Teaching to the Middle or to the Edges: Classroom Composition, Instructional Choices, and Their Impact on Student Achievement

Andrea Salvati [*]

August 13, 2021

Abstract

A growing body of research documents the importance of tailoring instruction to students' prior knowledge in fostering pupil achievement. Yet, few studies have tried to explore whether and how teachers adjust instruction in response to the composition of the classroom. This paper empirically investigates the relationship between instruction, classroom composition, and student achievement by developing and estimating a model of endogenous teaching decisions. Teachers choose how to allocate class time among topics that differentially impact the performance of students with heterogeneous levels of preparation. The model also allows teachers to choose effort, to vary in teaching ability, to value the achievement of students differently depending on their level of prior knowledge, and to have preferences over the alignment of instruction with state-level curriculum standards. I estimate the model using a unique dataset that combines school administrative data from five US school districts with a large set of instructional and teacher effectiveness measures. The results suggest that teachers attach a higher value to the achievement of students in the lower quantiles of the knowledge distribution. Moreover, I find that tailoring instruction to each student's level of initial knowledge would increase average achievement by 0.04SD. I further explore the policy implications of the estimated model by simulating the reassignment of students to classrooms based on their prior test scores performance (i.e., ability tracking). The results suggest that: *(i)* teachers respond to tracking by better tailoring instruction to students' initial knowledge; *(ii)* teachers assigned to lower tracks exert higher effort; *(iii)* the effect of tracking on achievement is heterogeneous across students, and its distribution depends on the way teachers are assigned to classrooms. Finally, additional simulation results suggest that existing educational standards are most suited to the level of preparation of students in the middle range of the knowledge distribution.

[*]Department of Economics, Rice University. Email: andrea.salvati@rice.edu. *Acknowledgements:* I am deeply indebted to my supervisors Kenneth Wolpin and Flavio Cunha for their guidance and encouragement. I also thank Rossella Calvi, Isabelle Perrigne, YingHua He, Jeremy Fox, Christopher Udry, Peter Arcidiacono, Bolun Li, Amir Kazempour, Ajinkia Keskar, and Daniel Prudencio for useful comments and feedback. All the errors, omissions, and interpretations are my own.

1. Introduction

A growing body of research documents the importance of tailoring instruction to students' skills in fostering achievement. Recent studies in educational psychology find significant positive effects of individualized instruction interventions on children's early grade reading and math skills (see e.g., Connor and Morrison, 2016; Connor et al., 2018). Moreover, increasing evidence suggests that mismatches between curricula and student preparedness are among the primary causes of children underachievement in developing countries (e.g., Glewwe et al., 2009; Pritchett and Beatty, 2012; Banerjee et al., 2016; Todd and Wolpin, 2018). In most countries, common attempts to prevent such misalignments often involve separating students into different groups or classrooms in order to better meet different instructional needs. In the US, about 75% of schools assigns 8^{th} grade students to math classes based on prior performances (a practice referred to as "ability tracking"), while about 70% of 4^{th} grade math teachers report sorting students by ability into instructional groups within the classroom (NAEP, 2015).[1] A common assumption underlying these practices is that teachers adjust the content, pace, and style of instruction in order to match the level of preparation of their students. Yet, very few studies have tried to investigate whether and how teachers actually choose their instruction in response to the composition of the classroom. The main goal of the present study is to fill this gap.

In the context of instructional decision-making, there is a general consensus that teachers choose instruction in order to improve their students' knowledge. Nevertheless, recent studies suggest that existing incentives (e.g., monetary bonuses), institutional constraints, or personal preferences could influence the way they get rewarded for the achievement of their students. In particular, teachers could end up attaching different values to the learning gains of different students and, as a result, choose to orient instruction towards specific segments of the classroom (see e.g., Duflo et al., 2011). Empirical and anecdotal findings from different education systems supports this hypothesis. For instance, increasing evidence shows that most developing countries tend to incentivize teachers and schools to focus their effort on the achievement of students in the upper quantiles of the distribution (see e.g., Glewwe et al., 2009; Duflo et al., 2011).[2] On the other hand, teachers in US schools are more likely to target below-proficiency students as a result of incentive-based policies like No Children Left Behind (NCLB) (e.g., Reback, 2008; Neal and Schanzenbach, 2010; Deming and Figlio, 2016). These

[1] A similar percentage of 4^{th} grade teachers applies ability grouping in reading classes, whereas no recent data has been published on tracking for 8^{th} grade English. Ability tracking is much less common in lower grades, with a prevalence of 30% and 5% in 4^{th} grade math and reading classes, respectively. On the other hand, there is some evidence of ability grouping in 8^{th} grade (24% in English in 2009, 52% in math in 2015). The prevalence of these practices has been steadily increasing in US public schools in the past decades, a trend likely driven by the growing diversity of the student body.

[2] In particular, the authors highlight the fact that given the high dropout rates of more disadvantaged students and the presence of monetary incentives based on student performances, teachers in developing countries are incentivized to focus on high-achieving students who are both less likely to leave school and more likely to perform well in high-stakes tests.

behavioral patterns can shape the relationship between classroom composition and instructional choices, and, in turn, affect the distributional impact of instruction on student learning by generating peer effects (see e.g., Sacerdote, 2011; Todd and Wolpin, 2018). From a policy standpoint, this implies that the behavioral response of teachers to the composition of the classroom is able to influence the effectiveness of education policies predicated on the assignment of students and teachers to classrooms (e.g., ability tracking). Finally, several other factors could also drive teachers' decision-making. For instance, teachers could choose to align the content and pace of instruction to specific curriculum standards (e.g., laid out by schools, districts, or state departments) or to follow personal preferences about what and how to teach.

This paper empirically investigates these mechanisms by developing and estimating a model of endogenous instructional decisions in a classroom setting. Teachers choose how to allocate class time among different topics and how much teaching effort to exert throughout the school year. Each student enters the grade with a level of baseline knowledge and exerts effort in school-related activities. A technology of knowledge formation links instructional time and effort, teacher ability, and student inputs to the production of end-of-year knowledge. A peculiar aspect of this technology is that the amount of time spent on each topic is allowed to have a different impact on learning depending on the student's readiness level. Moreover, the parametric specification adopted allows to pin down the class time allocations that would maximize the students' expected end-of-year achievement for each given level of prior knowledge. In particular, using the estimated parameters, I am able to compute the three different allocations of class time that would benefit the most students whose level of readiness fall, respectively, in the first, second, or third tercile of the distribution. Teachers choose time inputs and effort to maximize their expected utility, where the latter represents teacher's preferences over student achievement, class time allocation, and teaching effort. The utility function is specified in a way that allows teachers to value achievement differently depending on the student's level of baseline knowledge. This feature, combined with the knowledge technology, is able to generate the mechanism through which teachers optimally orient their instruction towards specific segments of the classroom. Finally, teachers bear a convex effort cost and have preferences over time spent teaching specific topics and over its alignment with the state-level curriculum standards. These standards are determined by the education authorities (i.e., the state-level department of education) and represent the content and pace of instruction that teachers are supposed to follow in order to (at least theoretically) attain student proficiency by the end of the grade.

The empirical part of this study uses data from the Measures of Teaching Effectiveness (MET) project, which merges school administrative data on test scores and other student characteristics from five US public school districts with a large set of measures of teacher ability, teacher effort, and detailed information on class time allocation across topics. I estimate the model through maximum simulated likelihood using a sub-sample of the data including 4^{th} grade math teachers. In particular, I exploit the availability of multiple measures to account for measurement error in end-of-year knowledge test scores and student-level and teacher-level inputs.

The results from the estimated model suggest that, while more prepared students tend to be more productive in learning new material, teachers attach higher rewards to the achievement of students with lower levels of initial knowledge. These estimates turn out to be a good characterization of the incentives provided by the US education system, especially as documented by evidence on the impact of recent incentive-based policies like NCLB (see e.g. Macartney et al., 2020; Deming and Figlio, 2016). Furthermore, my estimates show that the allocation of class time across different topics has a significantly different impact on end-of-year knowledge depending on the student's level of preparation. In particular, stronger students tend to benefit more from class time allocated to more advanced topics for 4^{th} grade (i.e., fractions and decimals) compared to less-prepared ones. I then exploit the estimated parameters of the technology to pin down the class time allocation vectors tailored to different levels of student baseline knowledge. I use these vectors to simulate a scenario in which each student receives the instruction tailored to her level of initial knowledge. Results suggest that end-of-year test scores in mathematics would increase by about 0.04σ.

Besides allowing for a more complete specification of knowledge accumulation process, the inputs included in the knowledge production function play a key role in controlling for factors potentially related to the non-random assignment of teachers to classrooms. Indeed, the latter is often considered the primary source of bias in the estimation of achievement production functions. To the extent that teacher assignments are based on prior test scores, teacher ability, or other observable characteristics, the inputs included are able to account for a wide range of potential confounding factors. Yet, assignment based on unobservables could still afflict the estimates. In order to check for the validity of the estimated model, I perform an out-of-sample validation exercise using data from the second year of the MET study in which teachers were randomly assigned to classrooms within each school. Specifically, I use the model estimates to predict measures of instruction and end-of-year knowledge in the second year sub-sample and then compare their summary statistics with the actual data.[3] Results show that, although teachers in the estimation sample were not randomly assigned, the model does a good job predicting second-year outcomes.

The estimated model allows me to perform counterfactual experiments that reveal the potential implications of accounting for teachers' endogenous response to the composition of the classroom. To this end, I run a counterfactual experiment where I re-assign students to classrooms based on their prior test scores, a practice commonly known in education as ability tracking. In particular, I

[3] Note that the simulated instructional time inputs cannot be compared to the actual data, as measures of class time allocation were not collected in the second year of the study.

simulate ability tracking under three alternative teacher assignment mechanisms: *(i)* random assignment, *(ii)* positive assortative matching (i.e., higher ability teachers to higher tracks), and *(iii)* negative assortative matching (i.e., higher ability teachers to lower tracks). I find that, in all three scenarios teachers respond to tracking by reallocating class time in a way that is better tailored to students' baseline knowledge. Moreover, teachers increase the amount of effort they exert when assigned to lower tracks, while the opposite occurs when assigned to the upper sections. Both these responses are directly implied by the increased classroom homogeneity generated by the tracking policy. In fact, while tracking allows teachers to better match the pace of instruction to the students' initial knowledge, it also separates students whose achievement is more rewarding (i.e., weaker students) from those whose performance is less so. As for the impact on student achievement, the effect of tracking on end-of-year knowledge depends significantly on the way teachers are assigned to classroooms. In particular, while both random assignment and negative assortative matching yield an increase in achievement gains in the low and middle parts of the distribution between 0.01σ and 0.04σ, with positive assortative matching there is a negligible increase in knowledge value-added for students in lower quantiles and an increase of about 0.05σ for high achieving students. The main novelty of these findings is that, on top of incorporating teachers' behavioral response to changes in classroom composition, they shed light on a new dimension that determines the distributional impact of tracking on student achievement, namely the teacher assignment mechanism.

In addition to classroom assignments practices, the development and implementation of curriculum standards has been one of the major focus of education policies in the past decades.[4] Conversely to the policy-maker expectations, various studies like Polikoff and Porter (2014) show that adhering to these standards does not always translate into higher achievement. In fact, one of the main drawbacks of this policy lies in its lack of flexibility, as these standards could impose a curriculum that is either too ambitious or too undemanding for the students entering a specific grade. To assess whether this is the case, I simulate a counterfactual scenario in which all teachers teach according to the state's curriculum standards. The results show that only students in the middle of the distribution would benefit from this policy. In contrast, students at the top and the bottom would bear a slight decrease in learning gains. Hence, this evidence suggests curriculum standards in the five states represented in the sample are tailored to students in the middle range levels of prior achievement.

The contribution of the present study spans several strands of the literature. First, this paper contributes to a recent literature on the impact of teacher incentives in US schools. For instance, Macartney et al. (2020) find that the implementation of NCLB in North Carolina created a peak of test scores growth around proficiency cutoffs, while Deming and Figlio (2016) report higher achievement

[4] In particular, the most prominent example of these policies is represented by the establishment of the Common Core Standards.

gains of low-achieving students in schools that were more likely to be marked as "low performing" under an accountability program in Texas. In both cases, the authors interpret these results in terms of teachers orienting instruction towards students at the margin or in the lower tail of the achievement distribution in response to the incentive programs. My results are in line with this evidence by confirming that teachers in US public schools attach higher rewards to achievement gains of students in lower quantiles. Second, my paper contributes to the literature on ability tracking. Despite a large number of empirical studies have tried to assess tracking practices, evidence on their effect on student outcomes is still mixed. For instance, Fu and Mehta (2018) find that tracking is beneficial only for high-achieving students while being detrimental for those assigned to lower tracks. Similar results are also found by Donaldson et al. (2017), who show that teachers assigned to lower tracks provide less emotional, organizational, and instructional support to students. On the other hand, by assessing a randomized experiment implemented in Kenyan schools, Duflo et al. (2011) find that tracking significantly increases the achievement of all students no matter their prior readiness level, and that these effects are likely driven by the behavioral response of teachers to the increased classroom homogeneity. My results are similar to Duflo et al. (2011) in that tracking allows teacher to better tailor instruction to the students' needs. However, differently from the existing literature, I find that the impact of tracking on student outcomes depends significantly on the way teachers are assigned to classrooms.

This paper also adds to a well-established literature on peer effects in the classroom (e.g. Brock and Durlauf, 2001; Sacerdote, 2011, for a review). As pointed out by Sacerdote (2011), there are a large number of channels through which peers can affect student outcomes. In particular, recent work by Duflo et al. (2011) and Todd and Wolpin (2018) highlight how peer spillovers can occur from the behavioral response of teachers to the distribution of student characteristics in the classroom. In a similar vein, using data from the MET project, Aucejo et al. (2018) find that different teaching practices have different effects on student achievement depending on the composition of the classroom. The present paper contributes to this literature by explicitly modeling teachers' response to classroom composition through the allocation of class time across topics. Finally, the present study contributes to a literature focusing on the relationship between instructional time inputs and student achievement. In particular, a common finding in education studies entails that the impact of time spent teaching more advanced topics is increasing in student baseline knowledge (Xue and Meisels, 2004; Engel et al., 2016). Nevertheless, Engel et al. (2013) show that teachers in early grades tend to spend too much time teaching basic topics already mastered by students when entering the grade. My findings corroborate this evidence in that: *i)* time allocated to more advanced topics is more productive for high-knowledge students and *ii)* tailoring instruction to students' needs has a positive effect on learning.

The rest of the paper is structured as follows. Section 2 describes the structure of the model and the specification of the knowledge formation technology. Section 3 analyses the identification of the model and estimation approach; Section 4 describes the data and reports descriptive statistics of the final sample; Section 5 discusses the estimation results as well as internal and external validation of the model. Finally, Section 6 analyses the counterfactuals and policy experiments, and Section 7 discusses concluding remarks.

2. A Model of Teacher's Instructional Decisions

This section presents a model which rationalizes teachers' instructional choices given their preferences and the technology of student knowledge production.

2.1 Environment and Teacher Choice Set

Consider a teacher t teaching in a class composed by N_t students, each of them indexed by i. The teacher is endowed with a level of ability A_t which affects the productivity of her instruction. Each student starts with a level of initial knowledge K_{0ti} and exerts learning effort h_{ti} (assumed to be exogenously determined).[5] The teacher is endowed with a total amount of class time $\overline{\tau}_t$ over the entire school year which can be allocated teaching J different topics. Time spent on each topic $j \in \{1,\ldots,J\}$ is denoted by $\tau_{tj} \in [0,\overline{\tau}_t]$, and the class time allocation vector is defined as $\boldsymbol{\tau}_t = (\tau_{t1},\ldots,\tau_{tJ})$, which satisfies the constraint $\sum_{j=1}^{J} \tau_{tj} = \overline{\tau}_t$. On top of class time allocation, the teacher chooses the amount of instructional effort to exert in class, which is assumed to be a non-negative scalar e_t. Both $\boldsymbol{\tau}_t$ and e_t are assumed to be pure public inputs, thus excluding the possibility of individualized instruction or within-classroom ability grouping practices.

2.2 Knowledge Production Technology

Student i end-of-year knowledge, K_{1ti}, is determined by the production function

$$K_{1ti} = \left[\delta_0 K_{0ti} + \delta_1 K_{0ti}^{\gamma_0} A_t^{\gamma_1} e_t^{\gamma_2} h_{ti}^{\gamma_3} \prod_{j=1}^{J} \tau_{tj}^{\eta_{jq}}\right] exp(\xi_{ti}), \tag{1}$$

where $1-\delta_0$ is the depreciation rate of knowledge, q denotes the $\overline{q}$-quantile of K_{0ti}, $\eta_{jq} \in [0,1]$ for each $j = 1,\ldots,J$, and ξ_{ti} is a normally distributed random shock, $\xi_{ti} \sim \mathcal{N}\left(0,\sigma_{\xi}^2\right)$, independent of all other inputs as well as of $\underline{K}_{0t} \equiv (K_{0t1},\ldots,K_{0tN_t})$ and $\underline{h}_{0t} \equiv (h_{0t1},\ldots,h_{0tN_t})$. Equation (1) is assumed to satisfy constant returns to scale (CRS) in time inputs conditional on the quantile q. Hence, $\sum_{j=1}^{J} \eta_{jq} =$

[5] In a newer version of this paper (which is currently a work in progress), I account for the endogeneity of student effort by building and estimating a coordination game between teachers and students. Yet, the inclusion of h_{ti} in the present version of the model serves primarily to (at least partially) control for unobserved family inputs and student non-cognitive skills in the achievement production function.

1 for each $q = 1, \ldots, \overline{q}$. The specification in (1) follows the original formulation of human capital production by Ben-Porath (1967) in that it posits a knowledge accumulation process where a flow of learning gains, known as knowledge value-added (i.e. the second term inside the brackets), is added to the level of existing stock of knowledge net of depreciation, $\delta_0 K_{0ti}$. The former captures the direct outcome of the learning process, where both instructional and non-instructional inputs are combined and transformed into additional knowledge.

The Cobb-Douglas specification of knowledge value-added is consistent with the theoretical and empirical literature on learning and cognitive achievement. First, this specification fits the intuitive idea that instruction is made of two complementary elements, namely the content (i.e. the specific topics covered in class) and the delivery of such content to the student. In the model, a teacher selects the content and pace of instruction by setting $\boldsymbol{\tau}_t$, while the delivery is governed by the exerted instructional effort e_t.[6] Furthermore, the quality of instruction delivery, in terms of its impact on students learning gains, depends on the ability of the teacher, A_t. The latter generally includes teacher skills like verbal ability and content knowledge, which are considered among the most important attributes of teaching effectiveness. However, the set of relevant attributes is potentially larger (see e.g. Darling-Hammond and Youngs, 2002; Andrew et al., 2005).[7] Second, equation (1) implies that the time inputs in $\boldsymbol{\tau}_t$ (i.e. time spent in a each topic j, τ_{tj}) are all complement to each other. This is consistent with the literature on learning trajectories of students in different subjects (see e.g. Kilpatrick et al., 2001, 2003, for a review of of the theoretical and empirical studies of instruction and learning in mathematics). [8] Third, the elasticity parameters $(\eta_{jq})_{j=1}^{J}$ are quantile-specific, thus allowing time inputs to be more or less productive depending on the level of i's initial knowledge, K_{0ti}. This is consistent with a vast literature in education which analyzes the benefits of the so called "differentiated instruction" (e.g. Tomlinson et al., 2003). Finally, the complementarity of teacher inputs with student baseline knowledge and student effort is consistent with the idea that the effectiveness of instruction depends on the student's preparation as well as on her engagement in school-related activities. In particular, the latter includes time spent studying the subject, amount of attention during classes, or class disruption. While the complementarity between instruction and K_{0ti} is well-documented in both economics and education, as in Bodovski and Farkas (2007), Engel et al. (2013), Engel et al. (2016),

[6] A discussion on the complementarity between content and delivery can be found in Agodini and Harris (2014).

[7] The importance of teaching ability in explaining achievement gains has been confirmed by a very influential study by Kane et al. (2013), who, using the same MET dataset employed in this paper, find that a very large set of research-based teaching effectiveness measures are able to strongly predict teachers' value-added.

[8] For example, learning how to compute the area of a rectangle can reinforce the understanding of integer multiplication. A potential drawback of the Cobb-Douglas specification in (1) is that it carries the strong assumption that, for any topic j, $\tau_{tj} = 0$ implies zero learning gains. Although relevant from a theoretical point of view, this assumption has no particular implication in the specific application of present study, as $\tau_{tj} = 0$ never occurs in the data. Moreover, the Cobb-Douglas specification has the desirable feature of allowing to point identify the time allocation vector tailored to each level of initial knowledge, as discussed in Section 2.2.1.

or Todd and Wolpin (2018), recent studies document both a causal relationship between pupil effort and achievement (e.g. Burgess, 2016) and an increase in teacher effort productivity when students are more engaged in learning inside and outside the classroom (see e.g. Todd and Wolpin, 2018).

2.2.1 Tailored Instruction

An implication of the specification in (1) is that it allows to find, for each specific student, the class time allocation that would maximize her expected end-of-year knowledge. Formally, this entails solving the following maximization problem

$$\max_{\boldsymbol{\tau}_t} E[K_{1ti}], \quad s.t. \quad \tau_{tj} \in [0, \overline{\tau}_t], j = 1, \ldots, J, \quad \sum_{j=1}^{J} \tau_{tj} = \overline{\tau}_t$$

where K_{1ti} is determined by equation (1) and the expectation is taken with respect to ξ_{ti} (i.e., the only unobserved input by the teacher). Taking the FOCs and rearranging we obtain the following closed form solution

$$\boldsymbol{\tau}_t^{*q} = (\overline{\tau}_t \eta_{1q} \ldots, \overline{\tau}_t \eta_{Jq}) = \overline{\tau}_t \boldsymbol{\eta}_q \tag{2}$$

where $\boldsymbol{\eta}_q = (\eta_{1q} \ldots, \eta_{Jq})$. Thus, $\boldsymbol{\tau}_t^{*q}$ represents the class time allocation *tailored* to the initial knowledge of students in the q^{th} quantile. This follows from a well-known property of the Cobb-Douglas with CRS and resources constraint, which implies that the optimal share of time allocated to each topic is given by the elasticity parameters $\boldsymbol{\eta}_q$. The tailored instruction $\boldsymbol{\tau}_t^{*q}$ is key in this model as it represents the channel through which teachers are able to target the instructional needs of specific students. In particular, teachers can orient instruction towards students at a specific quantile q by choosing a vector $\boldsymbol{\tau}_t$ closer to $\boldsymbol{\tau}_t^{*q}$.

2.3 Curriculum Standards

The majority of the US state departments of education adopt curriculum standards. In particular, these standards establish what students are supposed to know at the end of each grade and what teachers should teach in order to ensure students' proficiency. My model allows teachers to follow the state-level standards as defined by the vector $\boldsymbol{\varphi}_t = (\varphi_{t1}, \ldots, \varphi_{tJ})$, where each element φ_{tj} is the amount of class time the teacher is supposed to spend on topic j. Notice that standards could be different across schools depending on the state they are located. This implies that if two teachers t and t' are located in the same state, then $\boldsymbol{\varphi}_t = \boldsymbol{\varphi}_{t'}$.

2.4 Preferences

Teachers have preferences over their students' knowledge as well as chosen instruction over the school year.[9] In particular, preferences are represented by the utility function

$$U_t = \sum_{i=1}^{N_t} \omega_{ti} K_{1ti} - \frac{\alpha_0}{2} e_t^2 + \sum_{j=1}^{J} \left(\alpha_{1j} + \varepsilon_{tj}\right) \tau_{tj} + \frac{1}{2} \sum_{j=1}^{J} \alpha_{2j} \left(\tau_{tj} - \varphi_{tj}\right)^2. \tag{3}$$

Equation (3) is composed by four terms. The first one represents preferences over students' end-of-year knowledge, specified as a weighted average of the elements in $\underline{K}_{1t}$. The teacher attaches a (possibly different) value to each student's knowledge level, which is captured by the student-specific parameter ω_{ti}. The second terms in (3) represents teacher's effort cost, which is assumed to be quadratic in e_t and with $\alpha_0 > 0$. The third term, instead, captures teachers' preferences over the allocation of class time among classroom activities. Specifically, $(\alpha_{1k} + \varepsilon_{tk})$ is the marginal utility (cost) the teacher gets (bears) when increasing time spent on activity k (while holding $\underline{K}_{1t}$ fixed), with ε_{tk} a mean-zero preference shock and α_{1k} a parameter. Finally, the last term represents teacher's utility (cost) of deviating from the state curriculum standards $\boldsymbol{\varphi}_t$. This term allows to account for teacher's degree of adherence to the standards. In particular, $\alpha_{j2} < 0$ implies a general compliance of the teacher with the standards on topic j, while a positive value indicates a willingness to depart from φ_{tj}.

The student-specific parameters ω_{ti} follow the parametric specification

$$\omega_{ti} = \omega_1^q + Z'_{ti} \boldsymbol{\omega}_2. \tag{4}$$

where the first term on the RHS captures the part of ω_{ti} determined by i's baseline knowledge, which is represented by the quantile-specific parameter ω_1^q, and the last term allows the weight ω_{ti} to depend on other students' or teacher's characteristics Z_{ti}. Hence, this specification allows the teacher to weigh student achievement gains differently depending on her level of initial knowledge as well as according to other characteristics, like gender or race.

[9] Consistently with the literature on instructional effort choices, I assume that teachers do not account directly for students' future outcomes when making decisions (e.g. Barlevy and Neal, 2012; Macartney et al., 2015; Todd and Wolpin, 2018). In fact, the implicit assumption is that teachers care about students' future outcomes (like graduation, college enrollment, earnings etc.) only to the extent to which they are determined by knowledge produced during the school year they are teaching in.

2.5 Teacher's Optimization Problem

The teacher chooses effort e_t and class time allocation $\boldsymbol{\tau}_t$ in order to maximize the expected value of (3) subject to the choice variables constraints. Formally,

$$\max_{e_t, \boldsymbol{\tau}_t} E\left[\sum_{i=1}^{N_t} \omega_{ti} K_{1ti} - \frac{\alpha_0}{2} e_t^2 + \sum_{j=1}^{J} \left(\alpha_{1j} + \varepsilon_{tj}\right) \tau_{tj} + \frac{1}{2} \sum_{j=1}^{J} \alpha_{2j} \left(\tau_{tj} - \varphi_{tj}\right)^2 \right]$$
$$s.t. \quad e_t \geq 0, \quad \tau_{tj} \in [0, \overline{\tau}_t], \text{ for } j = 1, \ldots, J, \quad \sum_{j=1}^{J} \tau_{tj} = \overline{\tau}_t \tag{5}$$

where the expectation is taken with respect to K_{1ti}, conditional on observable inputs and curriculum standards. Substituting for $\tau_{tJ} = \overline{\tau}_t - \sum_{j=1}^{J-1} \tau_{tj}$ and taking the first-order conditions we obtain the following equations for an interior solution of optimal effort, e_t^*, and time spent on each activity $k = 1 \ldots, J-1$, τ_{tk}^*,

$$\gamma_2 e_t^{\gamma_2 - 1} \sum_{i=1}^{N_t} \omega_{ti} \delta_1 K_{0ti}^{\gamma_0} A_t^{\gamma_1} h_{ti}^{\gamma_3} \prod_{j=1}^{J} \tau_{tj}^{\eta_{jq}} - \alpha_0 e_t = 0 \tag{6a}$$

$$\sum_{i=1}^{N_t} \omega_{ti} \delta_1 K_{0ti}^{\gamma_0} A_t^{\gamma_1} e_t^{\gamma_2} h_{ti}^{\gamma_3} (\eta_{kq_i} \tau_{tk}^{-1} - \eta_{Jq} \tau_{tJ}^{-1}) \prod_{j=1}^{J} \tau_{tj}^{\eta_{jq}} +$$
$$+ (\widetilde{\alpha}_{1k} + \widetilde{\varepsilon}_{tk}) - \alpha_{2k} \left(\tau_{tk} - \varphi_{tk} - \tau_{tJ} + \varphi_{tJ}\right) = 0 \tag{6b}$$

where $\widetilde{\alpha}_{1j} \equiv (\alpha_{1j} - \alpha_{1J})$ and $\widetilde{\varepsilon}_{1j} \equiv (\varepsilon_{1j} - \varepsilon_{1J})$. Equation (6a) represents the optimality condition for e_t, which equates the marginal utility of students' expected end-of-year knowledge from a change in e_t with the marginal cost of effort. Notice that, given $\alpha_0 > 0$ and $e_t \geq 0$, equation (6a) holds as long as the weighted average in the first term is positive. Otherwise, the teacher optimally chooses to exert no effort by setting $e_t^* = 0$. This equation determines the relationship between instructional effort and the other inputs. In particular, it shows that teachers respond to the classroom distribution of initial knowledge, $\underline{K}_{0t} = (K_{0t1}, \ldots, K_{0tN_t})$, and how this relationship is governed by the interaction of the values attached to each student end-of-year knowledge, $(\omega_{sti})_{i=1}^{N_t}$, with the the other inputs determining the productivity of effort. Indeed, the more productive is effort in producing knowledge (holding α_0 fixed), the higher is the value of e_t^* the teacher chooses. These implications also characterize the relationship between instructional effort and class time allocation given the results obtained in 2.2.1. In fact, the closer is time allocation $\boldsymbol{\tau}_t$ to the value tailored to the students whose achievement teacher t finds most rewarding, the more effort she will exert.[10] The optimality condition for an interior solution of each time input τ_{tk}, $k = 1 \cdots, J-1$, is represented by equation (6b). Although the way

[10] Formally, if we define $q' \in \arg\max_q \{\omega_1^q\}_{q=1}^Q$ (i.e. the quantile of students whose achievement the teacher attaches the highest value), the lower is the distance between $\boldsymbol{\tau}_t$ and $\boldsymbol{\tau}_t^{*q'}$ (from (2)), the higher will be $\prod_{j=1}^J \tau_{tj}^{\eta_{jq}}$ and, in turn, e_t^*.

$\boldsymbol{\tau}_t^*$ is related to the value of other inputs and parameters is more complicated compared to the one with effort, similar mechanisms still apply. Indeed, the interaction between the achievement weights $(\omega_{ti})_{i=1}^{N_t}$ and the other production function parameters governs the relationship between class time allocation and the composition of the classroom. Finally, similarly to e_t^*, there is the possibility of corner solutions in $\boldsymbol{\tau}_t^*$, with either $\tau_{tk} = 0$ or $\tau_{tk} = \overline{\tau}_t$ for some topic $k = 1, \ldots, J$. As detailed in the next sections, corner solutions are not taken in consideration in the estimation as they are never selected by the teachers in the data.

3. Estimation

In the empirical specification of the model, both inputs and outputs are assumed to be latent factors measured with error. The factor model allows to correct for variables mis-measurement and the arbitrariness of their scales. This structure is in line with recent literature on child skills development (e.g. Cunha et al., 2010; Agostinelli and Wiswall, 2016) and similar to the specification employed by Todd and Wolpin (2018). This section describes the structure imposed to the latent factors as well as the system of measurement equations.

3.1 Latent and Measurement Structure

3.1.1 Latent Factors Structure

Each exogenously determined latent input $\theta \in \{A, K_0, h\}$ is assumed to depend linearly on a vector of initial conditions, X^θ, and on one or more random effects. Formally, the latent teacher's ability A_t is specified as

$$log(A_t) = X_t^A \boldsymbol{\beta}^A + \upsilon_t^A, \tag{7}$$

with υ_t^A being a teacher-level unobserved random component of teacher ability. Finally, student i's baseline knowledge K_{0ti} and individual effort h_{ti} are specified as

$$log(K_{0ti}) = X_{ti}^{K_0} \boldsymbol{\beta}^{K_0} + \upsilon_t^{K_0} + \zeta_{ti}^{K_0} \tag{8}$$

$$log(h_{ti}) = X_{ti}^{h} \boldsymbol{\beta}^{h} + \upsilon_t^{h} + \zeta_{ti}^{h}, \tag{9}$$

where $\upsilon_t^{K_0}$ and υ_t^h, and $\zeta_{ti}^{K_0}$ and ζ_{ti}^h are teacher/classroom-level, and student-level random effects, respectively.[11] Random effects at each separate level are allowed to be correlated across factors and are assumed to be orthogonal to the exogenous variables $X_{ti} \equiv (X_t^A, X_{ti}^{K_0}, X_{ti}^h)$, to each other, and to be mean zero and jointly normally distributed. Formally, $\boldsymbol{\upsilon}_t | X_{ti} \sim \mathcal{N}(\mathbf{0}, \Sigma_\upsilon)$, $\boldsymbol{\zeta}_{ti} | X_{ti} \sim \mathcal{N}(\mathbf{0}, \Sigma_\zeta)$, and $\boldsymbol{\upsilon}_t \perp$

[11] The model could also allow for school-level random effects. However, observations entailing only one teacher per school are quite frequent in the sample used for the estimation. As a result, separately identifying teacher and school-level effects would be a demanding task.

$\boldsymbol{\zeta}_{ti}$, where $\boldsymbol{\upsilon}_t \equiv (\upsilon_t^A, \upsilon_t^{K_0}, \upsilon_t^h)$ and $\boldsymbol{\zeta}_{ti} \equiv (\zeta_{ti}^{K_0}, \zeta_{ti}^h)$. Finally, the latent factors of effort e_t, time allocation $\boldsymbol{\tau}_t$, and end-of-year knowledge K_{1ti} are endogenously determined by the equations (6a), (6b), and (1), respectively.

3.1.2 Measurement Equations Structure

Teacher ability A_t, instructional effort e_t, and student effort h_{ti} are assumed to be latent factors measured with error by M_θ different measures, with $\theta \in \{A, e, h\}$. Dropping the subscripts to simplify the notation, let $Z^{\theta m}$ the m-th measure of latent θ, which is allowed to be either continuous or ordinal. Continuous measures are assumed to be linear in the log of the latent, that is

$$Z^{\theta m} = \mu_0^{\theta m} + \mu_1^{\theta m} log(\theta) + \varsigma^{\theta m}.$$

Ordinal measures $Z^{\theta m} \in \left\{\underline{Z}^{\theta m}, \ldots, \overline{Z}^{\theta m}\right\}$ are instead specified as step functions of a linear-in-log transformation of the latent θ, $Z^{\theta m *} = \mu_0^{\theta m} + \mu_1^{\theta m} log(\theta) + \varsigma^{\theta m}$. The linear-in-logs specifications of $Z^{\theta m}$ and $Z^{\theta m *}$ constrain latent values to be positive. The data provides only one measure for each of the remaining factors, namely baseline and end-of-year knowledge, K_{0ti} and K_{1ti}, and time inputs $\boldsymbol{\tau}_t$, which are therefore assumed to be measured without error. Furthermore, similarly to K_{0t} and given the nature of the measure of K_{1ti} in the data (i.e. a z-score), end-of-year knowledge is assumed to be measured as $Z^{K_1} = log(K_{1ti})$. Finally, I assume classical measurement error together with joint normality, that is $\boldsymbol{\varsigma}_{ti} \equiv (\varsigma_t^{A,m}, \varsigma_{ti}^{h,m}, \varsigma_t^{e,m}, \varsigma_{ti}^{K_1,m}) \sim \mathcal{N}(\mathbf{0}, \Sigma_\varsigma)$,, where Σ_ς is a diagonal variance-covariance matrix and $\boldsymbol{\varsigma}_{ti}$ with assumed orthogonal to all the observed and unobserved components of the latent factors.

3.1.3 Further Assumptions and Discussion

In order to bring the model to the data, it is necessary to first discuss some issues related to the measures available as well as to some necessary restriction to be imposed to the model. A first issue is given by the information available on instructional time inputs, as the MET data does not provide variables on $\boldsymbol{\tau}_t$ expressed in terms of time (e.g. hours, days, or weeks). Instead, data on class time allocation is available only in terms of fractions of total class time, $\boldsymbol{\tau}_t / \overline{\tau}_t$. In order to mitigate the potential consequences from the lack of information on $\overline{\tau}_t$, I allow the parameter δ_1 to be district-specific. This assumption seems particularly suited to the data, as there is evidence that schools participating in the MET study have to abide to a specific total number of school days and class hours determined by the school district (with only few exceptions). This implies that total class time $\overline{\tau}_t$ is going vary for the most part between and not within districts.[12] As for curriculum standards, Section

[12] If total class time is fixed to $\overline{\tau}_d$ across all schools within district d, equation (1) implies that the district-specific parameter employed in the empirical specification will be equal to $\delta_1 \overline{\tau}_d$.

2.3 points out that states do not actually provide the time variables $\boldsymbol{\varphi}_t$, but rather some documents which detail what skills a typical student is supposed to acquire in each subject by the end of each grade. Given that exact data on $\boldsymbol{\varphi}_t$ is not available, I will use information on the state test content collected by the MET study as a proxy of the standards.[13] Test content variables are also expressed as fractions, thus making them comparable to the curriculum data discussed above. The idea is that, to the extent that the state test is aimed at measuring students' proficiency, its content should reflect the educational standards set by the state. Moreover, there is evidence of alignment between tests and standards content, as shown by Polikoff et al. (2011).

For the empirical specification of the weights $(\omega_1^q)_{q=1}^{\overline{q}}$ and the elasticities $((\eta_{jq})_{j=1}^{J})_{q=1}^{\overline{q}}$, I use terciles, i.e. $\overline{q}=3$. In particular, in the rest of the paper I will often refer to the first, second, and third terciles as "low-level", "mid-level", and "high-level" baseline knowledge, respectively. An additional specification issue concerns then the variables to include as determinants of teacher's preferences, Z_{ti}, in equation (4). For this I follow recent empirical evidence on gender stereotypes (Carlana, 2019) and ethnicity role model effects (Gershenson et al., 2018) and include both teacher-level and student-level gender and race dummies. A third issue entails the inclusion of class size effects on K_{1ti}. To account for that, I follow Todd and Wolpin (2018) and allow the elasticity of effort to depend on N_t through the equation $\gamma_2 = \gamma_{20} + \gamma_{21} N_t$. Moreover, in order to ensure a solution for optimal effort e_t^*, I assume $\gamma_2 \in (0,2)$.[14] The parametric specification is then completed by imposing distributional assumptions on the the preference shocks $\widetilde{\boldsymbol{\varepsilon}}_t = (\widetilde{\varepsilon}_{t1}, \ldots, \widetilde{\varepsilon}_{tJ-1})$, which are assumed to be jointly normally distributed with mean zero, covariance matrix $\Sigma_{\widetilde{\varepsilon}}$, and orthogonal to the random effects $(\boldsymbol{\upsilon}_t, \boldsymbol{\zeta}_{ti})$ and measurement errors $\boldsymbol{\varsigma}_{ti}$.

A final concern is about the possibility of corner solutions in either exerted effort ($e_t^* = 0$) or in the time allocation choice ($\tau_{tk} = 0$ for some $k = 1\ldots, J-1$). In fact, a complete description of the model would necessitate an analysis of the conditions on the parameters, latent factors, and preference shocks values that give rise to each specific corner solution. However, since none of the measures of instruction used in my analysis are consistent with corner solutions, the empirical specification will employ the FOCs in (6a) and (6b) as the only conditions of optimality required to estimate the model.

3.2 Identification

Parameters of the exogenous latent factors equations (7), (8), and (9) and measurements equations are identified through observable determinants X_{ti} and multiple measurements of each latent

[13] The MET study actually provides data on the state standards content, whose variables are measured as fractions of "items" in the curriculum standards document about each single topic. As it is not clear whether the fraction of items in such documents is a good measure of the "weight" a state gives to each topic, using test content data seems a better option.

[14] In practice, in the estimation I impose the restriction $\gamma_{20} + \gamma_{21} \max_t\{N_t\} \in (0,2)$.

variable (upon necessary normalizations). Identification of the latent and measurement equations for the exogenous inputs (A_t, h_{ti}) follows the canonical arguments of structural equation modeling as in Goldberger (1972). Consider a latent factor $\theta \in \{A, h, K_0\}$, with subscripts dropped whenever it is not confusing to do so. First, I normalize the intercept and slope of an arbitrary measure $m = 1$ by setting $\mu_0^{\theta 1} = 0$ and $\mu_1^{\theta 1} = 1$. Given the orthogonality assumptions of both random effects and measurement errors, the latent equation parameters are identified by regressing the first measure $Z^{\theta 1}$ on X^θ, that is

$$\boldsymbol{\beta}^\theta = E[(X^\theta)' X^\theta]^{-1} E[(X^\theta)' Z^{\theta 1}].$$

Once the parameters $\boldsymbol{\beta}^\theta$ are known, the slopes and intercepts of the remaining measurement equations $m = 2, \ldots, M_\theta$ for $\theta \in \{A, h\}$ are identified as follows. First, regress each measurement $Z^{\theta m}$ on X^θ and obtain $\tilde{\boldsymbol{\mu}}^{\theta m} = E[(X^\theta)' X^\theta]^{-1} E[(X^\theta)' Z^{\theta m}]$, and then compute

$$\mu_1^{\theta m} = \tilde{\mu}_j^{\theta m} / \beta_j^\theta, \qquad \mu_0^{\theta m} = \tilde{\mu}_0^{\theta m} - \mu_1^{\theta m} \beta_0^\theta,$$

for an arbitrary j^{th} element of $\tilde{\boldsymbol{\mu}}^{\theta m}$, $j \geq 2$, and with β_0^θ being the latent factor equation constant. It is then possible to pin down Σ_υ, and Σ_ζ by computing the covariances between measures. In particular, the diagonal elements $\sigma^2_{\upsilon\theta}$, and $\sigma^2_{\zeta\theta}$ are obtained by

$$\sigma^2_{\zeta\theta} = Cov(Z_{ti}^{\theta 1}, Z_{ti}^{\theta m}) / \mu_1^{\theta m} - Var(X_{ti}^\theta \boldsymbol{\beta}^\theta) - \sigma^2_{\upsilon\theta},$$

where $m \geq 2$ if $\theta \in \{A, h\}$ and $m = 1$ if $\theta = K_0$; $(\overline{Z}_t^{\theta 1}, \overline{X}_t^\theta)$ are class-level means; and the last equation holds only for the student-level latent factors, h and K_0. Similarly, the off-diagonal elements $\sigma_{\upsilon\theta\theta'}$, and $\sigma_{\zeta\theta\theta'}$, for $\theta, \theta' \in \{A, h, K_0\}$, $\theta \neq \theta'$, are determined as

$$\begin{aligned}
\sigma_{\upsilon\theta\theta'} &= Cov(\overline{Z}_t^{\theta 1}, \overline{Z}_t^{\theta' 1}) - Cov(\overline{X}_t^\theta \boldsymbol{\beta}^\theta, \overline{X}_t^{\theta'} \boldsymbol{\beta}^{\theta'}) \\
\sigma_{\zeta\theta\theta'} &= Cov(Z_{ti}^{\theta 1}, Z_{ti}^{\theta' 1}) - Cov(X_{ti}^\theta \boldsymbol{\beta}^\theta, X_{ti}^{\theta'} \boldsymbol{\beta}^{\theta'}) - \sigma_{\upsilon\theta\theta'}.
\end{aligned}$$

Finally, the measurement error variances for teacher ability and student inputs measures are obtained as

$$\begin{aligned}
\sigma^2_{\varsigma Am} &= Var(Z_t^{Am}) - Var(X_t^A \boldsymbol{\beta}^A) - \sigma^2_{\upsilon A} & m &= 1, \ldots, M_A \\
\sigma^2_{\varsigma hm} &= Var(Z_{ti}^{hm}) - Var(X_{ti}^h \boldsymbol{\beta}^h) - \sigma^2_{\upsilon h} - \sigma^2_{\zeta h}, & m &= 1, \ldots, M_h.
\end{aligned}$$

To illustrate the sources of identification for the knowledge technology and preferences parameters, I will first analyze the case of no measurement error. First, independent variation in the inputs $(K_{0ti}, A_t, \boldsymbol{\tau}_t, h_{ti})$ and end-of-year knowledge K_{1ti} allows to pin down the knowledge production function parameters $(\delta_0, \delta_1, \gamma_0, \gamma_1, \gamma_3, ((\eta_{jq})_{j=1}^J)_{q=1}^3, \sigma_\xi)$. On the other hand, as teacher effort is a deterministic function of other inputs (as derived by (6a)), following the normalization $\alpha_0 = 1$, $(\gamma_{20}, \gamma_{21})$ are identified off data on e_t and K_{1ti} combined with variation in other inputs and class size. As for the utility function parameters, the main source of identification comes from data on instructional choices together with variation in classroom composition characteristics. Specifically, the weights parameters $((\omega_1^q)_{q=1}^3, \boldsymbol{\omega}_2)$ are identified off the effect of variation in $(\underline{K}_{0t}, \underline{Z}_t)$ on curriculum choices, $\boldsymbol{\tau}_t$, and exerted effort, e_t. Finally, the utility parameters $(\alpha_{1j}, \alpha_{2j})_{j=1}^J$ and the preference shocks covariance matrix $\Sigma_{\tilde{\varepsilon}}$ are identified off distributional moments of observed time inputs $\boldsymbol{\tau}_t$ and curriculum standards $\boldsymbol{\varphi}_t$ combined with the functional form restrictions imposed by the FOCs in (6b).

Turning to the latent factors model considered in this paper, the identification argument follows the one in Todd and Wolpin (2018). In particular, let the optimal instructional choices from (6a)-(6b) and the end-of-year knowledge production function (1) be represented as functions of the exogenous initial conditions, $(X_{ti}, \boldsymbol{\varphi}_t)$, and the random shocks $(\boldsymbol{\upsilon}_t, \boldsymbol{\zeta}_{ti}, \tilde{\boldsymbol{\varepsilon}}_t)$. Formally

$$e_t^* = a_e(X_{ti}, \boldsymbol{\varphi}_t, \boldsymbol{\upsilon}_t, \boldsymbol{\zeta}_{ti}, \tilde{\boldsymbol{\varepsilon}}_t) \tag{10a}$$

$$\tau_{tj}^* = a_{\tau_j}(X_{ti}, \boldsymbol{\varphi}_t, \boldsymbol{\upsilon}_t, \boldsymbol{\zeta}_{ti}, \tilde{\boldsymbol{\varepsilon}}_t), \qquad j = 1, \ldots, J \tag{10b}$$

$$K_{1ti} = a_{K_1}(X_{ti}, \boldsymbol{\varphi}_t, \boldsymbol{\upsilon}_t, \boldsymbol{\zeta}_{ti}, \tilde{\boldsymbol{\varepsilon}}_t). \tag{10c}$$

Consider now a system of equations that combines: *(i)* the exogenous latent factor equations from (7) and (9) with the measurement equations; *(ii)* the equations on endogenous effort and end-of-year knowledge, (10a) and (10c), with the measurements; and *(iii)* (8) and (10b), for $j = 1, \ldots, J$, assumed to be measured without error. This system is a measurement model for the latent factors $(\boldsymbol{\upsilon}_t, \boldsymbol{\zeta}_{ti}, \tilde{\boldsymbol{\varepsilon}}_t)$ analogous to (3.7) in Cunha et al. (2010). As a result, it is possible to invoke Theorem 2 from the same paper in order to identify both utility and production function parameters.

3.3 Likelihood Function

Estimation is carried out through simulated maximum likelihood (SML). Let Θ be the vector of all the parameters of the model to be estimated (including latent factor and measurement equations). The likelihood contribution of teacher/classroom t given Θ is given by the joint density of the latent factors measurements for the teacher and all N_t students, denoted by $\mathcal{M}_t$, conditional on the initial conditions $\mathcal{X}_t \equiv (X_{t1}, \ldots, X_{tN_t}, \boldsymbol{\varphi}_t)$. Denoting the simulated likelihood contribution of teacher t by $\hat{L}_t(\Theta|\mathcal{M}_t; \mathcal{X}_t)$, the likelihood function to be maximized over Θ is

$$\mathcal{L}(\Theta) = \prod_{t=1}^{T} \hat{L}_t(\Theta|\mathcal{M}_t; \mathcal{X}_t),$$

where T is the total number of classrooms/teachers in the sample. A complete description of likelihood function formulas and of the estimation procedure is reported in the Appendix.

4. Data

For the empirical part of this paper I use data from the Measurements of Effective Teaching (MET) project. This study was conducted in two years in six large US school districts and involved 2714 4^{th}-to-9^{th} grade Math and English Language Arts (ELA) teachers in 317 schools.[15] The main goal of this project was to assess the ability of a large set of research-based indicators of teacher quality to identify effective teachers. Moreover, in order to ensure validity of the estimates, teachers in the second year of the study were randomly assigned to classrooms within each school (while in the first year the assignment was performed as usual). The data collected by the MET study include detailed information on: *i)* teaching practices in the classroom from both video-recorded lessons and surveys taken by both teachers and students; *ii)* topics covered in end-of-grade state tests; *iii)* self-reported student information on own effort and home environment; *iv)* end-of-grade state tests scores in various subjects and teacher and student demographics from administrative district data. All these data provide the variables used to measure the latent inputs and outputs of the model. The estimation is carried out using a sub-sample of the first-year MET data including 4^{th} grade Math teachers who took the Survey of Enacted Curriculum (SEC).[16] Originally developed by Porter and Smithson (2001) to study the alignment of classroom instruction with curriculum standards and test content, this survey asks teachers to report their class time allocation throughout the school year across an exceptionally fine-grained array of different topics spanning all school grades. These answers were then converted and re-expressed as fractions of total class time.[17] The survey was conducted only in the first year of the MET study, thus not allowing me to exploit the second-year random assignment of teachers to estimate the model. Nevertheless, second-year data will be used to perform an out-of-sample validation exercise (see Section 5.3).

Table 1 reports descriptive statistics on class time allocation and curriculum state standards. Topics are aggregated in five different groups representing common areas of mathematics covered in 4^{th} grade classes, namely: place value, rounding, addition, and subtraction (group 1); multi-digit multiplication and division (group 2); shapes, angles, and geometry (group 3); fractions and decimals

[15]The school districts included in the study are: Charlotte-Mecklenbourg Schools (NC), Dallas Independent School District (TX), Denver Public Schools (CO), Hillsborough County Public Schools (FL), Memphis City Schools (TN), and the New York City Department of Education (NY).

[16]The full sample of teachers taking the SEC included 4^{th} and 8^{th} grade Math and ELA teachers. The choice of focusing on 4^{th} grade Math classes was driven purely by the higher sample size.

[17]As pointed out in Section 3.1.3, the MET data does not provide teachers' original answers on the actual time spent on different topics. As a reference point, Trends in International Mathematics and Science Study (TIMSS) reports that in 2015 teachers in the US spent on average 216 class hours.

(group 4); unit conversion and measurement (group 5).[18] On average, teachers split 3/4 of the school year evenly between teaching multi-digit multiplication and division (26%) fractions and decimals (25.5%) and unit conversion and measurement (24.2%). The remaining time is then largely devoted to geometry topics (17.5%), while only a smaller fraction of class time focuses on more basic topics like place value, rounding, addition, and subtraction of whole numbers (6.8%). Columns 2 to 5 show the variation of these time allocations across teachers. There is a 2 percentage points variation between the first and third tercile in the percentage of class time devoted to place value, rounding, addition, and subtraction, whereas the difference is about 7-8 percentage points for all other topics groups. These correspond to about 25% of the value taken by the mean for groups 2, 4 and 5, to about 30% for group 1, and to more than 40% for geometry topics (group 3). Similarly, the ratio between the standard deviation and the mean is around 30% for groups 1 and 3 and never above 25% for all other groups. The picture depicted by these statistics is that teachers vary mostly in the fraction of time they allocate on geometry topics and, to a lesser extent, on the more basic topics included in group 1.

[Table 1 here]

Panel B reports descriptive statistics on the content composition of the state curriculum standards as measured by the percentage of test items covering each topic group in the end-of-year grade 4 state test. As reported in Column 1, the curriculum standards averages are very similar to those of class time allocation, thus suggesting a potential alignment between the two. However, descriptive statistics in Panel C show that differences between standards and classroom instruction are in fact significant. Specifically, Column 1 shows that, for each topic, the average absolute deviation between these two is always greater than the standard deviation of class time allocation. This is particularly true for the geometry group, where the value of the mean absolute deviation is 40% larger than the standard deviation. A potential reason behind this misalignment with the standards could entail adjustments of teaching strategies to the composition of the classroom. As displayed in Panel D, classrooms could vary substantially in their composition as it pertains to students' level of math readiness. Columns 3-5 show that one fourth of the classrooms in the sample have a fraction of low-achieving (1^{st} tercile) students below 0.14 while another one fourth display values above 0.46. This range is a little higher for the fraction of high-achieving students, with 25% of the classrooms having 14.3% or less students performing on the higher part of the distribution and another 25% with 52% or more. On the other hand, classrooms tend to have more similar percentage of students with initial knowledge falling in the middle of the distribution, with the first and third quartiles being 24.1% and 41.7%, respectively.

[Table 2 here]

[18] The SEC allows class time to be allocated in 183 topics combined with five possible levels of "cognitive" demand, for a total of 915 cells. The choice of the five topic groups was inspired by the Common Core standards classification.

Table 2 displays descriptive statistics of the student and teacher characteristics determining student effort and teacher ability. Panel A shows that students in the sample are on average 9-10 years old (as expected for 4^{th} graders) and the gender ratio among them is almost 1:1, with a slightly higher percentage of females. The majority of the student population is black (43%) followed by white (26%) and Hispanic students (25%). About 6% of the students have been identified as gifted, while almost 9% are placed in special education programs (SpEd) due to learning disabilities and 17% are labeled as English language learners (ELL). As regards to students' socioeconomic status (SES) and family/home environment, almost half of the students in the estimating sample receive either free or reduced-price lunch and 88% of them has at least one computer at home. Finally, nearly half of the responding students possess at least 25 books in their bedroom, 40% of the students report to have always a quite place to study at home, and 69% has always a person at home who can help with homework. The sample is clearly not representative of the US student population, as it includes a much higher percentage of black and Hispanic students and a slightly lower percentage of students served by SpEd programs compared to national averages. As for teacher characteristics, Panel B shows summary statistics of the determinants of teacher ability. The majority of the teachers (83%) are generalist, i.e. teach both Math and ELA to the same classroom. As for characteristics related to their human capital, teachers in the sample have on average 6.5 years of teaching experience in the school district with a quite significant variability, and about half of the teachers possess a Master's degree.

[Table 3 here]

Table 3 provides descriptive statistics of the measures of used in the estimation of the model. Baseline and end-of-year knowledge are measured, respectively, by the 3^{rd} and 4^{th} grade math state test z-scores. In particular, these have been rescaled in order to reflect differences between districts and grade, and are expressed in terms of $3^{r}d$ grade z-scores.[19] The mean and standard deviation of the 3^{rd} grade z-scores in the final sample are 0.10 and 0.95, respectively, thus showing slightly higher and less variable performances in mathematics compared to overall MET sample, whose test scores have, by construction, mean 0 and standard deviation equal to 1. On the other hand, the 4^{th} grade state z-scores display an increase in the mean of 0.3σ with respect to the 3^{rd} grade level and a similar standard deviation of 0.95.

There are 11 measures for latent teacher ability, including three Classroom Assessment Scoring System (CLASS), two Framework for Teaching Mathematics (FFTM), and two Mathematical Quality of Instruction (MQI) scale scores from video-recorded lessons, as well as four measures from the stu-

[19]The rescaling was necessary since state test z-scores are computed by grade, subject, and district (i.e. state), thus making them incomparable across states and between 3^{rd} and 4^{th} grade. In particular, the rescaling involved using the BAM math test (administered by the MET project) z-scores district-level means to restore differences between districts, and the PIAT test grade-level means (computed using NLSY data) to allow between-grades comparisons.

dent survey. The CLASS scores are measured on a scale of 1-7 and include: the behavior management score, which evaluates teacher's ability to set clear behavior expectations, to prevent and redirect students misbehavior, and to obtain students' compliance; the content understanding score, which refers to both the depth of the lesson's content and the teacher's ability to help students in understanding the framework and key ideas of the topic taught; and the productivity score which measures teacher's level of preparation for the lesson as well as her ability to maximize learning time and to set clear routines and instructional expectations. As for the FFTM scores, measured on a scale of 1-4, the management of class procedures score measures the degree of smooth functioning of the classroom, whereas the management of student behavior score evaluates the teacher's ability to manage student conduct and to respond to their misbehavior. Differently from CLASS and FFTM measures, the MQI scores assess the pedagogical knowledge and preparation of the teacher necessary to teach mathematics. Specifically, the richness of mathematics score refers to the teacher's ability to explain mathematical ideas as well as to draw connections and illustrate different aspects of math concepts, while the mathematical knowledge for teaching (MKT) score measures the overall teacher's knowledge in the specific area of mathematics taught in 4^{th} grade. Finally, the last four measures are class-level averages of student evaluation scores on the teacher's ability to deliver instruction, where each single student evaluation uses a scale of 0-4. As seen, teachers have pretty good evaluations in terms of their ability to explain concepts (with averages higher than 3), while they tend to get lower ratings with respect to their ability to control the class behavior.

Similarly to the last four measures of teaching ability, the 8 measures of latent teacher effort come from student evaluations (again, on a scale of 0-4). In particular, these measures refer to the teacher's level of feedback and motivational support provided to the students, as well as to her level of effort in trying to avoid that students fall behind during the lesson. As shown by the third panel of Table 3, teachers average score is higher than 3 in five out of eight measures, while lower scores are reported in terms of teachers' effort to not waste time in class, to summarize the lesson, and to write feedback on homework and exams. Finally, measures of student effort are represented by the answers to 4 questions included in the student survey. Students report a quite high level of effort in school activities, with more than 40% of the sample declaring to always do their best work in class, to never give up when work gets hard, to never take it easy not trying to do their best, or to complete all the homework assigned.

[Table 4 here]

A necessary condition for the identification of the model's parameters entails the non-independence between measures of each latent variable. To this end, Table 4 reports the correlation matrices of the measures described in Table 3. Specifically, the correlation between continuous variables is measured by the Pearson's correlation coefficient, while the Pearson's chi-squared statistics is computed to assess the association between categorical variables. Almost all pairs of measures display statistically significant correlations, with the exception of the student-survey measures of teacher ability which more than half of the times are unrelated to the CLASS, FFTM, and MQI scale scores.

5. Estimation Results

5.1 Parameter Estimates

Table 5 reports the estimated parameters of the production and utility functions (1) and (3). Estimates of the measurements and exogenous latent factors equations are instead reported in Table B.3 of the Appendix. The knowledge production technology entails a depreciation rate of math knowledge between grades 3 and 4 of about 39% together with a positive effect of every input on end-of-year knowledge. The coefficients converting input units (δ_{1d}) are positive and very similar across districts 1 through 5. However, a LR test of $H_0 : \delta_{11} = \cdots = \delta_{15}$ rejects the null that all these parameters are equal. Column (3) shows the elasticity parameters of inputs. Baseline knowledge displays the largest elasticity with 0.45, followed by student effort with a value of about 0.37. Teacher ability has a positive effect on knowledge production, but its elasticity coefficient is imprecisely measured. Finally, teacher effort displays a small but statistically significant elasticity, taking a value of about 0.006 for an average classroom of 23 students. In particular, the negative estimate of γ_{21} suggests that teacher effort is less productive in larger classes, with a decrease in elasticity of about 0.0002 for each additional student.

Table 7 reports the elasticities of class time inputs conditional on initial knowledge tercile. For students in the first tercile, the most productive topics are those related to either measurement or fractions and decimals, with both elasticities being around 0.3, followed by geometry with 0.2 and multiplication and division with 0.16. The topics with the lowest productivity are those included in the first group (place value, rounding, addition, and subtraction), with an estimated parameter slightly below 0.04. On the other hand, the estimates for students in the second and third terciles show that class time allocated to multiplication and division or fractions and decimals is the most productive in terms of learning gains (with elasticities all around 0.3), followed by measurement topics (0.16 and 0.2 for second and third terciles), geometry (0.15 and 0.11), and finally place value, addition, and subtraction with 0.08 and 0.04 for second and third tercile students, respectively. Comparing the parameters in Table 7 across terciles for each topic group, it is possible to grasp the learning profiles of students with different levels of baseline knowledge. In particular, the most intuitive result is that allocating a higher share of class time to fractions and decimals, which are typically considered the most advanced topics in fourth grade, is going benefit more higher terciles. Similarly, multiplication and division topics seem to be more productive for students with higher baseline knowledge.

On the other hand, class time allocated to both measurement and geometry are significantly more productive for first tercile students, thus possibly implying that students with higher level of initial knowledge have already mastered these topics. Finally, the parameters on place value, addition, and subtraction are less easy to interpret, as they tend to be less productive for first tercile students despite being generally considered more basic topics.

[Table 5 here]

The utility function parameters are reported in Panel B of Table 5. As shown in Column (1), teachers attach the highest value to achievement gains of students with low baseline knowledge (ω_1^1) and the lowest to students with high initial knowledge (ω_1^3). Hence, teachers exhibit a higher preference for compensatory teaching aimed at fostering the learning gains of students starting with lower levels of math knowledge. In particular, a unit increase in math achievement from of a student with low initial knowledge rewards the teacher about 14 times more than the same increase for a student with a higher level of math readiness, and slightly more than 2 and a half times more than a student in the second tercile.[20] Column (3) shows the estimated parameters on teacher preferences over class time allocation. Each $\tilde{\alpha}_{1j} \equiv \alpha_{1j} - \alpha_{15}$ (for $j = 1, \ldots, 4$) captures the utility a teacher gets from reallocating 1% of class time from unit conversion and measurement to topics in group j, while holding fixed students' achievement and the distance between instruction and the state standards. The estimates suggest that teachers bear a disutility when reallocating time from measurement to fractions and decimals topics, whereas reallocation to all other topics has no statistically significant impact. Moreover, a likelihood ratio test fails to reject the null that $\tilde{\alpha}_{11} = \cdots = \tilde{\alpha}_{14} = 0$.[21] Finally, the estimated values of α_{2j} suggest that teachers bear a quite large utility cost from not aligning instruction to the curriculum standards for geometry, and to a lesser degree to those for place value, addition and subtraction, and measurement topics.

5.2 Within-Sample Model Fit

Table 8 compares the predicted means and standard deviations from the estimated model with the data. The model does a pretty good job in fitting the data, although the degree of the prediction precision is different across groups of simulated measures. First, the model is able to predict the mean and standard deviation of the 4^{th} grade math test score (representing the log of end-of-year knowledge) very well, with negligible differences between the data and the simulated values. As for the class time allocations across topics, the average simulated values are very similar to the respective data means, with the largest difference being in the time spent on measurement topics, whose

[20] Not surprisingly, a LR test rejects the null hypothesis that these weights are all equal to each other.

[21] Notice that $\tilde{\alpha}_{11} = \cdots = \tilde{\alpha}_{14} = 0$ is equivalent to $\alpha_{11} = \cdots = \alpha_{15}$, where each α_{1j} could well be different zero. Hence, the test fails to reject the null that teachers value time spent on each topic the same.

mean value is only 0.4 percentage points higher than in the data. On the other hand, the model tends to systematically over-estimate the standard deviations of class time allocation. In particular, while the predicted standard deviation for "place value..." topics is just slightly higher, the simulated distributions for all other topic groups yields a standard deviation up to 2.3 percentage points higher (i.e. for geometry topics). These differences do not seem substantial when compared to the mean values, although they are quite large relative to the sample standard deviations.

As for the teacher effort measures, there are negligible differences between the predicted means and the data means, whereas the simulations tend to slightly over-estimate the standard deviations. Table 8 shows that the model is also able to fit the ability measures pretty well, although with a lesser degree of precision when compared to teaching effort. Specifically, the simulation tend to over-estimate the means of CLASS behavioral management and productivity scores as well as the FFTM scores on class management, while all other measures are much closer to the data means. All standard deviations of the ability measures are predicted pretty well by the model, with the largest difference being with respect to the CLASS measures, whose values are systematically lower by about 0.03 score points. Finally, the last panel of Table 8 reports descriptive statistics of the student effort measures converted from Likert scale to a 0-4 score for the first 3 measures and a 1-5 scale for the last one ("How much homework..."). Comparing the predicted means with the data it is possible to see that the model underestimate the first and third measures ("I have done my best..." and "In this class, I take it easy...") by about 0.3 points, while overestimating the second and fourth measures by about 0.1 and 0.2 points, respectively. As for the standard deviations, the predicted values for the first two measures are 0.15 and 0.03 points higher, respectively, while for the third and fourth measures the model underestimate its value respectively by 0.15 and 0.1 points. The main reason behind these discrepancies are due to the fact that all measures of student effort are discrete with a single continuous latent variable, and the presence of measurement error could generate discontinuities in the simulated values. [22]

[Table 8 here]

Column (5) in Table 8 reports the share of total variance due to measurement error for teacher effort, teacher ability, and student effort measures. The importance of measurement error for teacher effort is quite variable, going from 42% for "Teacher asks if we are following..." to a maximum of 95% for "Teacher does not waste time...". All other teacher effort measures entail a degree of measurement error accounting between a half and two thirds of the total variation of the measure. An even larger range in the degree of measurement error is displayed by teacher ability measures. Indeed, these include a very precise measure in CLASS behavior management scale, with a share of variance due to

[22] B.4 in the Appendix reports the predicted student effort measures expressed with the original Likert scale.

measurement error of only 18%, as well as a set of very noisy measures with a level of noise signal ratio equal or higher than 90%, including MQI scores and three out of four measures based on the student survey. Finally, the measure of student effort with the lowest noise signal ratio is "In this class, I stop trying..." with almost 60%, while the poorest measure entails a level of measurement error of about 92%. In the remaining two measures of student effort, measurement error accounts for about two thirds of the total variation.

[Table 9 here]

A further analysis on the ability of the model to fit the data is given by the regression parameters shown in Table 9. The goal of this exercise is to assess whether the model is able to replicate a reduced form relationship between some endogenous variables (i.e. end-of-year knowledge and teacher effort) and the exogenous variables that determine the latent factors. Given the non-linearity of the expressions of both teacher effort and end-of-year knowledge, this is a quite challenging task. As shown by the first two columns of Table 9, only the coefficient on student gender display a different sign between the predicted and the actual end-of-year knowledge, and in both cases the estimate is not statistically significant at the 5% level. All other coefficients are very similar, with those on the simulated end-of-year knowledge generally displaying a lower magnitude in virtually every case. These results are not surprising, given that one would expect both 3^{rd} and 4^{th} grade test scores to have similar regression coefficients on the exogenous determinants, and since the (log) specification of K_{1ti} in (1) entails K_{0ti} entering linearly with a discount factor between zero and one. The values reported right under the coefficient estimates represent the boundaries of the 95% confidence interval. As seen, 24 out of 31 estimated coefficients from the simulated end-of-year knowledge fall into these confidence intervals. On the other hand, a more demanding task seems to be replicating the relationship between the measures of teaching effort and exogenous determinants. Columns 3 and 4 of Table 9 compare the regression coefficients of the "Teacher takes time to summarize the lesson" measure (which has the lowest degree of measurement error) on teacher characteristics and some measures of the classroom composition represented by the mean and standard deviation of students' baseline knowledge in the classroom. All the signs are same except for the dummy on the teacher being a subject matter generalist, who is anyway imprecisely estimated. Moreover, 5 out of 9 coefficients fall into the 95% intervals of the respective data estimates. Despite the non-linearities in the relationship between e_t, the classroom distribution of baseline knowledge, and other exogenous characteristics as described in (6a), the model seem to do a good job in replicating their linear relationship as estimated from the actual data.

5.3 Out-of-Sample Validation

I exploit the second-year data of the MET study to perform an out-of-sample validation of the model. In particular, this exercise entails using the estimated parameters to predict second-year outcomes (end-of-year knowledge, effort, and class time allocation) given the teacher and student initial conditions. The sub-sample includes all the teachers participating to the second year of the study who were randomly assigned to a classroom within the same school.[23] The randomization was performed by MET researchers in order to correct for the potential bias in the estimates of teacher fixed effects caused by the non-random assignment of teachers to classrooms, especially when the latter is based on unobservable student or teacher characteristics. In the theoretical framework of this paper, the non-random assignment of teachers in Year 1 of the study would bias the effect of the teacher inputs A_t, $\boldsymbol{\tau}_t$, and e_t on end-of-year knowledge if these were correlated with omitted inputs included in ξ_{ti} even after controlling for student effort and baseline knowledge. Therefore, this exercise allows to me to check indirectly whether the model specification is able to capture the variation underlying the teacher assignment mechanism. Table 10 compares descriptive statistics of selected variables from the first-year sample used for the estimation and the second-year one. There are several major differences between students and teachers in these two samples. First, students in the second-year sample display lower initial and end-of-year knowledge in math, with test scores being about 0.13σ lower than the first-year sample. Second, students are younger in the Year 2 sample, with an average age lower by about 0.6 (equivalent to about seven months). Third, teachers have on average one less year of experience teaching in the district. Fourth, classrooms now include, on average, one more student and their compositions are more skewed towards low-baseline knowledge ones. Indeed, while the fraction of 2^{nd} tercile students is very similar compared to Year 1, classrooms have on average 5 percentage points more students in the 1^{st} tercile and about 5 percentage points less students in the 3^{rd} tercile. All other student and teacher characteristics are very similar between the two samples.

[Table 11 here]

Table 11 compares predicted and actual means and standard deviations using the second year sample. As shown by the values reported, the model does a very good job in predicting all the outcomes outside of the sample used in estimation, and virtually the whole analysis on the discrepancies between data and predictions made in Section 5.2 holds in this case as well. Despite the absence of data on class time allocation in the second-year sample does not allow to assess the model performances in predicting them, it can be noticed that the model predicts an increase of about one percentage point in class time allocated to fractions and decimals and a decrease of the same amount

[23] Not all teachers participating to the study were randomly assigned to a classroom in Year 2. See the MET documentation for further details.

in geometry topics compared to the first year values. Similarly to Section 5.2, Table 12 reports the estimated coefficients of linear regressions of the endogenous variables on the latent factor determinants using both the actual and the simulated Year 2 sub-samples. Interestingly, the results show that the model performs slightly better out-of-sample than within the sample in predicting end-of-year knowledge while the opposite is true for teacher effort. Specifically, only 3 out of 28 estimated coefficients from the simulated end-of-year knowledge falls outside of the 95% confidence interval, and the only two parameters different sign are not statistically significant and entail the dummies on teacher's masters degree and on whether the teacher teaches both Math and English. On the other hand, the regression parameters of the predicted teacher effort on class-level mean and standard deviation of baseline knowledge do not fall in the 95% confidence interval, although they still have same sign compared to the data counterparts. All other estimated parameters are instead very similar between the data and the simulations.

[Table 12 here]

6. Counterfactual Analysis and Policy Experiments

6.1 Tailoring Instruction to Students' Initial Knowledge

Section 2.2.1 shows how the elasticities of time inputs $\boldsymbol{\eta}_q = (\eta_{1q} \ldots \eta_{Jq})$ can be interpreted as the shares of class time allocated to each topic group that, conditional on all other inputs, maximize the end-of-year knowledge of a student with a baseline knowledge in the q^{th} tercile. This implies that the vector $\boldsymbol{\tau}_q^{max} \equiv (\overline{\tau}_t \eta_{1q} \ldots \overline{\tau}_t \eta_{Jq})$, with $\overline{\tau}_t = 100$ (since expressed in percentage points), is the class time allocation *tailored* to students in the q^{th} tercile. In this section I assess the impact of tailoring instruction to students' knowledge on students' learning gains. Table 13 compares both class time allocation and student achievement under the baseline and counterfactual scenario. The first panel illustrates the differences between what teachers would do if they were tailoring instruction to each specific tercile and the actual class time allocation they choose in the data. It is possible to notice that, apart from unit conversion and measurement topics, the average allocations are always in between the minimum and the maximum values of the tailored instruction. Moreover, class time allocation tends to be for the most part skewed towards 1^{st} tercile students, which is consistent with the results on teachers' valuing more the achievement of weaker students. The only exception entails the first topic group, where the average percentage of class time lies in between the tailored instruction values of the 2^{nd} and 3^{rd} terciles, while being further away from the optimal allocation for 1^{st} tercile students.

The second and third panels of Table 13 show the simulated impact of tailoring instruction to each student in the sample, which entail imposing a different class time vector $\boldsymbol{\tau}_t$ to each student by setting it equal to the tailored instruction values. The effect of this counterfactual entails an overall increase in students test score performances at the end to the year of about 0.04σ, with the highest effect being on 1^{st} tercile students (with 0.048σ) and the lowest on 3^{rd} tercile students with 0.034σ. Although it is common to measure policy effects on student achievement by the effect on test scores, it is also interesting to see the effect on the actual amount of learning occurred during the school year. The third panel in Table 13 shows the impact of tailoring instruction to student knowledge on the end-of-year knowledge value added. Overall, the counterfactual yields an increase in learning gains of about 0.12σ, which again is the lowest 1^{st} tercile students with 0.057σ and highest for those falling in the 3^{rd} tercile with 0.21σ. The reason underlying the larger impact on high-baseline knowledge students is twofold. First, as teachers in the data focus more on 1^{st} tercile students, their class time allocation tend to be quite distant from the tailored instruction values to the 3^{rd} tercile. Hence, students in this tercile will benefit more compared to those in the lower part of the knowledge distribution, whose time inputs are already closer to their learning profiles. Second, the positive value taken by the estimated elasticity of baseline knowledge, γ_0, entails that the higher the level of initial knowledge, the more productive will be the inputs. As a result, the positive effect of tailoring instruction to initial knowledge is magnified to a higher degree for students in the higher part of the distribution.

[Table 13 here]

6.2 Ability Tracking and Teacher Assignment Mechanisms

The parameter estimates described in Section 3 suggest that instructional choices are associated to the composition of the classroom. However, neither the direction nor the magnitude of this relationship can be trivially determined by just looking at the parameter values, as many factors could contribute to the teacher's decision-making. For instance, an increase in the share of students in the classrooms with low initial knowledge could induce the teacher to exert more effort given the higher value she attaches to the achievement gains of these students. On the other hand, this effect could be offset by the fact that instruction is less productive for low-knowledge students (given that $\gamma_0 > 0$). Moreover, one should also account for teacher preferences over time spent teaching different topics, the alignment with curriculum standards, class size (which affects the marginal product of teacher effort), the level of teacher ability, as well as the within-classroom distribution of student effort and initial knowledge. In order to shed light on the implications of these mechanisms, I perform a counterfactual experiment involving the reassignment of students and teachers to classrooms known as ability tracking. In particular, tracking consists in assigning students to classrooms (often referred to as *tracks*) based on based on their prior test score performance. A major implication of this policy is that it creates more homogeneous classes, as students with similar levels of baseline knowledge are systematically assigned to the same track. Because of this systematic separation of students, there has been a large debate on the actual benefits of tracking for students outcomes. On the one hand, proponents of this policy claim that tracking benefits all students as it allows teachers to better tailor instruction to students' needs. On the other hand, the opponents argue that tracking generates segregation and might be detrimental for weaker students who would systematically be exposed to lower-achieving peers, and therefore exacerbate inequalities in achievement. Despite the numerous studies, empirical evidence on the effect of tracking on student achievement are still inconclusive and mostly dependent on the specific environment where it is applied. The theoretical framework of this paper is particularly suited to assess the impact of tracking on student achievement as it nests one of the main channels through which tracking can affect student achievement, namely teachers endogenously adjusting instruction in response to the composition of the classroom.

In order to have a full representation of all the 4^{th} grade teachers in each school, I perform this policy experiment using the full sample of 4^{th} grade math classes in the first year of the MET study.[24] In the simulation, I keep the same number of classrooms as the original data and I change class sizes to be the same within each school. Moreover, as there are multiple ways to track students, I choose to simulate the most extreme configuration of tracking, which consists in ordering students (within each school) from lowest to highest based on their 3^{rd} grade math test score and then assigning them sequentially to each classroom. In practice, schools can choose to implement tracking with different intensities depending on the importance they give to prior performance in the assignment of students to classrooms. One way to measure the degree of ability tracking in a school is by computing the share of total variance of baseline test scores due to between-classroom variation. In fact, schools that track students will display a higher share of between-classroom variation due to a lower within-classroom variance (given by the higher classroom homogeneity). At the extreme, the share of between-classroom variance can range from zero or close to zero to a maximum given by the value taken under the tracking configuration simulated in this section.[25] Table 14 reports descriptive statistics of tracking intensity in the sample at the school level. The comparison between the average share of between-classroom variation in the data with the tracking policy experiment configuration suggests that schools participating to the MET study apply a very low level of tracking in 4^{th} grade math classrooms. This is confirmed by the descriptive statistics on the ratio between these two shares (third row), which gives a scale of the degree of tracking relative to the most extreme case. With 1 being the maximum, schools display an average degree of tracking of about 0.12, with 75% of all schools scoring at or below 0.145.[26]

[24] I therefore include teachers not entering the estimation sample since they did not take the SEC survey (and therefore did not report their class time allocation).

[25] Its value can be zero whenever it is possible to allocate students such that the classroom-level test score averages are identical to each other. Clearly, this is possible only for special configurations of the within-school sample.

[26] To further assess the level of tracking in the data I simulated a scenario with random assignment of students to classrooms. The resulting values of instruction and achievement are very similar to the ones in the data.

Although ability tracking per se does not explicitly prescribe anything about how teachers are assigned to classroom, the latter is a potentially crucial aspect of this policy experiment. In practice, there is evidence that schools make these assignments in a non-random fashion, usually based on both teacher and student characteristics. For instance, Kalogrides et al. (2013) show that in a large school district in Florida schools tend to assign more educated and/or experienced teachers to high-achieving students, and either female or minority teachers to lower-achieving ones. In the present paper, I take a slightly different direction and use the level of teacher ability as a discriminant for their assignment to classrooms. In particular, I simulate the tracking policy with three alternative teacher assignment mechanisms: 1) random assignment to classrooms; 2) negative assortative matching (i.e. higher ability teachers to lower tracks); 3) positive assortative matching (i.e. higher ability teachers to upper tracks). Results of the simulated counterfactuals are shown in Table 15.[27] The values reported in the last three columns for teacher effort and class time allocation represent averages conditional on students baseline knowledge terciles. For instance, the first column shows the average class time allocation and teacher effort experienced by students with baseline knowledge in the first tercile. There are several patterns that emerge from these results. First, the effect of tracking on instructional choices is quite similar across teacher assignment mechanisms. Second, teachers respond to tracking by tailoring instruction to students' baseline knowledge. This can be seen by comparing the simulated results with the baseline time inputs and the respective tailored values in Table 13. Indeed, the average class time allocation across topics for each tercile is now much closer to the tailored values compared to the baseline values. Third, teacher exert more effort when assigned to lower tracks and less when assigned to high baseline knowledge students. In particular, under tracking students in the first tercile see an increase in teacher effort by about 1σ, while those in the third tercile see an decrease of the same magnitude. Students in the middle of the distribution experience only a small increase in teacher effort.

Fourth, the effect of tracking on achievement is positive on average, but its distribution across student terciles depends on the teacher assignment mechanism. Specifically, tracking increases the average 4^{th} grade test score by about 0.01σ with negative assortative matching and by 0.003σ with positive assortative matching. However, both assignment mechanisms have opposite effects on students in the tails of the initial knowledge distribution. Assigning the best teachers to the lower tracks yields a positive impact on 1^{st} tercile students of about 0.027σ and a negative effect on 3^{rd} tercile students of about 0.024σ. On the other hand, assigning the best teachers to the higher knowledge students and and the lower ability teachers to the lower tracks slightly benefits students in the 3^{rd} tercile ($+0.001\sigma$) at the detriment of students in the 1^{st} tercile who end up with a lower test score by

[27] Results are based on a total of 8,000 simulated classrooms. Given the missing information on class time allocation for a large part of the teachers in the sample, the baseline values are also simulated given the assignments of students and teachers to classrooms in the data.

about 0.012σ. Interestingly, students in the middle of the distribution always benefit from tracking, although the effect is higher under negative assortative matching. Indeed, end-of-year knowledge of 2^{nd} tercile students increases by 0.025σ and 0.015σ under negative and positive assortative matching, respectively. Finally, as one would expect the effects of tracking under random assignment of teachers lie in between the impacts of the two mechanisms, although their signs are the same as those under negative assortative matching. In particular, randomly assigning teachers to classrooms leads to an overall increase in test scores of 0.006σ, which disaggregated by student prior knowledge entails an increase in performances in the 1^{st} tercile and 2^{nd} terciles of 0.001σ and 0.02σ, respectively, and a slight decrease in 3^{rd} tercile students mean test scores of 0.005σ.

A joint interpretation of all these results allows to depict a clearer picture about the mechanisms at work. First, tracking drastically increases student homogeneity within each classroom, which induces teachers to better tailor instruction to students' initial knowledge. On the other hand, teacher effort adjusts differently depending on the track the teacher is assigned to. In particular, their preference for learning gains of weaker students drives them to increase effort when assigned to lower tracks and to decrease it in higher tracks. As a result, while students would benefit from classroom instruction being better suited to their preparedness, the change in teacher effort would end up benefiting lower-knowledge students and hurt the high-achieving ones. However, these effects could be either magnified or offset by the level of teacher ability. In particular, assigning the best teachers to lower tracks and the low-ability teachers to the higher tracks exacerbates the effect of tracking via the instruction adjustment, whereas positive assortative matching would offset and then invert the effects from the effort adjustment for students in the tails of the knowledge distribution. These results are in line with the implications derived by Duflo et al. (2011) regarding teachers better tailoring instruction to students' needs under tracking. However, my results are not able to give a univocal answer on whether tracking is has a positive or negative effect on students, since its impact on achievement at different levels of initial knowledge strictly depends on how teachers are assigned to classrooms.

6.3 The Impact of Curriculum Standards on Instruction and Achievement

Since the early 1990s, many education policies have been involved in the establishment of educational standards. The main objective of these standards is to establish a common benchmark for student proficiency across schools, as well as to provide teachers with guidelines on how to structure their curricula and pace of instruction. Despite their importance in the education policy agenda, empirical evidence on the effectiveness of curriculum standards in raising student achievement is still inconclusive. Hence, a first order question is whether the existing curriculum standards are set at a level which would actually foster student achievement. To do that, I simulate a counterfactual where I impose teachers to teach according to the standards in their state. Formally, I set $\boldsymbol{\tau}_t = \varphi_t$ for each s

and t. The simulation results are reported in Table 16. Comparing the counterfactual scenario with the status quo (where teachers are free to choose their time allocation) it is possible to see that teaching according to the state-level standards would be slightly detrimental for students in the 1^{st} and 3^{rd} terciles but quite beneficial for students in the middle of the knowledge distribution. Interestingly, the increase in achievement for 2^{nd} tercile students occurs despite the overall decrease in teacher effort. Hence, these results suggest that curriculum standards in the 5 states represented in the MET data are more suited to students with an average level of initial knowledge.

7. Conclusion

This paper explores the relationship between instruction, classroom composition, and student knowledge accumulation by developing and estimating a model of endogenous instructional decisions. Teachers try to maximize expected utility by choosing how much effort to exert in class as well as the content and pace of instruction through the allocation of class time among different topics. The model specifies a technology of knowledge production and allows teachers decisions to depend on how much they value student achievement. For the empirical part of the analysis I use a subsample of 4^{th} grade math classes from the first year of the MET project data. The estimation is carried out through maximum simulated likelihood. Estimates of the model suggest that teachers attach higher values to learning gains of students with lower levels of initial knowledge. These results are consistent with existing empirical evidence on teachers' behavioral response to incentive programs implemented in US schools. Moreover, students with different baseline knowledge display different learning profiles, as shown by the difference in the quantile-specific time elasticity parameters. The model fits the data very well, and an out-of-sample validation exercise using data from the second year of MET study suggests that the non-random assignment of teachers in the first year does not affect significantly the estimates. A first counterfactual analysis shows that student end-of-year knowledge can increase by 0.04σ on average if instruction were tailored to their baseline knowledge. Then, I explore the implications of accounting for the endogenous response of teachers to the classroom composition by performing a policy experiment involving ability tracking of students. Teachers respond to tracking by better tailoring instruction to the students readiness level, while also exerting more effort when assigned to lower tracks compared to higher tracks. On the other hand, the impact of tracking on student achievement depends critically on how teachers are assigned to classrooms. While my findings on teachers' behavioral response to tracking are similar to Duflo et al. (2011), their novelty lies in the fact that they take into account a factor often disregarded in the tracking literature, namely the teacher assignment mechanism. Finally, a second set of counterfactual experiments suggest that existing curriculum standards are mainly tailored to students in the middle of the baseline achievement distribution.

References

Agodini, Roberto and Barbara Harris, "How Four Elementary Math Curricula Perform Among Different Types of Teachers and Classrooms," Technical Report 2014.

Agostinelli, Francesco and Matthew Wiswall, "Identification of dynamic latent factor models: The implications of re-normalization in a model of child development," Technical Report, National Bureau of Economic Research 2016.

Andrew, Michael D, Casey D Cobb, and Peter J Giampietro, "Verbal ability and teacher effectiveness," *Journal of teacher education*, 2005, *56* (4), 343–354.

Aucejo, Esteban M, Patrick Coate, Jane Fruehwirth, Sean Kelly, and Zachary Mozenter, "Teacher effectiveness and classroom composition," Technical Report, CEPR Discussion Paper No. DP13166 2018.

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukherji, Marc Shotland, and Michael Walton, "Mainstreaming an effective intervention: Evidence from randomized evaluations of Teaching at the Right Level in India," Technical Report, National Bureau of Economic Research 2016.

Barlevy, Gadi and Derek Neal, "Pay for percentile," *American Economic Review*, 2012, *102* (5), 1805–31.

Ben-Porath, Yoram, "The production of human capital and the life cycle of earnings," *Journal of political economy*, 1967, *75* (4, Part 1), 352–365.

Bodovski, Katerina and George Farkas, "Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and instruction," *The Elementary School Journal*, 2007, *108* (2), 115–130.

Brock, William A and Steven N Durlauf, "Interactions-based models," in "Handbook of econometrics," Vol. 5, Elsevier, 2001, pp. 3297–3380.

Burgess, Simon M, "Human capital and education: The state of the art in the economics of education," Technical Report 2016.

Carlana, Michela, "Implicit stereotypes: Evidence from teachers gender bias," *The Quarterly Journal of Economics*, 2019, *134* (3), 1163–1224.

Connor, Carol McDonald and Frederick J Morrison, “Individualizing student instruction in reading: Implications for policy and practice,” *Policy insights from the behavioral and brain sciences*, 2016, *3* (1), 54–61.

— **, Michèle MM Mazzocco, Terri Kurz, Elizabeth C Crowe, Elizabeth L Tighe, Taffeta S Wood, and Frederick J Morrison**, “Using assessment to individualize early mathematics instruction,” *Journal of school psychology*, 2018, *66*, 97–113.

Cunha, Flavio, James J Heckman, and Susanne M Schennach, “Estimating the technology of cognitive and noncognitive skill formation,” *Econometrica*, 2010, *78* (3), 883–931.

Darling-Hammond, Linda and Peter Youngs, “Defining highly qualified teachers: What does scientifically-based research actually tell us?,” *Educational researcher*, 2002, *31* (9), 13–25.

Deming, David J and David Figlio, “Accountability in US education: Applying lessons from K-12 experience to higher education,” *Journal of Economic Perspectives*, 2016, *30* (3), 33–56.

Donaldson, Morgaen L, Kimberly LeChasseur, and Anysia Mayer, “Tracking instructional quality across secondary mathematics and English Language Arts classes,” *Journal of Educational Change*, 2017, *18* (2), 183–207.

Duflo, Esther, Pascaline Dupas, and Michael Kremer, “Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya,” *American Economic Review*, 2011, *101* (5), 1739–74.

Engel, Mimi, Amy Claessens, and Maida A Finch, “Teaching students what they already know? The (mis) alignment between mathematics instructional content and student knowledge in kindergarten,” *Educational Evaluation and Policy Analysis*, 2013, *35* (2), 157–178.

— **, — , Tyler Watts, and George Farkas**, “Mathematics content coverage and student learning in kindergarten,” *Educational researcher*, 2016, *45* (5), 293–300.

Fu, Chao and Nirav Mehta, “Ability tracking, school and parental effort, and student achievement: A structural model and estimation,” *Journal of Labor Economics*, 2018, *36* (4), 923–979.

Gershenson, Seth, Cassandra Hart, Joshua Hyman, Constance Lindsay, and Nicholas W Papageorge, “The long-run impacts of same-race teachers,” Technical Report, National Bureau of Economic Research 2018.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin, “Many children left behind? Textbooks and test scores in Kenya,” *American Economic Journal: Applied Economics*, 2009, *1* (1), 112–35.

Goldberger, Arthur S, "Structural equation methods in the social sciences," *Econometrica: Journal of the Econometric Society*, 1972, pp. 979–1001.

Kalogrides, Demetra, Susanna Loeb, and Tara Béteille, "Systematic sorting: Teacher characteristics and class assignments," *Sociology of Education*, 2013, *86* (2), 103–123.

Kane, Thomas J, Daniel F McCaffrey, Trey Miller, and Douglas O Staiger, "Have we identified effective teachers? Validating measures of effective teaching using random assignment," in "Research Paper. MET Project. Bill & Melinda Gates Foundation" Citeseer 2013.

Kilpatrick, J., W.G. Martin, D. Schifter, and National Council of Teachers of Mathematics, *A Research Companion to Principles and Standards for School Mathematics*, National Council of Teachers of Mathematics, 2003.

Kilpatrick, Jeremy, Jane Swafford, Bradford Findell, and National Research Council, *Adding It Up: Helping Children Learn Mathematics*, Washington, DC: The National Academies Press, 2001.

Macartney, Hugh, Robert McMillan, and Uros Petronijevic, "Incentive design in education: An empirical analysis," Technical Report, National Bureau of Economic Research 2015.

__ , __ , and __ , "Incentive Design in Education: A Distributional Analysis," Technical Report 2020.

Neal, Derek and Diane Whitmore Schanzenbach, "Left behind by design: Proficiency counts and test-based accountability," *The Review of Economics and Statistics*, 2010, *92* (2), 263–283.

Polikoff, Morgan S and Andrew C Porter, "Instructional alignment as a measure of teaching quality," *Educational Evaluation and Policy Analysis*, 2014, *36* (4), 399–416.

__ , __ , and John Smithson, "How well aligned are state assessments of student achievement with state content standards?," *American Educational Research Journal*, 2011, *48* (4), 965–995.

Porter, Andrew C and John L Smithson, "Defining, Developing, and Using Curriculum Indicators. CPRE Research Report Series.," Technical Report 2001.

Pritchett, Lant and Amanda Beatty, "The negative consequences of overambitious curricula in developing countries," Technical Report 293 2012.

Reback, Randall, "Teaching to the rating: School accountability and the distribution of student achievement," *Journal of public economics*, 2008, *92* (5-6), 1394–1415.

Sacerdote, Bruce, "Peer effects in education: How might they work, how big are they and how much do we know thus far?," in "Handbook of the Economics of Education," Vol. 3, Elsevier, 2011, pp. 249–277.

Todd, Petra and Kenneth I Wolpin, "Accounting for Mathematics Performance of High School Students in Mexico: Estimating a Coordination Game in the Classroom," *Journal of Political Economy*, 2018, *126* (6), 2608–2650.

Tomlinson, Carol Ann, Catherine Brighton, Holly Hertberg, Carolyn M Callahan, Tonya R Moon, Kay Brimijoin, Lynda A Conover, and Timothy Reynolds, "Differentiating instruction in response to student readiness, interest, and learning profile in academically diverse classrooms: A review of literature," *Journal for the Education of the Gifted*, 2003, *27* (2-3), 119–145.

Xue, Yange and Samuel J Meisels, "Early literacy instruction and learning in kindergarten: Evidence from the early childhood longitudinal studykindergarten class of 1998–1999," *American educational research journal*, 2004, *41* (1), 191–229.

Table 1: Class Time Allocation, State Curriculum Standards, and Classroom Composition

	Mean	St.Dev	p25	Median	p75
Panel A: Classroom instruction (% of total class time)					
Place value, rounding, addition, and subtraction	6.797	2.165	5.614	6.678	7.567
Multi-digit multiplication and division	25.981	6.246	21.967	25.229	28.899
Shapes, angles, and geometry	17.513	5.135	14.014	17.448	21.078
Fractions and decimals	25.546	5.367	22.362	25.464	29.427
Unit conversion and measurement	24.163	6.182	20.472	24.627	28.352
Panel B: Curriculum standards (% of test content)					
Place value, rounding, addition, and subtraction	7.967	2.632	5.466	6.525	9.801
Multi-digit multiplication and division	29.483	6.023	29.747	31.125	33.270
Shapes, angles, and geometry	15.910	5.157	10.075	18.705	20.570
Fractions and decimals	25.617	9.222	16.433	25.804	37.791
Unit conversion and measurement	21.024	7.851	13.398	19.219	23.418
Panel C: Instruction-standards discrepancy (relative to standards)					
Place value, rounding, addition, and subtraction	−0.075	0.365	−0.347	−0.095	0.144
Multi-digit multiplication and division	−0.077	0.302	−0.283	−0.160	0.073
Shapes, angles, and geometry	−0.265	0.681	−0.186	0.008	0.646
Fractions and decimals	0.140	0.491	−0.290	−0.002	0.544
Unit conversion and measurement	0.279	0.489	−0.143	0.286	0.585
Panel D: Classroom Composition					
Class size	23.29	4.97	20	23	26
Students baseline knowledge (3^{rd} grade test score):					
% low-level (1^{st} tercile)	30.20	21.80	14.30	26.30	46.40
% mid-level (2^{nd} tercile)	33.20	13.20	24.10	34.80	41.70
% high-level (3^{rd} tercile)	36.60	26.80	14.30	30.20	52.00

Table 2: Student and Teacher Characteristics

	Mean	Std.Dev		Mean
Panel A: Students (Obs. = 2532)				
Age	9.52	0.50	Gifted	0.06
Male	0.48		Special education (SpEd)	0.09
White	0.26		English language learner (ELL)	0.17
Black	0.43		Reduced price/free lunch	0.45
Hispanic	0.25			
N. books in bedroom:			N. computers at home:	
None	0.09		None	0.12
≥1 and ≤10	0.22		One	0.45
≥11 and ≤24	0.21		More than one	0.43
≥25	0.48			
Has person at home to help with homework:			Has no quiet place to study at home:	
Never	0.02		Never	0.40
Mostly not	0.03		Mostly not	0.12
Sometimes	0.09		Sometimes	0.16
Mostly	0.17		Mostly	0.12
Always	0.69		Always	0.2
Panel B: Teachers (Obs. = 101)				
Years of experience in the district	6.40	5.94		
Masters degree	0.53			
Teaches both Math and ELA (generalist)	0.83			

Table 3: Latent Factors Measures

Student knowledge:	Mean	Std.Dev			
3^{rd} grade math state test score (rescaled)	0.102	0.954			
4^{th} grade math state test score (rescaled)	0.306	0.949			
Teacher ability					
CLASS Behavior management (1-7 scale)	5.943	0.715			
CLASS Content understanding (1-7 scale)	4.137	0.481			
CLASS Productivity (1-7 scale)	5.918	0.555			
FFTM Management of class procedures (1-4 scale)	2.763	0.354			
FFTM Management of student behavior (1-4 scale)	2.840	0.344			
MQI Richness of mathematics (1-3 scale)	1.340	0.261			
MQI Mathematical knowledge for teaching (MKT) score (1-3 scale)	2.030	0.218			
Teacher explains clearly (survey 0-4 scale)	3.321	0.295			
Teacher controls class behavior (survey 0-4 scale)	2.251	0.437			
Teacher explains in orderly way (survey 0-4 scale)	3.180	0.300			
Teacher can explain in several ways (survey 0-4 scale)	3.216	0.295			
Teacher effort					
Teacher explains in another way if we do not understand (survey 0-4 score)	3.325	0.285			
Teacher pushes us to work hard (survey 0-4 score)	3.092	0.370			
Teacher does not waste time in class (survey 0-4 score)	2.664	0.385			
Teacher asks us if we understand the lesson (survey 0-4 score)	3.329	0.315			
Teacher asks us if we are following along (survey 0-4 score)	3.440	0.277			
Teacher writes feedback on our papers (survey 0-4 score)	2.887	0.387			
Teacher takes the time to summarize the lesson (survey 0-4 score)	2.813	0.480			
Teacher encourage us to do our best (survey 0-4 score)	3.533	0.257			
Student effort	Never	Mostly not	Sometimes	Mostly	Always
I have done my best quality work in this class	0.007	0.010	0.089	0.242	0.457
In this class, I stop trying when the work gets hard	0.488	0.119	0.103	0.046	0.048
In this class, I take it easy and do not try to do my best	0.427	0.096	0.090	0.066	0.119
	None	Some	Most	All	All plus extra
How much homework do you usually complete?	0.006	0.062	0.106	0.489	0.137

Table 4: Correlation Matrices: Latent Factors Measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Teacher ability (Pearson's correlation coeff.)											
CLASS Behavior management (1)	1.000										
p-value											
CLASS Content understanding (2)	0.383	1.000									
p-value	0.001										
CLASS Productivity (3)	0.810	0.495	1.000								
p-value	0.000	0.000									
FFTM Management of class procedures (4)	0.591	0.381	0.486	1.000							
p-value	0.000	0.001	0.000								
FFTM Management of student behavior (5)	0.680	0.338	0.450	0.753	1.000						
p-value	0.000	0.004	0.000	0.000							
MQI Richness of mathematics (6)	0.307	0.413	0.348	0.197	0.161	1.000					
p-value	0.009	0.000	0.003	0.102	0.182						
MQI Mathematical knowledge for teaching (MKT) score (7)	0.321	0.321	0.283	0.332	0.326	0.347	1.000				
p-value	0.007	0.007	0.018	0.005	0.006	0.003					
Teacher explains clearly (8)	0.119	0.101	0.179	0.155	0.074	0.055	−0.108	1.000			
p-value	0.325	0.407	0.139	0.201	0.545	0.649	0.374				
Teacher controls class behavior (9)	0.332	0.291	0.427	0.306	0.292	0.110	0.170	0.341	1.000		
p-value	0.005	0.014	0.000	0.001	0.014	0.364	0.159	0.000			
Teacher explains in orderly way (10)	0.199	0.061	0.199	0.025	0.039	0.127	0.007	0.587	0.325	1.000	
p-value	0.098	0.616	0.099	0.838	0.747	0.294	0.955	0.000	0.001		
Teacher can explain in several ways (11)	0.173	0.255	0.353	0.191	0.156	0.180	0.062	0.564	0.385	0.557	1.000
p-value	0.153	0.033	0.003	0.113	0.198	0.135	0.609	0.000	0.000	0.000	
Teacher effort (Pearson's correlation coeff.)											
Teacher explains in another way if we do not understand (1)	1.000										
p-value											
Teacher pushes us to work hard (2)	0.207	1.000									
p-value	0.041										
Teacher does not waste time in class (3)	0.150	0.319	1.000								
*p-value*e	0.140	0.001									
Teacher asks us if we understand the lesson (4)	0.365	0.360	0.167	1.000							
p-value	0.000	0.000	0.101								
Teacher asks us if we are following along (5)	0.409	0.380	0.016	0.609	1.000						
p-value	0.000	0.000	0.876	0.000							
Teacher writes feedback on our papers (6)	0.372	0.238	0.102	0.178	0.261	1.000					
p-value	0.000	0.018	0.319	0.079	0.009						
Teacher takes the time to summarize the lesson (7)	0.452	0.357	0.248	0.442	0.448	0.459	1.000				
p-value	0.000	0.000	0.014	0.000	0.000	0.000					
Teacher encourage us to do our best (8)	0.425	0.286	0.405	0.233	0.324	0.281	0.214	1.000			
p-value	0.000	0.004	0.000	0.021	0.001	0.005	0.035				
Student input (Pearson's χ^2)											
I have done my best quality work in this class (1)	.										
p-value	.										
In this class, I stop trying when the work gets hard (2)	240.90	.									
p-value	0.00										
In this class, I take it easy and do not try to do my best (3)	276.08	393.21	.								
p-value	0.00	0.00									
How much homework do you usually complete? (4)	143.23	108.17	92.01	.							
p-value	0.00	0.00	0.00								

Table 5: Production and Utility Functions Parameter Estimates

Panel A: Production Function Parameters					
Parameter	Value	Std.Err.	Parameter	Value	Std.Err.
	(1)	(2)		(3)	(4)
δ_0	0.6107	0.0552	γ_0	0.4471	0.0542
δ_{11}	0.0286	0.0028	γ_1	0.0451	0.0296
δ_{12}	0.0270	0.0029	γ_{20}	0.0104	0.0011
δ_{13}	0.0213	0.0024	γ_{21}	−0.0002	3.2E−05
δ_{14}	0.0262	0.0028	γ_3	0.3729	0.0517
δ_{15}	0.0317	0.0033	σ_ξ	0.5238	0.0077

Panel B: Utility Function Parameters					
Parameter	Value	Std.Err.	Parameter	Value	Std.Err.
	(1)	(2)		(3)	(4)
ω_1^1	25.9664	0.5340	$\widetilde{\alpha}_{11}$	0.0391	0.1686
ω_1^2	10.8909	0.1922	$\widetilde{\alpha}_{12}$	−0.1598	0.2297
ω_1^3	1.8399	0.3863	$\widetilde{\alpha}_{13}$	0.1726	0.5700
ω_{21}	0.8134	0.3863	$\widetilde{\alpha}_{14}$	−0.4565	0.1557
ω_{22}	1.0964	0.4283	α_{21}	0.0772	0.0223
α_{22}	0.0138	0.0122	α_{23}	0.3209	0.0856
α_{24}	−0.0040	0.0073	α_{25}	0.0448	0.0109

Table 6: Preference Shocks Variance-Covariance Matrix

	$\widetilde{\varepsilon}_{11}$	$\widetilde{\varepsilon}_{12}$	$\widetilde{\varepsilon}_{13}$	$\widetilde{\varepsilon}_{14}$
$\widetilde{\varepsilon}_{11}$	1.0000			
$\widetilde{\varepsilon}_{12}$	0.9277	1.2440		
	(0.1072)	(0.1078)		
$\widetilde{\varepsilon}_{13}$	−0.2023	−0.4772	7.7622	
	(0.2511)	(0.5500)	(0.3374)	
$\widetilde{\varepsilon}_{14}$	0.7654	0.9864	−0.4105	1.0852
	(0.1220)	(0.0895)	(0.0687)	(0.0834)

Table 7: Estimation Results: Elasticity of Class Time Inputs Parameters

Parameter	Topic group	Estimate (by tercile) $q=1$	$q=2$	$q=3$
η_{1q}	Place value, rounding, addition, and subtraction	0.0339	0.0824	0.0436
		(0.0070)	(0.0121)	(0.0109)
η_{2q}	Multi-digit multiplication and division	0.1665	0.3128	0.2938
		(0.0275)	(0.0402)	(0.0547)
η_{3q}	Shapes, angles, and geometry	0.2233	0.1524	0.1096
		(0.0465)	(0.0525)	(0.0734)
η_{4q}	Fractions and decimals	0.2716	0.2947	0.3545
		(0.0325)	(0.0399)	(0.0448)
η_{5q}	Unit conversion and measurement	0.3047	0.1577	0.1986

standard errors in parenthesis

Table 8: Within-Sample Model Fit

	Data		Model		
	Mean	Std.Dev.	Mean	Std.Dev.	$\sigma_{\text{noise}}/\sigma_{\text{total}}$
Knowledge measures:					
4^{th} grade math state test score (rescaled)	0.306	0.949	0.300	0.940	
Class time topic area (% of total class time):					
Place value, rounding, addition, and subtraction	6.786	2.192	6.681	2.265	
Multi-digit multiplication and division	25.880	6.217	26.066	7.823	
Shapes, angles, and geometry	17.617	5.124	17.486	7.436	
Fractions and decimals	25.498	5.379	25.692	6.856	
Unit conversions and measurement	24.218	6.240	24.627	7.180	
Teacher effort					
Teacher explains in another way if class does not understand	3.325	0.285	3.322	0.286	0.581
Teacher pushes students to work hard	3.092	0.370	3.101	0.383	0.699
Teacher does not waste time	2.664	0.385	2.653	0.417	0.955
Teacher asks questions to make sure students understand	3.329	0.315	3.330	0.327	0.478
Teacher asks if students are following along	3.440	0.277	3.440	0.284	0.417
Teacher writes feedback on paper	2.887	0.387	2.878	0.379	0.750
Teacher takes time to summarize lesson	2.813	0.480	2.829	0.500	0.511
Teacher encourages students to do their best	3.533	0.257	3.521	0.268	0.780
Teacher ability					
CLASS Behavior management scale	5.943	0.715	6.111	0.685	0.176
CLASS Content understanding scale	4.137	0.481	4.179	0.453	0.832
CLASS Productivity scale	5.918	0.555	6.030	0.539	0.320
FFTM Management of class procedures score	2.763	0.354	2.813	0.357	0.430
FFTM Management of student behavior score	2.840	0.344	2.924	0.346	0.443
MQI Richness of mathematics score	1.340	0.261	1.343	0.250	0.959
MQI Mathematical knowledge for teaching (MKT) score	2.030	0.218	2.043	0.222	0.899
Teacher explains clearly (0-4 score)	3.321	0.295	3.354	0.295	0.920
Teacher controls class behavior (0-4 score)	2.251	0.437	2.286	0.449	0.792
Teacher explains in orderly way (0-4 score)	3.180	0.300	3.195	0.293	0.910
Teacher can explain in several ways (0-4 score)	3.216	0.295	3.233	0.293	0.904
Student effort					
I have done my best quality work in this class	3.406	0.801	3.179	0.948	0.766
In this class, I stop trying when the work gets hard	0.817	1.214	0.904	1.240	0.594
In this class, I take it easy and do not try to do my best	1.190	1.512	0.935	1.364	0.726
How much homework do you usually complete?	3.862	0.814	4.048	0.710	0.919

Table 9: Within-Sample Model Fit: Explanatory Variables (OLS regression)

	End-of-Year Knowledge: 4^{th} Grade Math Test Score			Teacher Effort: Teacher Takes Time to Summarize the Lesson		
	Data		Model	Data		Model
	Estimate	95% CI	Estimate	Estimate	95% CI	Estimate
Classroom-level regressors:						
Baseline Knowledge: Class average	0.3550∗∗∗	(0.2550, 0.4550)	0.2640∗∗∗	−0.2620∗∗∗	(−0.3443, −0.1797)	−0.1640∗∗∗
Baseline Knowledge: Class Std.Dev.	−0.1510∗∗∗	(-0.1961, -0.1059)	−0.0810∗∗∗	0.1170	(−0.0241, 0.2581)	0.0520∗∗
Teacher experience in the district	0.0004	(-0.0055, 0.0063)	0.0010	−0.0120∗∗	(−0.0218, −0.0022)	−0.0040
Teacher has Master's degree	0.0001	(-0.0646, 0.0648)	0.0050	0.0360	(−0.1169, 0.1889)	0.0120
Teacher teaches both Math and ELA	0.1290∗∗	(0.0290, 0.2290)	0.060∗∗∗	−0.0110	(−0.0894, 0.0674)	0.0530
Curriculum standards (% of questions):						
Place value, rounding, addition, and subtr.	0.0300∗∗∗	(0.0143, 0.0457)	0.0540∗∗∗	−0.0540∗∗∗	(−0.0716, −0.0364)	0.0230∗∗∗
Multi-digit multiplication and division	−0.0250∗∗∗	(-0.0309, -0.0191)	−0.0190∗∗∗	0.0390∗∗∗	(0.0312, 0.0468)	0.0210∗∗∗
Shapes, angles, and geometry	−0.0380∗∗∗	(-0.0537, -0.0223)	−0.0410∗∗∗	0.013∗∗∗	(0.0052, 0.0208)	0.0110∗∗
Fractions and decimals	0.0004	(-0.0035, 0.0043)	0.0010	−0.0230∗∗	(−0.0289, −0.0171)	0.0020∗∗∗
Student-level regressors:						
Male	0.0120	(-0.0527, 0.0767)	−0.0270			
Gifted	0.7420∗∗∗	(0.6538, 0.8302)	0.6250∗∗∗			
Special education (SpEd)	−0.4030∗∗∗	(-0.5382, -0.2678)	−0.3380∗∗∗			
English language learner (ELL)	−0.0430	(-0.2116, 0.1256)	−0.0610∗			
Free or reduced price lunch	0.0120	(-0.1095, 0.1335)	−0.0004			
Black	−0.4810∗∗∗	(-0.6104, -0.3516)	−0.3140∗∗∗			
Hispanic	−0.2420∗∗∗	(-0.3518, -0.1322)	−0.1960∗∗∗			
Age	−0.1070∗∗∗	(-0.1736, -0.040)	−0.0860∗∗			
N. books in bedroom (Base = "None"):						
≥1 and ≤10	0.3130∗∗∗	(0.1190, 0.5070)	0.0930			
≥11 and ≤24	0.3380∗∗∗	(0.1851, 0.4909)	0.1630∗∗∗			
≥25	0.2980∗∗∗	(0.1138, 0.4822)	0.1890∗∗∗			
N. computers at home (Base = "None"):						
One	0.1040∗∗∗	(0.0334, 0.1746)	0.0850∗∗∗			
More than one	0.1940∗∗∗	(0.0980, 0.2900)	0.1460∗∗∗			
Has person at home to help with homework (Base = "Never"):						
Mostly not	0.4860∗∗∗	(0.2626, 0.7094)	0.3130∗∗∗			
Sometimes	0.1830∗∗	(0.0399, 0.3261)	0.1870∗∗∗			
Mostly	0.2770∗∗∗	(0.0869, 0.4671)	0.2870∗∗∗			
Always	0.1780∗∗∗	(0.0810, 0.2730)	0.2290∗∗∗			
Has no quiet place to study at home (Base = "Never"):						
Mostly not	0.0560	(-0.0577, 0.1697)	0.0430			
Sometimes	−0.1050∗∗	(-0.1912, -0.0188)	−0.0560			
Mostly	−0.1790∗∗∗	(-0.2652, -0.0928)	−0.1070			
Always	−0.1630∗∗∗	(0.0334, 0.1746)	−0.1370∗∗∗			
Constant	0.1601∗∗∗	(0.7386, 2.4634)	1.1960∗∗			

Levels of significance: * p<0.10, ** p<0.05, *** p<0.01

Table 10: Year 1 and Year 2 Samples Comparison - Selected Descriptive Statistics

	Year 1		Year 2	
	Mean	Std.Dev	Mean	Std.Dev
Obs.	2352		4452	
Student knowledge:				
3^{rd} grade math state test score (rescaled)	0.102	0.954	−0.023	0.991
4^{th} grade math state test score (rescaled)	0.306	0.949	0.194	0.961
Student characteristics:				
Age	9.52	0.50	8.91	0.81
Male	0.48		0.50	
White	0.26		0.23	
Black	0.43		0.46	
Hispanic	0.25		0.24	
Gifted	0.06		0.06	
Special education (SpEd)	0.09		0.11	
English language learner (ELL)	0.17		0.14	
Reduced price/free lunch	0.45		0.49	
Teacher characteristics:				
Obs.	177			
Years of experience in the district	6.40	5.94	5.42	4.59
Masters degree	0.53		0.46	
Teaches both Math and ELA (generalist)	0.83		0.82	
Classroom Composition				
Class size	23.29	4.97	24.15	6.29
Students baseline knowledge (3^{rd} grade test score):				
% low-level (1^{st} tercile)	30.20	21.80	35.20	25.70
% mid-level (2^{nd} tercile)	33.20	13.20	32.90	13.40
% high-level (3^{rd} tercile)	36.60	26.80	31.90	24.80

Table 11: Out-of-Sample Validation (Year 2 Data Sample Fit)

	Data		Model	
	Mean	Std.Dev.	Mean	Std.Dev.
Knowledge measures:				
4^{th} grade math state test score (rescaled)	0.194	0.961	0.188	0.940
Class time topic area (% of total class time):				
Place value, rounding, addition, and subtraction			6.512	2.135
Multi-digit multiplication and division			25.846	7.283
Shapes, angles, and geometry			16.613	7.096
Fractions and decimals			26.388	6.324
Unit conversions and measurement			24.642	6.842
Teacher effort				
Teacher explains in another way if class does not understand	3.328	0.299	3.345	0.314
Teacher pushes students to work hard	3.192	0.427	3.134	0.400
Teacher does not waste time	2.709	0.402	2.647	0.414
Teacher asks questions to make sure students understand	3.392	0.322	3.373	0.365
Teacher asks if students are following along	3.502	0.252	3.481	0.328
Teacher writes feedback on paper	2.959	0.468	2.904	0.410
Teacher takes time to summarize lesson	2.981	0.434	2.888	0.550
Teacher encourages students to do their best	3.600	0.289	3.544	0.279
Teacher ability				
CLASS Behavior management scale	5.803	0.512	6.273	0.656
CLASS Content understanding scale	4.120	0.496	4.233	0.448
CLASS Productivity scale	5.803	0.419	6.164	0.536
FFTM Management of class procedures score	2.691	0.346	2.879	0.349
FFTM Management of student behavior score	2.767	0.380	2.976	0.339
MQI Richness of mathematics score	1.353	0.263	1.359	0.258
MQI Mathematical knowledge for teaching (MKT) score	2.027	0.225	2.060	0.221
Teacher explains clearly (0-4 score)	3.324	0.269	3.374	0.296
Teacher controls class behavior (0-4 score)	2.211	0.506	2.342	0.443
Teacher explains in orderly way (0-4 score)	3.229	0.347	3.222	0.301
Teacher can explain in several ways (0-4 score)	3.311	0.292	3.262	0.288
Student effort				
I have done my best quality work in this class	3.409	0.815	3.184	0.944
In this class, I stop trying when the work gets hard	0.831	1.246	0.898	1.240
In this class, I take it easy and do not try to do my best	1.316	1.519	0.931	1.366
How much homework do you usually complete?	3.933	0.833	4.050	0.708

Table 12: Out-of-Sample Fit: Explanatory Variables (OLS regression)

	End-of-Year Knowledge: 4^{th} Grade Math Test Score			Teacher Effort: Teacher Takes Time to Summarize the Lesson		
	Data		Model	Data		Model
	Estimate	95% CI	Estimate	Estimate	95% CI	Estimate
Classroom-level regressors:						
Baseline Knowledge: Class average	0.2870∗∗∗	(0.0988,0.4752)	0.1740∗∗∗	−0.1360∗∗∗	(−0.2281,−0.0439)	−0.3250∗∗∗
Baseline Knowledge: Class Std.Dev.	−0.1150∗	(-0.2346, 0.0046)	−0.0220	0.0340	(−0.0483,0.1163)	0.1650∗∗∗
Teacher experience in the district	−0.0010	(-0.0147, 0.0127)	−0.0001	−0.0060	(−0.0276,0.0156)	0.0010
Teacher has Master's degree	0.0290	(-0.0455, 0.1035)	−0.0110	−0.0580	(−0.2403,0.1243)	−0.0030
Teacher teaches both Math and ELA	0.0140	(-0.1585, 0.1865)	−0.0230	−0.0270	(−0.1426,0.0886)	−0.0030
Student-level regressors:						
Male	0.0330	(-0.0454, 0.1114)	0.0180			
Gifted	0.6450∗∗∗	(0.4706, 0.8194)	0.604∗∗∗			
Special education (SpEd)	−0.3010∗∗	(-0.5323, -0.0697)	−0.2700∗∗∗			
English language learner (ELL)	−0.1700∗∗∗	(-0.2974, -0.0426)	−0.206∗∗∗			
Free or reduced price lunch	0.0100	(-0.0743, 0.0943)	0.0140			
Black	−0.4670∗∗∗	(-0.5532, -0.3808)	−0.3110∗∗∗			
Hispanic	−0.1520∗∗∗	(-0.2382, -0.0658)	−0.0880∗∗∗			
Age	−0.0870∗∗∗	(-0.1203, -0.0537)	−0.0610∗∗∗			
N. books in bedroom (Base = "None"):						
≥1 and ≤10	0.1830∗∗∗	(0.1164, 0.2496)	0.0860∗∗∗			
≥11 and ≤24	0.2240∗∗∗	(0.0868, 0.3612)	0.1370∗∗∗			
≥25	0.2890∗∗∗	(0.1969, 0.3811)	0.207∗∗∗			
N. computers at home (Base = "None"):						
One	0.162∗∗∗	(0.1267, 0.1953)	0.1500∗∗∗			
More than one	0.209∗∗∗	(0.1463, 0.2717)	0.1800∗∗∗			
Has person at home to help with homework (Base = "Never"):						
Mostly not	0.3330∗∗∗	(0.0958, 0.5702)	0.2190∗∗∗			
Sometimes	0.211	(-0.0634, 0.4854)	0.0950			
Mostly	0.1880	(-0.0648, 0.4408)	0.1090∗∗∗			
Always	0.1760	(-0.0494, 0.4014)	0.1210∗∗			
Has no quiet place to study at home (Base = "Never"):						
Mostly not	−0.0110	(-0.0424, 0.0204)	−0.0450∗∗∗			
Sometimes	−0.1040∗∗∗	(-0.1412, -0.0668)	−0.0860∗∗			
Mostly	−0.2140∗∗∗	(-0.3002, -0.1278)	−0.2080∗∗∗			
Always	−0.2080∗∗∗	(-0.2374, -0.1786)	−0.2130∗∗∗			
Constant	0.5170	(-0.1474, 1.1814)	0.4310	3.260∗∗∗	(3.1208,3.3992)	2.9800∗∗∗

Levels of significance: * p<0.10, ** p<0.05, *** p<0.01

Table 13: Impact of Tailoring Instruction to Student Baseline Knowledge

Class time allocation	Data	Tailored instruction: 1^{st} tercile	Tailored instruction: 2^{nd} tercile	Tailored instruction: 3^{rd} tercile
Place value, rounding, addit., and subtr.	6.786	3.339	8.241	4.358
Multi-digit multipl. and division	25.880	16.653	31.285	29.378
Shape, angles, and geometry	17.617	22.327	15.242	10.959
Fractions and decimals	25.498	27.161	29.466	35.447
Unit conversions and meas.	24.218	30.520	15.766	19.858

4^{th} grade math test score ($log(K_{1ti})$)	Baseline	Tailored Instruction	Difference
Baseline knowledge tercile:			
All terciles (SD = 0.940)	0.300	0.338	0.038
1^{st} tercile	−0.570	−0.525	0.045
2^{nd} tercile	0.159	0.196	0.037
3^{rd} tercile	1.106	1.138	0.032

Knowledge value-added ($K_{1ti} - \delta_0 K_{0ti}$)	Baseline	Tailored Instruction	Difference
Baseline knowledge tercile:			
All terciles (SD = 0.474)	0.731	0.790	0.059
1^{st} tercile	0.348	0.375	0.027
2^{nd} tercile	0.573	0.617	0.044
3^{rd} tercile	1.168	1.265	0.098

Table 14: Tracking Intensity Across Schools in Year 1 of MET Sample

	Mean	St.Dev	p25	Median	p75
Share of between-classroom variance out of K_{0ti} within-school variance					
Data (1)	0.074	0.089	0.016	0.039	0.099
Tracking policy experiment (2)	0.616	0.158	0.521	0.626	0.729
Tracking Intensity (ratio of (1) to (2))	0.123	0.158	0.027	0.062	0.145

Table 15: Impact of Ability Tracking on Classroom Instruction and Student Achievement

	Baseline			
	All terciles	1^{st} tercile	2^{nd} tercile	3^{rd} tercile
Teacher effort (SD = 0.123)	1.032	1.068	1.043	0.989
Class time allocation:				
Place value, round., addit., and subtr.	6.642	6.219	6.777	6.870
Multi-digit multipl. and division	26.072	24.166	26.190	27.591
Shapes, angles, and geom.	17.015	18.145	16.988	16.071
Fractions and decimals	25.805	25.924	25.470	26.037
Unit conversion and meas.	24.467	25.546	24.575	23.431
4^{th} grade end-of-year test score (SD = 0.937)	0.260	−0.602	0.163	1.099
	Ability Tracking with Teacher Random Assignment			
	All terciles	1^{st} tercile	2^{nd} tercile	3^{rd} tercile
Teacher effort	1.026	1.205	1.077	0.820
Class time allocation:				
Place value, round., addit., and subtr.	6.746	4.963	7.885	7.142
Multi-digit multipl. and division	26.174	21.407	28.401	28.051
Shapes, angles, and geom.	17.000	18.663	16.915	15.652
Fractions and decimals	25.854	26.477	25.266	25.906
Unit conversion and meas.	24.227	28.490	21.534	23.249
4^{th} grade end-of-year test score	0.266	−0.601	0.183	1.094
	Ability Tracking with Negative Assortative Matching			
	All terciles	1^{st} tercile	2^{nd} tercile	3^{rd} tercile
Teacher effort	1.029	1.226	1.076	0.812
Class time allocation:				
Place value, round., addit., and subtr.	6.701	4.928	7.831	7.097
Multi-digit multipl. and division	26.341	20.678	28.087	29.465
Shapes, angles, and geom.	16.915	19.433	17.267	14.398
Fractions and decimals	25.938	25.858	25.364	26.579
Unit conversion and meas.	24.107	29.103	21.451	22.461
4^{th} grade end-of-year test score	0.269	−0.575	0.188	1.075
	Ability Tracking with Positive Assortative Matching			
	All terciles	1^{st} tercile	2^{nd} tercile	3^{rd} tercile
Teacher effort	1.024	1.196	1.067	0.834
Class time allocation:				
Place value, round., addit., and subtr.	6.637	5.071	7.918	6.705
Multi-digit multipl. and division	26.528	20.694	28.541	29.534
Shapes, angles, and geom.	16.966	18.862	16.557	15.746
Fractions and decimals	25.778	26.355	25.144	25.914
Unit conversion and meas.	24.091	29.018	21.840	22.101
4^{th} grade end-of-year test score	0.263	−0.612	0.178	1.100

Table 16: Teaching According to Curriculum Standards

Baseline				
	All terciles	1^{st} tercile	2^{nd} tercile	3^{rd} tercile
Teacher effort (SD = 0.117)	1.009	1.066	1.026	0.950
4^{th} grade end-of-year test score (SD = 0.940)	0.300	−0.570	0.159	1.106
State Curriculum Standards				
	All terciles	1^{st} tercile	2^{nd} tercile	3^{rd} tercile
Teacher effort	1.002	1.058	1.019	0.944
4^{th} grade end-of-year test score	0.296	−0.596	0.168	1.105

Table 17: Impact of Curriculum Standards on Classroom Instruction and Student Achievement

	All terciles	1^{st} tercile	2^{nd} tercile	3^{rd} tercile
Baseline				
Teacher effort (SD = 0.117)	1.009	1.066	1.026	0.950
Class time allocation:				
Place value, round., addit., and subtr.	6.786	6.582	6.893	6.853
Multi-digit multipl. and division	25.880	26.028	26.056	26.393
Shapes, angles, and geom.	17.617	17.385	17.440	17.354
Fractions and decimals	25.498	25.736	25.346	25.445
Unit conversion and meas.	24.218	25.269	24.265	23.355
4^{th} grade end-of-year test score (SD = 0.940)	0.300	−0.570	0.159	1.106
Low-Level Curriculum Standards				
	All terciles	1^{st} tercile	2^{nd} tercile	3^{rd} tercile
Teacher effort	1.006	1.065	1.025	0.945
Class time allocation:				
Place value, round., addit., and subtr.	5.191	5.041	5.310	5.205
Multi-digit multipl. and division	22.067	20.789	22.047	23.084
Shapes, angles, and geom.	22.165	22.618	22.158	21.818
Fractions and decimals	23.746	23.603	23.757	23.850
Unit conversion and meas.	26.831	27.950	26.729	26.043
4^{th} grade end-of-year test score	0.297	−0.554	0.145	1.095
Mid-Level Curriculum Standards				
	All terciles	1^{st} tercile	2^{nd} tercile	3^{rd} tercile
Teacher effort	1.012	1.066	1.029	0.953
Class time allocation:				
Place value, round., addit., and subtr.	7.016	6.787	7.115	7.110
Multi-digit multipl. and division	26.961	25.374	26.845	28.306
Shapes, angles, and geom.	17.186	17.784	17.232	16.677
Fractions and decimals	26.225	26.032	26.193	26.405
Unit conversion and meas.	22.611	24.024	22.615	21.502
4^{th} grade end-of-year test score	0.303	−0.574	0.167	1.109
High-Level Curriculum Standards				
	All terciles	1^{st} tercile	2^{nd} tercile	3^{rd} tercile
Teacher effort	1.010	1.064	1.027	0.953
Class time allocation:				
Place value, round., addit., and subtr.	6.204	6.040	6.324	6.227
Multi-digit multipl. and division	27.624	26.001	27.500	29.003
Shapes, angles, and geom.	14.410	15.070	14.487	13.825
Fractions and decimals	26.945	26.715	26.903	27.161
Unit conversion and meas.	24.818	26.175	24.785	23.785
4^{th} grade end-of-year test score	0.302	−0.577	0.163	1.112

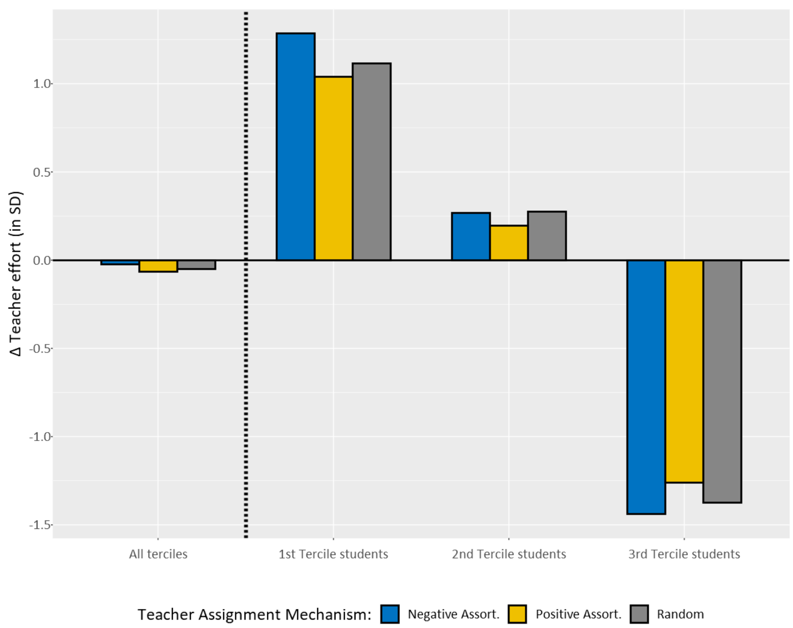

Figure 1: Impact of Ability Tracking with Teacher Ability-Based Assignment on Teacher Effort

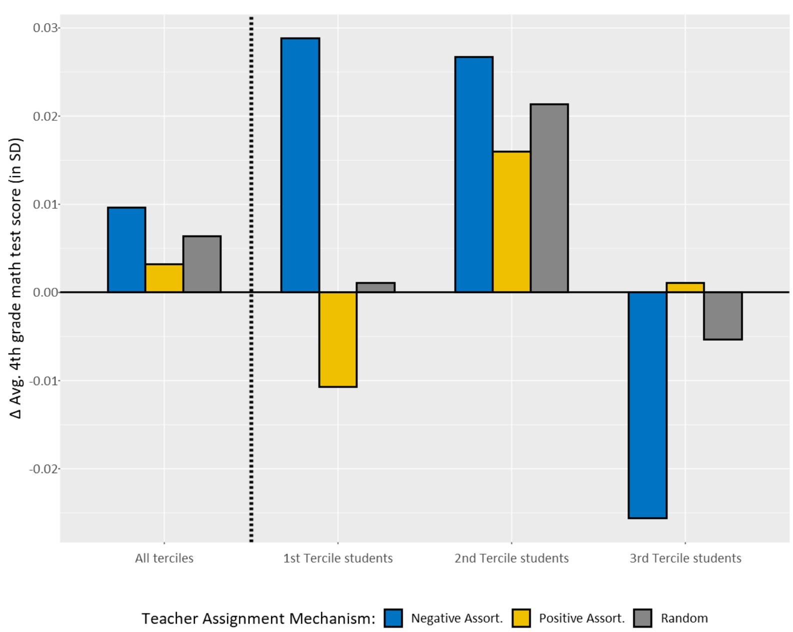

Figure 2: Impact of Ability Tracking with Teacher Ability-Based Assignment on Knowledge Value-Added

Appendix

A. Model Solution and Likelihood Function

Equations (6a) and (6b), for $k = 1,\dots,J-1$, represent the first order conditions for an interior solution of e_t and $\boldsymbol{\tau}_t$. Rearranging (6a) we can derive the closed form solution of teacher optimal effort conditional on $\boldsymbol{\tau}_t$ and other inputs

$$e_t^* = \left(\frac{\gamma_2}{\alpha_0}\sum_{i=1}^{N_t}\omega_{ti}\delta_1 K_{0ti}^{\gamma_0} A_t^{\gamma_1} h_{ti}^{\gamma_3}\prod_{j=1}^{J}\tau_{tj}^{\eta_{jq}}\right)^{\frac{1}{2-\gamma_2}} \tag{11}$$

Substituting (11) in (6b) and moving $\widetilde{\varepsilon}_{tk}$ to the LHS we can rewrite the FOCs for τ_{tk}, $k = 1\dots,J-1$ as

$$-\widetilde{\varepsilon}_{tk} = \sum_{i=1}^{N_t}\omega_{ti}\delta_1 K_{0ti}^{\gamma_0} A_t^{\gamma_1}\left(\frac{\gamma_2}{\alpha_0}\sum_{i=1}^{N_t}\omega_{ti}\delta_1 K_{0ti}^{\gamma_0} A_t^{\gamma_1} h_{ti}^{\gamma_3}\prod_{j=1}^{J}\tau_{tj}^{\eta_{jq}}\right)^{\frac{\gamma_2}{2-\gamma_2}} h_{ti}^{\gamma_3}(\eta_{kq_i}\tau_{tk}^{-1}-\eta_{Jq}\tau_{tJ}^{-1})\prod_{j=1}^{J}\tau_{tj}^{\eta_{jq}}+ \\ +\widetilde{\alpha}_{1k}-\alpha_{2k}\left(\tau_{tk}-\varphi_{tk}-\tau_{tJ}+\varphi_{tJ}\right) \tag{12}$$

For each student i taught by teacher t in school s, an observation in the data includes the measures of both exogenous and endogenous latent factors together with the initial conditions,

$$D_{ti} = \left((A_t^m)_{m=1}^{M_A}, (e_t^m)_{m=1}^{M_e}, (\phi_t^m)_{m=1}^{M_\phi}, (h_{ti}^m)_{m=1}^{M_h}, (K_{1ti}^m)_{m=1}^{M_{K_1}}, K_{0ti}, \boldsymbol{\tau}_t^*, X_{ti}\right),$$

with $X_{ti} = \left(X_t^A, X_t^\phi, X_{ti}^{K_0}, X_{ti}^h, \varphi_t, Z_{ti}\right)$. The vector of observations for class t is then $\underline{D}_t = (D_{t1},\dots,D_{tN_t})$. Define the vectors of random effects $\chi_{ti} \equiv (\boldsymbol{v}_t, \boldsymbol{\zeta}_{ti})$ and $\boldsymbol{\chi}_t = (\chi_{t1},\dots,\chi_{tN_t})$. In order to derive the likelihood contribution of class t, let's first assume that $\boldsymbol{\chi}_t$ is observed by the econometrician. Then, given the distributional assumptions and suppressing the subscripts for more clarity, the likelihood of the exogenous latent factor $\theta \in \{A,\phi,h\}$ is given by

$$\ell_\theta\big((\theta^m)_{m=1}^{M_\theta} | X^\theta, \boldsymbol{\chi}\big) = \prod_{m=1}^{M_\theta}\ell_\theta^m(\theta^m|\theta)$$

where $\ell_\theta^m(\cdot)$ is the likelihood of the m^{th} measure of θ. As for the endogenous variables, conditional on $\left(\underline{X}_t, \boldsymbol{\tau}_t, \boldsymbol{\chi}_t\right)$, optimal effort e_t^* is completely deterministic. Hence, the likelihood of the effort measures is just

$$\ell_e\big((e_t^m)_{m=1}^{M_e} | \underline{X}_t, \boldsymbol{\tau}_t, \boldsymbol{\chi}_t\big) = \prod_{m=1}^{M_e}\ell_e^m\big(e_t^m | X_t, \boldsymbol{\tau}_t, \boldsymbol{\chi}_t\big).$$

On the other hand, the likelihood of end-of-year knowledge K_{1ti} is given by

$$\ell_{K_1}\big((K_{1ti}^m)_{m=1}^{M_{K_1}}|\underline{X}_t,\boldsymbol{\chi}\big)=\left(\prod_{m=1}^{M_{K_1}}\ell_{K_1}^m(K_{1ti}^m|K_{1ti})\right)\ell_{K_1}^f\big(K_{1ti}|\underline{X}_t,\boldsymbol{\tau}_t,\boldsymbol{\chi}_t\big).$$

Finally, the conditional likelihood of $\boldsymbol{\tau}$ is derived from the FOCs (12), for $k=1\ldots,J-1$, and the distributional assumption on $\boldsymbol{\varepsilon}_t$. Specifically, expressing the LHS of (12) for all $k=1\ldots,J-1$ as a multivariate function $\widetilde{\boldsymbol{\varepsilon}}\big(\underline{X}_t,\boldsymbol{\chi}_t\big)$ we have that the likelihood of $\boldsymbol{\tau}$ is given by

$$\ell_{\boldsymbol{\tau}}\big(\boldsymbol{\tau}_t|\underline{X}_t,\boldsymbol{\chi}_t\big)=\left|\det J\left(\widetilde{\boldsymbol{\varepsilon}}\big(\underline{X}_t,\boldsymbol{\chi}_t\big)\right)\right|\times\ell_{\widetilde{\boldsymbol{\varepsilon}}}\left(\widetilde{\boldsymbol{\varepsilon}}\big(\underline{X}_t,\boldsymbol{\chi}_t\big)\Big|\underline{X}_t,\boldsymbol{\chi}_t\right)$$

where $J\left(\widetilde{\boldsymbol{\varepsilon}}\big(\underline{X}_t,\boldsymbol{\chi}_t\big)\right)$ is the Jacobian matrix of $\widetilde{\boldsymbol{\varepsilon}}\big(\underline{X}_t,\boldsymbol{\chi}_t\big)$ (with derivatives with respect to $(\tau_{t1},\ldots,\tau_{tJ-1})$) and $\ell_{\widetilde{\boldsymbol{\varepsilon}}}(\cdot)$ the likelihood of $\widetilde{\boldsymbol{\varepsilon}}_t$ (a multivariate normal with mean zero and covariance matrix $\Sigma_{\widetilde{\boldsymbol{\varepsilon}}}$).

As a result, the likelihood contribution of class t in school s conditional on $\big(\underline{X}_t,\boldsymbol{\chi}_t\big)$ is given by

$$\begin{aligned}
&L_t\Big(\Theta\Big|D_t,\boldsymbol{\chi}_t\Big)=\\
&\quad=\left[\prod_{i=1}^{N_t}\ell_{K_1}\big((K_{1ti}^m)_{m=1}^{M_{K_1}}|\underline{X}_t,\boldsymbol{\tau}_t,\boldsymbol{\chi}_t\big)\ell_h\big((h_{ti}^m)_{m=1}^{M_h}|X_{ti}^h,\boldsymbol{\chi}_t\big)\ell_{K_0}\big(K_{0ti}|X_{ti}^{K_0},\boldsymbol{\chi}_t\big)\boldsymbol{\Phi}\left(\boldsymbol{\zeta}_{ti}|\underline{X}_t,\boldsymbol{\upsilon}_t;\Sigma_\zeta\right)\right]\times\\
&\qquad\times\ell_e\big((e_t^m)_{m=1}^{M_e}|\underline{X}_t,\boldsymbol{\tau}_t,\boldsymbol{\chi}_t\big)\ell_A\big((A_t^m)_{m=1}^{M_A}|X_t^A,\boldsymbol{\chi}_t\big)\ell_\phi\big((\phi_t^m)_{m=1}^{M_\phi}|X_t^\phi,\boldsymbol{\chi}_t\big)\times\\
&\qquad\times\ell_{\boldsymbol{\tau}}\big(\boldsymbol{\tau}_t|\underline{X}_t,\boldsymbol{\chi}_t\big)\boldsymbol{\Phi}\left(\boldsymbol{\upsilon}_t|\underline{X}_t;\Sigma_\upsilon\right)
\end{aligned}\tag{13}$$

with $\boldsymbol{\Phi}(\cdot;\Sigma)$ a multivariate normal density with zero mean and covariance matrix Σ. Now since $\boldsymbol{\chi}_t$ is actually unobserved, we need to integrate it out, that is

$$L_t(\Theta|D_t)=\int L_t(\Theta|D_t,\boldsymbol{\chi})d\boldsymbol{\chi}.\tag{14}$$

Given R draws from the joint distribution of $\boldsymbol{\chi}_t$, denoted $(\hat{\boldsymbol{\chi}}_{tr})_{r=1}^R$, we can perform a Monte Carlo integration to approximate (14) and obtain our simulated likelihood contribution of class t

$$\begin{aligned}
\hat{L}_t=\frac{1}{R}\sum_{r=1}^R\Bigg\{&\left[\prod_{i=1}^{N_t}\ell_{K_1}\big((K_{1ti}^m)_{m=1}^{M_{K_1}}|\underline{X}_t,\boldsymbol{\tau}_t,\hat{\boldsymbol{\chi}}_{tr}\big)\ell_h\big((h_{ti}^m)_{m=1}^{M_h}|X_{ti}^h,\hat{\boldsymbol{\chi}}_{tr}\big)\ell_{K_0}\big(K_{0ti}|X_{ti}^{K_0},\hat{\boldsymbol{\chi}}_{tr}\big)\right]\times\\
&\times\ell_e\big((e_t^m)_{m=1}^{M_e}|\underline{X}_t,\boldsymbol{\tau}_t,\hat{\boldsymbol{\chi}}_{tr}\big)\ell_A\big((A_t^m)_{m=1}^{M_A}|X_t^A,\hat{\boldsymbol{\chi}}_{tr}\big)\ell_\phi\big((\phi_t^m)_{m=1}^{M_\phi}|X_t^\phi,\hat{\boldsymbol{\chi}}_{tr}\big)\times\ell_{\boldsymbol{\tau}}\big(\boldsymbol{\tau}_t|\underline{X}_t,\hat{\boldsymbol{\chi}}_{tr}\big)\Bigg\}
\end{aligned}\tag{15}$$

B. Additional Tables and Figures

Table B.1: Exogenous Inputs Equation Parameter Estimates

Determinants	Student effort (log)		Baseline knowledge (log)	
	Estimate	Std.Err.	Estimate	Std.Err.
Constant	0.0000	.	0.6021	0.3843
Male	−0.1973	0.0404	0.0167	0.0313
Gifted	0.2588	0.0851	0.8784	0.0705
Special education (SpEd)	−0.2091	0.0614	−0.4173	0.0574
English language learner (ELL)	−0.1838	0.0612	−0.0351	0.0575
Free/Reduced price lunch	−0.0012	0.0382	0.0118	0.0372
Black	−0.0167	0.0551	−0.4760	0.0512
Hispanic	−0.0278	0.0587	−0.2750	0.0569
Age	−0.0160	0.0328	−0.1023	0.0328
N. books in bedroom (base = None):				
≥1 and ≤10	0.1400	0.0599	0.1084	0.0594
≥11 and ≤24	0.1104	0.0620	0.2026	0.0614
≥25	0.2518	0.0615	0.2048	0.0564
Has quiet place to study at home:				
Mostly not	−0.1587	0.0593	0.1250	0.0534
Sometimes	−0.2623	0.0579	−0.0155	0.0464
Mostly	−0.2523	0.0612	−0.0962	0.0536
Always	−0.2052	0.0479	−0.1332	0.0428
N. computers at home (base = none)				
One	0.1459	0.0516	0.0705	0.0514
More than one	0.0530	0.0537	0.1959	0.0550
Has person at home to help with homework (base = never)				
Mostly not	−0.2913	0.1216	0.4161	0.1221
Sometimes	−0.1756	0.1111	0.2239	0.1083
Mostly	−0.0454	0.1032	0.3383	0.1031
Always	0.0613	0.0955	0.2379	0.0970

	Teacher ability (log)	
	Estimate	Std. Err.
Years of experience	0.0995	0.0328
Years of experience (squared)	−0.0051	0.0017
Master's degree	−0.0639	0.1303
Generalist teacher	−0.1421	0.1712
Constant	0.000	.

Table B.2: Exogenous Inputs Random Effects Covariances

Teacher-level Random Effects Covariance Matrix

	Baseline knowledge (log)	Student effort (log)	Teacher ability (log)
Baseline knowledge (log)	0.3863		
	(0.0195)		
Student effort (log)	0.0350	0.2524	
	(0.0037)	(.)	
Teacher ability (log)	−0.0144	0.0477	0.5756
	(0.0090)	(0.0085)	(0.0605)

Student-level Random Effects Covariance Matrix

	Baseline knowledge (log)	Student effort (log)
Baseline knowledge (log)	0.7336	
	(0.0160)	
Student effort (log)	0.0350	0.4678
	(0.0037)	(0.0494)

Standard errors in parenthesis

Table B.3: Measurement Equations Parameters

	Intercept (μ_{0m}^{y})		Slope (μ_{1m}^{y})		Meas.Error ($\sigma_{\varsigma ym}$)	
	Estimate	Std. Err.	Estimate	Std.Err.	Estimate	Std.Err.
Teacher effort (log)						
Teacher explains in another way if we do not understand	3.3222	0.0280	1.0000	.	0.2235	0.0081
Teacher pushes us to work hard	3.0964	0.0366	1.1359	0.2659	0.3121	0.0150
Teacher does not waste time in class	2.6466	0.0399	0.5240	0.2646	0.3913	0.0218
Teacher asks us if we understand the lesson	3.3319	0.0218	1.2787	0.2464	0.2284	0.0093
Teacher asks us if we are following along	3.3440	0.0275	1.1728	0.2194	0.1895	0.0068
Teacher writes feedback on our papers	2.8870	0.0378	1.0258	0.2679	0.3371	0.0170
Teacher takes the time to summarize the lesson	2.8226	0.0477	1.8912	0.3704	0.3542	0.0219
Teacher encourage us to do our best	3.5276	0.0263	0.6811	0.1847	0.2369	0.0084
Teacher ability (log)						
CLASS Behavior management scale	5.8455	0.1815	1.0000	.	0.3038	0.0242
CLASS Content understanding scale	4.1102	0.0687	0.2983	0.0761	0.4298	0.0265
CLASS Productivity scale	5.8462	0.1324	0.7154	0.0738	0.3358	0.0199
FFTM Management of class procedures score	2.6955	0.0815	0.4338	0.0505	0.2469	0.0099
FFTM Management of student behavior score	2.8022	0.0779	0.4148	0.0480	0.2342	0.0089
MQI Richness of mathematics score	1.3260	0.0294	0.0811	0.0438	2.2534	0.0091
MQI Mathematical knowledge for teaching (MKT) score	2.0123	0.0293	0.1134	0.0366	0.2091	0.0062
Teacher explains clearly	3.3137	0.0372	0.1341	0.0491	0.2816	0.0113
Teacher controls class behavior	2.2080	0.0711	0.3290	0.0709	0.3953	0.0226
Teacher explains in orderly way	3.1630	0.0383	0.1414	0.0496	0.2845	0.0115
Teacher can explain in several ways	3.2023	0.0382	0.1460	0.0482	0.2758	0.0108
Student effort (log)						
I have done my best quality work in this class	0.0000	.	1.0000	.	1.0000	.
Cutoff 1 ("Never"-"Mostly not")	−2.1655	0.6291				
Cutoff 2 ("Mostly not"-"Sometimes")	−1.8312	0.6277				
Cutoff 3 ("Sometimes"-"Mostly")	−0.8251	0.6319				
Cutoff 4 ("Mostly"-"Always")	0.2065	0.6323				
In this class, I stop trying when the work gets hard	0.0000	.	1.2150	0.2597	0.8228	0.0001
Cutoff 1 ("Never"-"Mostly not")	−1.5057	0.6263				
Cutoff 2 ("Mostly not"-"Sometimes")	−1.0904	0.6250				
Cutoff 3 ("Sometimes"-"Mostly")	−0.5112	0.6231				
Cutoff 4 ("Mostly"-"Always")	−0.0629	0.6225				
In this class, I take it easy and do not try to do my best	0.0000	.	0.9818	0.1920	0.8883	0.0001
Cutoff 1 ("Never"-"Mostly not")	−1.2581	0.6613				
Cutoff 2 ("Mostly not"-"Sometimes")	−0.9054	0.6611				
Cutoff 3 ("Sometimes"-"Mostly")	−0.5147	0.6596				
Cutoff 4 ("Mostly"-"Always")	−0.1852	0.6596				
How much homework do you usually complete?	0.0000	.	0.5206	0.1030	0.9699	0.0001
Cutoff 1 ("Never"-"Mostly not")	−2.8090	0.6095				
Cutoff 2 ("Mostly not"-"Sometimes")	−1.7126	0.6056				
Cutoff 3 ("Sometimes"-"Mostly")	−1.0552	0.6043				
Cutoff 4 ("Mostly"-"Always")	0.7980	0.6049				

Table B.4: Within-Sample and Out-of-Sample Model Fit of Student Effort Measures

Within-Sample Fit	Never		Mostly not		Sometimes		Mostly		Always	
	Data	Model	Data	Model	Data	Model	Data	Model	Data	Model
I have done my best quality work in this class	0.007	0.024	0.010	0.023	0.089	0.164	0.242	0.329	0.457	0.460
In this class, I stop trying when the work gets hard	0.488	0.566	0.119	0.157	0.103	0.148	0.046	0.066	0.048	0.063
In this class, I take it easy and do not try to do my best	0.427	0.606	0.096	0.115	0.090	0.112	0.077	0.070	0.119	0.096
	None		Some		Most		All		All plus extra	
	Data	Model	Data	Model	Data	Model	Data	Model	Data	Model
How much homework do you usually complete?	0.006	0.002	0.062	0.039	0.106	0.098	0.489	0.629	0.137	0.231

Out-of-Sample Validation (Year 2 Data Sample Fit)	Never		Mostly not		Sometimes		Mostly		Always	
	Data	Model	Data	Model	Data	Model	Data	Model	Data	Model
I have done my best quality work in this class	0.007	0.024	0.021	0.021	0.010	0.164	0.291	0.329	0.431	0.462
In this class, I stop trying when the work gets hard	0.611	0.570	0.145	0.153	0.115	0.148	0.062	0.064	0.067	0.064
In this class, I take it easy and do not try to do my best	0.482	0.609	0.131	0.116	0.128	0.109	0.106	0.069	0.153	0.097
	None		Some		Most		All		All plus extra	
	Data	Model	Data	Model	Data	Model	Data	Model	Data	Model
How much homework do you usually complete?	0.008	0.002	0.062	0.039	0.147	0.096	0.554	0.632	0.229	0.231

Table B.5: Year 2 Sample - Student and Teacher Characteristics

	Year 1		Year 2			Year 1	Year 2
	Mean	Std.Dev	Mean	Std.Dev		Mean	Mean
Panel A: Students							
Obs.	2352		4452				
Age	9.52	0.50	8.91	0.81	Gifted	0.06	0.06
Male	0.48		0.50		Special education (SpEd)	0.09	0.11
White	0.26		0.23		English language learner (ELL)	0.17	0.14
Black	0.43		0.46		Reduced price/free lunch	0.45	0.49
Hispanic	0.25		0.24				
N. books in bedroom:					N. computers at home:		
None	0.09		0.08		None	0.12	0.11
≥1 and ≤10	0.22		0.21		One	0.45	0.41
≥11 and ≤24	0.21		0.21		More than one	0.43	0.48
≥25	0.48		0.50				
Has person at home to help with homework:					Has no quiet place to study at home:		
Never	0.02		0.03		Never	0.40	0.36
Mostly not	0.03		0.06		Mostly not	0.12	0.12
Sometimes	0.09		0.13		Sometimes	0.16	0.16
Mostly	0.17		0.14		Mostly	0.12	0.11
Always	0.69		0.64		Always	0.21	0.25
Panel B: Teachers							
Obs.	177						
Years of experience in the district	6.40	5.94	5.42	4.59			
Masters degree	0.53		0.46				
Teaches both Math and ELA (generalist)	0.83		0.82				

Table B.6: Year 2 Sample - Latent Factors Measures

	Year 1		Year 2	
Student knowledge:	Mean	Std.Dev	Mean	Std.Dev
3^{rd} grade math state test score (rescaled)	0.102	0.954	−0.023	0.991
4^{th} grade math state test score (rescaled)	0.306	0.949	0.194	0.961
Teacher effort				
Teacher explains in another way if we do not understand (survey 0-4 score)	3.325	0.285	3.328	0.299
Teacher pushes us to work hard (survey 0-4 score)	3.092	0.370	3.192	0.427
Teacher does not waste time in class (survey 0-4 score)	2.664	0.385	2.709	0.402
Teacher asks us if we understand the lesson (survey 0-4 score)	3.329	0.315	3.392	0.322
Teacher asks us if we are following along (survey 0-4 score)	3.440	0.277	3.502	0.252
Teacher writes feedback on our papers (survey 0-4 score)	2.887	0.387	2.959	0.468
Teacher takes the time to summarize the lesson (survey 0-4 score)	2.813	0.480	2.981	0.434
Teacher encourage us to do our best (survey 0-4 score)	3.533	0.257	3.600	0.289
Teacher ability				
CLASS Behavior management scale	5.943	0.715	5.803	0.512
CLASS Content understanding scale	4.137	0.481	4.120	0.496
CLASS Productivity scale	5.918	0.555	5.803	0.419
FFTM Management of class procedures score	2.763	0.354	2.691	0.346
FFTM Management of student behavior score	2.840	0.344	2.767	0.380
MQI Richness of mathematics score	1.340	0.261	1.353	0.263
MQI Mathematical knowledge for teaching (MKT) score	2.030	0.218	2.027	0.225
Teacher explains clearly (survey 0-4 score)	3.321	0.295	3.324	0.269
Teacher controls class behavior (survey 0-4 score)	2.251	0.437	2.211	0.506
Teacher explains in orderly way (survey 0-4 score)	3.180	0.300	3.229	0.347
Teacher can explain in several ways (survey 0-4 score)	3.216	0.295	3.311	0.293
Student effort				
I have done my best quality work in this class				
Never	0.007		0.007	
Mostly not	0.010		0.021	
Sometimes	0.089		0.104	
Mostly	0.242		0.291	
Always	0.457		0.577	
In this class, I stop trying when the work gets hard				
Never	0.488		0.611	
Mostly not	0.119		0.145	
Sometimes	0.103		0.115	
Mostly	0.046		0.062	
Always	0.048		0.067	
In this class, I take it easy and do not try to do my best				
Never	0.427		0.482	
Mostly not	0.096		0.131	
Sometimes	0.090		0.128	
Mostly	0.066		0.106	
Always	0.119		0.153	
How much homework do you usually complete?				
None	0.006		0.008	
Some	0.062		0.051	
Most	0.106		0.147	
All	0.489		0.554	
All plus extra	0.137		0.229	

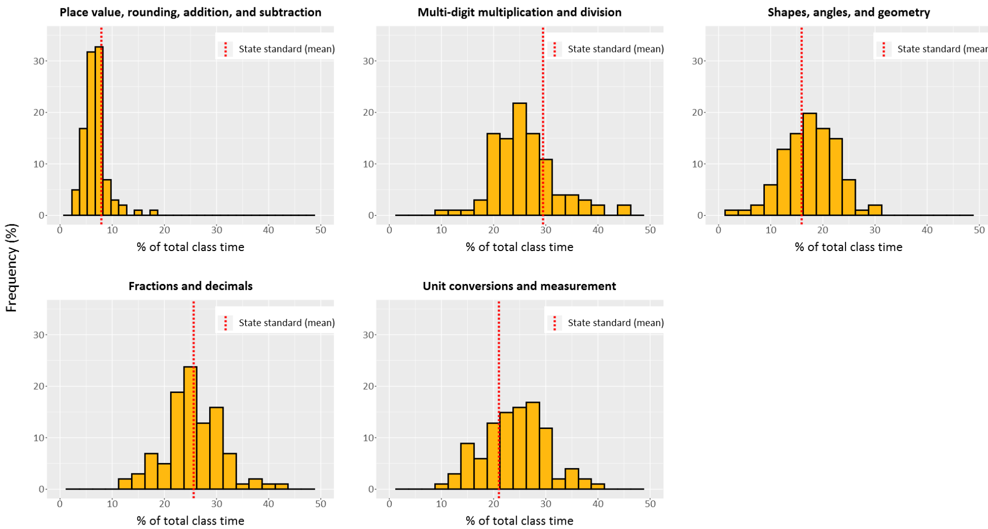

Figure B.1: Distribution of Class Time Allocations

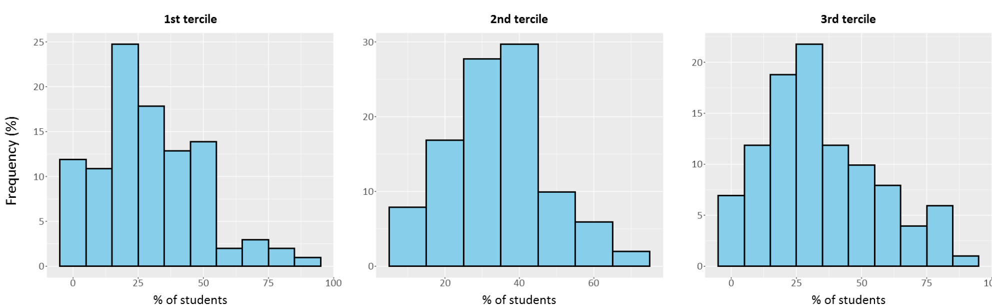

Figure B.2: Distribution of Classroom Composition

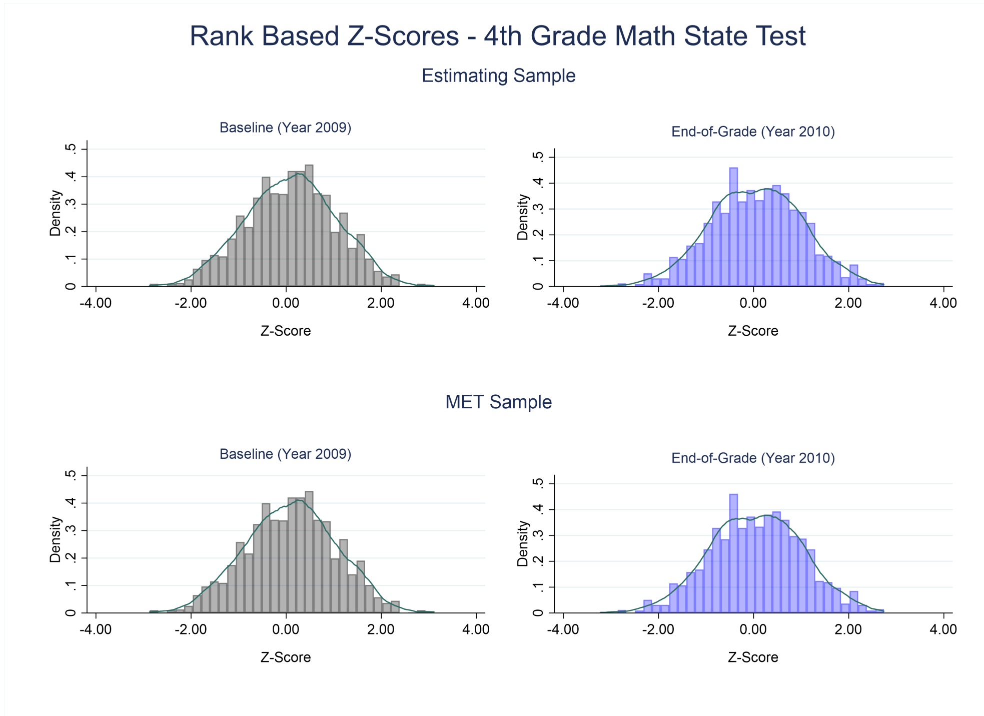

Figure B.3: Distributions of State Test Scores

Time on Topic		Grades K-12 Mathematics Topics	Expectations for Students in Mathematics				
<none>	1	**Number Sense/Properties/Relationships**	Memorize Facts/ Definitions/ Formulas	Perform Procedures	Demonstrate Understanding of Mathematical Ideas	Conjecture/ Generalize/ Prove	Solve Non-Routine Problems/Make Connections
⓪ ① ② ③	101	Place value	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③
⓪ ① ② ③	102	Whole numbers and integers	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③
⓪ ① ② ③	103	Operations	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③
⓪ ① ② ③	104	Fractions	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③
⓪ ① ② ③	105	Decimals	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③
⓪ ① ② ③	106	Percents	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③
⓪ ① ② ③	107	Ratios and proportions	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③
⓪ ① ② ③	108	Patterns	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③
⓪ ① ② ③	109	Real and/or rational numbers	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③
⓪ ① ② ③	110	Exponents and scientific notation	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③
⓪ ① ② ③	111	Factors, multiples, and divisibility	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③	⓪ ① ② ③

Figure B.4: A snippet of the Survey of Enacted Curriculum for mathematics.