A Partial Identification Approach to Identifying the Determinants of Human Capital Accumulation: An Application to Teachers

Nirav Mehta
University of Western Ontario *

November 23, 2024

Abstract

This paper views career growth in teacher quality through the lens of human capital theory to understand the roles of On-the-Job Training (OJT) and Learning-by-Doing (LBD) in human capital formation. If OJT is the primary determinant of human capital, incentive pay policies could create a *dynamic multitasking* problem, leading teachers to reduce their human capital investments, thereby lowering future student achievement. In contrast, teacher human capital and future achievement would both increase if LBD were the dominant force. To explore this, I develop explicit bounds on components of a human capital production function allowing for both channels, which I estimate using experimental variation from Glewwe et al. (2010), a teacher incentive pay experiment in Kenya. I find that LBD is present and also estimate an informative upper bound on the OJT component. This suggests that dynamic multitasking, while theoretically relevant, may have limited practical significance, at least in this context.

Keywords: partial identification, human capital, teacher quality, on-the-job training, learning-by-doing

JEL codes: I2, I28, J2, J24, J45, C1

*email: `nirav.mehta@uwo.ca`, web: `https://niravmehtax.github.io`. I thank Martin Luccioni and Tian Liu for research assistance, and Roy Allen, Lance Lochner, Rory McGee, Magne Mogstad, Seth Richards-Shubik, Lisa Tarquinio, and Camille Terrier for their comments. This paper also benefited from comments from the 2019 meeting of the CESifo Economics of Education Group. I gratefully acknowledge funding from the SSHRC Insight Development Grant Program. This paper makes use of publicly available data from Glewwe et al. (2010), which can be downloaded from `https://doi.org/10.3886/E113755V1`.

1 Introduction

Teacher quality, typically measured by a teacher's value added to student achievement, is an important determinant of student achievement and economic growth (Rivkin et al., 2005; Hanushek, 2011). While the research and policy focus has typically been on cross-sectional variation in teacher quality, researchers have also documented substantial growth in quality over teachers' careers (Harris and Sass, 2011; Wiswall, 2013; Papay and Kraft, 2015). For example, using data from North Carolina, Wiswall (2013) finds that the average growth in teacher quality over 35 years of experience is equal to almost one standard deviation of the cross-sectional teacher quality among novice teachers.

While the growth in quality over teachers' careers is clearly important, little is known about why it occurs. Economists have developed two main theories to explain how workers accumulate human capital, which is then used to produce output: *On-the-Job Training* (OJT) and *Learning-by-Doing* (LBD). In the "pure OJT" model, workers (in the current application, teachers) allocate work time away from production (e.g., classroom teaching) to invest in their human capital (e.g., teachers' professional development) (Becker, 1964; Ben-Porath, 1967). This model implies a tradeoff between current and future production because of the *multitask problem* (Hölmstrom and Milgrom, 1991), which emerges because OJT investments are unobserved and, therefore, noncontractible. This tradeoff is not present under the "pure LBD" model, wherein workers accumulate human capital via the act of production (Rosen, 1972; Weiss, 1972; Blinder and Weiss, 1976).[1]

The importance of these potential determinants of human capital has implications for the design of effective education policy, and, more generally, optimal employee compensation in the presence of market failures. For example, if OJT was the main driver of teacher human capital accumulation then an incentive pay scheme that increased current achievement could reduce long-run teacher quality, by diverting teachers from making human capital investments. This would not be the case if instead LBD were the dominant force. More generally, the importance of OJT versus LBD forces has also been shown to have implications for the design of tax policy (Heckman et al., 2002; Blandin, 2018) and even the steady state of an economy (Hansen and İmrohoroğlu, 2009).

Unfortunately, we do not know about the importance of OJT versus LBD forces among any type of worker, let alone teachers. Killingsworth (1982) shows there are no clear general testable implications of the OJT versus LBD theories of human capital accumulation when using standard observational data on workers. Additionally, as Heckman et al. (2002) discuss, identification in such contexts is further hampered by equilibrium re-adjustments that occur in the private sector.[2]

[1] For an example of an OJT-intensive job, consider being a student in school. For an example of an LBD-intensive job, consider "chicken sexing", where workers gain an ability to determine the sex of baby chickens through their experience sorting chicks (McWilliams, 2018).

[2] Belley (2017) uses "excess variation" in wage growth to test for the presence of a non-LBD determinant in human capital production, rejecting the specifications lacking non-LBD determinant for women; he is careful not to ascribe this variation to the presence of OJT. Researchers using the OJT framework (see, e.g., Haley, 1976; Heckman, 1976) typically circumvent the fact that OJT investments are unobserved by solving for the optimal investment path in a worker's dynamic program, which yields estimates of the importance of OJT in the technology of human capital

This has left researchers and policymakers lacking even basic, qualitative information about the presence of OJT versus LBD forces for workers of any type, and may also help explain why prior empirical research has adopted "pure" specifications, only allowing for either OJT or LBD as the force generating human capital.[3]

To fill this gap, this paper develops a framework that yields new findings about the roles played by OJT and LBD forces in the technology governing teacher human capital accumulation. In light of the prominent identification difficulties, I adopt a partial identification approach. The identified set contains all of the values of the OJT and LBD components of teacher human capital accumulation consistent with the data, which can then be projected onto the marginal identified set for either component. I estimate the identified set using data from an experimental intervention, Glewwe et al. (2010), who study a teacher performance pay scheme enacted over many sites across Kenya.[4] Labor prices are often fixed in the education sector because teachers are typically paid according to public salary schedules (Podgursky and Springer, 2011) and worker-specific output is measured, making teacher quality trajectories a particularly attractive context for discriminating between OJT and LBD, above and beyond the obvious policy interest in understanding the dynamics of teachers' human capital accumulation.

The starting point is the workhorse models that have become ubiquitous in empirical research over the last five decades. The value-added (VA) model (Hanushek, 1971; Murnane, 1975), which measures the contribution of different educational inputs to the production of student achievement, is used in the overwhelming mass of research on teacher quality (for more recent examples, see Kane et al., 2013; Chetty et al., 2014).[5] The literature studying human capital accumulation has from its inception typically used log-linear specifications when considering either OJT (Brown, 1976; Haley, 1976; Heckman, 1976) or LBD (Blinder and Weiss, 1976), and these specifications continue to dominate in more recent research (see, e.g., Heckman et al., 1998; Blandin, 2018; Taber et al., 2024).[6]

I extend the modal specifications in the literature to nest both OJT and LBD forces, meaning teacher human capital may depend on both OJT investment and production inputs. A teacher's quality, or their value added, depends on their human capital and production inputs. The resulting specification leverages experimentally estimated average treatment effects (ATEs). This guarantees identification is not driven by assumed nonlinearities in functional forms, while not being very

accumulation. See Kuruscu (2006) for an example using a different approach.

[3]This is also true of recent research studying heterogeneity in the returns to experience for teachers (Kraft and Papay, 2014) or other types of workers (see, e.g., Shaw and Lazear, 2008; Haggag et al., 2017). Here, too, the distinction between OJT versus LBD has implications for policy, as heterogeneous impacts of interventions would be exacerbated or attenuated, depending on how human capital was generated.

[4] See Deaton (2010) for a discussion of the external validity of results from RCTs. Also see Rosenzweig and Udry (2020), which studies external validity respect to the specific channel of aggregate shocks.

[5]See McCaffrey et al. (2003) and Hanushek and Rivkin (2012) for detailed discussions.

[6]As discussed by Willis (1985), these specifications for human capital production had already had a long history of use in labor economics as of three decades ago. These specifications are also consistent with those in the statistical literature spawned by Abowd et al. (1999). See Shaw (1989) for an empirical model of LBD using a different specification and see Fu et al. (2021) for an empirical model of OJT using a different specification.

restrictive compared to commonly used specifications. As is typically the case in the literature, such as in the canonical empirical treatments of OJT (Brown, 1976; Heckman, 1976) as well as in more recent research using the OJT framework (see, e.g., Kuruscu, 2006; Taber et al., 2024), I model investment and production inputs as unobserved.[7] This is a particularly good match for teachers because natural measures of human capital investments, such as formal professional development, or additional certification or education (e.g., Master's degrees), have been shown to be poor predictors of teacher quality (Hanushek, 2003; Hanushek and Rivkin, 2006; Harris and Sass, 2011; Podgursky and Springer, 2011; Jackson et al., 2014; Taylor, 2022), and the general lack of availability of data on how teachers allocate their work time (Hanushek and Rivkin, 2012).[8]

Given the literature's prevailing treatment of post-schooling OJT investments as unobserved, it might be helpful at this point to elaborate on the distinction between OJT and LBD activities. At their core, OJT investments capture any activity that reduces output today to increase output in the future. This tradeoff between current and future output is not necessarily present under LBD. The different dynamic implications of OJT and LBD allow me to distinguish between them without data on production or investment activities. The presence of the dynamic tradeoff not only identifies the different channels; it is also why researchers and policymakers would want to understand the importance of OJT versus LBD forces in the first place.[9]

Identification is possible in the current paper because, in addition to measuring effects of output-based incentives on student achievement during the intervention, Glewwe et al. (2010) also collected follow-up data on student achievement. The intuition can be outlined in the following example: suppose that achievement increased in the treatment group while the incentive scheme was in place and that the random assignment of the intervention was balanced (both are consistent with Glewwe et al. 2010). Then the increase in achievement during the intervention must have come from an increase in treatment-group teachers' production input. Assuming for simplicity there was no labor supply response (this is neither necessary, nor what I assume in my analysis), this increase in production would have come at the cost of reduced OJT investment. Further, a positive treatment effect on achievement after the intervention ended, beyond that accounted for by the persistence of the during-intervention increase in achievement, would suggest that teacher human capital had increased. Because production increased and OJT investment decreased, this would suggest the

[7] Kuruscu (2006), who considers general worker contexts, argues that it is more appropriate to treat on-the-job investments as unobserved because (at least) some training may not be observed by researchers. While this would complicate identification in the typical setting, the current paper's approach allows us to learn about the importance of OJT and LBD forces despite this fundamental data limitation.

[8] In his recent review of human capital in education, Burgess (2016) writes that "[w]e don't really know what effective teachers do that makes them effective," (p. 72).

[9] Although not essential for the OJT framework or this analysis, examples can clarify how teacher activities relate to production and investment inputs. For production-intensive activities, consider classroom teaching and providing student feedback. For investment-intensive activities, consider refining one's teaching practice, revising instructional materials, or developing new pedagogical tools. Treating inputs as unobserved avoids assumptions about their mapping to activities, and can also reveal whether the dynamic tradeoff exists. For example, if teachers are too inundated with classroom duties there might be no scope for making OJT investments, in which case the estimates would indicate no role for this force.

production input dominated in human capital production, pointing to the presence of the LBD component. On the other hand, a post-intervention treatment effect lower than that which could be explained by the persistence of the prior achievement effect would imply that teacher human capital had decreased, pointing to the presence of the OJT component. Put differently, the sign of the follow-up treatment effect on student achievement (net of the persistence of any achievement effects during the intervention) reveals whether production (corresponding to LBD), which must have increased, or investment (corresponding to OJT), which must have decreased, is the dominant force in the human capital technology. Without the follow-up data the identified set would be uninformative, leaving us in the typical case in which we could not separate OJT and LBD forces.

I find that the 95% confidence set for each parameter is informative. Specifically, the lower bound on the LBD component is greater than zero and the upper bound on the OJT component is lower than its uninformative level. I reject the "pure OJT" specification (in which LBD plays no role) at the 5% significance level, but I cannot reject the "pure LBD" specification, as the confidence set for the OJT component contains zero. When further imposing returns-to-scale assumptions (i.e., restrictions on the sum of the OJT and LBD components in human capital production), bounds are tighter. Under the strongest assumption, constant returns to scale, the 95% confidence set for the LBD component lies strictly above that for the OJT component, meaning the LBD component is likely larger than the OJT component.

This paper's framework allows me to make several contributions. First, it works with a human capital production function that nests both the OJT and LBD mechanisms, as opposed to the "pure" specifications used in prior empirical work. This allows for standard hypothesis testing about important qualitative features, in particular, the presence of either force in the human capital production function. Second, it uses a partial identification approach that features a closed-form solution for the identified set: it is a closed line segment, where I have analytical expressions for the end points. This characterization sharply contrasts with the literature's lack of a proof showing how to separately identify OJT and LBD forces, yet obtains quite naturally here. Moreover, the primary input to the identified set is an experimentally estimated ATE, sometimes viewed as "the gold standard" in causal inference. Thus, it is also easy to estimate the identified set and to do inference on it, which is rare in the partial identification literature (Imbens and Manski, 2004; Stoye, 2009; Tamer, 2010). By viewing the experimental results through the lens of human capital theory, I also add to the literature highlighting the need for interpretability of findings from the black box of randomized experiments (Deaton and Cartwright, 2018; Todd and Wolpin, 2023).[10]

Finally, in contrast to the aforementioned work estimating "pure" LBD or OJT models, the approach does not have to make assumptions regarding the optimality of input choices or outcomes observed in the data. Researchers with different goals, say, of examining heterogeneity in human capital profiles or understanding how teacher quality would change under counterfactual

[10] Additionally, by estimating a structural production technology we can learn more in terms of generalizable takeaways than we can from an atheoretical parsing of results of an experimental intervention, insofar as we have recovered "deep" parameters (as in Marschak 1953; see Rust 2013 for a more recent discussion).

incentive pay schemes, would have to make different (likely stronger) statistical assumptions or impose behavioral assumptions. These represent fruitful and complementary avenues for future research. That being said, there is a vast body of research focused on estimating production functions. As discussed by Griliches and Mairesse (1995), estimates of production functions can serve as a good starting point to answer many important questions. In addition to their intrinsic scientific merit, such estimates could help inform researchers and policymakers about important questions across myriad domains, for example, the effects of inputs to the production of children's' cognitive achievement (see, e.g., Todd and Wolpin, 2003; Cunha et al., 2010) to the effects of deregulation on aggregate firm productivity (Olley and Pakes, 1996) to how health care providers trade off quality and quantity (Grieco and McDevitt, 2017).

Somewhat separately from the aforementioned human capital literature, a small but substantively important set of papers examines other channels underlying teachers' improvement as they gain more experience. Ost (2014), who measures the returns to teachers' general and grade-specific experience, finds both to be important determinants of teacher quality growth. Cook and Mansfield (2016) extend this work to also allow for general and context-specific permanent components to teacher quality. The current paper nicely complements these papers by explicitly viewing teacher quality through the lens of the main conceptual frameworks for human capital, and by identifying and separating the OJT and LBD channels of human capital accumulation, which are not necessary to distinguish given these other papers' goals.[11] By highlighting and examining the importance of the dynamic impacts of incentive pay on teacher human capital accumulation, this paper also complements the extensive literature studying teacher quality more generally, recently discussed in Hanushek and Rivkin (2006); Jackson et al. (2014); Strøm and Falch (2020), and the sub-literature on teacher incentives (see, e.g., Hanushek and Raymond, 2005; Muralidharan and Sundararaman, 2011; Imberman and Lovenheim, 2015; Petronijevic, 2016; Macartney et al., 2018; Dinerstein and Opper, 2022; Taylor, 2022).[12]

2 Data and variables

Study design Glewwe et al. (2010) implemented a teacher incentive pay scheme in 100 primary schools in Kenya, which provided bonuses (in the form of in-kind prizes largely of consumption goods, e.g., suits, tea sets, and blankets) to teachers and headmasters at schools where students did well on standardized exams administered as part of the standard curriculum in Kenya.[13] The

[11] These papers may be viewed as part of the substantial empirical literature studying task-specific human capital for more general worker contexts. For recent examples, see Poletaev and Robinson (2008); Sanders (2010); Yamaguchi (2012); Robinson (2018). Sanders and Taber (2012) contains a detailed discussion of this and other extensions of the standard OJT human capital model.

[12] Although less tightly related, this paper also relates to the literature studying teacher labor markets (see, e.g., Dolton and Klaauw, 1999; Stinebrickner, 2001; Behrman et al., 2016; Tincani, 2021; Biasi, 2021; Bobba et al., 2021; Biasi et al., 2021).

[13] It is not clear that headmasters' having been incentivized would substantially affect the results. For example, if treated headmasters motivated their teaching staff to work harder, this would not affect the identified set if

scheme applied to students in grades 4-8, and bonuses were based on school-level averages to discourage competition between teachers. The treatment group (50 schools) was exposed to the incentive scheme in 1998 and 1999, which I refer to as the *active-treatment*, or *intervention* years. In addition to being observed then, outcomes for both the treatment and control groups were also observed for both a pre-treatment year (1997), and a post-active-treatment year (2000), constituting a (short) student panel. There were 7,492 students in the treatment schools and 8,226 students in the control schools in the pre-treatment year.[14] Next, I provide an overview of the variables used in the current analysis; please see Glewwe et al. (2010) for additional details about the data and Section 4 for a more detailed discussion of how I map the data to the model (presented in Section 3).

Variables I measure output using the incentivized standardized achievement test presented in the main results of Glewwe et al. (2010), which is an average of seven standardized subject-specific achievement tests, where standardization was performed with respect to the means and standard deviations of the control group. It is important to note that the experiment was balanced. In particular, mean test scores in the pre-active-treatment year were balanced, and teacher exit and entry also did not significantly differ between the control and treatment groups.

The data also include information about whether teachers were present during random, unannounced, site visits to schools, which I use to measure labor supply. Development economists commonly use attendance as a measure of labor supply, using unannounced site visits to avoid relying on administrative logs, which may overstate teacher attendance (Chaudhury et al., 2006, pp. 100–101). Chaudhury et al. (2006) document that teacher and health-worker absences are quite common in a variety of contexts (they use data from Bangladesh, Ecuador, India, Indonesia, Peru, and Uganda), and also report a finding that absences seemed to be widely distributed among workers.[15]

While attendance data may capture meaningful variation in labor supply, extensive-margin data may fall short as a measure of labor supply if there is substantial intensive-margin variation.[16] To get a sense of how important this issue might be in a development context, we can consider Duflo et al. (2012), who study how exogenous monitoring (via cameras with timestamps) may

achievement effects ultimately manifested through changes in teacher inputs. This highlights a strength of the current approach, that the derived bounds do not require one to rationalize changes in teacher behavior using a behavioral model. That being said, a researcher interested in the important, though somewhat different, question of predicting teachers' responses to another intervention would need to impose behavioral assumptions to understand how teacher inputs would be affected (e.g., specifying and estimating teachers' and headmasters' preferences).

[14] As in Glewwe et al. (2010), I exclude one of the districts (Teso), which did not offer the achievement test in the first year.

[15] For example, Chaudhury et al. (2006) write on p. 100 that "the majority of absences appear to be due to those who attend between 50 percent and 80 percent of the time, and the median teacher is absent 14 to 19 percent of the time." (In fact, the source for this calculation was Glewwe et al. (2003), the working paper version of Glewwe et al. (2010).)

[16] Even if school days are fixed in length, this alone would not rule out there being intensive-margin variation—for example, teachers might be present for only part of a school day.

affect teacher absences, finding that the vast majority of labor supply variation was along the extensive margin.[17] While their context differs from that of Glewwe et al. (2010), both papers consider teachers in developing countries, where policymakers are concerned about the high teacher absenteeism rates. Putting the above together, the probability that a teacher was present during site visits may reasonably be viewed as affine transformations of hours teachers were at work, a common measure of labor supply. I examine the sensitivity of my results with respect to my measure of labor supply in Section 4.4.2.

3 Model

The model provides the foundation for the partial identification of the components governing human capital accumulation. It relates output (i.e., student achievement) and teacher human capital to production and investment inputs. The specifications for the production of human capital and output enable a closed-form, constructive, identification proof and calculation of bounds using ATE estimates, meaning the analysis is informative about average behavior.[18] They are also not very restrictive either in terms of implications for behavior,[19] or compared with specifications commonly used in the literature.[20]

3.1 Environment

Let the periods be indexed by $t = 0, 1, 2$, where $t = 0$ is the pre-treatment period, $t = 1$ is the "active-treatment" period, in which teachers in the treatment group are offered output-based incentives, and $t = 2$ is the post-active-treatment period, where teachers in either treatment or control group no longer are offered the incentives.

[17] Duflo et al. (2012) measure absences in two ways: monthly random site visits to schools and by information submitted by each teacher. This information consisted of two pictures of the teacher with a least eight students, taken at least five hours apart on the same day (the cameras were modified to prevent alterations of photo timestamps). The two absence measures agreed for about 80% of the observations (i.e., the enumerator marked the teacher in question as "present" during a random site visit and the teacher provided the required input, or the enumerator marked the teacher as "absent" during the site visit, and the teacher did not provide the required input).

[18] While averages are not the only conceivable objects of policy interest, they certainly have garnered a vast amount of interest by researchers and policymakers, as they may average out unobserved heterogeneity and also may be useful for certain normative considerations (e.g., maximizing output).

[19] The human capital production function does not assume that growth in human capital is constant over time. Indeed, as the model does not assume optimality of teacher behavior, it allows for, e.g., decreasing OJT investments over the course of a teacher's career, which could generate the concave value-added profiles documented in the literature.

[20] The model developed in this section could be viewed as a log-linear approximation to a specification considering the behavior of a representative teacher and student in each of the control and treatment groups (see Appendix A), and is also consistent with (specifically, nested by) the commonly used translog specification used in the burgeoning literature using dynamic factor models to understand skill growth (see, e.g., Agostinelli and Wiswall, 2023; Del Bono et al., 2022; Freyberger, 2021). While the link developed in the appendix might provide useful context, it is in no way necessary to this paper's analysis.

The human capital of teacher i in period t, θ_{it}, is produced according to

$$\theta_{it} = \delta_\theta \theta_{it-1} + \delta_I I_{it-1} + \delta_h h_{it-1}, \tag{1}$$

where θ_{it-1} is teacher i's human capital last period, I_{it-1} is teacher i's investment in human capital last period, and h_{it-1} is teacher i's input to production (e.g., actively teaching, engaging in teaching preparation, providing feedback to students, etc.) last period. Output produced by teacher i teaching student j in period t, y_{ijt}, which is measured by performance on a standardized achievement test, follows a standard value-added specification (Hanushek, 1971; Murnane, 1975; Hanushek, 1979):

$$y_{ijt} = \underbrace{\theta_{it} + \beta_h h_{it}}_{q_{it}} + \beta_y y_{jt-1} + \epsilon_{ijt}, \tag{2}$$

where y_{jt-1} is student j's prior test achievement (meaning β_y measures the persistence of student knowledge), $\theta_{it} + \beta_h h_{it}$ is teacher i's value added to student j's achievement in period t (i.e., *quality*, q_{it}),[21] and ϵ_{ijt} is an ex post IID productivity shock, with $\mathrm{E}\left[\epsilon_{ijt}|q_{it}, y_{j,t-1}\right] = 0$.[22]

It is not uncommon to fix labor supply when operating within the "pure OJT" framework (see, e.g., Ben-Porath, 1967; Heckman et al., 1998; Kuruscu, 2006; Huggett et al., 2011), but doing so would hamstring the LBD force, so it is important to allow for labor supply responses in light of our goals. The teacher's input (or resource) constraint links production, h_{it}, investment, I_{it}, and leisure, l_{it}:

$$h_{it} + I_{it} + l_{it} = M. \tag{3}$$

The quantities h, I, and l could be viewed as shares of a teacher's endowment M of total time, or of a teacher's overall focus or potential total inputs respectively allocated to production, investment, and leisure. Under the latter interpretations, a teacher has a fixed "budget" of focus/potential total inputs, which can be allocated between h, I, and l; naturally, a teacher must be working to engage in production or investment. While I typically use the term "time" to be consistent with the human capital literature (see, e.g., Ben-Porath, 1967; Heckman, 1976; Killingsworth, 1982), it may be useful to keep these alternative interpretations in mind. I refer to $h_{it} + I_{it}$ as *total inputs*. In the standard case where the endowment is assumed to be fixed, total inputs are equal to *labor supply*, $M - l_{it}$, which I sometimes also refer to as *(total) work time*; eq. (3) then captures the notion that

[21] Note that the scale of θ is determined by y. This implicit normalization in eq. (2) has no bearing on the returns to teacher human capital, which are governed by δ_θ. I show below that δ_θ does not appear in the expressions for the bounds (see Agostinelli and Wiswall, 2023, for a discussion of some of the pitfalls to avoid when (point-)estimating the returns to scale for skills). Also, I abstract from student characteristics here; as will become clear, what matters is obtaining a reasonable estimate of β_y.

[22] Note that the tuple (i, j, t) implies a pair (g, t), where $g \in \{T, C\}$ indicates whether the observation belongs to the treatment (T) or control (C) group. In other words, I am effectively making the standard value-added modeling assumption that the conditional mean of ϵ_{ijt} is the same across the treatment and control groups, for the pre-, active-, and post-active-treatment periods. A body of research supports the view that controlling for prior achievement (as is done in a value-added model) does a reasonably good job of controlling for unobserved prior inputs (see, e.g., Kinsler, 2012; Chetty et al., 2014).

total inputs should not increase if leisure has increased. As is also standard, neither h_{it} nor I_{it} are observed, and, further, they have no inherent scale. However, we can identify labor supply because the data contain measures of it (or equivalently, of its complement, leisure l_{it}), which means the scale of total inputs is fixed by the scale of the measure of labor supply.[23]

Due to lack of nested specifications in the literature, the model specification, while perhaps natural, may nonetheless warrant discussion. As in the classic human capital frameworks, given labor supply, production increases are accomplished via reductions in OJT investment; I refer to this input reallocation as the *dynamic multitasking* margin. At the same time, production increases can also obtain from increasing labor supply, fixing investment. I discuss below how the model can in principle capture what many in the teacher incentives literature have viewed as an "effort" margin underlying responses to output-based interventions, and also examine the role potentially played by additional response margins, by analyzing an extended version of the model relaxing the fixed endowment (see Section 4.4.3).

I assume that $\delta_k \in \mathfrak{D}_k$, where $\mathfrak{D}_k = [0, \overline{\delta}]$, for $k = \theta, I, h$, and that $\beta_k \in \mathfrak{B}_k$, where $\mathfrak{B}_k = [0, \overline{\beta}]$, for $k = h, y$. The lower bound of zero for each parameter captures the natural assumption that inputs cannot have negative effects. The upper bound for each parameter ($\overline{\delta}$ or $\overline{\beta}$) is taken to be large; I discuss below how the specific values of $\overline{\delta}$ and $\overline{\beta}$ do not affect this paper's main findings.

I define the "pure OJT" specification for the human capital production function as $\delta_I > 0$ and $\delta_h = 0$. Analogously, in the "pure LBD" specification, we have $\delta_h > 0$ and $\delta_I = 0$. Some scenarios representing different possibilities for true values of $\delta := [\delta_h, \delta_I]'$ are illustrated in Figure 1 (bounds for the *identified set* for (δ_h, δ_I), which is denoted $D \subseteq \mathfrak{D}_h \times \mathfrak{D}_I$, are derived in Section 3.2). For example, the point labeled "pure OJT", on the horizontal axis, features a positive OJT component, with no LBD component, in contrast to the point labeled "pure LBD", on the vertical axis. The interior point, labeled "OJT and LBD both present", represents the possibility that teachers accumulate human capital via both OJT and LBD components; in contrast, the "pure" versions of the OJT and LBD human capital production functions are mutually exclusive.

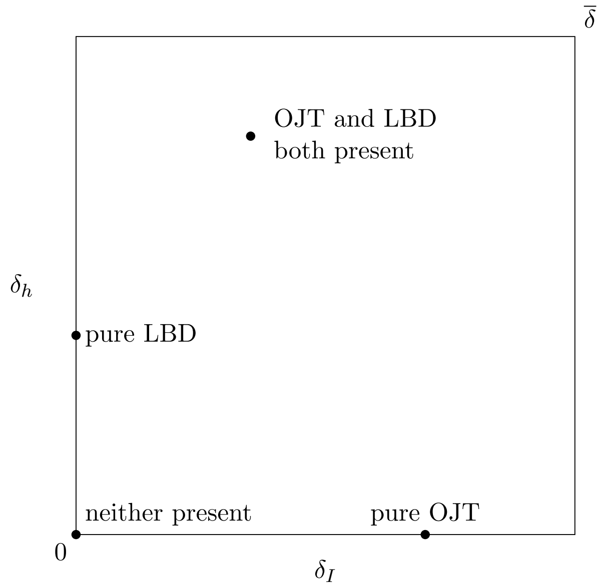
3.2 Derivation of bounds

This section develops bounds for (δ_h, δ_I) that only depend on period-specific estimates of the average treatment effects of the intervention on student achievement and on labor supply, and an estimate of the persistence of student knowledge (which could be estimated using the same dataset, or obtained from another source). To proceed, I first define the mean difference between the treatment and control groups for some variable z in period t as $\Delta z_t := \mathrm{E}\left[z_t^T\right] - \mathrm{E}\left[z_t^C\right]$, where $\mathrm{E}\left[z_t^g\right]$ denotes the expected value of z for group g in period t.[24] The bounds are derived under a maintained assumption of experimental balance (i.e., $\Delta z_0 = 0$ for $z = \theta, h, y, \epsilon$ and $\Delta y_{-1} = 0$),

[23] The value of the endowment, M (on the right side of eq. (3)), is immaterial so long as it is fixed, as it differences out in the analysis below.

[24] For $t \geq 1$, Δz_t corresponds to the average treatment effect (ATE) on z_t.

Figure 1: Examples of OJT and LBD specifications



which is consistent with the extensive balance tests conducted in Glewwe et al. (2010).[25] The bounds are sharp, meaning they contain only the values of (δ_h, δ_I) that cannot be rejected given the data. For example, the sharp bound on δ_I would increase the lower bound for δ_I as much as possible while still being consistent with the data; if the lower bound were greater than zero, we would reject the pure LBD specification.

First, note that in any period the constant endowment implies that

$$\Delta h_t + \Delta I_t + \Delta l_t = 0. \tag{4}$$

For the pre-treatment period, $t = 0$, the mean difference in achievement between the treatment and control groups is

$$\Delta y_0 = \underbrace{\Delta\theta_0}_{=0} + \beta_h \underbrace{\Delta h_0}_{=0} + \beta_y \underbrace{\Delta y_{-1}}_{=0} + \underbrace{\Delta\epsilon_0}_{=0} = 0, \tag{5}$$

i.e., a balanced experimental design implies there will be no average difference in the pre-treatment scores. Consistent with the balance tests in Glewwe et al. (2010), which did not reject experimental balance, I also assume $\Delta\epsilon_t = 0$ for $t = 1, 2$, and so I omit $\Delta\epsilon_t$ hereafter.

[25]For example, they find mean pre-active-treatment scores are not significantly different between the control and treatment groups; see eq. (5). See Deaton and Cartwright (2018) for a discussion of the usefulness of the balance assumption in analyzing experimental data.

For the active-treatment period, $t = 1$, we have

$$\Delta y_1 = \underbrace{\Delta\theta_1}_{=0} + \beta_h \Delta h_1 + \beta_y \underbrace{\Delta y_0}_{=0} = \beta_h \Delta h_1. \tag{6}$$

Equation (6) shows that the difference between treatment and control achievement in the active-treatment period can only come from the change in mean production time Δh_1, as the fact that θ_{it} depends on lagged inputs and balance between the treatment and control groups implies that $\Delta\theta_0 = 0$, while, as shown in eq. (5), balance between the treatment and control groups implies that $\Delta y_0 = 0$.

The mean difference in achievement between the treatment and control groups for the post-active-treatment period, $t = 2$, is

$$\begin{aligned}
\Delta y_2 &= \Delta\theta_2 + \beta_h \Delta h_2 + \beta_y \Delta y_1 \\
&= \delta_I \Delta I_1 + \delta_h \Delta h_1 + \beta_h \Delta h_2 + \beta_y \Delta y_1 \\
&= \delta_I \underbrace{[-\Delta l_1 - \Delta h_1]}_{=\Delta I_1,\text{ from eq. (4)}} + [\delta_h + \beta_h \beta_y]\Delta h_1 + \beta_h \Delta h_2 \\
&= -\delta_I \Delta l_1 + \left[\frac{\delta_h - \delta_I + \beta_h \beta_y}{\beta_h}\right] \underbrace{\Delta y_1}_{\beta_h \Delta h_1,\text{ from eq. (6)}} + \beta_h \Delta h_2,
\end{aligned} \tag{7}$$

which uses $\Delta\theta_1 = 0$ to go from the first to the second line. Equation (7) shows that, in general, only a locus of (δ_h, δ_I) will be identified, and, further, that many other variables appear in the same equation: the achievement production function parameters (β_h, β_y) and, even after using eq. (6) to eliminate Δh_1, the quantities $(\Delta y_1, \Delta y_2, \Delta l_1, \Delta h_2)$.[26] The upper bound on the parameter spaces for δ_h and δ_I, $\overline{\delta}$, is also unknown. I estimate $(\beta_y, \Delta y_1, \Delta y_2, \Delta l_1)$ in Section 4.1. However, Δh_2 is unobserved and it cannot be eliminated without additional assumptions, as was Δh_1 via eq. (6). In the following I will sometimes refer to the labor supply response during active treatment, $-\Delta l_1$, which is opposite of the change in leisure.

I now discuss how I arrive at values for $(\beta_h, \overline{\delta}, \Delta h_2)$. I show below that Δy_1 is significantly greater than zero (i.e., there was a positive effect of the intervention during the active-treatment period). In light of eq. (6) and the fact that $\beta_h \geq 0$, it is then reasonable to treat β_h as strictly positive, as the positive effect on achievement during the active-treatment period could only be rationalized by an increase in time allocated to production (i.e., $\Delta h_1 \geq 0$). Because the scale of β_h is not identified separately from $\overline{\delta}$, I fix $\beta_h = 1$ and $\overline{\delta} = 1$ hereafter.[27] Note, however, that the

[26] Note that δ_θ does not appear in eq. (7). Intuitively, this parameter measures how differences in teacher human capital emanating from differences in inputs from two periods ago affect production today; experimental balance means these differences are all zero, causing δ_θ to drop out.

[27] If the researcher viewed the model as a (log-linearized) approximation to a nonlinear model (see Appendix A for discussion), it would be natural to have $\beta_h = 1$ (i.e., the same as the scale of θ, which is consistent with the interpretation that a teacher's output equals her share of human capital allocated to production), and also to not allow δ_h or δ_I to exceed 1 (i.e., $\overline{\delta} = 1$).

specific values of β_h and $\overline{\delta}$ do not affect the main findings, such as whether bounds are informative or whether I can reject either the "pure OJT" or "pure LBD" specifications.[28]

Finally, I assume $\Delta h_2 = 0$, i.e., that production after the active-treatment period is the same in the control and treatment groups. While any potential income effects were dominated by substitution effects during active treatment (as discussed above, the data are consistent with $\Delta h_1 > 0$), the substitution effects would no longer be present afterwards. The structure of the incentive scheme, however, suggests that income effects, if present, would not be very large: the rewards were non-fungible prizes, which 24 of the 50 treatment schools received, and the prize values ranged from less than 2% to 4% of the typical teacher's annual salary (Glewwe et al., 2010, p. 208). While this does not preclude there being an average difference in production after active treatment, it also seems likely that Δh_2 would be small, if it were indeed different from zero.[29]

Assumption 1 summarizes the parameter values discussed thus far. Assumption 1(i) is maintained hereafter. Assumption 1(ii) corresponds to making no assumption on returns to scale for the human capital production function. Section 3.4 explores how stronger assumptions about the returns to scale would tighten the identified set.

Assumption 1 (Baseline, no returns to scale assumption)**.**

(i) $\beta_h = 1,\ \Delta h_2 = 0$

(ii) $(\delta_h, \delta_I) \in [0,1] \times [0,1]$.

Using the values obtained thus far, eq. (7) becomes

$$\Delta y_2 = [\delta_h - \delta_I + \beta_y]\, \Delta y_1 - \delta_I \Delta l_1,$$

which we can rearrange to express the LBD component δ_h in terms of the OJT component δ_I and the remaining parameters:

$$\delta_h = \left[\frac{\Delta y_2}{\Delta y_1} - \beta_y\right] + \delta_I \left[1 + \frac{\Delta l_1}{\Delta y_1}\right]. \tag{8}$$

Equation (8) characterizes the combinations of (δ_h, δ_I) consistent with $(\beta_y, \Delta y_1, \Delta y_2, \Delta l_1)$. The identified set for (δ_h, δ_I) collects the values of (δ_h, δ_I) satisfying eq. (8) that are also consistent with the desired assumption on the returns to scale (e.g., Assumption 1(ii)). We can find the identified set for either parameter by projecting the locus characterized by eq. (8) onto the relevant axis, e.g., D_h can be obtained by projecting D onto the δ_h-axis. It will be convenient to rewrite eq. (8) as

$$\delta_h = \pi_{\text{icept}} + \pi_{\text{slope}} \cdot \delta_I, \tag{9}$$

[28] While there are possible parameter values where the results would be affected by the choice of β_h relative to $\overline{\delta}$, the estimated parameters are far from this region.

[29] I analyze the case where $\Delta h_2 \neq 0$ in Appendix C.5. Bounds would be tighter if $\Delta h_2 < 0$, and the effect of allowing for $\Delta h_2 = \rho > 0$ would affect bounds in the same way as would the same increase in β_y.

where the intercept of the locus of permissible combinations of (δ_h, δ_I) is the composite parameter $\pi_{\text{icept}} := \frac{\Delta y_2}{\Delta y_1} - \beta_y$ and the slope of the locus is the composite parameter $\pi_{\text{slope}} := 1 + \frac{\Delta l_1}{\Delta y_1}$.

Note that, even with the parameter values obtained so far, the relationship between (δ_h, δ_I) still depends on $(\beta_y, \Delta y_1, \Delta y_2, \Delta l_1)$. Section 4.1 discusses estimation of these parameters.[30] Briefly, all but β_y are estimated using the experimental variation. Standard methods (treatment-year fixed effects in the production function, coupled with the assumption that $\mathrm{E}\left[\epsilon_{ijt}|q_{it}, y_{j,t-1}\right] = 0$) yield consistent estimates of β_y when using just control-group data.

3.3 Comparative statics with respect to $(\beta_y, \Delta y_1, \Delta y_2, \Delta l_1)$

To gain intuition for how and when the data yield identification, I examine comparative statics for the identified set for (δ_h, δ_I) with respect to different combinations of elements of $(\beta_y, \Delta y_1, \Delta y_2, \Delta l_1)$, under Assumption 1. I assume achievement increased during the intervention (i.e., $\Delta y_1 > 0$), which is consistent with Glewwe et al. (2010), and consider cases that can be mapped to different values of the composite parameters $(\pi_{\text{icept}}, \pi_{\text{slope}})$, as these ultimately determine the values of (δ_h, δ_I) that are consistent with the data.

No labor supply response; no net post-intervention achievement effect I start by considering a case that highlights the potential difficulty in separating OJT and LBD forces. Suppose there is no labor supply response (where $\Delta l_1 = 0$, or $\pi_{\text{slope}} = 1$) and further suppose that there are no post-intervention achievement effects, net of those explained by active-treatment achievement effects (i.e., $\frac{\Delta y_2}{\Delta y_1} - \beta_y = 0$, or $\pi_{\text{icept}} = 0$), which means the persistence of the positive effect of Δh_1 on Δy_1 entirely accounts for Δy_2. This scenario corresponds to case (a) in Table 1, and to the dashed, 45-degree, line in the left panel of Figure 2. From the first line of eq. (7), the lack of net post-intervention achievement effects means we have $\Delta\theta_2 = 0$, i.e., post-intervention teacher human capital is on average the same in the treatment and control groups. Further, the lack of a labor supply response means the increase in production implied by $\Delta y_1 > 0$ must reduce investment by the same amount $(\Delta I_1 = -\Delta h_1)$, meaning any increase in δ_I can satisfy eq. (8) by a concomitant increase in δ_h (i.e., $\delta_I = \delta_h$). Projecting the identified set onto each axis, we can see that the marginal identified set for either δ_I or δ_h (depicted by the dashed lines just outside that parameter's axis) has not shrunk at all, meaning the bounds would be uninformative in this case. Intuitively, any feasible value of, e.g., δ_I, can be rationalized by the same value for the coefficient on the other input to human capital (in this example, δ_h). However, this case also shows that non-identification is non-generic, as it depends on there being no labor supply response as well as there being no post-intervention achievement effect, net of that attributable to the active-treatment achievement effect.

[30] If, for example, no measures of Δl_1 were available, the derived bounds would be the union over those for all possible values of Δl_1. As I show below, informative bounds can be obtained both when $\Delta l_1 = 0$ and when $\Delta l_1 \neq 0$.

No labor supply response; negative net post-intervention achievement effect Next consider case (b), which differs from case (a) in that the post-intervention achievement effect is lower than what can be explained by the persistence of the active-treatment achievement effect (i.e., $\frac{\Delta y_2}{\Delta y_1} - \beta_y < 0$, or $\pi_{\text{icept}} < 0$), which means teacher human capital after active treatment is lower in the treatment group ($\Delta\theta_2 < 0$). This case is depicted by the dotted lines in the left panel of Figure 2. The bounds become informative in this case, as the negative effect on post-intervention human capital rules out very low values of δ_I and very high values of δ_h. Intuitively, if the net effect of increasing time spent on production on teacher human capital is negative after active treatment, knowing that $\Delta I_1 = -\Delta h_1$ (as there continues to be no labor supply response in this scenario) implies that investment dominates production in human capital generation, meaning the OJT parameter δ_I must be larger than the LBD parameter δ_h.

Positive labor supply response; negative net post-intervention achievement effect In cases (a) and (b), there was no average difference in labor supply between the treatment and control groups during the active-treatment period. Now consider case (c), where we start from case (b) but now suppose there was an increase in labor supply, meaning $\Delta l_1 < 0$ (i.e., $\pi_{\text{slope}} < 1$). Here, the reduction in investment must be smaller than the increase in production (i.e., $\Delta I_1 > -\Delta h_1$). As shown in the dash-dotted lines in the left panel, this rotates the locus defined by eq. (8) clockwise from the intercept (which was already negative, as the starting point was case (b)), increasing the lower bound on δ_I and decreasing the upper bound on δ_h.

Other cases The above cases are not exhaustive. For example, the signs of $\frac{\Delta y_2}{\Delta y_1} - \beta_y$ or Δl_1 could be opposite to those considered in cases (b) or (c), in which case the bounds would be different. In this vein, consider case (d), in which there is no labor supply response but now a positive post-intervention achievement effect, net of the effect on active-treatment achievement. This case, which is roughly consistent with the estimation results in Section 4.1, is depicted in the right panel of Figure 2 and sets the stage for showing how strengthening returns-to-scale assumptions can yield tighter bounds for the identified set.

3.4 Comparative statics with respect to increasing assumption strength

In this section I examine how the identified set is affected by assumptions on $\delta_h + \delta_I$, the returns to scale in the human capital production function. A researcher might find it natural to restrict the returns to scale, by assuming they are nonincreasing. This is Assumption 2 below. Even stronger, the researcher might believe it reasonable to assume constant returns to scale (Assumption 3). This exploration of how assumptions about $\delta_h + \delta_I$ affect the identified sets is in the spirit of the "worst-case" approach of Horowitz and Manski (2000), which examines the sensitivity of findings to stronger sets of assumptions, some of which have been made in the literature. It does not constitute an endorsement of making these stronger assumptions.

Assumption 2 (Nonincreasing returns to scale (NIRS)). *Assumption 1(i) and* $\delta_h + \delta_I \leq 1$.

Assumption 3 (Constant returns to scale (CRS)). *Assumption 1(i) and* $\delta_h + \delta_I = 1$.

The right panel of Figure 2 shows the additional information embedded in assumptions about the returns to scale for (δ_h, δ_I), starting with Assumption 1 (case (d)); these cases are summarized in the lower half of Table 1. Case (d2) further imposes the restriction that $\delta_h + \delta_I \leq 1$ (Assumption 2), which means permissible combinations of (δ_h, δ_I) lie in the south/west right triangle.[31] By examining the dotted lines in the right panel of Figure 2 that correspond to this case, we can see that the marginal identified sets are tighter than those in case (d). Intuitively, the tighter upper bound for δ_h, combined with the nonincreasing returns to scale assumed in case (d2), yields a tighter upper bound for δ_I. Case (d3) further imposes the restriction that $\delta_h + \delta_I = 1$ (Assumption 3), which affords point identification (interior solid point in the right panel of Figure 2).

Table 1: Bounds: Comparative statics and effects of returns-to-scale assumptions

Case	$(\delta_h, \delta_I) \in$	Δl_1	$\frac{\Delta y_2}{\Delta y_1} - \beta_y$	π_{icept}	π_{slope}	Notes on bounds
(a)	$\delta_h \leq 1,\ \delta_I \leq 1$	0	0	0	1	Uninformative
(b)	$\delta_h \leq 1,\ \delta_I \leq 1$	0	< 0	< 0	1	Informative
(c)	$\delta_h \leq 1,\ \delta_I \leq 1$	< 0	< 0	< 0	< 1	Tighter than (b)
(d)	$\delta_h \leq 1,\ \delta_I \leq 1$	0	> 0	> 0	1	Informative
(d2)	$\delta_h + \delta_I \leq 1$	0	> 0	> 0	1	Tighter than (d)
(d3)	$\delta_h + \delta_I = 1$	0	> 0	> 0	1	Point identification

Note: $\Delta y_1 > 0$ in all cases. Cases (a)-(c) correspond to comparative statics with respect to parameters in the identified locus, and cases (d)-(d3) correspond to comparative statics with respect to different returns-to-scale assumptions.

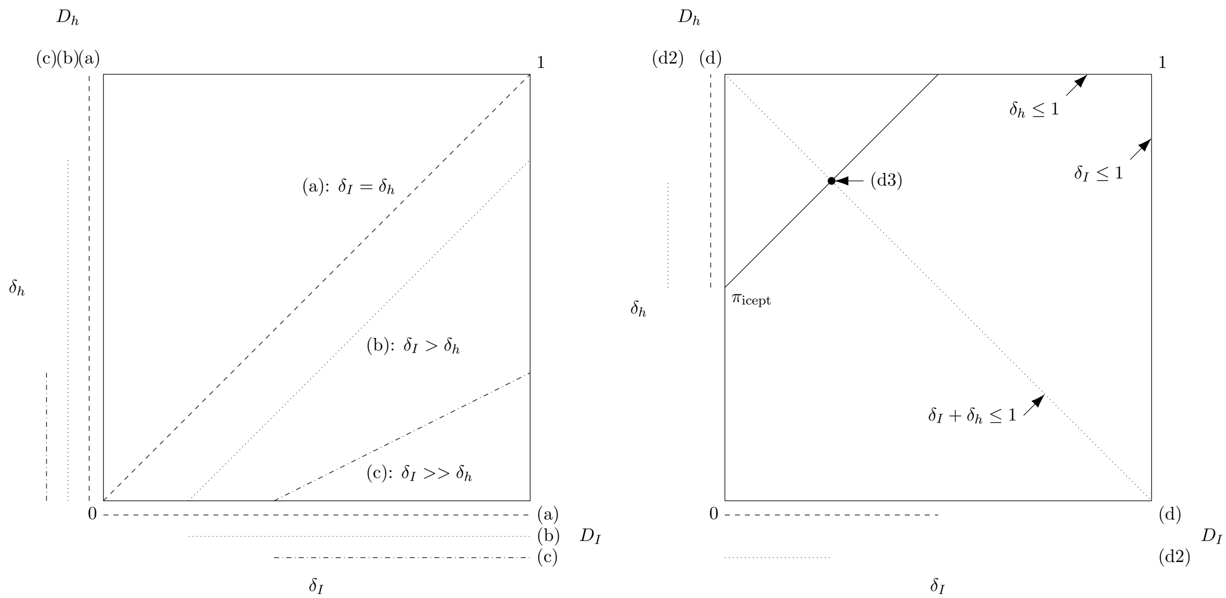
3.5 Expressions for marginal identified sets

This section ends by characterizing the marginal identified sets for δ_h and δ_I, which can be used to estimate the parameters' bounds and confidence sets. Under Assumption 1, i.e., when making no additional assumptions about returns to scale, we can derive the identified set for δ_h, D_h, by varying δ_I over its domain, resulting in

$$\delta_h \in [\max\{\pi_{\text{icept}}, 0\}, \min\{\pi_{\text{icept}} + \pi_{\text{slope}}, 1\}]. \tag{10}$$

[31] Note that this restriction in itself would not yield any information; without data (e.g., information about π_{icept}), the bounds on either parameter, obtained from projecting the south/west right triangle onto either axis, would be uninformative.

Figure 2: Illustration of example identified set cases



Note: Marginal identified sets for parameters are indicated by the lines outside their respective axes. Cases depicted in the figure are summarized in Table 1. The left panel corresponds to comparative statics, cases (a)-(c). The right panel corresponds to different returns-to-scale assumptions, cases (d)-(d3).

Solving eq. (9) instead for δ_I, we can analogously obtain D_I by varying δ_h over its domain:

$$\delta_I = \frac{\delta_h - \pi_{\text{icept}}}{\pi_{\text{slope}}} \Rightarrow \delta_I \in \left[\max\left\{\frac{-\pi_{\text{icept}}}{\pi_{\text{slope}}}, 0\right\}, \min\left\{\frac{1-\pi_{\text{icept}}}{\pi_{\text{slope}}}, 1\right\}\right]. \tag{11}$$

Invoking non-increasing returns to scale (Assumption 2) does not affect the lower bound of D_h, but does tighten its upper bound:

$$\begin{aligned}&\delta_h = \pi_{\text{icept}} + \pi_{\text{slope}} \cdot \delta_I \Rightarrow \delta_h \leq \pi_{\text{icept}} + \pi_{\text{slope}} \cdot [1-\delta_h] \Rightarrow \\ &\delta_h \in \left[\max\left\{\pi_{\text{icept}}, 0\right\}, \min\left\{\frac{\pi_{\text{icept}} + \pi_{\text{slope}}}{1+\pi_{\text{slope}}}, 1\right\}\right].\end{aligned} \tag{12}$$

We can analogously tighten the upper bound on D_I by applying Assumption 2:

$$\begin{aligned}&\delta_h = \pi_{\text{icept}} + \pi_{\text{slope}} \cdot \delta_I \Rightarrow [1-\delta_I] \geq \pi_{\text{icept}} + \pi_{\text{slope}} \cdot \delta_I \Rightarrow [1+\pi_{\text{slope}}] \cdot \delta_I \leq 1 - \pi_{\text{icept}} \\ &\delta_I \in \left[\max\left\{\frac{-\pi_{\text{icept}}}{\pi_{\text{slope}}}, 0\right\}, \min\left\{\frac{1-\pi_{\text{icept}}}{1+\pi_{\text{slope}}}, 1\right\}\right].\end{aligned} \tag{13}$$

Finally, imposing Assumption 3, we achieve point identification for both parameters:

$$\delta_h = \pi_{\text{icept}} + \pi_{\text{slope}} \cdot [1 - \delta_h] \qquad \Rightarrow \delta_h = \frac{\pi_{\text{icept}} + \pi_{\text{slope}}}{1 + \pi_{\text{slope}}} \tag{14}$$

$$[1 - \delta_I] = \pi_{\text{icept}} + \pi_{\text{slope}} \cdot \delta_I \qquad \Rightarrow \delta_I = \frac{1 - \pi_{\text{icept}}}{1 + \pi_{\text{slope}}}. \tag{15}$$

4 Empirical results

My estimation approach consists of two, distinct steps. In the first step (see Section 4.1), I estimate $(\Delta y_1, \Delta y_2, \beta_y, \Delta l_1)$, where I have a unique, closed-form, solution for the estimator of each component. In the second step (see Section 4.2), I use the estimates from the first step and the analytical characterization of the identified set to generate estimates of the identified set D and of the marginal identified sets for the parameters δ_h and δ_I, for each assumption about the returns to scale. Section 4.2 also describes how I create confidence sets for δ_h and δ_I.

4.1 Estimation step 1: Estimation of parameters determining the identified set

I pool the two active-treatment years to map the empirical application, which spans four years of data, to the three-period model structure. The pre-treatment year (1997) corresponds to model period $t = 0$. Both active-treatment years (1998 and 1999) are pooled into one period, corresponding to $t = 1$ in the model, and the post-active-treatment year (2000) corresponds to $t = 2$.[32] While pooling the active-treatment data has the virtue of providing a clean mapping from the data to the model, it also highlights the role played by the lack of an immediate payoff to OJT. Additionally, it is not obvious which value of $\hat{\beta}_y$ to use, as the pooled active-treatment data combine observations over two years.[33] I discuss this point just below.

Achievement Table 2 presents the achievement ATE estimates for the active-treatment period (Active $*$ treated) and the post-active treatment period (Post $*$ treated). I also include the treatment year as a regressor (Active-treatment, Post-active-treatment) to control for secular trends. The results for the treatment group in the active-treatment period indicate that the average treatment effect pooled over both active-treatment years, Δy_1, is positive and significantly different than zero (0.089, with a p-value less than 0.01). The next row indicates that student achievement in the post-active-treatment year, Δy_2, remained significantly higher (0.098, with a p-value less than 0.01) in the treatment group.[34] Through the lens of the model, this positive ATE could be

[32] The availability of only one year of post-intervention data means the results are based on a relatively short follow-up horizon. With multiple years of post-active-treatment data one could gauge longer-run effects on human capital. For example, a researcher with access to multiple years of follow-up data could pool those observations to estimate longer-run effects on human capital.

[33] The panel structure of the data is used only in the estimation of persistence β_y.

[34] While the current paper's estimating equations differ slightly from those in Glewwe et al. (2010), the point estimates in their Table 3 (averaging those from active-treatment years to render them comparable) are about 0.09-0.10, which are similar to those presented here.

Table 2: Estimates of ATEs on achievement

	Dependent variable:
	Test score y_{ijt}
Active-treatment ($t = 1$)	0.001
	(0.019)
Post-active-treatment ($t = 2$)	-0.001
	(0.017)
Active $*$ treated (Δy_1)	0.089
	(0.013)
Post $*$ treated (Δy_2)	0.098
	(0.025)
Constant	0.007
	(0.008)
Observations	26,537

Note: The t corresponds to a pooled period.

due to persistence of higher achievement from the active-treatment period and (potentially) higher teacher human capital for the treatment group in the post-active-treatment period. Establishing the relative importance of these effects is the goal of the next section.

Persistence Table 3 presents an estimate of the persistence component β_y using control group data, where the lagged score is that from the previous pooled period. I estimate the persistence component to be 0.544, and statistically greater than zero (with a p-value less than 0.01).[35] While this estimate is in the range of those in Andrabi et al. (2011), who estimated the persistence of a variety of cognitive skills using a dynamic panel data model applied to Pakistani schoolchildren,[36] researchers have also found that intervention-based changes in achievement fade out more quickly than those generated otherwise. For example, Jacob et al. (2010), who study the fade-out of different types of achievement gains, estimate a persistence value of 0.20 for teacher-induced gains in reading achievement (Jacob et al., 2010, Table 2).[37] Therefore, I conduct a robustness exercise for the case where $\beta_y = 0.20$ in Section 4.4.1.

[35] While this might strike some readers as a bit low (for example, Hanushek, 2011, assumed a baseline persistence value of 0.70 in his study of the effects of teacher quality on students lifetime earnings), this estimate is not driven by the use of pooled data; when instead using unpooled data the persistence estimate is 0.618.

[36] Allowing for both measurement error and unobserved student heterogeneity (in contrast with the specifications researchers have typically used to estimate achievement production functions), Andrabi et al. (2011) estimate the annual persistence of cognitive skills to range from 0.20 to 0.55, across a variety of subjects. Indeed, they argue that the upper part of their range may be too high.

[37] Math gains had an estimated persistence of 0.27.

Table 3: Estimate of the persistence component

	Dependent variable:
	Test score y_{ijt}
Lagged score $y_{j,t-1}$ (β_y)	0.544
	(0.010)
Constant	-0.047
	(0.008)
Observations	5,007

Note: These estimates are obtained from a regression of current test score on the lagged test score, run on the pooled data for pooled periods 1 and 2, for the control group. Specifically, the (period) t subindex on the test score y_{ijt} refers to a pooled period, meaning student j's lagged test score $y_{j,t-1}$ is the one from the previous pooled period.

Labor supply I now discuss how I obtain a value for the effect of the intervention on labor supply during the active-treatment period, $-\Delta l_1$. When modeling labor supply, the literature estimating OJT models of human capital accumulation typically treats labor supply as measurable (see, e.g., Brown, 1976; Heckman, 1976; Blandin, 2018; Fu et al., 2021). The Glewwe et al. (2010) data contain two candidate teacher attendance measures in each period: the fraction of teachers in attendance at the school during random, unannounced, site visits by the research team that period and the fraction of teachers present in their classrooms during the site visits.[38] Because the labor supply measure needs to capture total inputs, it seems more appropriate to use the share of teachers present at the school (for example, teachers could make investments outside the classroom). Table 4 shows the treatment effect on the share of teachers in attendance, -0.017, is not significantly different from zero.[39] It is important to note that, even though the estimated effect on labor supply is not significantly different from zero at any conventional significance level, I use the estimated labor supply response (taking into account its associated uncertainty) when estimating bounds and confidence sets. I examine how sensitive the results are with respect to my measure of labor supply in Section 4.4.2.

4.2 Estimation step 2: Estimating the bounds and Estimation of confidence sets

I provide an overview here of how I estimate the bounds and confidence sets for δ_h and δ_I; see Appendix B for more detail. For each assumption about the returns to scale, given a first-step estimate of $(\Delta y_1, \Delta y_2, \beta_y, \Delta l_1)$, I construct an estimate of the identified set, $\hat{D}$, which is a closed

[38] As discussed in Section 2, researchers working in development contexts commonly measure teachers' labor supply using attendance data. I use school-period-level averages because teacher-level data on either measure were not available.

[39] I also include the treatment year as a regressor here to control for secular trends. This finding is consistent with Glewwe et al. (2010), who find no evidence that teachers in the treatment group on average altered their school attendance (see Panel A of their Table 5).

Table 4: Estimate of ATE on teacher attendance

	Dependent variable:
	Present at school in period t
Active-treatment ($t = 1$)	0.012
	(0.025)
Post-active-treatment ($t = 2$)	0.052
	(0.025)
Active $*$ treated ($-\Delta l_1$)	-0.017
	(0.029)
Post $*$ treated	-0.011
	(0.041)
Constant	0.833
	(0.014)
Observations	202

Note: Dependent variable is school-level average share of site visits during which teachers were in attendance. The t corresponds to a pooled period.

line segment in $[0,1]^2$. Specifically, the parameters characterizing the identified set, π_{icept} and π_{slope}, can be estimated using $\widehat{\pi}_{\text{icept}} := \widehat{\frac{\Delta y_2}{\Delta y_1} - \beta_y}$ and $\widehat{\pi}_{\text{slope}} := \widehat{1 + \frac{\Delta l_1}{\Delta y_1}}$, which are computed using plug-in estimators. I can then generate estimates of the marginal identified sets by projecting $\hat{D}$ onto each marginal dimension (in practice, I use the explicit expressions for the marginal identified sets, derived in Section 3.5).

To simulate the joint distribution of $(\widehat{\pi}_{\text{icept}}, \widehat{\pi}_{\text{slope}})$, taking into account the variability of the inputs to the plug-in estimators, I bootstrap the joint distribution of $(\widehat{\Delta y}_1, \widehat{\Delta y}_2, \widehat{\beta}_y, \widehat{\Delta l}_1)$, where in each bootstrap replication, the individual elements are estimated as described in Section 4.1.[40] Estimated bounds for δ_h and δ_I are the average across the simulated marginal identified sets for that parameter. I construct confidence sets to contain δ_h and δ_I at the pre-specified significance level (I use a 95% confidence level).[41] By projecting the identified set onto the marginal dimensions, I avoid the problem of overly conservative confidence sets described by Kaido et al. (2019).

My estimation approach may bear a certain similarity with indirect inference, where the parameters of interest are estimated by fitting parameters from auxiliary models (here, these would be the estimates of $(\Delta y_1, \Delta y_2, \beta_y, \Delta l_1)$). However, as I explain in Appendix B.2, my approach would more properly be viewed as being related to the method of moments.

[40] I bootstrap using 100,000 replications of the data, stratified by treatment status.

[41] See Imbens and Manski (2004) for a discussion of the distinction between the confidence set for a parameter vs. the confidence set for a parameter's identified set.

4.3 Estimation results

Figure 3 illustrates the estimated bounds and 95% confidence sets for δ_h and δ_I under the different assumptions about $\delta_h + \delta_I$. The table below presents the corresponding estimates and also reports bound widths. Starting with the LBD parameter in panel (a), we can see that when we do not impose a returns-to-scale-type assumption ("None"), the estimated upper and lower bounds (thick, black, line) for δ_h are informative, and the lower bound is greater than zero at the 95% confidence level (thin, grey, line), leading us to reject the pure OJT specification (in which $\delta_h = 0$). Similarly, in panel (b) we can see that the estimated upper and lower bounds for δ_I are informative, and that the 95% confidence set for δ_I does not contain the upper bound when no assumption about returns to scale is made ("none").

At this point it would be natural to ask how reasonable it would be to surmise that LBD could underlie the observed patterns, in light of the fact that LBD operates via changes in production time during active treatment h_1, coupled with the previous finding that there was not a significant change in labor supply during active treatment. This is fine for two reasons. First, while the identified set for δ_I does include zero, it also includes positive values; such parameter values would naturally yield positive control-group levels of OJT investments that could decrease in the active-treatment period. Second, by using the estimates in Table 4 we can see that the 95% confidence interval for the change in labor supply ($-\Delta l_1$) is $[-0.074, 0.04]$, which includes increases of 9% over the pre-intervention mean of 0.833.

Imposing non-increasing returns to scale ("NIRS" on the horizontal axis in each panel) tightens the estimated upper bound for δ_h, although it does not affect the 95% confidence set, because the estimated lower bound, which governs the lower bound of the confidence set, is unchanged. However, imposing NIRS does tighten the estimated upper bound and upper bound on the confidence sets for δ_I. Intuitively, high values of both δ_h and δ_I are no longer mutually feasible under NIRS, lowering the upper bounds for both parameters. Consequently, we can see in the accompanying table that the width of the estimated bounds falls by about one half for both parameters when imposing NIRS (e.g., from 0.429 to 0.231 for δ_h), and the same is true of the width of the 95% confidence set for δ_I. Finally, imposing constant returns to scale ("CRS" on the horizontal axis in each panel) yields point identification for both parameters, resulting in confidence sets with the smallest width. The estimated bounds are of course zero width in this case.

Overall, under all the assumptions about returns to scale, the achievement ATE in the post-active-treatment period is larger than would be accounted for by the positive ATE in the active-treatment period (caused by an increase in h_1, an input to contemporaneous student achievement) and persistence of this increased student achievement.[42] That is, the OJT component, operating through $\Delta I_1 < 0$, is dominated by the LBD component, operating through $\Delta h_1 > 0$. We can reject

[42] Note the effect does not need to be stronger in the post period—what matters is whether the impact was stronger than could have been explained by persistence of impact from the previous year. That is, the informative bounds obtained here do not require $\Delta y_2 > \Delta y_1$. To see this, suppose, for simplicity, that $\Delta l_1 = 0$. In this case, the bounds would be qualitatively similar for $\Delta y_2/\Delta y_1 > \beta_y$, which would be possible even when $\Delta y_2 \leq \Delta y_1$.

the "pure OJT" specification in which $\delta_h = 0$ across all returns-to-scale assumptions at the 95% confidence level. However, we cannot reject that $\delta_I = 0$ in any of the returns-to-scale assumptions at the 95% confidence level; this means we cannot reject the "pure LBD" specification. Further, if we were willing to assume CRS, we could infer that δ_h was greater than δ_I, as the 95% confidence sets under CRS do not overlap; the same cannot be said under either of the weaker assumptions. In addition to these substantive results about determinants of human capital, the different results across Assumptions 1-3 demonstrate the identifying power of assumptions on the returns to scale.

4.4 Additional checks and robustness

4.4.1 Persistence

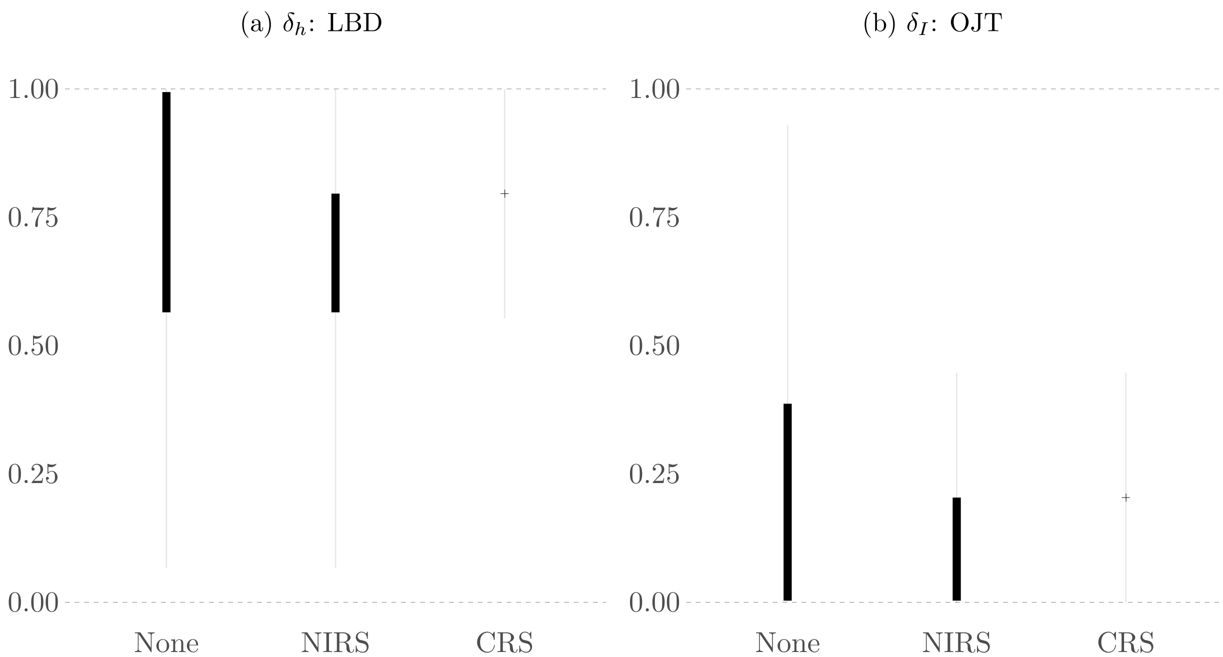
As discussed in Section 4.1, there is reason to believe the estimates of achievement persistence β_y, used in the baseline results presented just above, may be too high. Therefore, I recomputed the bounds using a persistence value of $\beta_y = 0.20$, which is the estimate of the persistence of teacher-induced gains in reading from Jacob et al. (2010). The results, presented in Appendix C.1, show the bounds are all (weakly) tighter than under the baseline results, and all intervals are strictly smaller (i.e., have lower width). Intuitively, the post-active-treatment achievement ATE is higher than can be explained by the persistence of the active-treatment achievement ATE, pointing to the presence of an LBD component. Lowering the value of β_y would only make this effect more prominent.

4.4.2 Measure of labor supply

Notwithstanding researchers' use of extensive-margin measures of labor supply such as attendance data to understand teachers' output across a variety of development contexts, there could still be a concern that my measure of labor supply might not fully capture teachers' relevant responses, especially in light of the fact that my measure—whether a teacher was present during a random site visit to the school—actually exhibits an (insignificant) *negative* response. While the potential mismeasurement of the labor supply response would only affect the estimated bounds insofar as the changes in total inputs substantially differed between the control and treatment groups, it would be prudent to examine the sensitivity of my results to my measure of labor supply.

I have conducted several robustness exercises regarding the change in active-treatment labor supply, $-\Delta l_1$. First, I derive bounds for the case where there is no labor supply response (in contrast to the slight negative response used in the main, baseline, results). As shown in Appendix C.2, this tends to rotate the (δ_h, δ_I) loci clockwise from their baseline values, lowering the upper bound for δ_h and increasing the upper bound for δ_I. Neither effect is very large, and the bounds and confidence sets for δ_h and δ_I continue to remain informative. I have also recomputed the bounds when using the share of teachers who were present in their classrooms during site visits as the measure of labor supply. This variable could be especially useful because according to

Figure 3: Estimated bounds and confidence sets



Notes: The left panel depicts estimated bounds (▬) and 95% confidence sets (—) for δ_h, under Assumptions 1, 2, and 3 (denoted on the bottom axis via "None", "NIRS", and "CRS", respectively). The right panel depicts the analogous results for δ_I.

	Assumption	Estimated bounds			95% confidence set		
Parameter	about $\delta_h + \delta_I$	Min.	Max.	Width	Min.	Max.	Width
δ_h	1: None	0.565	0.994	0.429	0.067	1	0.933
	2: NIRS	0.565	0.796	0.231	0.067	1	0.933
	3: CRS	0.796	0.796	0	0.553	1	0.447
δ_I	1: None	0.003	0.387	0.384	0	0.930	0.930
	2: NIRS	0.003	0.204	0.201	0	0.447	0.447
	3: CRS	0.204	0.204	0	0	0.447	0.447

Notes: Assumption 1 corresponds to no assumption about the returns to scale for $\delta_h + \delta_I$, and Assumptions 2 and 3 respectively correspond to non-increasing and constant returns to scale.

this measure, labor supply increased during active treatment (however, note this increase was not different from zero at any conventional significance level), perhaps mimicking an increase in "effort" or the total inputs chosen by teachers (I make this connection explicit via a model extension in the next subsection). As discussed in Appendix C.3, the bounds and confidence sets for δ_h and δ_I continue to remain informative when using this alternative measure of labor supply. Finally, I have recomputed the results under an alternative assumption where I use the change in labor supply implied by an external estimate of the labor supply response (from Attanasio et al., 2018), applied to the intervention. This assumption nicely complements the case where I measure labor supply using teachers' classroom presence, because the labor supply response is unambiguously positive in this case. As discussed in Appendix C.4, the estimated bounds and confidence sets for δ_h and δ_I continue to remain informative. Overall, the results are robust across a variety of different types of assumptions about the labor supply response.

4.4.3 Accounting for other teacher response margins

As Taylor (2022) discusses in his recent handbook chapter, researchers have conjectured that increases in teacher quality due to incentive pay may be driven in part by increases in teacher "effort", a non-contractible input to student achievement (see, e.g., Muralidharan and Sundararaman, 2011; Macartney et al., 2018; Dinerstein and Opper, 2022). My framework can accommodate this concept, through the production input. Through the lens of the model, during active treatment, increases in student achievement y_1 occur via increases in production h_1. Like "effort", the production input can generate variation in achievement even without variation in labor supply, and, empirically, changes in the production input are inferred from changes in output, in a manner similar to how "effort" responses are connected to output in the teacher incentives literature.

The meaningful distinction between my framework and the teacher quality literature lies not in whether my framework captures "effort", but rather, in the dynamic implications of incentivizing output. What the teacher incentive literature has deemed an "effort" margin could very well reflect a reduction in OJT investments, reducing teachers' future human capital and future student achievement (fixing labor supply, we would have $\Delta I_1 = -\Delta h_1$). This is the dynamic multitasking channel. On the other hand, if output increased due to, say, increased productivity of the production input, there may be a smaller or even no associated reduction in OJT investment (again fixing labor supply, we would have $\Delta I_1 > -\Delta h_1$). That is, the identification of the forces underlying human capital accumulation depends on the value of ΔI_1 implied by Δh_1.

A small extension to the model can help organize our thinking on this matter. First recall that the teacher's input constraint eq. (3), in which the endowment M is fixed, shows that the increase in h_1 implied by the increase in achievement could occur via reductions in l_1 and/or I_1. Now consider the more general case in which the endowment's side of the input constraint may be augmented:

$$h_t + I_t + l_t = M + b_t, \tag{16}$$

where the bonus time $b_t \geq 0$ allows for an increase in the *effective endowment*, $M + b_t$. For example, if teachers took less downtime to more actively engage in teaching, or spent their lunch breaks planning lessons (instead of being social with colleagues), this could be captured as an increase in b_t. Using the augmented input constraint, the mean difference in the investment input between the treatment and control groups during active treatment would be

$$\Delta I_1 = -\Delta h_1 - \Delta l_1 + \Delta b_1. \tag{17}$$

If $\Delta b_1 = 0$ (as in the baseline analysis), then $\Delta I_1 = -\Delta h_1 - \Delta l_1$, i.e., if labor supply is constant a change in h_1 could only obtain via the opposite change in OJT investment I_1. More generally, eq. (17) shows that a change in effective endowment (Δb_1) would have the same effect on the relationship between the change in OJT investment (ΔI_1) and the change in production (Δh_1)—and, therefore, on the bounds for δ—as would the same change in labor supply ($-\Delta l_1$).

Because the central goal of this paper is to use the human capital perspective to understand the dynamics of teacher quality, it seems most natural to start by assuming the effective endowment was fixed (i.e., $\Delta b_1 = 0$). However, the interpretation of the results would appear tenuous if driven by the categorical exclusion of b_1. Crucially, without some idea about what Δb_1 might look like, we would lack empirical guidance about whether the primary margin of adjustment was the dynamic multitasking channel or a change in the effective endowment. Therefore, I examine the role potentially played by changes in the effective endowment.

There are a couple of ways to examine how important responses along the effective endowment margin might be for interpreting this paper's results. First, and most directly, we can check for changes in observed variables that could provide information about potential changes in the effective endowment in the augmented input constraint, eq. (16), during active treatment. Such changes could manifest via activities that increased the productivity of work time (e.g., being more actively engaged in teaching) and/or by reducing slack by squeezing more usable time out of a fixed amount of work time (e.g., doing some lesson planning during lunch breaks). If we found evidence that $\Delta b_1 > 0$, this would cast doubt on the interpretation that the increase in achievement, Δy_1 (generated by an increase in h_1), obtained solely via reductions in I_1 (conditional on labor supply). Glewwe et al. (2010) examine several potential behavioral channels that might underlie their estimated achievement treatment effects. They find no evidence of treatment responses in several measures of teacher behavior, such as whether the teacher was using a blackboard or instructional aid, and even a subjective measure of teacher "energy level" that ranges from 1-5, with 5 being the most energetic (Glewwe et al., 2010, Table 5). While these might not capture every possible margin of response,[43] many of the same measures have been used to assess teachers' input provision in work

[43] For example, these teacher behavior measures would not measure inputs exerted at home, such as developing lessons or marking quizzes. However, if teachers' at-home inputs had indeed exhibited large responses to the intervention, one might also have expected to see their behaviors at school to have changed. Altonji et al. (2005) provide intuition for why this kind of check might be useful. The idea is that, if measured variables in a data set are chosen randomly from the full set of variables, the amount of selection on observed variables would be informative about the

focused on carefully measuring teacher behavior (Duflo et al., 2012, p. 1267). The lack of results along these dimensions suggests potential productivity increases/reductions in slack (corresponding to increases in the effective endowment) might not have been a primary margin of adjustment in the current context.

Some of the robustness exercises might also be informative on this point. Specifically, sensitivity with respect to the labor supply response can help us think about how the proposed effort responses might manifest, because, as shown by eq. (17), an increase in labor supply (i.e., $\Delta l_1 < 0$) would have the same effect on the bounds as would an increase in the effective endowment. As discussed in Section 4.4.2, the estimated bounds remain informative even when there is an increase in labor supply due to the intervention, suggesting that increases in the effective endowment would not appreciably affect the results.

4.5 Discussion

The finding that LBD, at least in part, explains growth in teacher quality means that, on average, teachers improve by teaching their students, in the current context. Of course, teachers might also improve by making OJT investments, as the confidence sets for the OJT component include strictly positive values. While this might seem quite intuitive, economists to date had not been able to identify the force within human capital theory behind why teacher quality increases with experience.

Returning to the specific context underlying this paper's estimates, Glewwe et al. (2010) assessed that the intervention may have increased "teaching to the test", not more general student knowledge. They surmised this because, in contrast to the results for the incentivized exam (used in the current paper), student achievement for a non-incentivized exam that covered similar material did not significantly increase in the treatment group during active treatment.[44] That being said, it is important to note that the associated change in teacher human capital was not limited to a one-time change contained by the active-treatment period, as it did show up in the post-active-treatment period, meaning that the current paper does identify a force generating a form of teacher human capital. However, just as the above assessment of Glewwe et al. (2010) warrants a modicum of caution when interpreting the findings of that paper, it does so for those of this paper too.

Related, while I use the incentivized test to most clearly demonstrate the current methodology for separately identifying OJT and LBD, it could be the case that the increase in production, even if directed toward "incentivized output", also increased broader, socially desirable, measures of human capital (i.e., the LBD channel made teachers better in general, even if the "doing" was incentivized). Unfortunately, post-active-treatment outcomes on the low-stakes exam were not

amount of selection on unobserved variables.

[44] If there were persistent improvements in culture of learning or effort emanating from "warm glow", it would be natural to also expect these to also have manifested in the low-stakes test scores during active treatment. The lack of discernible effects along this margin suggests these alternative potential causes might not be the main drivers underlying the achievement effects.

available in the data (Glewwe et al., 2010, p. 211). That being said, the approach developed in this paper could be applied to a context in which researchers have collected follow-up data on exams capturing meaningful variation in student achievement.

5 Conclusion

I develop a framework nesting the OJT and LBD forces of human capital accumulation, and derive theoretical bounds for OJT and LBD components. The developed bounds are sharp, and yield novel information about the presence and relative importance of the forces generating human capital. I estimate the bounds using publicly available data from Glewwe et al. (2010), who designed and implemented an experiment featuring an output-based teacher incentive scheme in Kenya. The estimated bounds are informative, and allow us to reject the "pure OJT" model, meaning the data are consistent with the presence of an LBD component to teacher human capital accumulation. This suggests the dynamic multitasking problem inherent to the "pure OJT" model may be tempered by the presence of an LBD component to human capital accumulation, at least, in the context of the application considered in this paper.

Overall, this paper constitutes an important step towards thinking about the design of effective educational policy that targets teachers. Performance pay is increasingly common in education (Pham et al., 2021) as well as in the public sector more generally (Ahmad et al., 2024). It is important to note then that this paper's framework could also be applied to other contexts, in education and otherwise, in which there were outcome-based incentives and a follow-up measure of output. Results from such future research would be comparable those in the current paper because they would correspond to parameters in a structural production function, and by comparing the different estimates we would further enrich our understanding of the process of human capital accumulation. More generally, my framework suggests we should interpret the results of incentive pay programs with their potential dynamic implications on agents' human capital in mind. Because follow-up data on outcomes of interest are not always available for analysis, a simple and practical corollary of this paper then is that researchers should collect follow-up data on output-based interventions.[45]

In light of the well-known identification difficulties, it may be surprising that we can learn something new about human capital accumulation, even under the transparent and relatively simple approach taken here. A very promising, complementary, tack would be the structural econometric approach, which would require different assumptions (some stronger), but could then also answer other important questions about the importance of OJT and LBD forces in teachers' human capital accumulation. Such an approach would also be well-suited to rationalize the observed patterns in the data, and would also permit simulation of behavior and outcomes under counterfactual incentive

[45] For example, Glewwe et al. (2010) did not collect follow-up data on low-stakes exams. Another potential candidate dataset would be the publicly available data from Muralidharan and Sundararaman (2011), an experimental evaluation of teacher incentive pay in Andhra Pradesh, India, that found meaningful increases in student learning. Unfortunately, follow-up data were not available, precluding application of the approach developed here.

schemes. It could potentially yield other benefits, such as allowing for heterogeneity in teacher human capital accumulation trajectories and, thus, heterogeneity in the growth of teacher quality.

APPENDIX

A Relationship to a log-linear specification

This section illustrates one way in which the linear technologies (1)-(2) relate to nonlinear specifications. The illustration considers a representative teacher, teaching a representative student, in each of the control and treatment groups; therefore I suppress the teacher and student subscripts in this section. I maintain the assumption of balance of the experimental design.

Consider the following production function for teacher human capital, denoted here as κ_t:

$$\kappa_t = \kappa_{t-1}^{\gamma_\kappa}[\kappa_{t-1}\iota_{t-1}]^{\gamma_\iota}[\kappa_{t-1}\zeta_{t-1}]^{\gamma_\zeta}, \tag{A18}$$

where κ_{t-1} is the teacher's human capital last period (which may depreciate), ι_{t-1} is the share of the teacher's human capital last period spent on OJT investment, and ζ_{t-1} is the share of the teacher's human capital last period spent on production. The parameters of interest, respectively representing the OJT and LBD components of human capital accumulation in (A18), are $(\gamma_\iota, \gamma_\zeta) \in [0,1]^2$.[46]

Using $\tilde{}$ to denote the log of a variable, we can write the log-linearized version of the human capital production function, eq. (A18):

$$\tilde{\kappa}_t = [\gamma_\kappa + \gamma_\iota + \gamma_\zeta]\tilde{\kappa}_{t-1} + \gamma_\iota\tilde{\iota}_{t-1} + \gamma_\zeta\tilde{\zeta}_{t-1}. \tag{A19}$$

Let w_t measure the cognitive skill of the student in period t, which is produced according to[47]

$$w_t = \kappa_t^{\lambda_\kappa}\zeta_t^{\lambda_\zeta}w_{t-1}^{\lambda_w}, \tag{A20}$$

which, in logs, is

$$\tilde{w}_t = \lambda_\kappa\tilde{\kappa}_t + \lambda_\zeta\tilde{\zeta}_t + \lambda_w\tilde{w}_{t-1}. \tag{A21}$$

This equation is a log-linearized value added specification for cognitive achievement, where the value added to log achievement is $\lambda_\kappa\tilde{\kappa}_t + \lambda_\zeta\tilde{\zeta}_t$.

As before, the bounds on the parameters of interest will depend on the ATEs for achievement and labor supply. With a representative teacher and student, we have $\Delta\tilde{z}_t = \tilde{z}_t^T - \tilde{z}_t^C$ for $z = \kappa, \iota, \zeta, w$.

For the pre-treatment period, $t = 0$, the mean difference in achievement between the treatment

[46]Note that the unit interval is conservative, as it allows for parameter values that could yield explosive growth and would thus likely be ruled out if estimating this model. As is the case with any model, this specification is meant to serve as an approximation to reality. If taken literally (and augmented with additional assumptions to rationalize input choices) then a pure LBD specification would likely yield an optimal OJT investment of $\iota_{t-1} = 0$, which would imply $\kappa_t = 0$. Even here, however, even small positive values of γ_ι would avoid this problem.

[47]I have not included an error because there is a representative student in each group.

and control groups is

$$\Delta\tilde{w}_0 = \lambda_\kappa \underbrace{\Delta\tilde{\kappa}_0}_{=0} + \lambda_\zeta \underbrace{\Delta\tilde{\zeta}_0}_{=0} + \lambda_w \underbrace{\Delta\tilde{w}_{-1}}_{=0} = 0, \tag{A22}$$

i.e., a balanced experimental design implies there will be no average difference in the pre-treatment average scores, as was also the case in Section 3.2.

For the active-treatment period, $t = 1$, we have

$$\Delta\tilde{w}_1 = \lambda_\kappa \underbrace{\Delta\tilde{\kappa}_1}_{=0} + \lambda_\zeta \Delta\tilde{\zeta}_1 + \lambda_w \underbrace{\Delta\tilde{w}_0}_{=0} = \lambda_\zeta \Delta\tilde{\zeta}_1. \tag{A23}$$

Similar to eq. (6), eq. (A23) shows that the difference between treatment and control achievement in the active-treatment period can only come from the change in production, $\Delta\tilde{\zeta}_1$.

The mean difference in achievement between the treatment and control groups for the post-active-treatment period, $t = 2$, is

$$\begin{aligned} \Delta\tilde{w}_2 &= \lambda_\kappa \Delta\tilde{\kappa}_2 + \lambda_\zeta \Delta\tilde{\zeta}_2 + \lambda_w \Delta\tilde{w}_1 \\ &= \lambda_\kappa \gamma_\iota \Delta\tilde{\iota}_1 + \lambda_\kappa \gamma_\zeta \Delta\tilde{\zeta}_1 + \lambda_\zeta \Delta\tilde{\zeta}_2 + \lambda_w \Delta\tilde{w}_1, \end{aligned} \tag{A24}$$

which uses $\Delta\theta_1 = 0$ to go from the first to the second line. Assuming for simplicity that the intervention had no effect on labor supply, we have

$$\Delta\zeta_t + \Delta\iota_t = 0, \tag{A25}$$

i.e., the effect on OJT investment is opposite that on production shares. Substituting using eqs. (A23) and (A25) and maintaining the assumption that post-active treatment production shares will not be different between the control and treatment groups (i.e., $\Delta\tilde{\zeta}_2 = 0$), eq. (A24) becomes

$$\Delta\tilde{w}_2 = \left[\frac{\lambda_\kappa \gamma_\zeta - \lambda_\kappa \gamma_\iota + \lambda_\zeta \lambda_w}{\lambda_\zeta}\right] \underbrace{\Delta\tilde{w}_1.}_{\lambda_\zeta \Delta\tilde{\zeta}_1,\ \text{from eq. (A23)}} \tag{A26}$$

The last step is to obtain values for the relevant quantities in (A26), $\lambda_\kappa, \lambda_\zeta, \lambda_w, \Delta\tilde{w}_1, \Delta\tilde{w}_2$. Analogous to the baseline analysis, I set $\lambda_\kappa = 1$ and $\lambda_\zeta = 1$; both of these are effectively normalizations,[48] and these parameters having the same value is consistent with a teacher's value added being the share of her human capital allocated to production. Next consider the remaining parameters, $\lambda_w, \Delta\tilde{w}_1, \Delta\tilde{w}_2$. It is well known that test scores measuring, e.g., cognitive skill, have no inherent

[48] λ_κ fixes the scale of pre-treatment teacher human capital, which is unobserved and yields no direct testable implications; this normalization was also made in the model in the main text. Setting $\lambda_\zeta = 1$ may not strictly be a pure normalization, as a null ATE on achievement in the active-treatment period could stem from $\lambda_\zeta = 0$ and/or $\Delta\tilde{\zeta}_1 = 0$. However, both of these parameters must be nonzero to match the positive ATE in the data; given this, then, setting $\lambda_\zeta = 1$ is innocuous.

scale, meaning any monotonic (increasing) transformations (e.g., logarithms) are also valid measures (see, e.g., Cunha and Heckman, 2008). Given the earlier argument that qualitative differences in inputs drive the analysis, and the known sensitivity of achievement tests to monotonic transformations (Bond and Lang, 2013), it is reasonable to use the estimator of β_y as an approximation for λ_w. In a similar way, we can use Δy_t, which corresponds to the estimated difference in value added between the treatment and control groups, to measure $\Delta \tilde{w}_t$, the difference in achievement for the students respectively representing the treatment and control groups. Putting all of this together, the bounds obtained for (δ_I, δ_h) would also apply to $(\gamma_\iota, \gamma_\zeta)$.

B Estimation details

Our goal is to obtain estimates of bounds and confidence sets for the parameters δ_h and δ_I. Let the vector $\delta := [\delta_h, \delta_I]'$ collect the parameters of interest and let the vector $\gamma := [\Delta y_1, \Delta y_2, \beta_y, \Delta l_1]'$ collect the other parameters in eq. (8), which characterizes the values of δ consistent with γ. I will work in terms of the equivalent formulation in eq. (9), which expresses the permissible values of δ in terms of the composite parameters $\pi := [\pi_{\text{icept}}, \pi_{\text{slope}}]'$, which are known, continuous, functions of γ. For convenience, I reproduce eq. (9) here, explicitly denoting the dependence of the π parameters on γ:

$$\delta_h = \pi_{\text{icept}}(\gamma) + \pi_{\text{slope}}(\gamma) \cdot \delta_I.$$

Let $\{w_i\}_{i=1}^n$ denote the data used to estimate γ (e.g., student achievement in different years, the share of teachers present during an unannounced site visit, treatment status, etc.), where w_i denotes the ith row of data. Let us denote the true values of the parameters using δ^0 and γ^0, with D^0 denoting the identified set for δ^0.[49] Finally, let us use δ_k, where $k = h, I$, to denote the projection of δ onto the dimension k, and D_k to denote the projection of the identified set for δ onto the dimension k, which returns the (marginal) identified set for δ_k.

The estimator $\hat{\gamma}$ solves the sample analogue of the population moment condition

$$\mathrm{E}\left[v(\gamma^0, w_i)\right] = 0, \tag{B27}$$

where the vector-valued function v stacks the moment conditions based on the "normal equations" for the separate regressions used to estimate the elements of γ.[50] Note that γ is exactly (not over-) identified when using eq. (B27), so I estimate it via separate regressions.

Let m collect the relevant information about the model structure for identifying δ (meaning eq. (9)), and let $r(\delta) \geq 0$ collect the inequality constraints describing the bounds on the parameter

[49] As is standard, also assume D^0 is not empty.

[50] Strictly speaking, the estimators of the parameters in γ are a subset of estimators of the parameters identified by a larger, exactly identified, system (e.g., intercept of achievement equation, which does not enter the expression for the identified set for δ) and data $\{w_i\}_{i=1}^n$. The estimators of the four-parameter subvector γ can be obtained by pre-multiplying estimators from the larger system with a selection matrix. Note each of the elements of γ are identified by exactly one moment condition in v, so the parameters in this set are all exactly identified.

space for δ (i.e., $\delta \in [0,1]^2$), as well as the assumption about returns to scale (e.g., no additional restrictions on δ, as in Assumption 1, or $\delta_h + \delta_I \leq 1$, as in Assumption 2).[51] Note that m depends on both δ and γ, while r depends only on δ. Using our new notation, elements in the identified set for δ solve

$$D(\gamma; m, r) = \{\delta \in \mathbb{R}^2 : m(\delta, \gamma) = 0 \ \wedge \ r(\delta) \geq 0\}, \tag{B28}$$

where

$$\begin{aligned}
m(\delta, \gamma) &= [1, \ \delta'] \times \begin{bmatrix} \pi_{\text{icept}}(\gamma) \\ -1 \\ \pi_{\text{slope}}(\gamma) \end{bmatrix} \\
\pi_{\text{icept}}(\gamma) &= \frac{\Delta y_2}{\Delta y_1} - \beta_y \\
\pi_{\text{slope}}(\gamma) &= 1 + \frac{\Delta l_1}{\Delta y_1}.
\end{aligned}$$

Inspection of eq. (B28) reveals that $D(\gamma; m, r)$ will be a closed line segment in $[0,1]^2$, as it is the intersection of the line defined by $m(\delta, \gamma) = 0$ and the convex sets satisfying $r(\delta) \geq 0$. The identified set $D(\gamma; m, r)$ is characterized entirely by the value of the parameter γ and the assumptions embodied in m and r. Because the distribution of w_i has no bearing on the identified set, outside of its effect operating through γ, there is nothing to estimate if γ is known.

Figure B4 illustrates the relationship between $(\gamma, \pi, m(\cdot), r(\cdot))$ and the identified set $D(\gamma)$ under Assumption 1 (where I have suppressed the dependence of D and its derivative objects on (m, r) to reduce clutter). For example, changes in γ will shift and/or rotate the line defined by $m(\delta, \gamma) = 0$ in a continuous manner, via $\pi(\gamma)$.

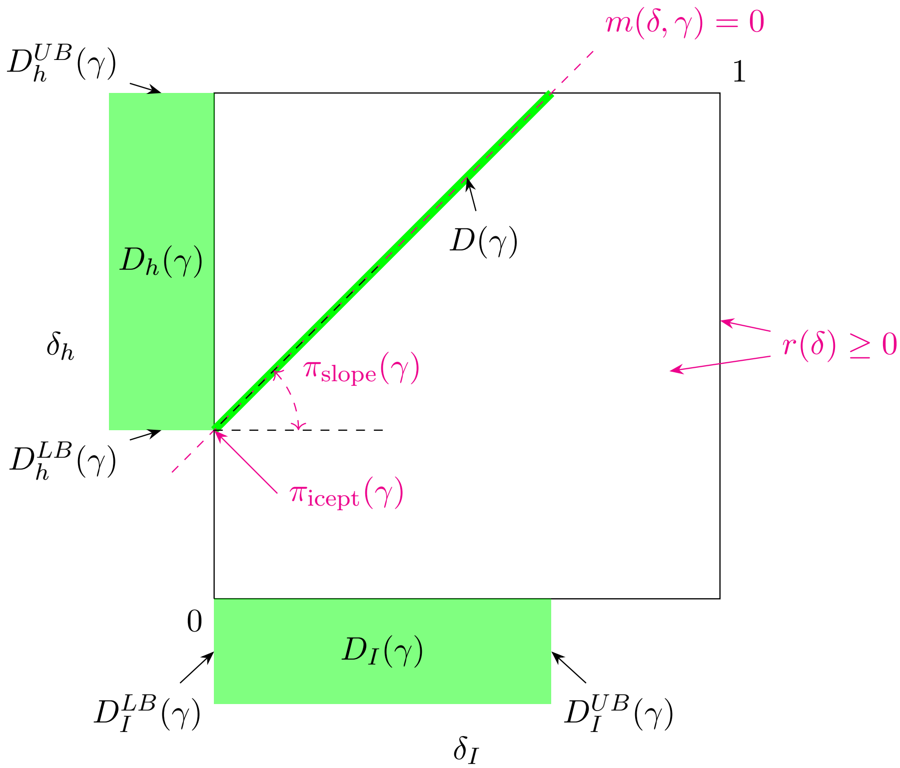
Standard arguments establish that $\hat{\gamma}$ will be asymptotically normally distributed around γ^0. Because the π are continuous and differentiable functions of γ, the delta method can then describe the asymptotic distribution of $\hat{\pi}$. Given an estimate of γ, the estimate for the identified set, $\hat{D} := D(\hat{\gamma}; m, r)$, collects the values of δ satisfying eq. (B28), using $\hat{\gamma}$ in place of γ. Estimates of the marginal identified sets, $\hat{D}_k$, can then be obtained by projecting $\hat{D}$ onto the axis for parameter k.

B.1 Implementation

Now I describe how I construct the confidence sets for the δ_k, given an assumption about the returns to scale, $r(\delta)$. Let $\hat{\pi}^s$ denote a draw from the bootstrap distribution of $\hat{\pi}$ (where I stratify

[51] Specifically, when there is no assumption about returns to scale (Assumption 1), $r(\delta) = r_1(\delta)$, which contains only the bounds on δ, i.e., $r_1(\delta)' = [\delta_h, 1-\delta_h, \delta_I, 1-\delta_I]$. Under nonincreasing returns (Assumption 2), $r(\delta)' = [r_1(\delta)', r_2(\delta)]$, where $r_2(\delta) = 1 - \delta_h - \delta_I$. The equality constraint under constant returns to scale (Assumption 3) can be written using two inequality constraints, only one of which is not redundant in light of Assumption 2. In this case then, we have $r(\delta)' = [r_1(\delta)', r_2(\delta), r_3(\delta)]$, where $r_3(\delta) = \delta_h + \delta_I - 1$.

Figure B4: Identified set $D(\gamma)$ under Assumption 1



Notes: Figure depicts the identified set for δ, $D(\gamma)$, and projections of the identified set onto the axes, $D_k(\gamma)$, suppressing the dependence on (m, r) to reduce clutter. The identified set is defined in eq. (B28), and takes as arguments the locus defined by $m(\delta, \gamma)$ and the restrictions encoded in $r(\delta) \geq 0$ when $r(\delta) = r_1(\delta)$ (i.e., Assumption 1).

by treatment status). For each $\hat{\pi}^s$, I perform a grid search over δ_k, retaining those values satisfying eq. (B28) as the estimate $\hat{D}_k^s$. Because D (hence, D's projection onto either dimension) is convex, I only need to perform a grid search for the endpoints in each dimension (e.g., $\hat{D}_h^{LB}, \hat{D}_h^{UB}$).[52] I then generate the interval of the desired size (i.e., 95% confidence level) by collecting all the values of δ_k that are in at least 95% of the $\hat{D}_k^s$ (Canay et al., 2024).[53]

[52]In practice, I use the explicit expressions for the marginal identified sets, derived in Section 3.5.

[53]See Andrews (2000) for a discussion of using the bootstrap when the parameter is on the boundary of the parameter space. I used `R` 4.2.2 (R Core Team, 2022) for all the analysis and output generation. I used the `stargazer` package (Hlavac, 2022) for regression tables and the `tikzDevice` package (Sharpsteen and Bracken, 2023) for the some of the figures.

B.2 Relationship to other estimation approaches

Method of moments My estimation procedure is related to the following method of moments problem:

$$(\hat{\delta}_{MM}, \hat{\gamma}_{MM}) \in \underset{\Omega(\delta,\gamma)}{\arg\min} \quad \left[\frac{1}{n}\sum_{i=1}^{n} v(\gamma, w_i)\right]' \times \left[\frac{1}{n}\sum_{i=1}^{n} v(\gamma, w_i)\right],$$

where the parameter space $\Omega(\delta, \gamma)$ encodes the constraints $m(\delta, \gamma) = 0$ and $r(\delta) \geq 0$ (I use the identity matrix as the (diagonal) weighting matrix, so I suppress it in the expression). The estimate $\hat{\gamma}_{MM}$ from the above will be unique with probability approaching 1, and the values of $\hat{\delta}_{MM}$ that minimize the above are collected in $\hat{D}_{MM}$. The parameter space Ω may seem a bit nonstandard due to the interdependence generated via $m(\delta, \gamma) = 0$, but it will be convex as it is the intersection of the line characterized by $m(\delta, \gamma) = 0$, which is convex, and the set satisfying $r(\delta) \geq 0$, which is convex under each assumption about the returns to scale. Note that the above could also be written in terms of a moment-inequality model (Tamer, 2010; Canay et al., 2024).

The estimates from my approach and the method-of-moments one above will differ when the value of $\hat{\gamma}$ returns an empty identified set (i.e., $\hat{D} = \emptyset$). Using my approach, an empty $\hat{D}$ would be evidence against the model. In contrast, when using the method-of-moments-based estimates an empty $\hat{D}$ would change the estimate $\hat{\gamma}$ to satisfy the constraints in $\Omega(\delta, \gamma)$. Under the standard assumption that D^0 is not empty, the estimates from my approach and the method-of-moments-based approach will be similar, asymptotically.

Indirect inference My approach does not involve indirect inference, as I do not find values of δ to match the parameters of an auxiliary model. Rather, I have a unique, closed-form, solution for the estimator of γ, and then eq. (9) provides an analytical characterization of the values of δ consistent with the estimate of γ. The identified set for δ is the subset of these values satisfying the relevant assumption about the returns to scale, $r(\delta) \geq 0$, and will be a closed line segment in $[0, 1]^2$. This could in turn be used to obtain the marginal identified sets for the parameters δ_h and δ_I via projection onto the relevant axis. Given the exact (point) identification of γ and the analytical characterizations of the identified sets for δ and its components, little would likely be gained from using indirect inference (see, e.g., the discussion in Jiang and Turnbull, 2004, p. 250).

C Robustness

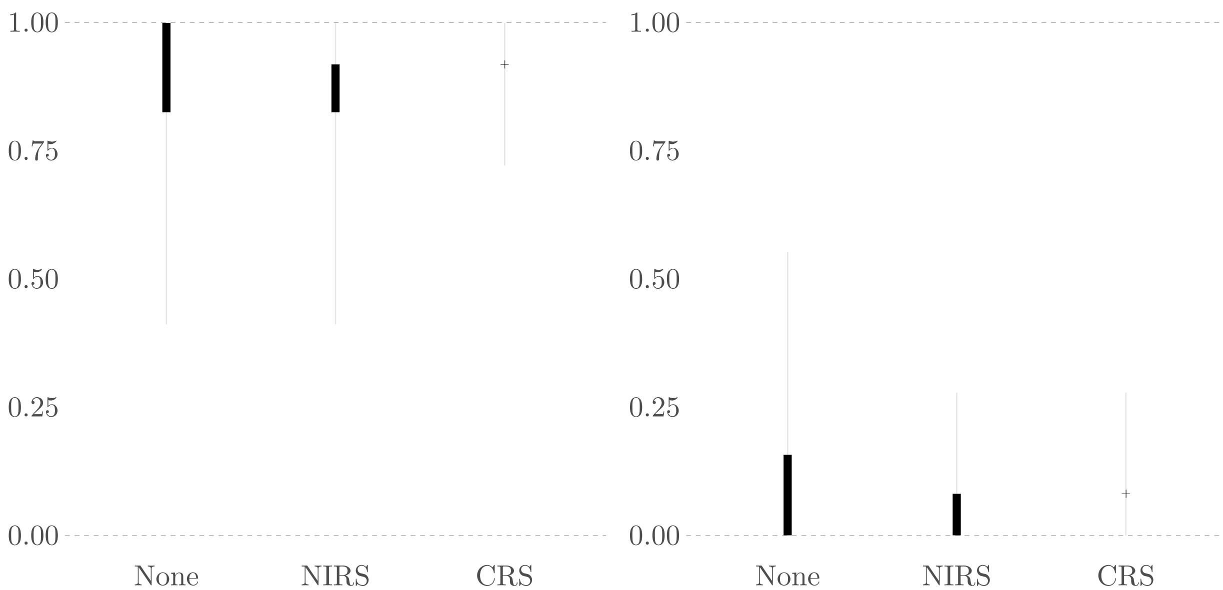
C.1 Persistence parameter

Figure C5 and the associated table present estimated bounds and confidence sets when using a persistence value of $\beta_y = 0.20$, the estimated persistence of teacher-induced reading gains from Jacob et al. (2010), Table 2. The results in Figure C5 show the bounds are all weakly tighter than under the baseline results, which used the estimate of β_y, and all bounds for intervals are strictly smaller (i.e., have lower width).

By increasing the lower bound for δ_h and decreasing the upper bound for δ_I, these results reinforce the finding that LBD is present in the human capital technology, while the role of OJT is further limited. In fact, the 95% confidence set for δ_h lies strictly above the 95% confidence set for δ_I under non-increasing returns to scale, while this was only true under the strongest assumption (constant returns to scale) for the baseline results.

Figure C5: Robustness: Estimated bounds and confidence sets when $\beta_y = 0.20$

(a) δ_h: LBD (b) δ_I: OJT



Notes: The left panel depicts estimated bounds ( ) and 95% confidence sets ( ) for δ_h, under Assumptions 1, 2, and 3 (denoted on the bottom axis via "None", "NIRS", and "CRS", respectively), as well as the additional assumption that $\beta_y = 0.20$. The right panel depicts the analogous results for δ_I.

Parameter	Assumption about $\delta_h + \delta_I$	Estimated bounds Min.	Max.	Width	95% confidence set Min.	Max.	Width
δ_h	1: None	0.825	0.999	0.174	0.412	1	0.588
	2: NIRS	0.825	0.919	0.094	0.412	1	0.588
	3: CRS	0.919	0.919	0	0.722	1	0.278
δ_I	1: None	0	0.157	0.157	0	0.553	0.553
	2: NIRS	0	0.081	0.081	0	0.278	0.278
	3: CRS	0.081	0.081	0	0	0.278	0.278

Notes: Assumption 1 corresponds to no assumption about the returns to scale for $\delta_h + \delta_I$, and Assumptions 2 and 3 respectively correspond to non-increasing and constant returns to scale. These results have been calculated under the additional assumption that $\beta_y = 0.20$.

C.2 No labor supply response

Figure C6 and the associated table present estimated bounds and confidence sets under the assumption of no labor supply response (meaning $\Delta l_1 = 0$). Relative to the measure used in computing the baseline results, assuming a null effect would correspond to a (slight, insignificant) increase in the labor supply response. This would tend to rotate the (δ_h, δ_I) loci clockwise from their baseline values, lowering the upper bound for δ_h and increasing the upper bound for δ_I (as shown by the point estimates) in the table. That being said, neither effect is very large, and the estimated bounds and confidence sets for δ_h and δ_I would remain informative.

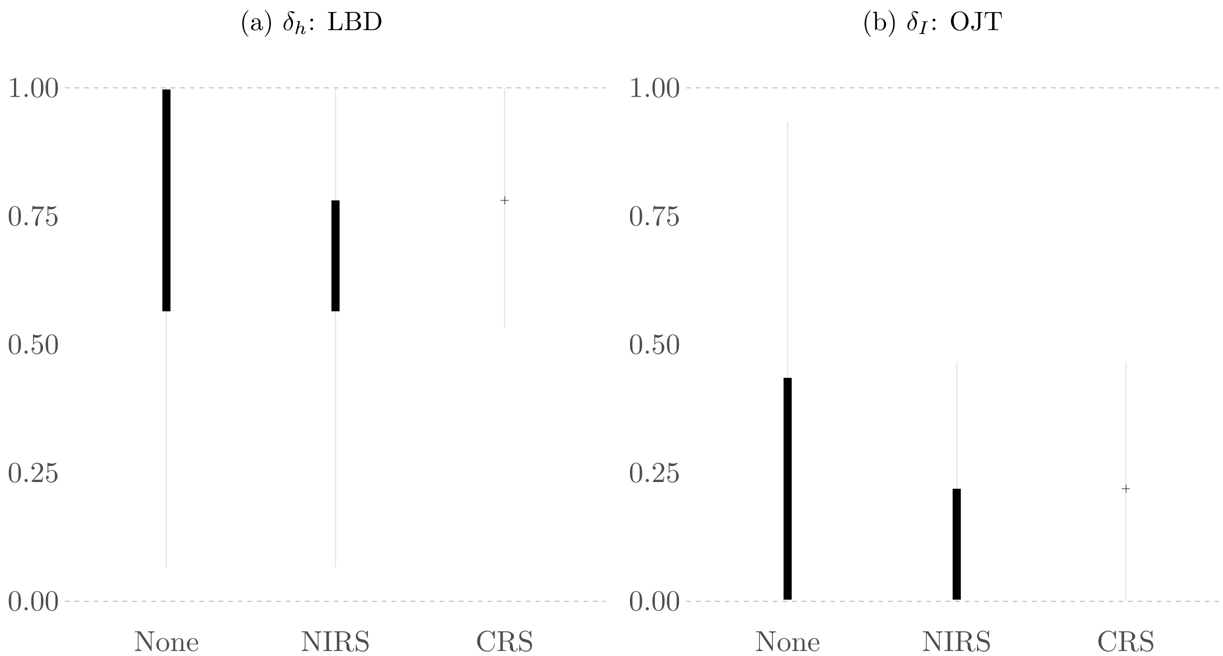
C.3 Labor supply is measured by teachers' presence in classrooms

Figure C7 and the associated table present estimated bounds and confidence sets when labor supply is measured using the share of site visits during which teachers were present in their *classrooms* (the baseline results used the share of such visits during which teachers were present at their *schools*). Table C5 presents treatment effects on both pooled attendance measures. The outcome in column (1) is the average share of teachers present at the school during the site visits (the labor supply measure used in the baseline results) while the outcome in column (2) is the average share of teachers present in the classroom during the site visits. As shown in column (2), the point estimate of the labor supply response $-\Delta l_1$ using this measure is slightly positive (though, as with the measure used for the baseline results, not distinguishable from zero at any conventional significance level). As was the case when there was no labor supply response (presented in Appendix C.2), this would tend to rotate the (δ_h, δ_I) loci clockwise from their baseline values, lowering the upper bound for δ_h and increasing the upper bound for δ_I (as shown by the point estimates) in the table. That being said, as was also the case when there was no labor supply response, neither effect is very large, and the estimated bounds and confidence sets for δ_h and δ_I would continue to remain informative.

C.4 Labor supply response is based on an externally estimated labor supply elasticity from Attanasio et al. (2018)

External estimates of labor supply elasticities can also help provide guidance about the size of potential labor supply responses. We may be more confident of results if, in addition to being robust to different measures of labor supply (e.g., measuring labor supply using teachers' classroom presence, as I did in Appendix C.3) they are also robust to the use of estimates identified using entirely different variation. Additionally, as discussed in Section 2, while my main labor supply measure may do a good job of capturing extensive-margin variation (in my context, teachers' being present or absent at school on any given day), it could, in principle, miss responses along the intensive margin. External estimates can then be especially useful for assessing the robustness of my results if they are informative about total (i.e., inclusive of the intensive margin) labor supply responses.

Figure C6: Robustness: Estimated bounds and confidence sets when there is no labor supply response



Notes: The left panel depicts estimated bounds (▬) and 95% confidence sets (—) for δ_h, under Assumptions 1, 2, and 3 (denoted on the bottom axis via "None", "NIRS", and "CRS", respectively), as well as the additional assumption that there is no labor supply response ($\Delta l_1 = 0$). The right panel depicts the analogous results for δ_I.

Parameter	Assumption about $\delta_h + \delta_I$	Estimated bounds Min.	Estimated bounds Max.	Estimated bounds Width	95% confidence set Min.	95% confidence set Max.	95% confidence set Width
δ_h	1: None	0.565	0.997	0.432	0.067	1	0.933
	2: NIRS	0.565	0.781	0.216	0.067	1	0.933
	3: CRS	0.781	0.781	0	0.533	1	0.467
δ_I	1: None	0.003	0.435	0.432	0	0.933	0.933
	2: NIRS	0.003	0.219	0.216	0	0.467	0.467
	3: CRS	0.219	0.219	0	0	0.467	0.467

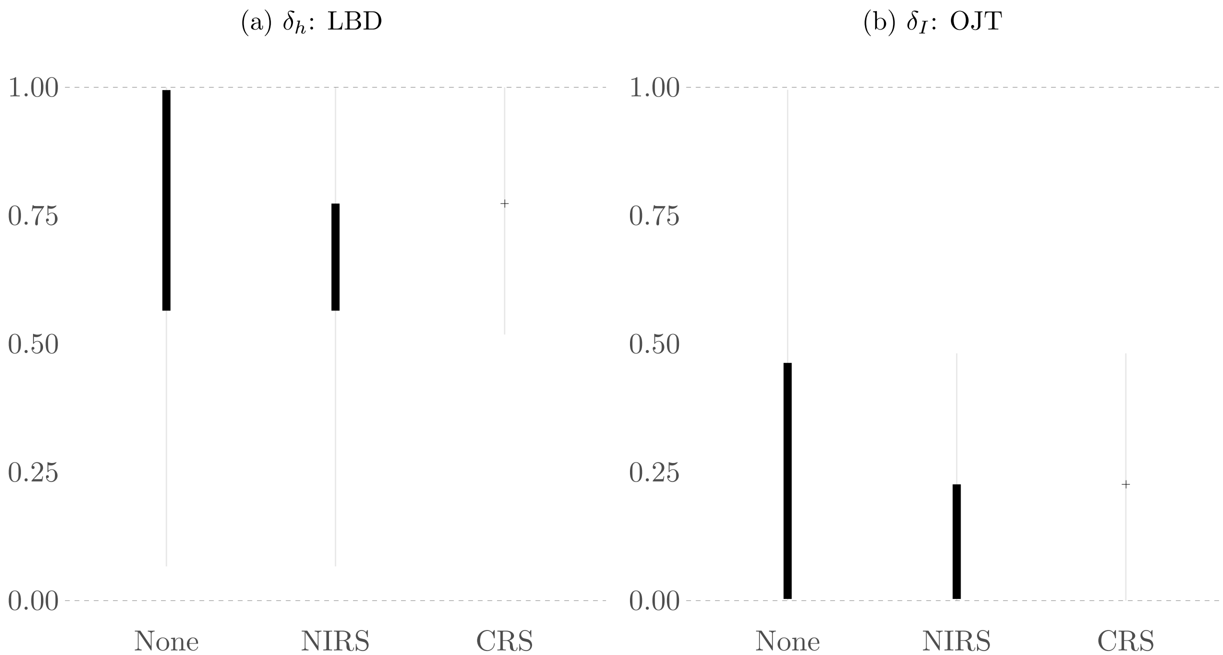
Notes: Assumption 1 corresponds to no assumption about the returns to scale for $\delta_h + \delta_I$, and Assumptions 2 and 3 respectively correspond to non-increasing and constant returns to scale. These results have been calculated under the additional assumption that there is no labor supply response ($\Delta l_1 = 0$).

Table C5: Estimate of ATE on teacher labor supply, measured by present at school vs. present in the classroom

	Dependent variable:	
	Present at school in period t	Present in classroom in period t
	(1)	(2)
Active-treatment ($t = 1$)	0.012	−0.032
	(0.025)	(0.050)
Post-active-treatment ($t = 2$)	0.052	−0.187
	(0.025)	(0.050)
Active ∗ treated ($-\Delta l_1$)	−0.017	0.006
	(0.029)	(0.058)
Post ∗ treated	−0.011	0.042
	(0.041)	(0.082)
Constant	0.833	0.699
	(0.014)	(0.030)
Observations	202	196

Note: Dependent variable in column (1) is school-level average share of site visits during which teachers were in attendance at the school, and dependent variable in column (2) is school-level average share of site visits during which teachers were present in the classroom. The t corresponds to a pooled period.

Figure C7: Robustness: Estimated bounds and confidence sets when labor supply is measured by teachers' presence in classrooms



Notes: The left panel depicts estimated bounds (▬) and 95% confidence sets (—) for δ_h, under Assumptions 1, 2, and 3 (denoted on the bottom axis via "None", "NIRS", and "CRS", respectively), using the fraction of site visits during which teachers were present in their classrooms as the measure of labor supply. The right panel depicts the analogous results for δ_I.

	Assumption	Estimated bounds			95% confidence set		
Parameter	about $\delta_h + \delta_I$	Min.	Max.	Width	Min.	Max.	Width
δ_h	1: None	0.565	0.995	0.430	0.067	1	0.933
	2: NIRS	0.565	0.773	0.208	0.067	1	0.933
	3: CRS	0.773	0.773	0	0.518	1	0.482
δ_I	1: None	0.003	0.463	0.460	0	0.995	0.995
	2: NIRS	0.003	0.227	0.224	0	0.482	0.482
	3: CRS	0.227	0.227	0	0	0.482	0.482

Notes: Assumption 1 corresponds to no assumption about the returns to scale for $\delta_h + \delta_I$, and Assumptions 2 and 3 respectively correspond to non-increasing and constant returns to scale. These results have been calculated using the fraction of site visits during which teachers were present in their classrooms as the measure of labor supply.

In a recent paper seeking to understand the literature's varied estimates of labor supply elasticities, Attanasio et al. (2018) use data from the US Consumer Expenditure Survey to estimate a flexible model of female labor supply, and then use their estimates to compute heterogeneous labor supply elasticities along both the intensive and extensive margins. In their context, the intensive margin captures typical hours worked per week and the extensive margin captures whether a woman worked at all in a given quarter. I focus on Attanasio et al. (2018)'s estimates of intensive-margin elasticities because, while their extensive-margin responses may be important for understanding female labor supply (especially in their general worker context), the quarter-based extensive-margin responses seem less relevant for my application (which considers full-time teachers). That is, their intensive-margin responses capture the desired responses in total labor supply when there is negligible (quarterly) extensive-margin variation, as should be the case in my context.[54]

Attanasio et al. report an estimated median intensive-margin Marshallian female labor supply elasticity of 0.18 (Attanasio et al., 2018, Table VII). At the sample mean value of labor supply of 0.833 (see Table 4), this elasticity would imply that a 3% increase in the wage would increase labor supply by 0.0045.[55] Accordingly, Figure C8 and the associated table present estimated bounds and confidence sets under the assumption $\Delta l_1 = -0.0045$ (i.e., leisure decreased, so labor supply increased, by 0.0045). As was the case when measuring labor supply using teachers' classroom presence, using $\Delta l_1 = -0.0045$ tends to rotate the (δ_h, δ_I) loci clockwise from their baseline values, lowering the upper bound for δ_h and increasing the upper bound for δ_I (as shown by the point estimates) in the table. The estimates are very similar to those computed when measuring labor supply using teachers' classroom presence, and the estimated bounds and confidence sets for δ_h and δ_I continue to remain informative. While the concordance of the results may not be surprising given the similarity of the change in labor supply across the two scenarios, the robustness of the findings with respect to very different types of alternative assumptions about the labor supply response should increase confidence in this paper's findings.

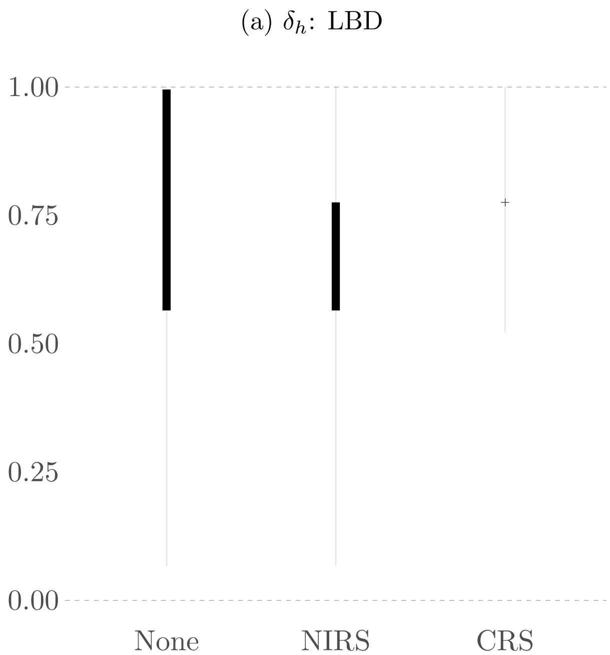
C.5 Allowing for post-active-treatment effects on production

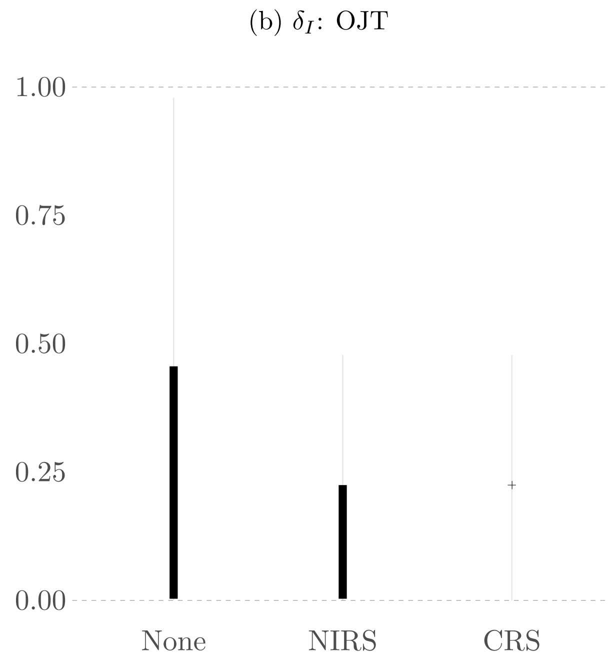
Here I explore how relaxing the assumption of no post-active-treatment effects on production (i.e., $\Delta h_2 = 0$) could potentially affect the results. To do this, I start from the end of eq. (7), but replace $\Delta h_2 = 0$ in Assumption 1(i) with $\Delta h_2 = \rho \Delta h_1$, where ρ is a parameter governing the magnitude of the affect on production in the post-active-treatment period. Under the natural assumption that Δh_2 is smaller in absolute magnitude than Δh_1, we would have $\rho \in [\underline{\rho}, \bar{\rho}]$, where $-1 < \underline{\rho} < 0 < \bar{\rho} < 1$. Working out the locus of permissible values of (δ_h, δ_I) under this alternative

[54] As I noted in Section 2, absences tend to be widely distributed among teachers.

[55] As discussed in Section 3.2, the prize value ranged from less than 2% to 4% of the typical teacher's salary. I use the intermediate value of 3% for simplicity.

Figure C8: Robustness: Estimated bounds and confidence sets when labor supply response is based on the estimated median Marshallian elasticity reported in Attanasio et al. (2018), Table VII





Notes: The left panel depicts estimated bounds (▬) and 95% confidence sets (—) for δ_h, under Assumptions 1, 2, and 3 (denoted on the bottom axis via "None", "NIRS", and "CRS", respectively), as well as the additional assumption that the labor supply response equals the increase in labor supply from a 3% increase in wages, evaluated at the estimated median Marshallian elasticity reported in Attanasio et al. (2018), Table VII. The right panel depicts the analogous results for δ_I.

Parameter	Assumption about $\delta_h + \delta_I$	Estimated bounds Min.	Estimated bounds Max.	Estimated bounds Width	95% confidence set Min.	95% confidence set Max.	95% confidence set Width
δ_h	1: None	0.565	0.995	0.430	0.067	1	0.933
	2: NIRS	0.565	0.775	0.210	0.067	1	0.933
	3: CRS	0.775	0.775	0	0.522	1	0.478
δ_I	1: None	0.003	0.456	0.453	0	0.979	0.979
	2: NIRS	0.003	0.225	0.222	0	0.478	0.478
	3: CRS	0.225	0.225	0	0	0.478	0.478

Notes: Assumption 1 corresponds to no assumption about the returns to scale for $\delta_h + \delta_I$, and Assumptions 2 and 3 respectively correspond to non-increasing and constant returns to scale. These results have been calculated under the additional assumption that the labor supply response equals the increase in labor supply from a 3% increase in wages, evaluated at the estimated median Marshallian elasticity reported in Attanasio et al. (2018), Table VII.

assumption, we have

$$
\begin{aligned}
\Delta y_2 &= -\delta_I \Delta l_1 + \left[\delta_h - \delta_I + \beta_h \beta_y\right] \beta_h \Delta h_1 + \overbrace{\rho \Delta h_1}^{\Delta h_2}, \\
&= -\delta_I \Delta l_1 + \left[\delta_h - \delta_I + \beta_y\right] \Delta y_1 + \rho \Delta y_1, \\
&\Rightarrow \\
\delta_h &= \left[\frac{\Delta y_2}{\Delta y_1} - [\beta_y + \rho]\right] + \delta_I \left[1 + \frac{\Delta l_1}{\Delta y_1}\right],
\end{aligned}
$$

where the second line obtains from eq. (6) and the maintained scale assumption $\beta_h = 1$. Having $\rho < 0$ would tighten the bounds, while positive values of ρ have the same effect as increases in β_y: they lower the intercept of the locus defined in eq. (8). Because the baseline estimate of β_y is arguably too high (see the discussion in Section 4.1, where I discuss how a value a bit less than half of the baseline estimate might be more in line with research on fade-out of intervention-based achievement changes), the baseline results could be viewed as already having accounted for a range of reasonable positive values of ρ.

References

Abowd, J. M., F. Kramarz and D. N. Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, 67(2):251–333, 1999.

Agostinelli, F. and M. Wiswall, "Estimating the Technology of Children's Skill Formation," *Journal of Political Economy*, 2023.

Ahmad, F., J. Lyu, E. Stazyk and M. Higgs, "Back to the Future Revisited: A Systematic Literature Review of Performance-Related Pay in the Public Sector," *International Public Management Journal*, pp. 1–25, 2024.

Altonji, J. G., T. E. Elder and C. R. Taber, "Selection On Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 113(1):151–184, 2005.

Andrabi, T., J. Das, A. I. Khwaja and T. Zajonc, "Do Value-Added Estimates Add Value? Accounting for Learning Dynamics," *American Economic Journal: Applied Economics*, 3(3):29–54, 2011.

Andrews, D. W., "Inconsistency of the Bootstrap When a Parameter Is on the Boundary of the Parameter Space," *Econometrica*, pp. 399–405, 2000.

Attanasio, O., P. Levell, H. Low and V. Sánchez-Marcos, "Aggregating Elasticities: Intensive and Extensive Margins of Women's Labor Supply," *Econometrica*, 86(6):2049–2082, 2018.

Becker, G. S., *Human Capital*, Columbia University Press, New York, 1964.

Behrman, J. R., M. M. Tincani, P. E. Todd and K. I. Wolpin, "Teacher Quality in Public and Private Schools Under a Voucher System: The Case of Chile," *Journal of Labor Economics*, 34(2):319–362, 2016.

Belley, P., "Understanding Wage Growth: Estimating and Testing Learning-by-Doing," *Journal of Human Capital*, 11(2):213–246, 2017.

Ben-Porath, Y., "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 75(4):352–365, 1967.

Biasi, B., "The Labor Market for Teachers under Different Pay Schemes," *American Economic Journal: Economic Policy*, 13(3):63–102, 2021.

Biasi, B., C. Fu and J. Stromme, "Equilibrium in the Market for Public School Teachers: District Wage Strategies and Teacher Comparative Advantage," Tech. rep., National Bureau of Economic Research, 2021.

Blandin, A., "Learning by Doing and Ben-Porath: Life-cycle Predictions and Policy Implications," *Journal of Economic Dynamics and Control*, 90:220–235, 2018.

Blinder, A. S. and Y. Weiss, "Human Capital and Labor Supply: A Synthesis," *Journal of Political Economy*, 84(3):449–472, 1976.

Bobba, M., T. Ederer, G. Leon-Ciliotta, C. Neilson and M. G. Nieddu, "Teacher Compensation and Structural Inequality: Evidence from Centralized Teacher School Choice in Perú," 2021.

Bond, T. N. and K. Lang, “The Evolution of the Black-White Test Score Gap in Grades K–3: The Fragility of Results,” *Review of Economics and Statistics*, 95(5):1468–1479, 2013.

Brown, C., “A Model of Optimal Human-Capital Accumulation and the Wages of Young High School Graduates,” *Journal of Political Economy*, 84(2):299–316, 1976.

Burgess, S. M., “Human Capital and Education: The State of the Art in the Economics of Education,” *Available at SSRN 2769193*, 2016.

Canay, I., G. Illanes and A. Velez, “A User’s Guide for Inference in Models Defined by Moment Inequalities,” *Journal of Econometrics*, 2024.

Chaudhury, N., J. Hammer, M. Kremer, K. Muralidharan and F. H. Rogers, “Missing in Action: Teacher and Health Worker Absence in Developing Countries,” *Journal of Economic Perspectives*, 20(1):91–116, 2006.

Chetty, R., J. N. Friedman and J. E. Rockoff, “Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates,” *American Economic Review*, 104(9):2593–2632, 2014.

Cook, J. B. and R. K. Mansfield, “Task-specific Experience and Task-specific Talent: Decomposing the Productivity of High School Teachers,” *Journal of Public Economics*, 140:51–72, 2016.

Cunha, F. and J. J. Heckman, “Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation,” *Journal of Human Resources*, 43(4):738–782, 2008.

Cunha, F., J. J. Heckman and S. M. Schennach, “Estimating the Technology of Cognitive and Noncognitive Skill Formation,” *Econometrica*, 78(3):883–931, 2010.

Deaton, A., “Instruments, Randomization, and Learning about Development,” *Journal of Economic Literature*, 48(2):424–455, 2010.

Deaton, A. and N. Cartwright, “Understanding and Misunderstanding Randomized Controlled Trials,” *Social Science & Medicine*, 210:2–21, 2018.

Del Bono, E., J. Kinsler and R. Pavan, “Identification of Dynamic Latent Factor Models of Skill Formation with Translog Production,” *Journal of Applied Econometrics*, 37(6):1256–1265, 2022.

Dinerstein, M. and I. M. Opper, “Screening with Multitasking: Theory and Empirical Evidence from Teacher Tenure Reform,” Tech. rep., National Bureau of Economic Research, 2022.

Dolton, P. and W. v. d. Klaauw, “The Turnover of Teachers: A Competing Risks Explanation,” *Review of Economics and Statistics*, 81(3):543–550, 1999.

Duflo, E., R. Hanna and S. P. Ryan, “Incentives Work: Getting Teachers to Come to School,” *American Economic Review*, 102(4):1241–1278, 2012.

Freyberger, J., “Normalizations and Misspecification in Skill Formation Models,” *arXiv preprint arXiv:2104.00473*, 2021.

Fu, C., H.-C. L. Lin and A. Tanaka, “College Loans and Post-Schooling Skill Accumulation,” *Working Paper*, 2021.

Glewwe, P., N. Ilias and M. Kremer, “Teacher Incentives,” *NBER Working Paper No. 9671*, 2003.

—, "Teacher Incentives," *American Economic Journal: Applied Economics*, 2(3):205–227, 2010.

Grieco, P. L. and R. C. McDevitt, "Productivity and Quality in Health Care: Evidence from the Dialysis Industry," *Review of Economic Studies*, 84(3):1071–1105, 2017.

Griliches, Z. and J. Mairesse, "Production Functions: The Search for Identification," 1995.

Haggag, K., B. McManus and G. Paci, "Learning by Driving: Productivity Improvements by New York City Taxi Drivers," *American Economic Journal: Applied Economics*, 9(1):70–95, 2017.

Haley, W. J., "Estimation of the Earnings Profile from Optimal Human Capital Accumulation," *Econometrica*, pp. 1223–1238, 1976.

Hansen, G. D. and S. İmrohoroğlu, "Business cycle fluctuations and the life cycle: How important is on-the-job skill accumulation?" *Journal of Economic Theory*, 144(6):2293–2309, 2009.

Hanushek, E. A., "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data," *American Economic Review*, 61(2):280–288, 1971.

—, "Conceptual and Empirical Issues in the Estimation of Educational Production Functions," *Journal of Human Resources*, pp. 351–388, 1979.

—, "The Failure of Input-Based Schooling Policies," *Economic Journal*, 113(485):F64–F98, 2003, ISSN 1468-0297.

—, "The Economic Value of Higher Teacher Quality," *Economics of Education Review*, 30(3):466–479, 2011.

Hanushek, E. A. and M. E. Raymond, "Does School Accountability Lead to Improved Student Performance?" *Journal of Policy Analysis and Management*, 24(2):297–327, 2005.

Hanushek, E. A. and S. G. Rivkin, "Teacher Quality," *Handbook of the Economics of Education*, 2:1051–1078, 2006.

—, "The Distribution of Teacher Quality and Implications for Policy," *Annual Review of Economics*, 4(1):131–157, 2012.

Harris, D. N. and T. R. Sass, "Teacher Training, Teacher Quality and Student Achievement," *Journal of Public Economics*, 95(7):798–812, 2011.

Heckman, J. J., "A Life-Cycle Model of Earnings, Learning, and Consumption," *Journal of Political Economy*, 84(4):S11–44, 1976.

Heckman, J. J., L. Lochner and R. Cossa, "Learning-by-Doing vs. On-the-Job Training: Using Variation Induced by the EITC to Distinguish Between Models of Skill Formation," *NBER Working Paper Series*, p. 9083, 2002.

Heckman, J. J., L. Lochner and C. Taber, "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents," *Review of Economic Dynamics*, 1(1):1–58, 1998.

Hlavac, M., *stargazer: Well-Formatted Regression and Summary Statistics Tables*, Social Policy Institute, Bratislava, Slovakia, 2022, r package version 5.2.3.

Hölmstrom, B. and P. Milgrom, "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization*, 7:24–52, 1991.

Horowitz, J. L. and C. F. Manski, "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association*, 95(449):77–84, 2000.

Huggett, M., G. Ventura and A. Yaron, "Sources of Lifetime Inequality," *American Economic Review*, 101(7):2923–2954, 2011.

Imbens, G. W. and C. F. Manski, "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72(6):1845–1857, 2004.

Imberman, S. A. and M. F. Lovenheim, "Incentive Strength and Teacher Productivity: Evidence from a Group-Based Teacher Incentive Pay System," *Review of Economics and Statistics*, 97(2):364–386, 2015.

Jackson, C. K., J. E. Rockoff and D. O. Staiger, "Teacher Effects and Teacher-related Policies," *Annual Review of Economics*, 6(1):801–825, 2014.

Jacob, B. A., L. Lefgren and D. P. Sims, "The Persistence of Teacher-Induced Learning," *Journal of Human Resources*, 45(4):915–943, 2010.

Jiang, W. and B. Turnbull, "The Indirect Method: Inference Based on Intermediate Statistics—A Synthesis and Examples," *Statistical Science*, 19(2):239–263, 2004, ISSN 0883-4237, doi: 10.1214/088342304000000152.

Kaido, H., F. Molinari and J. Stoye, "Confidence Intervals for Projections of Partially Identified Parameters," *Econometrica*, 87(4):1397–1432, 2019.

Kane, T. J., D. F. McCaffrey, T. Miller and D. O. Staiger, "Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment," *Research Paper. MET Project. Bill & Melinda Gates Foundation*, 2013.

Killingsworth, M. R., "'Learning by Doing' and 'Investment in Training': A Synthesis of Two 'Rival' Models of the Life Cycle," *Review of Economic Studies*, 49(2):263–271, 1982.

Kinsler, J., "Assessing Rothstein's Critique of Teacher Value-Added Models," *Quantitative Economics*, 3(2):333–362, 2012.

Kraft, M. A. and J. P. Papay, "Can Professional Environments in Schools Promote Teacher Development? Explaining Heterogeneity in Returns to Teaching Experience," *Educational Evaluation and Policy Analysis*, 36(4):476–500, 2014.

Kuruscu, B., "Training and Lifetime Income," *American Economic Review*, 96(3):832–846, 2006.

Macartney, H., R. McMillan and U. Petronijevic, "Teacher Performance and Accountability Incentives," *National Bureau of Economic Research Cambridge, MA*, 2018.

Marschak, J., "Economic Measurements for Policy and Prediction," pp. 293–322, 1953.

McCaffrey, D. F., J. Lockwood, D. M. Koretz and L. S. Hamilton, *Evaluating Value-Added Models for Teacher Accountability. Monograph.*, ERIC, 2003.

McWilliams, J., “The Lucrative Art of Chicken Sexing,” *Pacific Standard*, 2018.

Muralidharan, K. and V. Sundararaman, “Teacher Performance Pay: Experimental Evidence from India,” *Journal of Political Economy*, 119(1):39–77, 2011.

Murnane, R. J., *The Impact of School Resources on the Learning of Inner City Children*, Cambridge, MA: Ballinger, 1975.

Olley, G. S. and A. Pakes, “The Dynamics of Productivity in the Telecommunications Equipment Industry,” *Econometrica*, 64(6):1263–1297, 1996.

Ost, B., “How Do Teachers Improve? The Relative Importance of Specific and General Human Capital,” *American Economic Journal: Applied Economics*, 6(2):127, 2014.

Papay, J. P. and M. A. Kraft, “Productivity Returns to Experience in the Teacher Labor Market: Methodological Challenges and New Evidence on Long-term Career Improvement,” *Journal of Public Economics*, 130:105–119, 2015.

Petronijevic, U., “Incentives and Productivity in the Economics of Education,” Ph.D. thesis, University of Toronto (Canada), 2016.

Pham, L. D., T. D. Nguyen and M. G. Springer, “Teacher Merit Pay: A Meta-Analysis,” *American Educational Research Journal*, 58(3):527–566, 2021.

Podgursky, M. and M. Springer, “Teacher Compensation Systems in the United States K-12 Public School System,” *National Tax Journal*, 64(1):165–192, 2011.

Poletaev, M. and C. Robinson, “Human Capital Specificity: Evidence from the Dictionary of Occupational Titles and Displaced Worker Surveys, 1984–2000,” *Journal of Labor Economics*, 26(3):387–420, 2008.

R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022.

Rivkin, S. G., E. A. Hanushek and J. F. Kain, “Teachers, Schools, and Academic Achievement,” *Econometrica*, 73(2):417–458, 2005, ISSN 1468-0262.

Robinson, C., “Occupational Mobility, Occupation Distance, and Specific Human Capital,” *Journal of Human Resources*, 53(2):513–551, 2018.

Rosen, S., “Learning and Experience in the Labor Market,” *Journal of Human Resources*, pp. 326–342, 1972.

Rosenzweig, M. R. and C. Udry, “External Validity in a Stochastic World: Evidence from Low-Income Countries,” *Review of Economic Studies*, 87(1):343–381, 2020.

Rust, J., “The Limits of Inference with Theory: A Review of Wolpin (2013),” (2013), 2013.

Sanders, C., “Skill Uncertainty, Skill Accumulation, and Occupational Choice,” *Working Paper*, 2010.

Sanders, C. and C. Taber, “Life-Cycle Wage Growth and Heterogeneous Human Capital,” *Annu. Rev. Econ.*, 4(1):399–425, 2012.

Sharpsteen, C. and C. Bracken, *tikzDevice: R Graphics Output in LaTeX Format*, 2023, R package version 0.12.6.

Shaw, K. L., "Life-cycle Labor Supply with Human Capital Accumulation," *International Economic Review*, pp. 431–456, 1989.

Shaw, K. L. and E. P. Lazear, "Tenure and Output," *Labour Economics*, 15(4):704–723, 2008.

Stinebrickner, T. R., "A Dynamic Model of Teacher Labor Supply," *Journal of Labor Economics*, 19(1):196–230, 2001.

Stoye, J., "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77(4):1299–1315, 2009.

Strøm, B. and T. Falch, "The Role of Teacher Quality in Education Production," in "The Economics of Education," pp. 307–319, Elsevier, 2020.

Taber, C., A. Seshadri and X. Fan, "Estimation of a Life-Cycle Model with Human Capital, Labor Supply, and Retirement," *Journal of Political Economy*, 132(1):48–95, 2024, doi: 10.1086/726232.

Tamer, E., "Partial Identification in Econometrics," *Annual Review of Economics*, 2(1):167–195, 2010.

Taylor, E. S., "Teacher Evaluation and Training," *Handbook of the Economics of Education*, 7, 2022.

Tincani, M. M., "Teacher Labor Markets, School Vouchers and Student Cognitive Achievement: Evidence from Chile," *Quantitative Economics*, 12:173–216, 2021.

Todd, P. E. and K. I. Wolpin, "On the Specification and Estimation of the Production Function for Cognitive Achievement," *Economic Journal*, 113(485):F3–F33, 2003.

—, "The Best of Both Worlds: Combining Randomized Controlled Trials with Structural Modeling," *Journal of Economic Literature*, 61(1):41–85, 2023.

Weiss, Y., "On the Optimal Lifetime Pattern of Labour Supply," *Economic Journal*, 82(328):1293–1315, 1972.

Willis, R. J., *Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions*, Economics Research Center/NORC, 1985.

Wiswall, M., "The Dynamics of Teacher Quality," *Journal of Public Economics*, 100:61–78, 2013.

Yamaguchi, S., "Tasks and Heterogeneous Human Capital," *Journal of Labor Economics*, 30(1):1–53, 2012.