2.1 Structure

Consider a class, denoted by j, with a single teacher and N_j students. Each student, n, begins with an initial level of knowledge, K_{0nj}, $n = 1, ..., N_j$ and chooses a level of learning effort, ε_{nj}. The teacher has instructional ability a_t and chooses instructional effort ε_{tj}, a pure public input (the same for each student). Student and teacher effort augment knowledge only if the levels exceed some minimum threshold level, denoted as $\underline{\varepsilon}_s$ and $\underline{\varepsilon}_t$. Letting $\widetilde{\varepsilon}_n = \varepsilon_n - \underline{\varepsilon}_s$ and $\widetilde{\varepsilon}_t = \varepsilon_t - \underline{\varepsilon}_t$ be student and teacher levels of above-minimum effort, end-of-year knowledge for student n, K_n, (dropping the j subscript) is produced according to

$$K_n = \delta K_{0n} \cdot (1 + \kappa a_t^{\gamma_0} \widetilde{\varepsilon}_n^{\gamma_1} \widetilde{\varepsilon}_t^{\gamma_2}), \tag{1}$$

where $\delta K_{0n} \geq 0$ is the level of knowledge achieved if either the student or teacher chooses minimum effort ($\widetilde{\varepsilon}_n = 0$ or $\widetilde{\varepsilon}_t = 0$) and κ is a normalization that converts units. The Cobb-Douglas component of (1) represents the proportionate increase in knowledge due to student and teacher effort over that level produced with minimum student or teacher effort.[22]

In the value-added specification given in (1), K_{0n}, knowledge in the previous grade, is taken to be a sufficient statistic for initial ability and for all prior inputs of student and teacher effort. A nice feature of our specification is that it is consistent with a cumulative specification in which current knowledge depends on all past inputs and initial ability.[23] In particular, letting g denote grade level, the knowledge produced in grade g is

$$K_{gn} = \omega_g A_{0n} \cdot \prod_{l=1}^{g} (1 + \kappa a_{tg}^{\gamma_0} \varepsilon_{ng}^{\gamma_1} \varepsilon_{tg}^{\gamma_2}),$$

where A_{0n} is the students pre-school ability and ω_g is one minus the depreciation rate of initial ability. Note that $\omega_g A_{0n}$ is the level of knowledge a student would have at the end of grade g if the student were to supply only minimum effort in every grade. Dividing K_{gn} by $K_{g-1,n}$ leads to the value-added specification given in (1) with $\delta = \omega_g / \omega_{g-1}$.

Each student faces a variable cost of effort, c_n, and a fixed (start-up) effort cost, g_n.[24] Students

---

[22]If, for example, $\kappa a_t^{\gamma_0} \widetilde{\varepsilon}_n^{\gamma_1} \widetilde{\varepsilon}_t^{\gamma_2} = .5$, K_n would be 50 percent greater than the level of knowledge produced with minimum effort.

[23]Value-added specifications are often adopted in the literature to circumevent the need for histoical data on inputs and initial abilty.

[24]There are two reason for introducing the student fixed cost. First, the existence of a fixed cost may rationalize

maximize their utility from knowledge net of effort cost:

$$U_n(\widetilde{\varepsilon}_n) = \theta_n K_n - \frac{c_n}{2}(\widetilde{\varepsilon}_n)^2 - g_n I(\widetilde{\varepsilon}_n > 0), \tag{2}$$

where θ_n is the (constant) marginal utility of knowledge and $I(\cdot)$ is an indicator function that is one if the argument is positive and zero otherwise. Students supply above-minimum effort, $\widetilde{\varepsilon}_n > 0$, if and only if

$$U_n(\widetilde{\varepsilon}_n) > U_n(0) = \theta_n \delta K_{0n}$$

and minimum effort otherwise.

The teacher is assumed to care about the total amount of knowledge produced in the class. Given a variable effort cost, c_t, and a fixed cost, g_t, the teacher maximizes[25]

$$U_t(\widetilde{\varepsilon}_t) = \theta_t \sum_{n=1}^{N} K_n - \frac{c_t}{2}(\widetilde{\varepsilon}_t)^2 - g_t I(\widetilde{\varepsilon}_t > 0). \tag{3}$$

Similar to the students, the teacher supplies above minimum effort, $\widetilde{\varepsilon}_t > 0$, if and only if

$$U_t(\widetilde{\varepsilon}_t) > U_t(0) = \theta_t \delta \sum_{n=1}^{N} K_{0n} \tag{4}$$

and minimum effort otherwise.

Assuming that all student primitives, $K_{0n}, \theta_n, c_n, g_n$ for all $n = 1, ...N$ and teacher primitives, a_t, θ_t, c_t and g_t, are public information, the reaction functions for the Nash equilibrium game are:

$$
\begin{aligned}
\widetilde{\varepsilon}_n &= (\delta K_{0n})^{\frac{1}{2-\gamma_1}} (\gamma_1 \kappa a_t^{\gamma_0} \theta_n c_n^{-1})^{\frac{1}{2-\gamma_1}} \widetilde{\varepsilon}_t^{\frac{\gamma_2}{2-\gamma_1}} \text{ if } U_n(\widetilde{\varepsilon}_n) > U_n(0) \\
&= 0 \text{ if } U_n(\widetilde{\varepsilon}_n) \leq U_n(0)
\end{aligned}
\tag{5}
$$

---

the ALI experimental results, in particular the perhaps puzzling finding that there is an effect of teacher incentives only when combined with student incentives. If a substantial number of students are supplying only minimum effort because of the fixed cost, teacher incentives may induce increased teacher effort only if students supply above minimum effort, which can be achieved by paying students for performance. Second, and only anecdotally, teachers in the T2 treatment were unenthusiastic when informed of the teacher bonus, because they felt that performance pay for them would do nothing to overcome the major obstacle, namely how to motivate the students. On the other hand, the teachers in the T1 treatment were enthusiastic, even though they were not provided monetary incentives, precisely because students received incentives.

[25] We allow for a fixed cost for the teacher in the presentation of the model for completeness, although we assume it to be zero in the implementation.

$$\widetilde{\varepsilon}_t \;=\; \left(\gamma_2 \kappa a_t^{\gamma_0} \theta_t c_t^{-1}\right)^{\frac{1}{2-\gamma_2}} \left(\sum_{n=1}^{N} I(\widetilde{\varepsilon}_n > 0)(\delta K_{0n})\widetilde{\varepsilon}_n^{\gamma_1}\right)^{\frac{1}{2-\gamma_2}} \quad \text{if } U_t(\widetilde{\varepsilon}_t) > U_t(0) \tag{6}$$

$$\;=\; 0 \text{ if } U_t(\widetilde{\varepsilon}_t) \leq U_t(0)$$

for $n = 1, ..., N$.[26] As seen from (6), only the effort levels of students who supply above-minimum effort affect the teacher effort level. Given the technology (1), the marginal product of teacher effort is zero for students who put in only minimum effort.

If student n and the teacher both supply above minimum effort, then the unique solution to (5) and (6) has a closed form given by:

$$
\begin{aligned}
\widetilde{\varepsilon}_n^* \;=\;& \gamma_1^{\frac{2-\gamma_2}{4-2(\gamma_1+\gamma_2)}} \gamma_2^{\frac{\gamma_2}{4-2(\gamma_1+\gamma_2)}} \kappa^{\frac{2}{4-2(\gamma_1+\gamma_2)}} a_t^{\frac{2\gamma_0}{4-2(\gamma_1+\gamma_2)}} \left(\theta_t c_t^{-1}\right)^{\frac{\gamma_2}{4-2(\gamma_1+\gamma_2)}} \left(\delta K_{0n}\right)^{\frac{1}{2-\gamma_1}} \left(\theta_n c_n^{-1}\right)^{\frac{1}{2-\gamma_1}} \times \\
& \left(\sum_{n=1}^{N} I(\widetilde{\varepsilon}_n > 0)(\delta K_{0n})^{\frac{2}{2-\gamma_1}} \left(\theta_n c_n^{-1}\right)^{\frac{\gamma_1}{2-\gamma_1}}\right)^{\frac{\gamma_2}{4-2(\gamma_1+\gamma_2)}}
\end{aligned}
\tag{7}
$$

$$
\begin{aligned}
\widetilde{\varepsilon}_t^* \;=\;& \left(\gamma_1\right)^{\frac{\gamma_1}{4-2(\gamma_1+\gamma_2)}} \left(\gamma_2\right)^{\frac{2-\gamma_1}{4-2(\gamma_1+\gamma_2)}} \kappa^{\frac{2}{4-2(\gamma_1+\gamma_2)}} a_t^{\frac{2\gamma_0}{4-2(\gamma_1+\gamma_2)}} \left(\theta_t c_t^{-1}\right)^{\frac{2-\gamma_1}{4-2(\gamma_1+\gamma_2)}} \times \\
& \left(\sum_{n=1}^{N} I(\widetilde{\varepsilon}_n > 0)(\delta K_{0n})^{\frac{2}{2-\gamma_1}} \left(\theta_n c_n^{-1}\right)^{\frac{\gamma_1}{2-\gamma_1}}\right)^{\frac{2-\gamma_1}{4-2(\gamma_1+\gamma_2)}} .
\end{aligned}
\tag{8}
$$

As seen, a student's effort depends not only on own attributes, but also, through the teacher's effort decision, on the attributes of the other students in the class.[27]

2.2 Equilibrium Characterization

As noted in the introduction, this model can have multiple equilibria. The equilibrium characterization depends on the configuration of student and teacher fixed costs as follows.

1. No student or teacher fixed cost: $g_n = 0$ for all n, $g_t = 0$

If there are no fixed costs, then equations (7) and (8) constitute an equilibrium with $I(\widetilde{\varepsilon}_n > 0) = 1$ for all $n = 1, ..., N$. That is, all students and teachers put in above-minimum effort. The

---

[26] The perfect information assumption would seem to be a reasonable approximation in the classroom setting, where students and teachers interact on a daily basis over the school year.

[27] In this sense, there are student peer effects on end-of-year knowledge, although they all operate through the teacher's effort decision. Introducing direct peer effects in the knowledge production function, although desirable, is beyond the scope of the current paper.

equilibrium is not unique, because there is also an equilibrium in which all students and the teacher choose minimum effort. Given the production function (1), if all students and the teacher choose minimum effort ($\widetilde{\varepsilon}_n = 0$ for all n and $\widetilde{\varepsilon}_t = 0$,) there would be no incentive for any single student (or even all students as a group) or for the teacher to deviate as the marginal product of effort for any student or for the teacher is zero. These are the only two equilibria.

2. Positive fixed cost for teacher, zero fixed cost for all students: $g_t > 0$, $g_n = 0$ *for all* n

Depending on the size of the teacher fixed cost, there will be either one or two equilibria. If the fixed cost is such that the teacher chooses minimum effort, then all students choose minimum effort and that is the only equilibrium. If, on the other hand, the teacher chooses above-minimum effort, then there is an additional equilibrium in which all students choose above-minimum effort.

3. Positive student fixed cost: $g_n > 0, g_t \geq 0$

The configuration of potential equilibria differs substantially when there is a positive fixed cost for students, $g_n > 0$. In the case that the teacher supplies minimum effort, so will all of the students. Alternatively, assume that the teacher supplies above-minimum effort, i.e., that at least one student supplies above-minimum effort and the teacher's fixed cost is not large enough for (4) not to be satisfied. In that case, the number of potential equilibria, defined by the number of students supplying minimum effort, is 2^N. Determining the set of equilibria that arise for a given composition of students and teacher within a class (defined by $K_{0n}, \theta_n, c_n, g_n, \theta_t, c_t, a_t$) requires checking, for each of the 2^N possible effort configurations, whether any of the N students or the teacher would deviate, given the effort choice of all of the other students and the teacher.

As noted, any configuration in which the teacher chooses minimum effort and any student chooses above-minimum effort cannot be an equilibrium. Thus, we restrict attention to configurations in which the teacher supplies above-minimum effort. All such configurations are fully described by $I(\widetilde{\varepsilon}_n > 0) \times \widetilde{\varepsilon}_n^*$ for all $n = 1, .., N$ (that is, either $\widetilde{\varepsilon}_n = 0$ or $\widetilde{\varepsilon}_n = \widetilde{\varepsilon}_n^*$) and by $\widetilde{\varepsilon}_t^*$. To determine whether any particular configuration is an equilibrium, we need to check whether the teacher would deviate, that is choose minimum effort, and whether any student will deviate, that is, whether any student assigned minimum effort would prefer above-minimum effort and any student assigned above-minimum effort would choose minimum effort. Without any restrictions on

the parameter space, all 2^N configurations would need to be checked to determine the equilibrium set.

However, it is possible to derive sufficient conditions under which the number of equilibria is reduced to at most $N + 1$ and, thus, only that number of configurations would need to be checked. Recall that the minimum effort condition for the student is

$$\widetilde{\varepsilon}_n \quad > \quad 0 \text{ if } \widetilde{\lambda}_n = U_n(\widetilde{\varepsilon}_n^*) - U_n(0) > 0, \tag{9}$$

$$\widetilde{\varepsilon}_n \quad = \quad 0 \text{ if } \widetilde{\lambda}_n = U_n(\widetilde{\varepsilon}_n^*) - U_n(0) \leq 0. \tag{10}$$

where $\widetilde{\lambda}_n$ is evaluated at above-minimum effort for the teacher ($\widetilde{\varepsilon}_t = \widetilde{\varepsilon}_t^*$). Upon substituting (7), this condition can be written as

$$\widetilde{\lambda}_n(E) = \Psi(\delta K_{0n}\theta_n c_n^{-1}; a_t, \theta_t c_t^{-1}, \kappa, \gamma_1, \gamma_2) Z_E^{\frac{\gamma_2}{2-(\gamma_1+\gamma_2)}} - g_n c_n^{-1} \lessgtr 0, \tag{11}$$

where

$$\Psi(\cdot) = (\delta K_{0n}\theta_n c_n^{-1})^{\frac{2}{2-\gamma_1}} \kappa^{\frac{2}{4-2(\gamma_1+\gamma_2)}} a_t^{\frac{2\gamma_0}{2-(\gamma_1+\gamma_2)}} \gamma_1^{\frac{\gamma_1}{2-(\gamma_1+\gamma_2)}} (\gamma_2)^{\frac{\gamma_2}{2-(\gamma_1+\gamma_2)}} (\theta_t c_t^{-1})^{\frac{\gamma_2}{2-(\gamma_1+\gamma_2)}} (1 - .5\gamma_1),$$

$$Z_E = \sum_{n=1}^{N} I(\widetilde{\varepsilon}_n > 0)(\delta K_{0n})^{\frac{2}{2-\gamma_1}} (\theta_n c_n^{-1})^{\frac{\gamma_1}{2-\gamma_1}}$$

and where E corresponds to a potential equilibrium with a particular configuration of students supplying minimum effort.

As seen in (11), the value of $\widetilde{\lambda}_n$ depends on the particular configuration E only through Z_E. Note that $\Psi(\cdot) > 0$ (assuming $\gamma_1 < 2$) and $Z_E^{\frac{\gamma_2}{2-(\gamma_1+\gamma_2)}} > 0$. Denote E_N as the configuration in which *all* of the N students are assigned above-minimum effort. Order the students according to their value of $\widetilde{\lambda}_n(E_N)$, with student of order one having the lowest value, student of order two having the next lowest, etc. In general, the ordering of students will not be the same in other potential equilibria. However, because Z_E is the same for all students and is a sufficient statistic for any given potential equilibrium configuration, there exist conditions under which the ordering of students is invariant under all potential configurations. A sufficient condition is that the ratio of the student's fixed to variable cost $(g_n c_n^{-1})$ is the same for all students. The ordering in that case is fully determined by the value of $\delta K_{0n}\theta_n c_n^{-1}$. A necessary and sufficient condition for the order invariance in the case that $g_n c_n^{-1}$ varies in the class is that $g_n c_n^{-1}$ be inversely related to $\delta K_{0n}\theta_n c_n^{-1}$.

In both cases, a student's order among the N students in the class corresponds to their order in the sequence of $\widetilde{\lambda}(E_N)'s$, denoted as $\widetilde{\lambda}_1 \leq \widetilde{\lambda}_2 \leq ... \leq \widetilde{\lambda}_N$.

Order invariance implies that instead of 2^N potential equilibria, there are only $N + 1$; namely where none of the students choose above-minimum effort (E_0), only the highest order student chooses above-minimum effort (E_1), only the two highest choose above-minimum effort (E_2), etc. We can denote the potential set of equilibria to be checked as $\{E_0, E_{1,...}, E_N\}$, where the subscript indicates the number of ordered students with above-minimum effort.[28] In the rest of the discussion, order invariance is assumed.

2.3 Determining the set of equilibria

Consider the E_N configuration where all students exert above minimum effort. For that to be an equilibrium, none of the students nor the teacher will want to defect, that is, choose minimum effort. It is straightforward to check condition (4) to see whether the teacher wants to defect. If the teacher chooses above minimum effort, then, given the ordering, we now need to check only whether the student with the lowest value of $\widetilde{\lambda}_n$, the student with value $\widetilde{\lambda}_1$, would choose to defect. If not $(\widetilde{\lambda}_1 > 0)$, then no student with a higher value will defect, and E_N constitutes an equilibrium. If the lowest value student does defect, then E_N does not constitute an equilibrium.

Even if E_N is an equilibrium, there may be other equilibria in which some students choose minimum effort. Consider the candidate equilibrium E_{N-1} in which the lowest $\widetilde{\lambda}_n$ value student chooses minimum effort, $\widetilde{\varepsilon}_1 = 0$, and all other students choose above minimum effort, $\widetilde{\varepsilon}_n = \widetilde{\varepsilon}_n^* > 0$ for all $n > 1$. Assume that the teacher also optimally supplies above minimum effort. In that case student and teacher effort satisfy

$$\widetilde{\varepsilon}_1 = 0 \tag{12}$$

$$\widetilde{\varepsilon}_n = \widetilde{\varepsilon}_n^* = \gamma_1^{\frac{2-\gamma_2}{4-2(\gamma_1+\gamma_2)}} \gamma_2^{\frac{\gamma_2}{4-2(\gamma_1+\gamma_2)}} \kappa^{\frac{2}{4-2(\gamma_1+\gamma_2)}} a_t^{\frac{2\gamma_0}{4-2(\gamma_1+\gamma_2)}} (\theta_t c_t^{-1})^{\frac{\gamma_2}{4-2(\gamma_1+\gamma_2)}} (\delta K_{0n})^{\frac{1}{2-\gamma_1}} (\theta_n c_n^{-1})^{\frac{1}{2-\gamma_1}} \times$$

$$\left(\sum_{n=2}^{N} (\delta K_{0n})^{\frac{2}{2-\gamma_1}} (\theta_n c_n^{-1})^{\frac{\gamma_1}{2-\gamma_1}} \right)^{\frac{\gamma_2}{4-2(\gamma_1+\gamma_2)}} , n > 1$$

---

[28] The reason that no other configurations can be equilibria is that no student with a lower order would choose above-minimum effort without all of the higher order students also choosing above-minimum effort. A lower order student in a configuration in which such a student was assigned above-minimum effort and a higher order student was assigned minimum effort would always defect to minimum effort.

$$\widetilde{\varepsilon}_t^* = (\gamma_1)^{\frac{\gamma_1}{4-2(\gamma_1+\gamma_2)}}(\gamma_2)^{\frac{2-\gamma_1}{4-2(\gamma_1+\gamma_2)}}\kappa^{\frac{2}{4-2(\gamma_1+\gamma_2)}}a_t^{\frac{2\gamma_0}{4-2(\gamma_1+\gamma_2)}}(\theta_t c_t^{-1})^{\frac{2-\gamma_1}{4-2(\gamma_1+\gamma_2)}} \times$$
$$\left(\sum_{n=2}^{N}(\delta K_{0n})(\theta_n c_n^{-1})^{\frac{\gamma_1}{2-\gamma_1}}\right)^{\frac{2-\gamma_1}{4-2(\gamma_1+\gamma_2)}}. \tag{13}$$

The teacher, in configuration E_{N-1}, would supply less effort than in E_N, because the teacher cannot affect the effort level of student $n = 1$. Thus, each student of order $n > 1$ would also supply less effort. For the E_{N-1} configuration to be an equilibrium, it is necessary that the teacher and no student want to defect. Assume that the teacher does not want to defect. In terms of the students, we need to check the minimum effort condition only for the lowest order student $(n = 1)$ and the second lowest student $(n = 2)$. For student $n = 1$, the optimal above-minimum effort level that would be chosen given the effort level of the teacher and the other students is given by

$$\widetilde{\varepsilon}_1^* = \gamma_1^{\frac{2-\gamma_2}{4-2(\gamma_1+\gamma_2)}}(\gamma_2)^{\frac{\gamma_2}{4-2(\gamma_1+\gamma_2)}}\kappa^{\frac{2}{4-2(\gamma_1+\gamma_2)}}a_t^{\frac{2\gamma_0}{4-2(\gamma_1+\gamma_2)}}(\theta_t c_t^{-1})^{\frac{\gamma_2}{4-2(\gamma_1+\gamma_2)}}(\delta K_{0n})^{\frac{1}{2-\gamma_1}}(\theta_1 c_1^{-1})^{\frac{1}{2-\gamma_1}} \times$$
$$\left(\sum_{n=2}^{N}(\delta K_{0n})^{\frac{2}{2-\gamma_1}}(\theta_n c_n^{-1})^{\frac{\gamma_1}{2-\gamma_1}}\right)^{\frac{\gamma_2}{4-2(\gamma_1+\gamma_2)}}$$

Then, E_{N-1} is an equilibrium if and only if

$$\widetilde{\lambda}_1(E_{N-1}) < 0, \tag{14}$$
$$\widetilde{\lambda}_2(E_{N-1}) > 0,$$

where the $\widetilde{\lambda}'s$ are given in (11) with $Z_E = \sum_{n=2}^{N}(\delta K_{0n})^{\frac{2}{2-\gamma_1}}(\theta_n c_n^{-1})^{\frac{\gamma_1}{2-\gamma_1}}$. Because of the ordering, if it is optimal for student $n = 2$ not to choose minimum effort, $\widetilde{\lambda}_2(E_{N-1}) > 0$, it will also be optimal for higher ordered students not to do so.

The reason that the E_N configuration, where all students and the teacher choose above-minimum effort, and the E_{N-1} configuration, where all students, except for the student with the lowest $\widetilde{\lambda}_n$, choose above-minimum effort, can both be equilibria is the following. Because teacher effort is higher in E_N, it does not pay for student $n = 1$ to choose minimum effort, whereas in E_{N-1}, the lower teacher effort induced by the minimum effort of student $n = 1$ makes that choice by student $n = 1$ optimal. On the other hand, student $n = 2$ (and higher order students) optimally chooses above minimum effort even at the lower level of teacher effort in E_{N-1}.

14

It is also possible that E_N is an equilibrium, but that E_{N-1} is not. For that to be the case, if must be either that student $n = 1$ prefers above-minimum effort even when teacher effort is lower (in E_{N-1}) that is, that $\widetilde{\lambda}_1(E_{N-1}) > 0$ or that not only does student $n = 1$ prefer minimum effort when teacher effort is lower $(\widetilde{\lambda}_1(E_{N-1}) < 0)$, but student $n = 2$ does also, that is, that $\widetilde{\lambda}_2(E_{N-1}) < 0$. It is also possible that E_N is not an equilibrium, but that E_{N-1} is an equilibrium. In that case, student $n = 1$ prefers minimum effort in both cases $(\widetilde{\lambda}_1(E_N) < 0$ and $\widetilde{\lambda}_1(E_{N-1}) < 0)$ and student $n = 2$ prefers above minimum effort in both cases $(\widetilde{\lambda}_2(E_N) > 0$ and $\widetilde{\lambda}_2(E_{N-1}) > 0)$. Finally, neither would be an equilibrium if student $n = 1$ prefers minimum effort in both cases $(\widetilde{\lambda}_1(E_N) < 0$ and $\widetilde{\lambda}_1(E_{N-1}) < 0)$ and student $n = 2$ prefers minimum effort at least in E_{N-1} $(\widetilde{\lambda}_2(E_{N-1}) < 1)$.

The same analysis can be repeated to check for other candidate equilibria in sequence, $E_{N-2}, ..., E_0$. By considering one potential equilibrium at a time, it is possible to identify the entire set of equilibria for any given model parameters. As already noted, E_0 is always an equilibrium. To generalize, the effort levels in the E_{N-m} candidate equilibrium are given by

$$
\widetilde{\varepsilon}_1 = 0, \widetilde{\varepsilon}_2 = 0, ..., \widetilde{\varepsilon}_m = 0 \tag{15}
$$

$$
\widetilde{\varepsilon}_n = \widetilde{\varepsilon}_n^* = (\gamma_1)^{\frac{2-\gamma_2}{4-2(\gamma_1+\gamma_2)}} (\gamma_2)^{\frac{\gamma_2}{4-2(\gamma_1+\gamma_2)}} \kappa^{\frac{2}{4-2(\gamma_1+\gamma_2)}} a_t^{\frac{2\gamma_0}{4-2(\gamma_1+\gamma_2)}} (\theta_t c_t^{-1})^{\frac{\gamma_2}{4-2(\gamma_1+\gamma_2)}} (\delta K_{0n})^{\frac{1}{2-\gamma_1}} (\theta_n c_n^{-1})^{\frac{1}{2-\gamma_1}} \times
$$
$$
\left(\sum_{n=m+1}^{N} (\delta K_{0n})^{\frac{2}{2-\gamma_1}} (\theta_n c_n^{-1})^{\frac{\gamma_1}{2-\gamma_1}} \right)^{\frac{\gamma_2}{4-2(\gamma_1+\gamma_2)}} , n > m,
$$

$$
\widetilde{\varepsilon}_t^* = (\gamma_1)^{\frac{\gamma_1}{4-2(\gamma_1+\gamma_2)}} (\gamma_2)^{\frac{2-\gamma_1}{4-2(\gamma_1+\gamma_2)}} \kappa^{\frac{2}{4-2(\gamma_1+\gamma_2)}} a_t^{\frac{2\gamma_0}{4-2(\gamma_1+\gamma_2)}} (\theta_t c_t^{-1})^{\frac{2-\gamma_1}{4-2(\gamma_1+\gamma_2)}} \times
$$
$$
\left(\sum_{n=m+1}^{N} (\delta K_{0n})^{\frac{2}{2-\gamma_1}} (\theta_n c_n^{-1})^{\frac{\gamma_1}{2-\gamma_1}} \right)^{\frac{2-\gamma_1}{4-2(\gamma_1+\gamma_2)}} . \tag{16}
$$

In determining whether E_{N-m} is an equilibrium, we need to check whether student $n = m$ would defect, that is, choose above-minimum effort, and, if not, whether student $n = m + 1$ would defect, that is, choose minimum effort. For the former, we have

$$
\widetilde{\varepsilon}_m^* = (\gamma_1)^{\frac{2-\gamma_2}{4-2(\gamma_1+\gamma_2)}} (\gamma_2)^{\frac{\gamma_2}{4-2(\gamma_1+\gamma_2)}} \kappa^{\frac{2}{4-2(\gamma_1+\gamma_2)}} a_t^{\frac{2\gamma_0}{4-2(\gamma_1+\gamma_2)}} (\theta_t c_t^{-1})^{\frac{\gamma_2}{4-2(\gamma_1+\gamma_2)}} (\delta K_{0n})^{\frac{1}{2-\gamma_1}} (\theta_m c_m^{-1})^{\frac{1}{2-\gamma_1}} \times
$$
$$
\left(\sum_{n=m+1}^{N} (\delta K_{0n})^{\frac{2}{2-\gamma_1}} (\theta_n c_n^{-1})^{\frac{\gamma_1}{2-\gamma_1}} \right)^{\frac{\gamma_2}{4-2(\gamma_1+\gamma_2)}} . \tag{17}
$$

Assuming the teacher would not defect, E_{N-m} is an equilibrium if and only if

$$
\widetilde{\lambda}_m(E_{N-m}) < 0, \tag{18}
$$
$$
\widetilde{\lambda}_{m+1}(E_{N-m}) > 0.
$$

15

where the $\widetilde{\lambda}'s$ are given in (11) with $Z_E = \sum_{n=m+1}^{N} (\delta K_{0n})^{\frac{2}{2-\gamma_1}} (\theta_n c_n^{-1})^{\frac{\gamma_1}{2-\gamma_1}}$.

The equilibrium in which all students and the teacher supply above-minimum effort Pareto dominates all of the others. To see why, consider two equilibria that differ by whether a single student, say student of order j, optimally supplies minimum effort or optimally supplies above-minimum effort. In the latter case, all of the students supplying above minimum effort, those of order $j+1$ and higher, are better off, because the response of the teacher is to supply more effort. All of the students of order lower than j receive the same utility and so are no worse off. Student j, who optimally supplies more effort in the second equilibrium, is also better off, because utility must be above the utility received with minimum effort in order that it be an equilibrium. Finally, the teacher is better off when student j supplies more effort.

3   Estimation

As previously described, order invariance of the students in terms of the utility difference between above-minimum and minimum effort greatly reduces the number of potential equilibria and makes it computationally feasible to determine the set of equilibria. We will therefore assume that the fixed-to-variable cost ratio $g_n c_n^{-1} = \widetilde{g}_n$ does not vary for students within the same class, which gives order invariance. Also because θ_n and c_n always appear as $\theta_n c_n^{-1}$ and θ_t and c_t as $\theta_t c_t^{-1}$ in the determination of student and teacher optimal effort levels, we define $\widetilde{\theta}_n = \theta_n c_n^{-1}$ and $\widetilde{\theta}_t = \theta_t c_t^{-1}$. Finally, because g_t and c_t appear as $g_t c_t^{-1}$ in the teacher's decision about whether to supply above-minimum effort, we let $\widetilde{g}_t = g_t c_t^{-1}$.

Heterogeneity among students within a class potentially arises from differences in initial knowledge and preferences, K_{0n} and $\widetilde{\theta}_n$. Heterogeneity among students across classes in the same school and across schools arises from class- or school-wide differences in $K_{0n}, \widetilde{\theta}_n$ and $\widetilde{g}_n$. Heterogeneity among teachers within the same school or among schools arises from differences in $\widetilde{\theta}_t$ and a_t. In what follows and in the estimation, we assume that $\widetilde{g}_t = 0$.[29] We also allow for technology differences across schools. Specifically, the production function parameter, δ, is assumed to be school-specific and randomly drawn (from our perspective).

---

[29] We assume that teachers are sufficiently monitored so that they never choose minimum effort (unless all students choose minimum effort).

3.1 Latent factor structure

Student and teacher characteristics as well as student and teacher effort and end-of-year knowledge can at best only be imperfectly measured. We therefore treat $K_{0n}, \widetilde{\theta}_n, \widetilde{\theta}_t, \widetilde{g}_n$ and a_t as latent factors that are measured with error. To be concrete, let beginning knowledge of a student n enrolled in school h and assigned to class j depend on a set of exogenous initial conditions (X) and on school-, class- and individual-level error components:

$$K_{onjh} = X_{njh}^{K_0}\beta^{K_0} + \xi_h^{K_0} + \mu_{jh}^{K_0} + \omega_{njh}^{K_0}. \tag{19}$$

The first error component, $\xi_h^{K_0}$, allows for unobserved school-level differences in student initial knowledge, the second, $\mu_{jh}^{K_0}$, for unobserved class-level differences within a school and the third, $\omega_{njh}^{K_0}$, for idiosyncratic within-class differences. Student preference and fixed costs follow a similar error structure

$$\widetilde{\theta}_{njh} = X_{njh}^{\theta_n}\beta^{\theta_n} + \xi_h^{\theta_n} + \mu_{jh}^{\theta_n} + \omega_{njh}^{\theta_n}, \tag{20}$$

and

$$\widetilde{g}_{njh} = \beta_0^{g_n} + \xi_h^{g_n} + \mu_{jh}^{g_n}. \tag{21}$$

Note that the order invariance assumption described in the previous section requires that there is no within-class error component nor varying X's in $\widetilde{g}_{njh}$. All error components in (19, (20) and (21) are assumed to be mean zero, orthogonal to each other and to observed characteristics.

With respect to teachers, instructional ability is parameterized as

$$a_{tjh} = X_{tjh}^{a_t}\beta^{a_t} + \xi_h^{a_t} + \mu_{th}^{a_t}, \tag{22}$$

where $\xi_h^{a_t}$ represents school-level differences in teacher instructional ability and $\mu_{th}^{a_t}$ idiosyncratic within-school differences.[30] Similarly, teacher preferences are

$$\widetilde{\theta}_{tjh} = X_{tjh}^{\theta_t}\beta^{\theta_t} + \xi_h^{\theta_t} + \mu_{th}^{\theta_t}, \tag{23}$$

where the school- and teacher-level error components are assumed to be orthogonal to each other and to observable characteristics.[31]

---

[30] A given teacher may have multiple classes, in which case, the teacher-specific error component would apply to all of the classes. Notice that the class designation j is superfluous given the teacher designation.

[31] All of the student and teacher latent factors are censored from below at zero.

Student and teacher effort are the outcomes of the effort game, which are fully determined by the latent primitives. End-of-year knowledge is determined by the primitives that enter the production function (1) and by the chosen student and teacher effort. We also treat student and teacher effort and end-of-year knowledge as latents that are measured with error, as described below.[32]

3.2 Measurement equations

In terms of the latent primitives, there are assumed to be M^j measures for $j = K_{0n}, \widetilde{\theta}_{njh} \ a_{tjh}, \widetilde{\theta}_{tjh}$. The measurement equations are given by

$$K_{0n}^m = \alpha_{0n}^{K_0 m} + \alpha_{1n}^{K_0 m} K_{on} + \varsigma_{njh}^{K_{om}}, m = 1, ..., M^{K_0} \tag{24}$$

$$\widetilde{\theta}_{njh}^m = \alpha_{0n}^{\theta m} + \alpha_{1n}^{\theta m} \widetilde{\theta}_{njh} + \varsigma_{njh}^{\theta m} \text{ for } m = 1, ..., M^{\theta_n}, \tag{25}$$

$$\widetilde{\theta}_{tjh}^m = \alpha_{0t}^{\theta m} + \alpha_{1t}^{\theta m} \widetilde{\theta}_{tjh} + \varsigma_{tjh}^{\theta m} \text{ for } m = 1, ...,, M^{\theta_t} \tag{26}$$

$$a_{tjh}^m = \alpha_{0t}^{a m} + \alpha_{1t}^{a m} a_{tjh} + \varsigma_{tjh}^{a m} \text{ for } m = 1, ..., M^{a_t}. \tag{27}$$

where the different measurements of each of the latent factors are denoted with an m superscript. We do not have measures of $\widetilde{g}_n$, so it is treated as a random unobservable.

There are M_n^ε measures of student effort and M_t^ε measures of teacher effort. The effort measurement equations take the form

$$\varepsilon_{njh}^m = \alpha_{0n}^{\varepsilon m} + \alpha_{1n}^{\varepsilon m} \varepsilon_{njh} + \varsigma_{njh}^{\varepsilon m} \text{ for } m = 1, ..., M_n^\varepsilon, \tag{28}$$

$$\varepsilon_{tjh}^m = \alpha_{0t}^{\varepsilon m} + \alpha_{1t}^{\varepsilon m} \varepsilon_{tjh} + \varsigma_{tjh}^{\varepsilon m} \text{ for } m = 1, ..., M_t^\varepsilon. \tag{29}$$

There is one measure of end-of-year knowledge, a test score T_{njh}:

$$T_{njh} = K_{njh} + \varsigma_{njh}^T, \tag{30}$$

where K_{njh} is determined by (1). We refer to the effort levels and end-of-year knowledge as endogenous latent factors.

An observation consists of (i) measures of the effort levels of the N_{jh} students in each class j of school h, their end-of-year test scores and measures of their initial knowledge and preferences and

---

[32]This methodology allows us to reduce the dimension of the effective input space to one for both the student and the teacher. See Cunha and Heckman (2008) for further discussion.

(ii) the measures of the effort level, preference and ability of the teacher in each class. Denote the observation set for class j in school h as

$$O_{jh} = \{\varepsilon_{njh}^m, \varepsilon_{tjh}^m, T_{njh}, K_{0njh}^m, \theta_{njh}^m, \theta_{tjh}^m, a_{tjh}^m\},$$

for all measures and classrooms.

3.3 Likelihood

Let $\Omega_{jh}^1 = \{X_{njh}^{K_0}, X_{njh}^{\theta_n}, X_{tjh}^{a_t}, X_{tjh}^{\theta_t}\}$ denote the vector of observable characteristics of the students and teacher in class j of school h, $\Omega_h^2 = \{\xi_h^{K_0}, \xi_h^{\theta_n}, \xi_h^{a_t}, \xi_h^{\theta_t}, \xi_h^{g_n}, \kappa_{0h}\}$ the vector of school-level unobservables, $\Omega_{jh}^3 = \{\mu_{jh}^{K_0}, \mu_{jh}^{\theta_n}, \mu_{jh}^{g_n}, \mu_{th}^{a_t}, \mu_{th}^{\theta_t}\}$ the vector of within-school class- and teacher-level unobservables, $\Omega_{jh}^4 = \{\omega_{njh}^{K_0}, \omega_{tjh}^{\theta_n}\}$ the vector of within-class student-level unobservables and $\Omega_{jh}^5 = \{\varsigma_{njh}^{K_{om}}, \varsigma_{njh}^T, \varsigma_{njh}^{\varepsilon m}, \varsigma_{tjh}^{\varepsilon m}, \varsigma_{njh}^{\theta m}, \varsigma_{tjh}^{\theta m}, \varsigma_{tjh}^{a m}\}$ the vector of measurement errors.

Estimation is carried out by simulated maximum likelihood. Concretely, let the unobservables in Ω_{jh}^i $(i = 2, 3, 4)$ each have joint distribution F_i with variance-covariance matrix Λ_i. Denote the joint distribution of the measurement errors as F_5 with variance-covariance matrix, Λ_5. The likelihood contribution for students $n = 1, .., N_{jh}$ in class j of school h is the joint density of O_{jh}, that is, the measured efforts of students and teachers, students' end-of-year (ALI) test scores, measured student preferences and measured teacher abilities and preferences. For now, we ignore the restrictions and normalizations that are necessary for identification.

The estimation procedure is as follows:

1. Choose a set of parameter values:

$\{\kappa, \gamma_0, \gamma_1, \gamma_2, \delta, \beta^{K_0}, \beta^{\theta_n}, \beta^{\theta_t}, \beta^{a_t}, \beta^{g_n}, \Lambda_2, \Lambda_3, \Lambda_4, \Lambda_5, \Gamma\}$,

where Γ denotes the α_0 and α_1 parameters in the measurement error equations.

2. Draw school-level shocks, Ω_h^2, for each school, $h = 1, ..., H$, class- and teacher-level shocks, Ω_{jh}^3, in each school and student-level shocks, Ω_{jh}^4, for each class.

3. Given the shocks drawn in (2) and the set of observable variables (X), calculate each student's value of K_{onjh}, $\widetilde{\theta}_{njh}$, and $\widetilde{g}_{njh}$, and each teacher's value of a_{tjh} and $\widetilde{\theta}_{tjh}$.[33]

4. Solve for the set of equilibria for each class and for all schools for each of the $d = 1, ...D$

---

[33] Recall that the characteristics of students and of the teacher within a given class, both those observable to the researcher (Ω_{jh}^1) and those unobservable to the researcher $(\Omega_h^2, \Omega_{jh}^3, \Omega_{jh}^4)$, are public information.

draws. Each equilibrium is characterized by the optimal student and teacher effort levels, and implied end-of-year knowledge.

5. For each equilibrium and for each draw (d), and given the joint measurement error distribution, calculate the joint likelihood of observing all of the measured variables for the students and the teacher, which is given by the joint density of measurement errors. Denote the density value for the ith (ordered) equilibrium and the dth draw for classroom j in school h as $\widehat{f}_{ijh}(d)$ for $d = 1, ..., D$.

6. We assume that the equilibrium that is chosen for each draw from the set of equilibria, $E(d)$, comes from a multinomial distribution.[34]

The overall likelihood for class j in school h is the weighted average over the likelihoods for each draw and for each potential equilibrium effort configuration, namely

$$L_{jh} = \frac{1}{D} \sum_d \sum_i I(E_{ijh} \in E_{jh}(d)) \pi_{ijh} \widehat{f}_{ijh}(d) \tag{31}$$

For each draw d, the set of equilibria may differ. Thus, the probability that a particular effort configuration, E_{ijh}, is the selected equilibrium is given by the product of an indicator for whether that configuration is an equilibrium and the probability of selecting that configuration from the set of equilibria, π_{ijh}. The parameters of the equilibrium selection rule, $\pi_i(\cdot)$, are jointly estimated with the rest of the parameters of the model.

7. Repeat for all J_h classes in school h and over all $h = 1, , , .H$ schools. The likelihood over the entire sample is

$$\prod_{h=1}^{H} \prod_{j=1}^{J_h} L_{jh} \tag{32}$$

8. Repeat 1-7, maximizing (32) over the parameter vector given in step 1.

3.4   Identification

It is difficult to formally demonstrate identification as the model is currently specified, because of the non-linearities in the production function and in the derived student and teacher effort functions. To simplify, we consider the case of a class with a single student and a production function that is strictly linear in logs (that is, Cobb-Douglas). In addition, we assume there are no fixed costs.

---

[34] As further described below, the probability that equilibrium $E_i \in E(d)$ is chosen, π_i, may be a function of the characteristics of E_i, for example, the fraction of students choosing minimum effort.

In this case, there are only two possible equilibria, one where the student and the teacher choose above-minimum effort and one in which they both choose minimum effort. To simplify further, suppose that the equilibrium selection rule is that the equilibrium with above-minimum effort, being Pareto dominant, is selected with probability one.

The production function in this example is

$$\log K_n = \log \kappa + \delta \log K_{0n} + \gamma_0 \log a_t + \gamma_1 \log \widetilde{\varepsilon}_n + \gamma_2 \log \widetilde{\varepsilon}_t, \tag{33}$$

where now $\widetilde{\varepsilon}_n = \varepsilon_n / \underline{\varepsilon}_s$ and $\widetilde{\varepsilon}_t = \varepsilon_t / \underline{\varepsilon}_t$. Solving the effort game between the (single) student and teacher, equilibrium student and teacher effort are given by

$$
\begin{aligned}
\log \widetilde{\varepsilon}_n &= \nu_{0n}(\gamma_1, \gamma_2) + \frac{1}{2 - (\gamma_1 + \gamma_2)} \log \kappa + \frac{\delta}{2 - (\gamma_1 + \gamma_2)} \log K_{0n} + \frac{\gamma_0}{2 - (\gamma_1 + \gamma_2)} \log a_t \\
&\quad + \frac{2 - \gamma_2}{4 - 2(\gamma_1 + \gamma_2)} \log \theta_n + \frac{\gamma_2}{4 - 2(\gamma_1 + \gamma_2)} \log \theta_t, \tag{34}
\end{aligned}
$$

$$
\begin{aligned}
\log \widetilde{\varepsilon}_t &= \nu_{0t}(\gamma_1, \gamma_2) + \frac{1}{2 - (\gamma_1 + \gamma_2)} \log \kappa + \frac{\delta}{2 - (\gamma_1 + \gamma_2)} \log K_{0n} + \frac{\gamma_0}{2 - (\gamma_1 + \gamma_2)} \log a_t \\
&\quad + \frac{\gamma_1}{4 - 2(\gamma_1 + \gamma_2)} \log \theta_n + \frac{2 - \gamma_1}{4 - 2(\gamma_1 + \gamma_2)} \log \theta_t, \tag{35}
\end{aligned}
$$

where the ν_0's are functions of γ_1 and γ_2[35]. Upon substituting equilibrium student and teacher effort ((34) and (35)) into the production function (33), we get end-of-year knowledge as a function of the exogenous determinants of student end-of-year knowledge and student and teacher effort, namely,

$$
\begin{aligned}
\log K_n &= \nu_{0K}(\gamma_1, \gamma_2) + \frac{2}{2 - (\gamma_1 + \gamma_2)} \log \kappa + \frac{2\delta}{2 - (\gamma_1 + \gamma_2)} \log K_{0n} + \frac{2\gamma_0}{2 - (\gamma_1 + \gamma_2)} \log a_t \\
&\quad + \frac{\gamma_1}{2 - (\gamma_1 + \gamma_2)} \log \theta_n + \frac{\gamma_2}{2 - (\gamma_1 + \gamma_2)} \log \theta_t. \tag{36}
\end{aligned}
$$

To motivate the identification problem, suppose that a perfect measure of both student and teacher effort were available for multiple student-teacher pairs, but that there is no measure for either teacher instructional ability or student initial knowledge. Under that scenario, the problem in estimating (33) is that the effort game implies that both student and teacher effort will depend on unobserved teacher ability and on student initial knowledge. One approach would be to find instruments for student and teacher effort (which would also be ameliorative if student and teacher effort were measured with error). In the model as specified, preference shifters (X^θ_{njh} and X^θ_{tjh})

---

[35] Student and teacher variable costs are normalized to one.

that do not also affect student initial knowledge or teacher instructional ability would be valid instruments. However, if initial knowledge and teacher ability are generated through prior effort decisions that would also depend on preferences, and if preferences have some permanence, then exclusion restrictions of this kind would not be appropriate. The effort levels of other students in the class would also not be valid instruments, because they depend, in the model, on the initial knowledge and preferences of all students in the class (see (7)) as well as on teacher ability. Similarly, the effort levels of other teachers in the school would not be valid if teachers are not randomly allocated to schools.

The measurement structure previously detailed provides an alternative identification strategy. To fix ideas, first assume that we have perfect measures of K_{0n}, a_t, θ_n and θ_t and one measurement ($m = 1$) each for student and teacher effort and for K_n (see (30)). Given the linear in logs form, modify (28) and (29) so that measurement errors are proportional and impose the normalizations $\alpha_{0n}^{\varepsilon 1} = \alpha_{0t}^{\varepsilon 1} = 0$ and $\alpha_{1n}^{\varepsilon 1} = \alpha_{1t}^{\varepsilon 1} = 1$. Assuming that the reported measures of effort are inclusive of minimum effort, the measures of student and teacher effort are:

$$\log \varepsilon_n^1 = \log \underline{\varepsilon}_s + \log \widetilde{\varepsilon}_n + \varsigma_n^{\varepsilon 1} \tag{37}$$

$$\log \varepsilon_t^1 = \log \underline{\varepsilon}_t + \log \widetilde{\varepsilon}_t + \varsigma_t^{\varepsilon 1}. \tag{38}$$

It is easily seen that $\gamma_1, \gamma_2, \delta, \gamma_0$ are (over-)identified. Summing the coefficients on $\log \theta_n$ and $\log \theta_t$ in any of the three equations, (34), (35) or (36), provides an estimate of $\gamma_1 + \gamma_2$, which combined with the separate coefficients on each variable identifies the rest of the parameters.[36] Then, given that the constant term in (36) is identified from test scores, κ is identified and the minimum effort parameters, $\underline{\varepsilon}_s$ and $\underline{\varepsilon}_t$, are identified from the effort measurement equations.[37]

Suppose, instead, that the measures of K_{0n}, a_t, θ_n and θ_t are not perfect, but follow the structure posited previously, modified as above so that measurement errors are multiplicative. In that case, the parameters are identified as long as there are at least two measures of each with independent measurement errors. To illustrate, suppose that there are two measures of $\log \theta_n$, $\log \theta_n^1 = \alpha_0^1 + \alpha_1^1 \log \theta_n + \varsigma_n^1$ and $\log \theta_n^2 = \alpha_0^2 + \alpha_1^2 \log \theta_n + \varsigma_n^2$. If we solve for $\log \theta_n$ using $\log \theta_n^1$ with

---

[36] The measurement error variances of the student and teacher effort measures and the test score are also identified.

[37] Identification of the minimum effort levels, the constant term in the effort measurement equations, is unusual. Identification is achieved because the theory implies that the same parameters govern the knowledge production function and the effort supply equations.

the normalizations that $\alpha_0^1 = 0$ and $\alpha_1^1 = 1$, substitute it into (36) and then use the second measure as an instrument for the first measure, we can consistently estimate the parameter on $\log \theta_n$ in (36). A similar argument can be made for consistently estimating the parameters associated with $\log K_{0n}, \log a_t$ and $\log \theta_t$ in (36). The production function parameters can then be recovered as previously argued.[38] The same argument could have been applied to the effort equations (34) and (35). Thus, the production function parameters are overidentified from cross-equation restrictions. Given orthogonality of all measurement errors with each other and with the observable determinants, the variances and covariances of the latents are identified from standard arguments (e.g., Goldberger (1972)).

Extending these arguments to the non-linear setting with multiple students in a class and to multiple equilibria is not straightforward. Our conjecture is that they lead to further overidentifying restrictions, although the argument is heuristic. The non-linearity appears in two places, in the production function and in the derived equilibrium effort functions. With respect to the latter, that basic difference between the linear representation in (34) and (35) and that of the exact model, (7) and (8), is the term $\sum_{n=1}^{N} I(\widetilde{\varepsilon}_n > 0)(\delta K_{0n})^{\frac{2}{2-\gamma_1}}(\theta_n c_n^{-1})^{\frac{\gamma_1}{2-\gamma_1}}$. That term implies that the initial knowledge and student preference of all students in the class affect the effort levels of each student and the teacher. It thus adds variables, but no new parameters, presumably aiding identification. With respect to the production function, the term δK_{0n} reflects the end-of-year knowledge for students who supply minimum effort (which we have shown to be identified). The existence of equilibria in which some or all students supply minimum effort should aid in identifying δ.

4   Empirical Implementation

We estimate the model using data on the measures of latent model primitives, measures of student and teacher effort and a measure of end-of-year knowledge. As part of the ALI experiment, in addition to the curriculum-based end-of-year tests, extensive surveys were administered to both students and teachers that included questions that measure the latent factors in the model. Table one provides a list of the variables used in the estimation and a categorization according to their

---

[38]Given at least one determinant (an X variable) of $\log \theta_n$ and the cross-equation restrictions in the model, it is not necessary to normalize the slope coefficient in the first measurement equation. The same is true for $\log \theta_t$, but not for $\log K_n$ or $\log a_t$.

respective latent factors. As seen in the table, as is required for identification, each of the latent factors that determines effort levels and end-of-year knowledge has at least two measures and one determinant.[39] The determinants of the latent factors (the X's) are background information on students and teachers collected in the surveys.[40]

We estimate the model using data on control-group students in the 10th grade in the third (and last) year of the ALI program. We choose that subsample because there are more and better measures of the latents available in the student and teacher surveys in the third year and because all of those students have taken the national ninth grade ENLACE in the previous year. The academic year is divided into two semesters. It is not possible to estimate the model accounting for multiple semesters and for compositional changes within classes. To avoid this problem, we base the estimation on class assignments in the second semester.[41] We use data on the latent measures from the second semester (separate surveys were administered at the end of each semester) both with respect to the students and the teacher.[42]

Estimation is carried out with the distributional assumptions that Λ_2 (the school-level covariance matrix) and Λ_4 (the student idiosyncratic covariance matrix) are both joint normal. To aid in identification, we place some *a priori* zero covariance restrictions on Λ_2; specifically, the non-zero covariances include the following pairs: (K_0, θ_n), (K_0, g_n), (K_0, a_t) and (θ_t, a_t). Because of the small number of classes and (10th grade) teachers within each school, we were unable to identify

[39] Table 3 below provides more detail.

[40] Missing values for height (217 observations), primary school grade point average (376 observations) and parental education (21 observations) are imputed from regressions. The regression imputation for height is done separately for boys and girls and includes age and dummies for schools. The regression (probit) for parental education includes school dummies. The regression for primary school grade point average includes age, gender, parental education and school dummies. In each simulation draw of the model, we also draw from the residual variances of these regressions to obtain imputed values.

[41] Most students stay together in both semesters, although some drop out between semesters and the same teacher does not always stay with the same group of students, even when the composition of the class does not change. On average, 76 percent of the students were in the same class together in both semesters. In about one-half of the classes, over 90 percent of the students in their second semester class were also together in the same class in the first semester. In 35 percent of the second semester classes, all of the students had the same teacher in the first semester. In 38 percent, none of the students had the same teacher as in the first semester.

[42] The exceptions are the first three measures of student preferences reported in table 1, which are based on questions from the first semester survey.

the covariance matrix of class (or teacher) - level unobservables.[43] We therefore set Λ_3 to zero. Thus, all between-class variation within schools in teacher instructional ability and preferences is due to systematic differences in observables across classes as is also the case for student initial knowledge and preferences.

It is also necessary to specify an equilibrium selection rule. We specified the probability that an equilibrium is selected as a multinomial logit that depends on equilibrium characteristics, in particular on the fraction of students choosing minimum effort, an indicator variables for whether the equilibrium is the one in which all students choose minimum effort and for whether the equilibrium is the one that is pareto dominant (the equilibrium in which the most students supply above-minimum effort).[44]

4.1 Estimation Results

4.1.1 Parameter Estimates

Table 2 presents estimates of the production function parameters (and standard errors) along with selected summary statistics of the predicted latent variables.[45] All of the production function parameters are precisely estimated. The parameters indicate that the minimum effort level of end-of-year knowledge (K_n) increases with initial knowledge (K_{0n}) that is, $\delta > 0$, and that the marginal products of teacher and student effort are decreasing in class size ($\gamma_{11}, \gamma_{21} < 0$). Minimum effort for students is estimated to be the equivalent of 1.64 hours per week of time spent studying math and minimum effort for teachers the equivalent of 0.78 hours per week of class preparation time.

Mean initial knowledge is equivalent to a (standardized) score of 529.0 on the 9th grade Enlace and initial knowledge has a standard deviation of 44.9.[46] Mean end-of-year knowledge (K_n) is

---

[43]In 50 percent of the schools there was only one 10th grade teacher and in 36 percent only two. Further, there was only one class in 36 percent of the schools and only two in 62 percent of the schools. It should be noted that teacher effort measures are not class-specific for teachers with multiple classes. In forming the likelihood, we assume that, where a teacher has multiple classes, the teacher reports effort for a randomly selected class, where all classes are given equal probability.

[44]This specification most closely resembles that in Ackerberg and Gowrisankaran (2006).

[45]The entire set of parameters and their standard errors are shown in appendix table A.1.

[46]In terms of the second measure of initial knowledge, mean initial knowledge is equivalent to a grade in the ninth year mathematics course of 82.4 percent.

equivalent to a standardized score of 496.6 on the ALI test and has a standard deviation of 54.5.[47] Mean student effort (above the minimum) is equivalent to 1.82 hours per week of math study time. Mean teacher effort (above the minimum) is equivalent to 3.19 hours per week spent on class preparation. There is considerable variation in above-minimum student effort, with a coefficient of variation of .47, but much less variation in above-minimum teacher effort for which the coefficient of variation is .22. Student-specific primitives vary significantly more than those of teachers. The coefficient of variation in student preferences is .47, but only .17 for teacher preferences and the coefficient of variation in teacher instructional ability is only .03.

Our estimate of the equilibrium selection probability function implies that the equilibrium in which all students supply minimum effort is selected only when it is the only equilibrium and that the pareto dominant equilibrium is selected more often than the other equilibria.[48] That outcome is estimated to arise in only 0.6 percent of classes. In 45.7 percent of the classes, at least one student supplies minimum effort and in the remaining 53.7 percent of classes, all of the students supply above-minimum effort. The average fraction of students in a class who are at the minimum effort level is .097.[49]

4.1.2 Model Fit

Table 3 compares actual and predicted statistics for each latent variable measure. It also shows the fraction of the measure's variance that reflects variation in the latent (one minus the fraction due to measurement error). The measures are categorized according to their corresponding latent. The first row shows the fit for the ALI test, the measure of end-of-year (10th grade) math knowledge

---

[47]Based on an analysis of answer sheets, Behrman et. al. (2012) report that 3.7 percent of students in the control group engaged in copying. We account for this in our estimation by including an indicator variable in the test score measurement equation for whether a student was identified as a copier. Our estimate indicates that copiers increased their test scores by 14.2 points on average. The predicted mean test score incorporating cheating is 497.1 (see table 3).

[48]Including the fraction of students choosing minimum effort as an additional determinant of the equilibrium selection probability did not improve the fit of the model and so was omitted from the final specification. Thus, all equilibria other than the the pareto dominant equilibrium and the all minimum-effort equilibrium have equal probability.

[49]We discuss below the relevance of these results for providing an explanation for the treatment effects estimated in Behrman et. al. (2012).

(K_n). As seen, the model fits the mean and standard deviation well. According to the model estimates, 29.8 percent of the variance in the ALI test is accounted for by the variation in math knowledge (and the residual by measurement error). In comparison, 18.6 and 16.9 percent of the variance of the two measures of initial knowledge (K_{0n}), the ninth grade Enlace score and the student's grade in their ninth grade mathematics class, is accounted for by the variation in initial math knowledge.[50]

The fit for the measures of both the other exogenous and endogenous latent factors is generally quite good, although measures vary in their precision with respect to the corresponding latent. For example, the actual mean hours per week of study time devoted to mathematics, one of the measures of student effort, is 3.90, while the predicted mean is 3.83 hours; the actual standard deviation of the measure is 3.20 hours and the prediction is 3.24 hours. Moreover, hours of study time is censored from below at zero and from above at 10; the actual and predicted proportions at zero hours are 19.7 and 20.7 and at 10 hours, 8.9 and 6.0.[51] Hours of study time is, however, a noisy measure of student effort, with the noise component of the measure accounting for 93 percent of its variance. Another measure of student effort, the percent of time the student reports paying attention in class, which can take on values within the ranges 0-24 percent, 25-49 percent, 50-74 percent and 75-100 percent, is also fit well; the actual percentages are 5.3, 12.6, 35.0, and 47.3 while the predicted percentages are 6.0, 13.6, 35.4, 45.0. The measure is less noisy than hours of study time, with the variance component of the latent factor accounting for 12.8 percent of the total variance of the measure.[52] The fit for the measures of teacher effort are similarly good and the noise component of the measures of similar magnitudes. The least noisy measure of teacher effort is the number of hours per week spent on class preparation, for which the noise component is 91.1 percent of the total variance. With respect to student preferences, the three measures have

[50] Regressions of the ENLACE score and the 9th grade mathematics class grade on their observable determinants (listed in table 1) yield R-squares of 5 and 11 percent.

[51] As noted in table 1, hours per week of study time was modified to account for the "quality" of study time. Students who reported that they always texted or chatted on line while doing their homework (20 percent of the sample) were assigned a study time of zero hours. The ALI test score of those students is about 28 standardized points lower for any reported number of weekly hours of study time.

[52] Measures that are ordered categorical (or binary) are treated as themselves coming from underlying continuous latent variables. The proportion of variance that the measure explains of the model latent (in this case the preference for knowledge, θ_n) is with respect to the continuous latent of the measure.

degrees of precision of 28.9, 13.4 and 1.7 percent, while the two measures of teacher preferences have degrees of precision of 62.1 and 38.7 percent. The two measures of teacher instructional ability, based on student reports as was also the case for teacher preferences, have degrees of precision that are much lower, only 2.4 and 0.14 percent.

It is important to recognize that the extent of total variation accounted for by measurement error is in itself not critical for assessing the performance of the model or its ability to account for knowledge acquisition. As seen in table 2, all of the parameters of the production function are estimated with precision, indicating that the variance in the latents, student initial knowledge, teacher ability, student and teacher effort are sufficiently large, although the measures of the latents are noisy.

4.1.3 Characteristics of the Knowledge Production Function

Table 4 summarizes quantitatively the features of the production function for end-of-year knowledge. Each row shows the effect of *ceteris parabus* changes in each of the determinants of end-of-year knowledge, ranging from a two standard deviation decrease (relative to the mean) to a two standard deviation increase. When each determinant is varied, the other determinants are held constant at mean values. For reference, knowledge is 498.7 standardized points evaluated at the mean of the knowledge determinants. As seen in table 4, an increase in initial knowledge from two standard deviations below the mean to two standard deviations above would increase 10th grade math knowledge by 169.6 points or about 3 standard deviations. In contrast, increased teacher instructional ability (within the sample range) is considerably less productive; the same size change would increase 10th grade knowledge by 5.6 points, or about .1 of a standard deviation.

With respect to effort, the comparable (four standard deviation) change in student effort would increase 10th grade math knowledge by 14.7 points (.27 sd). However, there is a relatively large change in knowledge when a student moves above the minimum effort threshold; a change in student effort from minimum effort to two standard deviations below the mean, approximately a one standard deviation increase in effort, would increase math knowledge by 11.4 points (.21 sd). With respect to teacher effort, the change from two standard deviations below the mean to two standard deviations above the mean increases knowledge by only 2.4 points (.04 sd).

The last two rows of the table show the effects of changing class size and technology (δ).

Increasing class size from 16 to 52 students would reduce knowledge by only 3.1 points (.06 sd). Finally, improving a school's technology from two standard deviations below the mean of δ to two standard deviations above the mean would increase student knowledge by 110.8 points (2.0 sd), almost two-thirds of the gain that would occur from the same size change in student initial knowledge.

4.1.4 Accounting for Low Performance

The results in table 4 do not provide a complete picture as to the causes of low performance, because changes in student initial knowledge or in teacher ability will affect student and teacher effort, and table 4 holds effort levels constant. In addition, the impact of changing student and teacher preferences cannot be determined solely from the production function estimates. To account fully for low performance, we perform a series of counterfactuals where we change the composition of the classes in the control schools in terms of student and teacher primitives. In each case, we set a student or teacher primitive to be at least two standard deviations above the mean relative to its actual value. The results are reported in table 5.

The first column shows the baseline mean student and teacher effort levels and student end-of-year knowledge (both standardized and raw score). The second column shows those same statistics for the case in which initial knowledge is at least two standard deviations above the mean for all of the students in each class.[53] As seen, there are small effects on effort, reflecting the relatively low payoff to effort in producing knowledge; student effort increases by the equivalent of .19 hours per week of study time and teacher effort by .34 hours per week of class preparation time. End-of-year knowledge increases by 85.8 standardized points (1.57 sd).

The second column shows the impact of increasing teacher instructional ability so that all teachers are at least two standard deviations above the mean. The effect on student and teacher effort is similar to that for the change in student initial knowledge. However, because the marginal product of teacher ability is low, student knowledge increases only by 3.7 standardized points (.07 sd). Thus, improving teacher instructional ability, at least within the range of abilities in the data, is not a viable mechanism to improve student knowledge.

---

[53]Each of the counterfactuals was obtained by setting the particular latent to be two standard deviations above the mean whenever the latent was drawn below that level.

The counterfactual simulations related to changing preferences also lead to small effects on end-of-year knowledge. Unlike student initial knowledge and teacher ability, the effect of changing preferences on knowledge arises only through the effect on effort. As seen, setting the lower bound on student preferences at two standard deviations above the mean increases average student effort by the equivalent of 1.07 hours of study time per week and average teacher preparation time by the equivalent of .38 hours. Even with these increases in effort, however, average end-of-year knowledge increases by only 4.7 standardized points (about .09 sd). Setting the lower bound on teacher preferences has essentially no effect on average student effort and increases average teacher effort by .55 hours per week of class preparation time. The overall effect on student knowledge is essentially nil, only 0.5 standardized points. The last column shows the effect on end-of-year knowledge if all schools adopted the technology of the most productive schools (as measured by a value of δ that is two standard deviations above the mean). The effect of this change is a small increase in student and teacher effort. However, end-of-year knowledge increases by 56.2 standardized points (1.03 sd).

Tenth year math knowledge is measured in standardized ALI test score points, a relative scale. To get a better idea of what these changes mean in terms of an "absolute" scale, we can translate standardized points to the percentage correct raw score. As seen the raw score in the baseline is 37.8 percent. Using the raw score measure, the change above in student initial knowledge implies a change in 10th year math knowledge equivalent to a raw score of 47.6 percent (approximately an additional 8 questions correct out of a total of 79). Thus, even if all students were in the top 2.5 percent in terms of the current distribution of initial knowledge, on average less than one-half of the questions would be correctly answered.

These counterfactuals demonstrate that: (i) the initial preparation of students in terms of their incoming math knowledge has a large impact on 10th grade math performance, as measured by the standardized score, but a modest impact when measured by the raw score; (2) within the range of instructional abilities and preferences of teachers observed in the data, employing "better" teachers would have a small impact on student end-of-year knowledge; (3) increasing the enjoyment that students receive from acquiring math knowledge has a considerable impact on student effort and a smaller impact on teacher effort, but induces only a small increase in student end-of-year knowledge; (4) schools differ substantially in the technology of knowledge production and that can

have a quantitatively important effect on end-of-year knowledge.

5   Conclusions

This paper developed and estimated a strategic model of the joint effort decisions of students and teachers in a classroom setting to understand the reasons for low mathematics performance of Mexican high school students on curriculum-based examinations. The model allows for student and teacher heterogeneity in preferences for knowledge acquisition and in student initial mathematics preparation and teacher instructional ability. Student and teacher effort are assumed to be complementary inputs, which, with strategic behavior, leads to the existence of multiple Nash equilibria including one in which all students and the teacher supply "minimal" effort. In addition, students face a fixed cost of supplying effort above the minimum, which leads to additional potential equilibria in which some fraction of students in a class supply minimal effort. We showed that as long as the fixed cost is the same for students within a class (for example, related to the physical environment of the classroom), the number of equilibria cannot exceed one plus the class size, which makes both the solution of the model and estimation tractable.

Survey data of students and teachers collected as part of the ALI project provide multiple measures of student and teacher effort, student and teacher preferences, student initial knowledge and teacher ability, all of which we treat as latent variables. An end-of-year curriculum based test provides a measure of 10th year mathematics knowledge. A simulation-based maximum likelihood estimation procedure is used to recover the parameters of the knowledge production function as well as parameters governing the latent variable and measurement structures.

Estimation results indicate that the most important factor accounting for low performance is the lack of sufficient prior mathematics preparation. Based on the production function estimates, a *ceteris parabus* increase in student effort from its mean to one standard deviation above the mean, an increase equivalent to almost one extra hour per week of time devoted to studying math, would increase end-of-year knowledge by only .04 of a standard deviation. A similar increase in teacher effort would increase knowledge by less than .01 of a standard deviation. In contrast, a one standard deviation increase in student initial knowledge increases end-of-year knowledge by .78 of a standard deviation.

We also used the model to perform counterfactual experiments that change the composition of classes in terms of student and teacher primitives, incorporating the implied optimal changes in student and teacher effort. Increasing student initial knowledge in all classes to be at least two standard deviations above the mean increases end-of-year knowledge by 1.6 standard deviations above the mean, a ten percentage point increase in the raw score from 38 to 48 percent correct. In addition, the effect of giving all schools a technology that is two standard deviations above the mean would increase end-of-year knowledge by over one standard deviation, a 6.5 percentage point increase in the raw score. On the other hand, changes in student or teacher preferences or in teacher ability led to no more than a .09 standard deviation change in end-of-year knowledge.

An implication of these results is that there is a mismatch between the 10th grade mathematics curriculum and the mathematics preparation of incoming high school students. Because of this mismatch, increasing effort *per se* of either students or teachers does not lead to substantial increases in end-of-year knowledge. It thus appears that simply having a rigorous curriculum will not by itself improve knowledge.[54] Increasing the level of preparation in mathematics coming into 10th grade can significantly increase relative performance, even given the low productivity of effort. However, given the production technology, our estimates imply that even with a significant improvement in incoming knowledge (measured on a relative scale), less than one-half of the 10th grade curriculum, as measured by the ALI test, would be mastered on average.

Lastly, the introduction of this paper described the results of the ALI experiment, which provided performance-based monetary incentives to students and/or teachers. Although the model was estimated only on the control group, the results provide some insights about the results of the ALI experiment. One of the puzzling findings is that teacher incentives were only effective in combination with student incentives. One rationale for incorporating a fixed cost of student effort into our knowledge production function was that it created the potential for equilibria in which large fractions of students choose minimum effort. In that case, teacher effort might only respond to incentives when students were also offered incentives that induced them to choose above-minimum effort, potentially providing an explanation for the pattern of ALI treatment effects. Although our production function estimates results imply that inducing students not to choose minimum effort

---

[54] Although the results could reflect a lack of curriculum coverage by teachers, eighty percent of teachers, accounting for 85 percent of the students, report that they cover 75 percent or more of the curriculum.

can have a considerable impact on end-of-year knowledge, only a small proportion of students (10 percent) in a class on average are estimated to choose minimum effort. The estimated fixed costs are therefore not large enough to be an important explanation for the pattern of treatment effects under the ALI experiment.

Our estimates also reveal that the productivity of student and teacher effort are low. Given that, the differences between the treatment and control groups in the reported effort levels of students and teachers are not to be large enough to explain the observed differences in the ALI examination scores. For example, based on the number of hours per week studying math, student effort is, on average, .7 of a standard deviation higher for T3 than for the controls. That difference by itself would lead only to an increase in end-of-year knowledge of about .03 of a standard deviation, much less than the increase observed under the experiment.

Finally, our estimates showed that there are important differences across control schools in the knowledge production technology. In fact, the T3 treatment effect is of similar magnitude to what would arise if all schools in T3 had a level of productivity that was two standard deviations above the mean level of the control schools. This finding is consistent with the ALI incentive program having induced improvements in the productivity of treatment schools without aggregating school-specific causes of low performance.[55]

---

[55]Standard measures of school-level instructional infrastructure such as computer and library resources do not differ across treatment and control schools.

References

[1] ACKERBERG, DANIEL A. AND GAUTAM GOWRISANKARAN (2006): "Quantifying Equilibrium Network Externalities in the ACH Banking Industry," *The RAND Journal of Economics* 27 (3): 738-761.

[2] AUSTEN-SMITH, DAVID AND ROLAND G. FRYER (2005): "An Economic Analysis of Acting White," *Quarterly Journal of Economics* 120 (2): 551-583.

[3] BAJARI, PATRICK, HAN HONG AND STEPHEN P. RYAN (2010): "Identification and Estimation of a Discrete Game of Complete Information," *Econometrica* 78: 1529-1568.

[4] BEHRMAN, JERE R., SUSAN W. PARKER, PETRA E. TODD AND KENNETH I. WOLPIN (2013): "Aligning Learning Incentives of Students and Teachers: Results from a Social Experiment in Mexican High Schools," PIER working paper 13-004.

[5] BJORN, PAUL A., QUANG H. VUONG (1984): "Simultaneous Equations Models for Dummy Endogenous Variables: A Game Theoretic Formulation with Application to Labor Force Participation," SSWP 537, California Institute of Technology.

[6] CARD, DAVID AND LAURA GIULIANO (2011): "Peer Effects and Multiple Equilibria in the Risky Behavior of Friends," *Review of Economic and Statistics*, forthcoming.

[7] CUNHA, FLAVIO AND JAMES J. HECKMAN (2008): "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources* 43: 738-782.

[8] CUNHA, FLAVIO, JAMES J. HECKMAN AND SUSANNE M. SHENNACH (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica* 78: 883-931.

[9] DE PAULA, AUREO (2009): "Inferences in a Synchronization Game with Social Interactions," *Journal of Econometrics* 148: 56-71

[10] DE PAULA, AUREO (2012): Econometric Analysis of Games with Multiple Equilibria," CEMAP working paper CWP29/12.

[11] DUFLO, ESTHER, PASCALINE DUPAS AND MICHAEL KREMER (2008): "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya, NBER working paper No. 14475.

[12] DUFLO, ESTHER, REMA HANNA AND STEPHEN RYAN (2012): "Incentives Work: Getting Teachers to Come to School," *American Economic Review*, 102: 1241-1278.

[13] DURLAUF, STEVEN N. AND WILLIAM A. BROCK (2001): "Interaction-based models, in HANDBOOK OF ECONOMETRICS, VOL. 5, eds. James J. Heckman and Edward Leamer, Ch. 54, 3297-3380.

[14] GOLDBERGER, ARTHUR S. (1972): "Structural Equation Methods in the Social Sciences," *Econometrica* 40: 979-1001.

[15] HANUSHEK, ERIC A. (1994): MAKING SCHOOLS WORK, Brookings Institution: Washington D.C.

[16] KREMER, MICHAEL, EDWARD MIGUEL AND REBECCA THORNTON (2009): "Incentives to Learn, NBER working paper #10971.

[17] LAZEAR, EDWARD (2001): "Educational Production, *Quarterly Journal of Economics*, 116 (3): 777-803.

[18] JIA, PANLE (2008): "What Happens When Wal-Mart Comes to Town; An Empirical Analysis of the Retail Discounting Industry," *Econometrica* 76: 1236-1316.

[19] JORESKOG, KARL G. AND ARTHUR S. GOLDBERGER (1975): "Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable," *Journal of the American Statistical Association* 70: 631-639.

[20] PRITCHETT, LANT AND AMANDA BEATTY (2012): "The Negative Consequences of Overambitious Curricula in Developing Counties," CENTER FOR GLOBAL DEVELOPMENT Working Paper 293.

[21] SWEETING, ANDREW (2009): "The Strategic Timing Incentives of Commercial Radio Stations: An Empirical Analysis using Multiple Equilibria," *The RAND Journal of Economics* 40 (4): 710-742.

[22] VIVES XAVIER (2005): "Complementarities and Games: New Developments," *Journal of Economic Literature* 43: 437-479.

Table A.1
Parameter Estimates

Production Function

Parameter	Value	s.e.	Parameter	Value	s.e.
δ	.900	.441	γ_0	2.09	.139
σ_δ	.050	.020	γ_1	.308	.117
κ	.051	.004	γ_{11}	-.0017	1.71E-5
ε_s	1.64	.110	γ_2	.197	5.60E-4
ε_t	.780	.051	γ_{21}	-.0023	2.38E-6

Primitives: Standard Deviation(σ) and Correlations(ρ)

Parameter	Value	s.e.	Parameter	Value	s.e.
School-level			School-level		
σ_ξ^{k0}	21.8	961	$\rho(\xi^{k0}, \xi^{\theta n})$	.471	.002
σ_ξ^{at}	.009	.00046	$\rho(\xi^{k0}, \xi^{gn})$	-.075	.00014
$\sigma_\xi^{\theta n}$	.191	.008	$\rho(\xi^{k0}, \xi^{at})$	.372	.0031
$\sigma_\xi^{\theta t}$	.017	.002	$\rho(\xi^{at}, \xi^{\theta t})$	.901	.030
$\beta_0^{gn}, \sigma_\xi^{gn}$	.001, 7.52	1.13,.558			
Student-level			Student-level		
σ_ξ^{k0}	4.41	.373	$\rho(\xi^{k0}, \xi^{\theta n})$	.855	.0011
$\sigma_\xi^{\theta n}$	.100	.0072			

Measurements

Parameter	Value	s.e.	Parameter	Value	s.e.
K0: Initial knowledge ENLACE score			a_t : Teacher ability Teacher always knows the subject		
Constant	0.00	-	Constant	0.0	-
Slope	1.00	-	Slope	1.0	-
Variance	8947	3582	Variance	.021	.0015
			Teacher always has control of class		
Math 9th year grade					
Constant	3.26	38.6	Constant	.357	26.2
Slope	.0094	.0028	Slope	.308	3.77
Variance	.836	.335	Variance	.040	7.01
Θ_n: Student preferences			Θ_t: Teacher preferences Teacher always cares if student learns		
Likes math					
Constant	0.00	-	Constant	0.0	-
Slope	1.20	.076	Slope	5.82	.410
Cutoff 1	-.852	.161	Variance	.019	.0076
Cutoff 2 – Cutoff 1	.776	15.0			
Cutoff 3 – Cutoff 2	2.03	2.60			
Variance	1.00	-			

Hours/week studies non-math subjects			**Teacher wants students to pay attention**		
Constant	0.00	-	Constant	.056	6.36
Slope	5.73	47.77	Slope	4.82	8.67
Variance	23.1	9.96	Variance	.020	.312
Student does only what is expected					
Constant	0.00	-			
Slope	.401	.193			
Cutoff 1	-1.27	6.94			
Cutoff 2 - Cutoff 1	1.61	2.12			
Cutoff 3 - Cutoff 2	.556	.349			
Cutoff 4 - Cutoff 3	.940	1.00			
Variance	1.00	-			
e_n: Student effort			**e_t: Teacher effort**		
Hours/week studies math			**Hours/week of class preparation**		
Constant	0.00	-	Constant	0.00	-
Slope	1.00	-	Slope	1.00	-
Variance	17.1	3.40	Variance	5.28	9.88
Days attend math Class			**Hours/week tutoring outside of class**		
Constant	80.5	17.1	Constant	2.23	35.6
Slope	3.28	1.37	Slope	.267	313
Variance	11.7	6.30	Variance	8.15	12.6
Percent of time pays attention in class			**Gave only multiple choice tests (Y or N)**		
Constant	0.00	-	Constant	-.078	505.2
Slope	.453	.073	Slope	-.106	89.8
Cutoff 1	-.854	.184	Variance	1.00	-
Cutoff 2 - Cutoff 1	.761	19.4			
Cutoff 3 - Cutoff 2	1.06	13.8			
Put much effort into ALI exam (Y or N)			**Prepared students for ALI test (Y or N)**		
Constant	-1.16	3.17	Constant	-1.62	.160
Slope	.364	.351	Slope	.288	53.5
Variance	1.00	-	Variance	1.00	-
Attempted difficult ALI questions (Y or N)					
Constant	-.172	.985			
Slope	.325	2.18			
Variance	1.00	-			

T: ALI Test Score		
Constant	0.00	-
Slope	1.00	-
Variance	7503	229.3

<u>Determinants (X's)</u>

Parameter	Value	s.e.	Parameter	Value	s.e.
K0: Initial knowledge			a_t : Teacher ability		
Gender	-13.4	.905	Has education degree	.017	2.78
Height	.410	.022	Has masters degree	-.004	.0058
Age	-6.61	2.05	Years teaching	.0039	.0014
Primary school gpa	40.2	4.75	Years teaching square	-.00016	.000012
Parents education less than high school	-.858	.054	Constant	.835	.054
Constant	228.5	12.0	Θ_t:Teacher preferences		
			Has education degree	.009	.0024
Θ_n: Student preferences			Has masters degree	.020	.0021
Gender	-.006	.0068	Years teaching	.0015	4.07
Primary school gpa	1.34	.086	Years teaching square	-.00007	.0000086
Parents education less than high school	.287	.026	Constant	.129	.079
Constant	-3.11	.176			
T: ALI test score					
Copier	14.2	.953			
Equilibrium selection probability					
Dummy all students minimum effort	-12.0	9.09			
Dummy pareto dominant equilibrium	1.16	.289			

Table 1
Measures and Determinants of Latent Variables.

Student Entering Knowledge (K_{0n})
   Determinants: gender, height, primary school average, parental education, age.
   Measures: 9th grade mathematics ENLACE score, 9th grade mathematics class grade.

Student Knowledge Preference (θ_n)
   Determinants: gender, primary school average, parental education.
   Measures: extent to which likes math generally, hours per week study non-math subjects, frequency
            skips math, extent to which does only what is required.

Student Effort (e_n)
   Measures: hours per week study math[a], percent time pay attention in class, extent of effort on ALI
            test, attempt to answer difficult questions on ALI test.

Teacher Instructional Ability (a_t)
   Determinants: has education degree, has masters degree, teaching experience.
   Measures: fraction of students who say that teacher "always" knows subject well,
            fraction of students who say teacher "always" has good control of class.

Teacher Preference for Student Knowledge (θ_t)
   Determinants: has education degree, has masters degree, teaching experience.
   Measures: fraction of students who say that teacher "always" cares that they learn the material,
            fraction of students who say teacher "always" cares that they pay attention in class.

Teacher Effort (e_t)
    Measures: hours per week spent planning for classes, hours per week help students outside of class,
            gave only multiple choice tests, prepared students for ALI test.

Student Final Knowledge (K_n)
   Measure: ALI test score.

_____

   a. Interacted with one minus a dummy variable equal to one if the student always chatted and/or
       texted while doing homework, zero otherwise.

Table 2
Production Function Parameter Estimates and Predicted Summary Statistics

Parameter	Value (s.e.)		Parameter	Value (s.e.)	
κ	.051	(.004)	γ_1	.308	(.117)
δ	.900	(.441)	γ_{11}	-.0017	(1.71E-5)
$\underline{\varepsilon}_s$	1.64	(.110)	γ_2	.197	(5.60E-4)
$\underline{\varepsilon}_t$	.780	(.051)	γ_{21}	-.0023	(2.38E-6)
γ_0	2.09	(.139)			

Predicted Summary Statistics[a]

Statistic	Value	Statistic	Value
μ_{K0}	529.0	μ_K	496.6
σ_{K0}	44.9	σ_K	54.5
$\mu_{\theta n}$	.701	μ_{en}	1.82
$\sigma_{\theta n}$	.327	σ_{en}	.848
μ_{at}	.838	μ_{et}	3.19
σ_{at}	.025	σ_{et}	.689
$\mu_{\theta t}$	.132		
$\sigma_{\theta t}$	.023		

Statistic	Value
Fraction of classes in which:	
All students supply minimum effort	.006
Some students supply minimum effort	.457
No students supply minimum effort	.537
Average fraction of students in a class supplying minimum effort	.097

a. Based on 2000 simulations of each class

<div align="center">

Table 3
Model Fit

</div>

	Actual		Predicted		
	μ	σ	μ	σ	$\sigma^2_{True} / \sigma^2_{Total}$
Measure of 10th Year Math Knowledge (K)					
ALI Test	500.5	99.8	497.1	102.3	.298
Measures of Student Effort (e_n):					
Hours/week study math	3.90	3.20	3.83	3.24	.070
Percent equal 0	19.7	-	20.7	-	
Percent equal 10	8.9	-	6.0	-	
Days attended math class	79.6	1.10	79.8	0..79	.571
Percent of time pay attention in class					.128
0 – 24	5.3	-	6.0	-	
25 – 49	12.6	-	13.6	-	
50 – 74	35.0	-	35.4	-	
75 – 100	47.3	-	45.0	-	
Puts much effort into ALI test (%)	33.1	-	31.6	-	.087
Tried to answer difficult questions on ALI test (%)	67.3	-	65.7	-	.070
Measures of Teacher Effort					
Hours/week of class preparation	3.99	2.31	3.97	2.40	.089
Hours/week tutor outside of class	3.05	2.94	3.08	2.92	.004
Gave only multiple choice tests (%)	37.4	-	33.9	-	.005
Prepared students for ALI test (%)	21.3	-	24.5	-	.038
Measures of Student Initial Knowledge (K_0)					
Ninth grade Enlace standardized score	530.9	104.2	529.0	102.5	.186
Ninth grade mathematics grade	8.26	1.03	8.23	1.01	.169
Measures of Student Preference for Knowledge (θ_n)					
Hours/week studies non-math subjects	4.62	3.48	4.35	3.60	.289

Table 3 (continued)

Likes math (%)					.134
Never	5.2	-	5.7	-	
Almost never	13.1	-	13.9	-	
Sometimes	65.6	-	65.4	-	
Always	16.1	-	15.0	-	
Student does only what is expected					.017
Strongly agree	6.2	-	6.2	-	
Agree	45.1	-	46.3	-	
Neither agree or disagree	20.8	-	20.5	-	
Disagree	21.4	-	20.9	-	
Strongly disagree	6.6		6.1	-	
Measures of Teacher Preferences for Student Knowledge (θ_t) Fraction of students in class who respond that:					
Teacher always cares if they learn	.779	.172	.757	.176	.621
Teacher always wants students to pay attention in class	.693	.179	.689	.174	.387
Measures of Teacher Instructional Ability (a_t) Fraction of students in class who respond that:					
Teacher always knows the subject	.818	.159	.827	.131	.024
Teacher always has control of class	.608	.204	.612	.196	.0014

Table 4
Properties of the Knowledge Production Function:
Predicted Knowledge at Alternative Values of Student and Teacher Inputs[a]

	$\mu-2\sigma$	$\mu-\sigma$	$\mu+\sigma$	$\mu+2\sigma$
K_0	413.7	456.1	540.9	583.3
e_n	487.5[b]	495.2	500.8	502.5
e_t	497.1[b]	497.9	499.0	499.5
a_t	495.8	497.1	499.9	501.4
N^c	500.1	499.3	497.8	497.0
δ	443.1	470.8	526.2	553.9

a. Knowledge evaluated at means is 498.5.

b. Knowledge at minimum student or teacher effort is 476.1.

c. Class sizes are 16,25,43,52. Mean class size is 34.

Table 5
The Effect of Changes in Student Initial Knowledge, Teacher Instructional Ability, Student and Teacher Preferences, and Technology on 10th Grade Mathematics Knowledge[a]

	Baseline	ΔK_0	Δa_t	$\Delta \theta_n$	$\Delta \theta_t$	$\Delta \delta$
e_n	1.82	2.01	1.98	2.89	1.84	1.95
e_t	3.19	3.53	3.47	3.57	3.74	3.45
K_n						
Std. Score	496.6	582.4	500.3	501.3	497.1	552.8
Raw Score (%)	37.8	47.6	38.3	38.4	37.9	44.3
Change from Baseline in SD	-	1.57	.068	.086	.009	1.03

a. All students at least two standard deviations above the sample mean for each change.