The Lost Human Capital

Teacher Knowledge and Student Achievement in Africa*

Tessa Bold, Deon Filmer, Ezequiel Molina, Jakob Svensson

Latest version August 2017

We use unique data from representative surveys from seven countries representing a substantial proportion of the population of sub-Saharan Africa to assess the effect of teacher content knowledge on student outcomes. We show that after four years of schooling, the majority of students fail to master tasks covered in the second year curriculum. This result, adjusting for the cumulative process of knowledge acquisition and imperfect persistence in learning between grades, can partly be explained by the fact that many teachers struggle with tasks that their students should master in lower primary. Had all students been taught by teachers deemed to master the lower secondary curriculum—a minimum official criterion in the countries in the sample—our estimates imply that students would have acquired an additional three quarters of a curriculum-scaled year of schooling after four years and consequently that the observed gap in effective education after four years would have been reduced by one third. The study highlights the huge shortcomings in teacher quality in Africa and the lost human capital as a consequence.

---

* We are grateful to the many researchers, survey experts and enumerators who have worked on the country surveys that made this paper possible. We would like to thank Peter Fredriksson and David Strömberg for extensive discussions and suggestions. We appreciate comments from seminar participants at Stockholm School of Economics and the World Bank. We are grateful to the World Bank, and in particular the Service Delivery Indicators Trust Fund (funded in large part by the Hewlett Foundation) for supporting the research. We also thank the Riksbankens Jubileumsfond (RJ) for financial support. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.
[a]IIES, Stockholm University and CEPR, tessa.bold@iies.su.se; [b]The World Bank, dfilmer@worldbank.org; [c]The World Bank, ezequielmolina@worldbank.org; [d]IIES, Stockholm University and CEPR, jakob.svensson@iies.su.se.

1. Introduction

In many low income countries, children learn little from attending school. Four out of five students in Mozambique and Nigeria, for example, after more than three years of compulsory language teaching, cannot read a simple word of Portuguese and English, respectively. In India, only one in four fourth grade student manages tasks—such as basic subtraction—that is part of the curriculum for the second grade and roughly half of the students in Uganda, after three years of mathematics teaching, cannot order numbers between 0-100.[1]

A growing body of evidence—based on teacher value-added and experimental studies—suggests that teacher quality, broadly defined, is a key determinant of student learning. Less is known, however, about what specific dimensions of teacher quality matter and how teachers perform along these dimensions.

In Bold et al. (2017), teacher quality in Africa is quantified along three core quality dimensions: Time spent teaching, teachers' knowledge of the subject they are teaching, and teachers' pedagogical skills. Here we take the next logical step and attempt to causally assess the impact on student learning of one of these: teachers' content subject knowledge.

Using unique data of over 5,000 teachers and 20,000 students collected from nationally representative surveys in Kenya, Nigeria, Mozambique, Senegal, Tanzania, Togo, and Uganda—which together represent close to 40 percent of the region's total population—we first document how far along the official curriculum children have progressed after almost four years of schooling; i.e., how many curriculum-scaled years of schooling, or effective years of schooling, they have acquired. We present a simple statistical model of cumulative knowledge acquisition, accounting for imperfect persistence in learning between grades. Placing mild restrictions on the correlation between observed and unobserved teacher knowledge over time, we then exploit within-student within-teacher variation to estimate both the contemporaneous effect of teacher content knowledge on student achievement as well as the extent of fade out of the teachers' impact in earlier grades.

We show that teacher content knowledge has a large and significant contemporaneous effect on student performance. Our preferred specification implies that a 1 standard deviation (SD) increase in teacher content knowledge; i.e. in teachers' curriculum-scaled years of education, increases students' effective years of schooling by 0.09 SD in the short run (after one year of teaching). This implies that moving a student from the 5th to the

[1] The estimates for Mozambique, Nigeria, and Uganda are derived using the data we present in this paper. The estimate for India is from ASER (2013).

95th percentile of the teacher effective years of education distribution increases students' effective years of schooling by 0.27 SD in one year. These effects, however, fade out relatively fast over time, with approximately 50 percent of the short-run effect persisting between grades, yielding a persistence parameter similar to those estimated using value-added models (see Kane and Staiger 2008; Jacob, Lefgren, and Sims 2010; Rothstein 2010; and Andrabi et al., 2011).

Using the estimated structural parameters—the parameter capturing the contemporaneous effect of teacher content knowledge on student learning and the persistence in learning parameter—we then calculate the learning achievements of students in a series of counterfactual experiments. Specifically, we quantify how many more effective years of schooling students would have acquired if they had been taught by teachers that master the lower secondary curriculum—a minimum official criterion in the countries in the sample—but also a level of knowledge few primary school teachers display.

We find that had students been taught by teachers deemed to master the lower secondary curriculum, they would have acquired 0.67−0.84 additional years of effective schooling, implying that raising teacher content knowledge to the lower bar for primary teachers in Africa would in itself, holding teacher effort and pedagogical skills constant (and currently at low levels), reduce the observed effective education gap after four years by 30-38 percent.

Methodologically, our identification strategy extends the within-student across subject comparisons exploited in Dee (2005; 2007) and the within-student within-teacher across subjects comparison proposed by Metzler and Woessmann (2012).[2] Specifically, we exploit the fact that we have access to test score data for both the current teachers and previous year's teacher. This allows us to estimate relatively tight bounds on both the contemporaneous effect of teacher content knowledge on student achievement as well as the extent of fade out of the teachers' impact in earlier grades. Most importantly, by combining these structural estimates, it allows us also estimate the cumulative effect of teacher knowledge on student achievement.[3]

[2] Metzler and Woessmann (2012) use a contemporaneous specification relating grade six student achievement test score measure to contemporaneous teacher score measure using Peruvian data. Clotfelter et al. (2010) and Lavy (2015) also exploit within-student across subject variation to assess impact of teacher credentials and instruction time, respectively, on student achievement, while Bietenbeck et al. (2017) exploit within-student within-teacher comparison to assess the impact of teacher knowledge and textbook provision on student test scores.

[3] As noted in Todd and Wolpin (2003), the underlying education production technology in the contemporaneous specification is justified if students are taught by the same teacher throughout their schooling history or that the

We also use newly collected representative micro data from several countries. Unlike many other large scale data collection efforts, which assess students (and teachers) using multiple choice items and thus introduce additional chance variation in test scores, we use data collected using one-on-one tests (for students) and test scores (for teachers) derived from mock student tests marked by teachers. Importantly, given the design of the two tests, we can map each teacher's (and student's) knowledge onto grade-specific (curriculum) knowledge and thus estimate how far students have progressed after four years of studies, and what level of the curriculum teachers master.

Our work is related to a growing literature on the impact of teacher quality. Several studies, primarily from the U.S, but more recently also from middle income countries (Ecuador and Pakistan), demonstrate the importance of teachers using a value-added approach, with effect sizes; i.e. the effect on student performance from a one standard deviation improvement in teacher value added, ranging from 0.1-0.2SD (Rockoff, 2004; Rivkin et al., 2005; Aaronson et al., 2007; Chetty et al., 2014; Araujo et al., 2016; and Bau and Das, 2017). Our results confirm the importance of teachers, although using a different empirical methodology, and demonstrate the external validity of the value added findings to a low income environment where both the average quality of teachers is low but also the variation in measured teacher quality is likely much greater than in previous studies, and especially studies from the U.S. Our findings also complement a growing experimental literature on teacher quality, showing that teacher effort, broadly defined, can be raised, by providing financial incentives tied either to attendance or student performance (Duflo et al., 2012; Muralidharan and Sundararaman, 2011), or by exploiting the operation of dynamic incentives; i.e., contract teacher programs (Duflo et al., 2015; Bold et al., 2013; Muralidharan and Sundararaman, 2013), leading to improved learning outcomes which can be substantial.[4] Here we focus on teacher content knowledge, noting that there is an important evidence gap in the lack of well-identified studies on the impact of teacher knowledge on learning outcomes in developing countries (see Glewwe and Muralidharan, 2015).

Another common finding in the value added literature is that standard teacher characteristics—experience, education, among others—explain very little of the differences in

effect from previous years teaching completely fades out or that current teacher skills are unrelated to previous teacher skills. Neither of these assumptions, however, are likely to empirically valid in most settings.

[4] There is also a growing literature on the provision of supplemental remedial education, and on automated teaching through computer-aided learning programs or scripted lesson plans, reviewed in, for example, Glewwe and Muralidharan (2015), showing large effects (in the former case), or more mixed results (in the latter case) on student learning outcomes, that at least indirectly speaks to the importance of teacher quality.

teacher quality. This may not be so surprising, especially in a developing country setting, since these characteristics appear largely uncorrelated with key dimensions of teaching many would argue are important, such as the effort exerted, or motivation, of the teacher (Chaudhury et al., 2006), teacher's knowledge of the subject she is teaching and how well she teaches that subject (Bold et al., 2017). More recently, however, research on teacher effectiveness has examined characteristics that are typically not collected by school administrators, including content knowledge (Rockoff et al., 2011, Bau and Das, 2017) and measures of how teachers organize teaching and provide instructional support (Kane and Staiger, 2012; Araujo et al., 2016). Specifically, Bau and Das (2017) find that higher content knowledge is associated with significantly and quantitatively larger positive effect on teacher value added. Their preferred estimates imply that a 1 SD increase in teacher test scores raise student test scores by 0.07 SD. We find, exploiting within-student within-teacher variation to estimate the causal effect of student knowledge an effect size of similar magnitude (0.09 SD) in the short run.

We proceed by first providing a short description of the data we use. In Section 3, we turn to providing summary statistics on both student and teacher content knowledge. In Section 4 we present a statistical model of cumulative knowledge acquisition, describe how we attempt to estimate the causal effect of teacher knowledge on student performance, and discuss the identifying assumptions. Section 5 presents specification and placebo tests followed by the main results. Finally, Section 6 concludes with a short discussion of the implications of our findings.

2. Data and context

We use data from the Service Delivery Indicators (SDI)—an ongoing Africa-wide program with the aim of collecting informative and standardized measures of what primary teachers know, what they do, and what they have to work with. The SDI program—piloted in Tanzania and Senegal in 2010 (Bold et al., 2010, 2011)—grew out of concern about poor learning outcomes observed in various student tests as well as evident shortcomings, most clearly (and perhaps most damagingly) manifested at the school level, in fast-expanding systems of education.

To date, the SDI program has collected data, including from the two pilot countries, from a total of seven countries (eight surveys): Kenya (2012), Mozambique (2014), Nigeria (2013), Senegal (2010), Tanzania (2010, 2014), Togo (2013), and Uganda (2013). In each country, representative surveys of between 150 and 760 schools were implemented using a multistage, cluster-sampling design. Primary schools with at least one fourth-grade class formed the sampling frame. The samples were designed to provide representative estimates for teacher effort, knowledge, and skills in public primary schools, broken down by urban and rural location. For four of the six non-pilot surveys, representative data were also collected for private primary schools. Across the eight surveys, the SDI collected data on 2,600 schools, over 21,000 teachers and 24,000 students in Sub-Saharan Africa (see Bold et al., 2017, for details of the sample).

The surveys collected a broad set of school, teacher, and student specific information, with an approach that relies as much as possible on direct observation rather than on respondent reports. Data were collected through visual inspections of fourth-grade classrooms and the school premises, direct physical verification of teacher presence by unannounced visits, and teacher and student tests. Bold et al. (2017) document how African teachers perform along three core quality dimension: Time spent teaching, teachers' knowledge of the subject they are teaching, and teachers' pedagogical skills. Table 1 reports summary statistics on time spent teaching and teachers' pedagogical skills (see Bold et al., 2017, for details). Teachers, on average, are absent from class 44% of the time and about half of that classroom absence is due to teachers not at all being at the school during regular teaching ours. As a result, while the scheduled teaching time for fourth graders is relatively long—5 hours and 25 minutes—the actual time students are taught is about half that time (2 hours and 46 minutes). Pedagogical knowledge is low, with one in ten teachers deemed to have minimum pedagogy knowledge, and even fewer teachers are judged to properly manage to assess students learning progression and shortcoming.

In each school, ten students were sampled from a randomly selected grade 4 classroom. The choice to test students that had completed the third grade was made with the following objectives in mind: on the one hand a desire to assess cognitive skills at young ages when these are most malleable; and on the other hand a desire to assess the learning outcomes of students who have completed at least some years of schooling and to assess language learning at a time when all children would have had lessons in the official language of their country (English in Nigeria and Uganda, English and Swahili in Kenya and Tanzania, French in Senegal and Togo, and Portuguese in Mozambique). In each school, the students' current, and to the extent possible, previous language and mathematics teacher were selected for testing.[5] Overall, in more than 80 percent of the schools surveyed (and for 65 percent of the

[5] In five of eight surveys, teachers at higher grades were also sampled.

students), data was collected on both the current and previous teachers, i.e. the teacher in grade 4 and in grade 3.

The student test was designed as a one-on-one evaluation, with enumerators reading instructions aloud to students in their mother tongue. This was done in order to build up a differentiated picture of students' cognitive skills; i.e., oral one-to-one testing allows one to test whether a child can solve a mathematics problem even when his/her reading ability is so low that he/she would not be able to attempt the problem independently.

The language test, which evaluated ability in English (Kenya, Nigeria, Tanzania, and Uganda), French (Senegal and Togo), or Portuguese (Mozambique), ranged from simple tasks that tested letter and word recognition to a more challenging reading comprehension test. The mathematics test ranged in difficulty from recognizing and ordering numbers, to the addition of one- to three-digit numbers, to the subtraction of one- and two-digit numbers, and to the multiplication and division of single-digit numbers. In both language and mathematics the tests spanned items from the first four years of the curriculum.[6]

In contrast to other approaches to assess teachers' knowledge, where teachers take exams, teachers were asked to mark (or "grade") mock student tests in language and in mathematics. This method of assessment has two potential advantages. First, it aims to assess teachers in a way that is consistent with their regular teaching activities—namely, marking student work. Second, by using a different mode of assessment for teachers compared to students, it recognizes teachers as professionals. In the analysis, we use data on language knowledge of those teachers who teach language, and data on mathematics knowledge of those teachers who teach mathematics.

Both the language and mathematics tests for teachers covered items starting at Grade 1 level (simple spelling or grammar exercises, addition and subtraction) and included items up to the upper primary level (Cloze passages to assess vocabulary and reading comprehension, interpretation of information in a diagram and/or a graph and more advanced math story problem). Both the student and the teacher test have good reliability, with a reliability ratio (estimated by Cronbach's alpha) above 0.8 in both subjects on the student side and above 0.85 in both subjects on the teacher side.[7]

[6] The teacher and student subject tests were designed by experts in international pedagogy and validated against 13 Sub-Saharan African primary curricula (Botswana, Ethiopia, Gambia, Kenya, Madagascar, Mauritius, Namibia, Nigeria, Rwanda, Seychelles, South Africa, Tanzania, and Uganda). See Johnson, Cunningham and Dowling (2012) for details. A few items in the tests also measured grade 5 knowledge.

[7] Cronbach's alpha is defined as the square of the correlation between the measured test score and the underlying metric. A Cronbach alpha of 1 would indicate that the test is a perfect measure of the underlying metric (though

3. Teacher and student knowledge

Analysis of student assessments typically involve transforming raw score data, such as the fraction correct, into a specific test score measure. In many cases, including in well-known publicly available data sets, this transformation is based on test specific scales and effect sizes are measured in standard deviations of these transformed data (see Jacob and Rothstein, 2016).[8] As normalization of student test scores removes some of the underlying information in the raw data, and as the standard deviation of a test score can be sensitive to the range of question difficulty on the test, we instead construct our main test score measure using the school curriculum as a yardstick.[9] That is, we use the raw scores to determine the grade level of proficiency of students and teachers and label it as effective years of schooling (for students) and effective years of education (for teachers) acquired. Appendix A provides details on how these curriculum-scaled years of schooling/education are defined.

The transformation into effective years of schooling, we argue, makes sense here since the test covered test items from grade 1 up to the end of primary school and since the samples by construction are nationally representative. The measure has at least three advantages. First, the transformed data points are informative in themselves. Second, using effective years of schooling/education as the test score enables us to compare students and teachers using the same scale. Third, it allows one to extrapolate beyond effective years of schooling observed in the sample in a meaningful way. As a robustness test, however, we also report and compare the results using the curriculum-scaled years of schooling measure with two other transformations of the raw scores; the fraction correct on the student and teacher test, and scores rescaled using item response analysis.[10]

---

not necessarily of student/teacher knowledge). As a rule of thumb, values between 0.8-0.9 are considered as good.

[8] The National Assessment of Educational Progress (NAEP) test program in the US, for example, which assesses students in grades 4, 8, and 12, reports scale scores, ranging from roughly 100 to 400 with standard deviations around 30 (Jacob and Rothstein, 2016). The NAEP also reports discrete proficiency categories (basic, proficient, and advanced). The Southern and Eastern Africa Consortium for Monitoring Educational Quality's (SACMEQ) test of grade 6 students, with data collected in 1995, 2000, and 2007, report scale scores with a mean of 500 and standard deviation 100 across students participating in the second wave (2000).

[9] Both the teacher and student subject tests were validated against a large set of Sub-Saharan African primary curricula (see footnote 6). Thus, all items in the student tests covered items in the first four grades in the countries surveyed. The teacher test covered items from grade 1 to 7. We use the Kenyan curriculum to link test items to specific grade levels. This choice should be kept in mind when making cross-country comparisons, which is not the focus here (our empirical specification uses variation across students), as there might be some, albeit likely small, variations across countries in which grade each subject item was introduced.

[10] Item response theory is a method to estimate a respondent's underlying ability/latent trait based on their answers to a series of items, in our case an estimate of the student's (teacher's) knowledge based on the pattern of correct/incorrect questions on the test. To do so, IRT specifies a parametric model for the probability of a correct answer given the test-takers latent trait and properties of the item. While models vary in the precise parameterization, they generally share the following features: the probability of a correct answer is decreasing in

Table 2 and Figure 1 show the percentage of students at each effective year of schooling level for language and mathematics. On average, students have 1.5 effective years of schooling in language and mathematics after three and half years of studies. That is, the median mathematics student, after completing approximately three and half years of schooling, does not master the second grade curriculum in mathematics. Comparing across countries, the average student in Kenya (the top performer) has acquired two and a half years of effective years of schooling after three and a half years of schooling, while the average student in Mozambique (the bottom performer) has acquired only 0.4 years of curriculum-scaled years of schooling (see Figure 2).

Breaking down the summary score, a quarter of fourth grade students have not acquired any effective years of schooling and one-third have not acquired any effective years of schooling in mathematics. 15% and 21% have acquired one year of curriculum-scaled years of schooling in language and mathematics, respectively. Less than a quarter of the students have three years or more of effective years of schooling in language and mathematics.

Table 2 shows how the average scores and scores constructed with item response theory line up with the curriculum-scaled years of schooling measure. There is a strong positive correlation between the different definitions of the scores, above 90% in all cases. While the raw score does not take the difficulty of the questions into account, the IRT and the effective years of schooling are two different methods to scale scores by difficulty. The IRT is essentially a data-driven approach which classifies a question as easy or difficult on the basis of how many teachers (or students) were able to answer it. The effective years of schooling classify a question as easy or difficult based on where on the curriculum it is located. It thus speaks to the validity of the curriculum-scaled years of schooling score to see a high correlation between the two measures.

In Table 4 and Figure 3, we show the percentage of teachers at each quality adjusted year of education. On average, teachers have 3.5 years of quality adjusted education in language and 3.7 years of quality adjusted education in mathematics.[11] Again there are large

---

the difficulty of an item and increasing in the ability/latent trait of the test-taker (see Jacob and Rothstein, 2016). To estimate the item parameters for the student test, we specify a 2-parameter logistic model which describes each item by its difficulty and the extent to which it discriminates between students of different ability. To the teacher test, we apply a partial credit model, which allows for items that are scored on an ordinal (but not necessarily binary) scale. Given the estimated item parameters and patterns of correct/incorrect answers, we can then construct a measure of the underlying student and teacher subject knowledge.

[11] While these numbers are low, they are consistent with the alternative measures of teacher knowledge presented in Bold et al. (2017), who calculate that two thirds of teachers across Sub-Saharan Africa have subject knowledge equivalent to a fourth grader, defined as mastering 80% of the material covering grade 1 to 4 on the

differences across countries, with Kenyan mathematics teachers (on average) having 5.7 effective years of education (top performer) and mathematics teachers in Togo having only 1.9 years of quality adjusted education (see Figure 4).

Looking at the distribution, strikingly, just over a third of the teachers master at best the second grade curriculum and 90% of teachers have quality adjusted education at or below five years in language and at or below six years in mathematics. Also in the case of teacher knowledge, there is a high correlation between the different score aggregates (see Table 4).

4. A statistical model

In this section we lay out a statistical model for cognitive achievement that assumes that children's achievement, as measured by test performance after t years of school, is the outcome of a cumulative process of knowledge acquisition.

We first present the model in a general form and then use it to highlight the assumptions we make, given the structure of our data, in order to estimate the causal effect of teacher quality on student learning.

Let $y_{ijt,k}$ be student i's achievement in school j after t years of schooling (or in grade t, in subject k. As in Todd and Wolpin (2003) we view knowledge acquisition as a production process in which current and past inputs are combined with an individual's innate ability (or motivation), denoted as ω_{ij}, to produce a cognitive outcome that can be measured. These inputs include school-supplied inputs, parent-supplied inputs, and teacher-supplied inputs, some of which could vary by subject and some of which do not. Specifically, let $S_{ijt,k} = \{s_{ijt,k}, \bar{s}_{ijt}\}$ and $P_{ijt,k} = \{p_{ijt,k}, \bar{p}_{ijt}\}$; i.e., each input vector consists of a vector of subject-specific and subject-invariants inputs, the latter indicated by a bar over the variable. Further, assume the vector of teacher-supplied inputs is $T_{ijt,k} = \{x_{ijt,k}, c_{ijt,k}, \bar{x}_{ijt}, \bar{c}_{ijt}\}$, where $x_{ijt,k}$ is the subject content knowledge of the teacher teaching student i in school j at t (or grade t) in subject k, $c_{ijt,k}$ is a vector of other subject-by-teacher characteristics/skills, and $\bar{x}_{ijt}$ and $\bar{c}_{ijt}$ are the corresponding subject-invariant terms.

test. Our definition here is in one respect more stringent since a teacher has to score, for example, all grade 4 questions correctly in order to be categorized as having grade 4 knowledge (rather than 80% of questions covering grade 1-4). On the other hand, we do not require, again using grade 4 as an example, that the teacher also manages all tasks covering grade 1-3 in order to be categorized as having grade 4 knowledge. Adopting a slightly more lenient approach that allows for some margin of error at each grade level increases teachers' average effective years of education to 4.5 years. The regression results are left intact with this more lenient definition (results available upon request). Nevertheless, we use the stricter measure here as it makes for a more transparent definition of the teacher test score.

Let $T_{ij,k}(t)$, $S_{ij,k}(t)$, $P_{ij,k}(t)$ denote input histories up to t years of schooling (or for students in grade t), then allowing for measurement error in test scores, denoted by $\varepsilon_{ijt,k}$, the production function, after t years of schooling (or students in grade t) is

(1) $$y_{ijt,k} = F\left[T_{ij,k}(t), S_{ij,k}(t), P_{ij,k}(t), \omega_{ij}, \varepsilon_{ijt,k}\right].$$

Equation (1) highlights the two main problems we face in estimating the causal (cumulative) effect of teacher content knowledge on student learning. First, students' innate ability, and several school and parent-supplied inputs, are inherently unobservable and may be correlated with x_{ij}, for instance if better students sort into schools with better teachers. Second, student achievement in grade (or year) $t_n > 1$ is a function of the whole history of the quality of teaching; i.e., x_n, x_{n-1},…x_1. We deal with these issues in two ways. First, we exploit variation in student and teacher knowledge within students and within teachers. Second, we add additional structure to how teacher content knowledge and student learning evolve over time and utilize the fact that we have test score data both for students' current teachers and the teachers who taught them in the previous year to estimate bounds on the total cumulative effect of teacher knowledge after four years of education. Below we describe these two steps in more details.

4.1. Within-student within-teacher variation

Linearizing the production function and exploiting the fact that we observe test scores for two subjects, k and k', we can take the first-difference version of equation (1). If the coefficients on all inputs are subject-invariant; i.e., a one unit increase in, for example, effective years of education, has the same marginal effect on test scores in subject k as in subject k', this within-student transformation ensures that all subject-invariant unobserved heterogeneity at the school and parent level is removed.[12]

In our sample, about half of the students are taught by a class teacher; i.e., a teacher teaching both subjects in a given grade in the current and previous year (although not necessarily the same teacher in both years). If we restrict attention to students who were taught by such class teachers, we can further remove any teacher-specific, subject-invariant heterogeneity in each of these years, yielding the following specification for fourth year students,

(2) $$\Delta y_{ij4} = y_{ij4,k} - y_{ij4,k'} = \alpha_0 + \alpha_4 \Delta x_{ij4} + \alpha_3 \Delta x_{ij3} + \alpha_2 \Delta x_{ij2} + \alpha_1 \Delta x_{ij1} + \epsilon_{ij}$$

[12] As reported in section 5, we cannot reject the hypothesis that the effect is the same in the two subjects.

where $\Delta x_{ijt} = x_{ijt,k} - x_{ijt,k'}$ is the difference in teacher subject content knowledge across the two subjects, and where the error term ϵ_{ij} subsumes all the remaining unobservable inputs; i.e.,

(3) $$\epsilon_{ij} = \Delta\omega_{ij} + \sum_{t=1}^{4}\left[\theta_t^S \Delta s_{ijt} + \theta_t^P \Delta p_{ijt} + \theta_t^T \Delta c_{ijt|f_k=f_{k'}}\right] + \sum_{t=1}^{2} \theta_t^{T\prime} \Delta C_{ijt|f_k \neq f_{k'}} + \Delta\varepsilon_{ijt}$$

where we distinguish between unobserved teacher skills and characteristics for students who were taught by the same teacher f in both subjects in year 1 and 2, $\theta_t^T \Delta c_{ijt|f_k=f_{k'}}$,, and for those who were taught by different teachers in language and mathematics, $\sum_{t=1}^{2} \theta_t^{T\prime} \Delta C_{ijt|f_k \neq f_{k'}}$. In the former case, only subject-specific variation enters the error term, while in the latter both subject-specific and subject-invariant variation may matter; i.e., $C_{ijt} = \left[c_{ijt,k}, \bar{c}_{ijt}\right]$. With access to complete content knowledge data on teachers, estimation of the within-student, within-teacher specification (2) with OLS would recover the causal effect of teacher knowledge on student performance if the variation in teacher knowledge across subjects (Δx_{ijt}) and the error term (ϵ_{ij}) are orthogonal. As this is our key identifying assumption we consider next what it implies.

Consider first the differenced ability/motivation term $\Delta\omega_{ij}$ in (3). Note that $\Delta\omega_{ij} \neq 0$ only if students have subject specific abilities/motivations. If that is the case, our identifying assumption rules out that students systematically sort, based on these subject-specific abilities, into schools with subject-specific teacher knowledge. For example, our assumption would be invalid if students in lower primary with relatively higher motivation for mathematics sort into schools (or classrooms) with relatively more knowledgeable mathematics teachers. It also places some restrictions on what parents and schools do. For example, while our identifying assumption does allow for parents (or the school) to respond to their children's low mathematics aptitude by providing additional teaching (or hire a private tutor), they cannot do this to compensate for insufficient teacher mathematics knowledge.[13] More generally, while differential (across subjects) supply of school (Δs_{ijt}) and parental (Δp_{ijt}) inputs may occur across schools and students, and may be correlated with various school and student characteristics, our maintained assumption is that these differential input flows are uncorrelated with the variation in teacher content knowledge across subjects. In the context of lower primary schooling in Africa, these assumptions appear reasonable and we provide

[13] Using data from kindergarten students in Ecuador, Araujo et al. (2016) find that while parents recognize better teachers, they do not change their behaviors to take account of differences in teacher quality. Note that our identifying assumption here is even weaker. We assume parents do not respond to differential (across subjects) differences in the quality of the teacher.

additional evidence in support of them in section 5. However, the assumptions remain fundamentally untestable without complete data on inputs histories.

There are reasons to believe that the potential bias, if any, due to the omission of variation in teacher skills and characteristics across subjects that arises from the student not being taught by a class teacher; i.e., a teacher teaching both subject, in year 1 and 2, $\sum_{t=1}^{2} \theta_t^{T\prime} \Delta C_{ijt|f_k \neq f_{k'}}$, in equation (2), will be small. Specifically, the structure of primary school in the countries we study is such that students tend to have one teacher who teaches both language and mathematics in lower primary (grades 1-3) while subject teachers, who specialize in either language or mathematics, become progressively more common as students move to upper primary and secondary. In other words, if a student is taught by a class teacher in both mathematics and language in grade t, it makes it likely, we would argue, that the student was taught by a class teacher who teaches both subjects also in $t-1$, while the opposite transition may not hold. This pattern is also clear from the data. Specifically, while 60% of the students taught by a teacher teaching both subjects in grade 3 are also taught by subject-specific teachers in grade 4, more than 90% of the students that are taught by a class teacher who teaches both subjects in grade 4 also had a class teacher in grade 3. Although we do not know who taught the students in grade 1 and 2, we argue that these data patterns suggest that our core sample (which is restricted to students who had class teachers in grade 3 and 4) is unlikely to contain many students who were taught by teachers specializing in their subject in grade 1 and 2. Indeed, if all students that have class teachers in grade 3 also have class teachers in grade 1 and 2 (though again not necessarily the same teacher as in grade 3), then the within student transformation of the data and the restriction of the sample to students who have only one teacher in both subjects in grade 3 (and in grade 4) is sufficient to remove the nuisance term $\sum_{t=1}^{2} \theta_t^{T\prime} \Delta \bar{c}_{ijt}$.

Finally, consider next the variation in Δc_{ijt}; i.e., in other (unobservable) characteristics or skills that vary by subject. For example, a teacher, teaching both subjects, may be more motivated to teach a subject she masters relatively well, or possibly put more effort into teaching if she is less knowledgeable of the subject. To the extent these additional subject-specific traits are systematically correlated with teacher subject-specific content knowledge, α_t needs to be reinterpreted slightly more broadly; i.e. as the impact of teacher content knowledge and other unmeasured teacher subject-specific teaching traits correlated with it.

4.2. The cumulative effect of teacher knowledge on student learning

In equation (2), the difference in student learning outcomes in grade 4, denoted by Δy after dropping student, school, and time subscripts, is a function of the students' teachers' knowledge over their full schooling history. In the data, however, we do not observe Δx_2 and Δx_1. Thus, instead of equation (2), what we can estimate with the data is

$$\Delta y = \beta_0 + \beta_4 \Delta x_4 + \beta_3 \Delta x_3 + \mu \ . \tag{4}$$

Clearly, even under the identifying assumption discussed above, i.e., $\text{cov}\left(\Delta x_{ijt}, \epsilon_{ij}\right) = 0$, we cannot recover the structural coefficients of interest from (4) as long as there is correlation between subject differences in teacher test scores across grades and their effect on current student scores does not decay completely. Nevertheless, under mild conditions, knowledge of current and previous test scores in grade 3 and 4 is sufficient to estimate a range for the contemporaneous effect of teacher content knowledge on student achievement, as well as the extent of fade out of the teachers' impact in earlier grades and, most importantly, the cumulative effect of teacher knowledge on student learning resulting from combining these.

To see this, we start by introducing the possibility that student learning, as measured by test scores, partly fades out over time.[14] Specifically, we parameterize imperfect persistence of achievement in the statistical model by assuming that test scores decay geometrically. We further assume that the contemporaneous effect of teacher knowledge is constant across years, $\alpha_j = \alpha$, that is, the effect of teacher knowledge is not affected by the age at which it is applied. With these assumptions, we can rewrite (2) as

$$\Delta y = \alpha_0 + \alpha \Delta x_4 + \alpha\gamma \Delta x_3 + \alpha\gamma^2 \Delta x_2 + \alpha\gamma^3 \Delta x_1 + \epsilon \tag{5}$$

where γ—the parameter that links achievement across periods—captures the degree of persistence, and where, for convenience, we have dropped all i and j subscripts.

Finally, we assume stationarity of the test score distribution over time; i.e., $\text{var}(\Delta x_t) = \text{var}(\Delta x_{t'}) = \text{var}(\Delta x)$.[15] Given the assumptions above, the OLS estimator β_4 and β_3 in (4) can be written as (see derivation in the appendix):

$$\text{plim}\ \hat{\beta}_4 = \alpha + \alpha\gamma^2 \left(\frac{\rho_{2,4} - \rho_{2,3}\rho}{1-\rho^2}\right) + \alpha\gamma^3 \left(\frac{\rho_{1,4} - \rho_{1,3}\rho}{1-\rho^2}\right), \tag{6}$$

[14] Recent research, using data from both developed and developing countries, suggest that student learning, as measured by test scores, fades rapidly. Kane and Staiger (2008); Jacob, Lefgren, and Sims (2010); and Rothstein (2010), for example, show that teacher effects dissipate by between 50-80% over one year. Similar patterns are also observed in a number of education experiments (see Andrabi et al., 2011, for a discussion and references).

[15] This assumption hold in our data for test scores in grade 3 and 4. The assumption is, however, not crucial, but simplifies the algebra somewhat.

and

(7) $$\text{plim } \hat{\beta}_3 = \alpha\gamma + \alpha\gamma^2\left(\frac{\rho_{2,3}-\rho_{2,4}\rho}{1-\rho^2}\right) + \alpha\gamma^3\left(\frac{\rho_{1,3}-\rho_{1,4}\rho}{1-\rho^2}\right),$$

where $\rho_{t,t'} = \frac{\text{cov}(\Delta x_t, \Delta x_{t'})}{\text{var}(\Delta x)}$ and where ρ is estimated from the linear projection of current test scores on previous test scores (differenced across subjects); i.e., from equation (8),

(8) $$\Delta x_4 = \rho_0 + \rho_{43}\Delta x_3 + v_{4,3}\,.$$

Imposing the stationary assumption implies that $\rho_{43} = \rho_{34} = \rho$ and that the regression coefficient equals the correlation coefficient.

To recover the structural parameters α and γ from these reduced form estimates, we need to know how the observed (Δx_4, Δx_3) and the omitted (Δx_2, Δx_1) teacher scores are correlated across grades. Rather than imposing particular values, we take an agnostic approach and allow the correlation coefficients, $\rho_{t,t'}$, to vary freely in a mildly, and we would argue realistically, restricted space. We then estimate the full distribution of possible cumulative effects of teacher knowledge on student learning. Specifically, we assume that all $\rho_{t,t'} \geq 0$. Second, we assume that $\rho_{t,t'}$ is decreasing in $|t - t'|$. In other words, the further apart are any two sets of grades, the lower the correlation between teacher knowledge. Finally, we assume that $\rho_{t,t-1}$ is decreasing in t. Note that this implies that no $\rho_{t,t-1}$ is smaller than the estimated correlation between test scores (differenced across subjects) in grade 3 and 4; i.e., ρ.

These restrictions are motivated by the typical pattern of transitions of teachers and their grades through primary school coupled with the assumption that these transition patterns are the main (though not necessarily exclusive) drivers of correlations across grades. In particular, there is a substantial share of teachers who transition with their grade each year (restriction (i)), any grade t is more likely to be taught by their teacher in grade $t-1$ than teachers in earlier grades (restriction (ii)), and teachers in earlier grades are more likely to transition together with their grade than teachers in higher grades (restriction (iii)).[16]

Given the restrictions on the correlation patterns of omitted and included teacher knowledge, we can further show that the total cumulative effect is effectively bounded by two easily interpretable data generating processes and that the range of estimated values is systematically related to the properties of these two limiting cases.

[16] Note that implicit in this argument is the assumption that teacher knowledge is essentially fixed (or changes only very little year by year. In other words, in a context in which teacher in-service training is rare, measuring the knowledge of a teacher who taught in year $t-1$ in year t, provides a good measure of the teacher knowledge input in year $t-1$.

The first case corresponds to a situation where students are taught by the same teacher in lower primary (i.e. grade 1, 2 and 3) and then some students change teacher only once they reach grade 4. This implies that $\rho_{t,t'} = 1$, and $\rho_{4,t'} = \rho\ \forall t, t' < 4$. This case effectively provides a lower bound on the range of cumulative effects we estimate. In this scenario, the coefficients on the teacher knowledge variables suffer only from collinearity in the regressors. All that is necessary for solving for the coefficients of interest are the functional form assumptions about the education production function in equation (5), namely a constant instantaneous effect of teacher knowledge α coupled with geometric decay (and of course the identification assumptions). In this case:

(6a) $$\text{plim } \hat{\beta}_4 = \alpha$$

(7a) $$\text{plim } \hat{\beta}_3 = \alpha\gamma + \alpha\gamma^2 + \alpha\gamma^3$$

In the second case, the relationship of teacher knowledge (differenced across subjects) follows an AR(1) process,

(8b) $$\Delta x_t = \rho_0 + \rho\Delta x_{t-1} + v_{t,t-1}\ \forall t,$$

with $\text{cov}\left(\Delta x_t, v_{t',t'-1}\right) = 0\ \forall t, t'$ and hence $\rho_{t,t'} = \rho^{|t-t'|}$.[17] This case effectively provides an upper bound on the range of cumulative effects we estimate. Such a data generating process would, for example, describe a situation in which a fraction ρ of teachers is re-allocated in each year and teacher transitions are the sole source of correlations of subject differences in test scores across grades. In this scenario

(6b) $$\text{plim } \hat{\beta}_4 = \alpha$$

(7b) $$\text{plim } \hat{\beta}_3 = \alpha\gamma + \alpha\gamma^2\rho + \alpha\gamma^3\rho^2$$

Beyond the values for $\rho_{t,t'}$, there are three unknowns in, the variant versions of, equations (6), (7), and (8), namely α, γ, ρ, which can be uncovered from the estimated coefficients $\hat{\beta}_3$, $\hat{\beta}_4$ and $\hat{\rho}$.[18]

4.3. Inference

To make inference about the structural parameters of interest, $\widehat{\boldsymbol{\theta}} = \left\{\hat{\alpha}, \widehat{\alpha\gamma}, \widehat{\alpha\gamma^2}, \widehat{\alpha\gamma^3}\right\}$, we need to estimate their standard errors. The asymptotic variance-covariance matrix is given by $V = \sigma_\epsilon^2 (E(\Delta\tilde{x}'\Delta\tilde{x}))^{-1}/N$, where $\Delta\tilde{x} = \{\Delta\tilde{x}_4, \Delta\tilde{x}_3, \Delta\tilde{x}_2, \Delta\tilde{x}_1\}$ is the matrix of demeaned test score subject differences across four years. Note that each of the two terms in the variance-

[17] Strictly speaking, it is the teacher knowledge that a student is exposed to in grade t, that follows an AR(1) process.

[18] In the estimation, we also adjust $\hat{\beta}_4$ for the fact that test scores are measured half way through year 4, while the structural model measures the impact of teaching after a full year of teaching in each grade.

covariance matrix depends on population moments for which we have no sample analogue. That is, we cannot replace, for example $E(\Delta\tilde{x}'\Delta\tilde{x})_{3,2} = Cov(\Delta x_3, \Delta x_2)$, by its sample analogue simply because we do not observe Δx_2. To circumvent this problem, we show in the Appendix that under the assumptions above, each of the terms can be rewritten as a function of population moments whose sample analogue we do observe.[19]

5. Results

In Table 6, we begin to explore the relationship between teacher and student knowledge. We start in columns (1) and (2) by regressing, for each subject separately, student achievement (effective years of schooling in that subject) on teacher subject knowledge (effective years of education in that subject), controlling only for a set of country fixed effects. For both subjects, there is a large positive association and the estimates across subjects are similar. Column (3) pools the data, thus implicitly restricting the coefficients on the teacher subject knowledge to be the same across subjects. In column (4), we include past teacher knowledge (also in effective years of education), resulting in a fall in the estimated coefficient on current teacher knowledge by 30 percent (as can be seen by comparing with the findings in column (5) which reports the results of the contemporaneous specification in the sample with both current and past teachers; i.e. the sample used in column (4)).

In column (6), we introduce student fixed effects to control for sorting of students to schools (or teachers) on the basis of subject invariant characteristics, as well as other unobserved student, and subject invariant, characteristics.[20] In this specification, the effect of teacher knowledge on student test scores can only be driven by differences between the two subjects. The results suggests that part of the association in column (1) is driven by better students sorting into better schools, but the point estimate is still large and significant.[21]

[19] Note that applying the delta method to the non-linear functions of $\hat{\beta}_4$ and $\hat{\beta}_3$ is not possible here. First, with the exception of the limiting cases, there are no easy closed form solutions for the parameters of interest in terms of the reduced form coefficients. Second, the delta method would not provide the full variance-covariance matrix (only its diagonal) that is needed to test hypotheses about the total cumulative effect.

[20] Following Ashenfelter and Zimmerman (1997), we test the restriction implied by the fixed effect model; i.e., that the (reduced form) effects of teacher content knowledge on student content knowledge, in a given grade, are the same across subjects ($\beta_{4,math} = \beta_{4,language}$ and $\beta_{3,math} = \beta_{3,language}$), by rewriting our main specification as a correlated random effects model, using all available variables that vary by students as controls. We cannot reject the hypotheses that the effects of teacher content knowledge is the same in the two subjects for both grade 3 teachers, $\chi^2(\beta_{3,math} = \beta_{3,language}) = 0.63$ [Prob $> \chi^2 = 0.43$], and grade 4 teachers, $\chi^2(\beta_{4,math} = \beta_{4,language}) = 0.74$ [Prob $> \chi^2 = 0.39$], thus providing support for our fixed effect specification.

[21] Note that the fixed effects specification tends to inflate existing measurement error, so the smaller effect size could also be a consequence of the decreased signal to noise ratio in this specification.

It is still possible that the association reported in column (6) is driven by teaching activities that vary across subjects because students are not taught by the same teacher. Hence, in column (7), our preferred specification, we also introduce teacher fixed effects for the current and previous teacher. The results change only slightly, suggesting that other activities and skills of teachers that do not vary across subjects are not correlated with the variation in teacher subject knowledge in a way that would affect the estimate.

5.1. Robustness checks of the reduced form estimates

To recap the identification assumptions stated in Section 4, two things have to be true for our preferred specification, reported in column (7) in Table 6, to be interpreted as causal: (i) there must not be other factors (at teacher level or otherwise) that drive both student and teacher subject differences in knowledge; (ii) there is no sorting by students and teachers on the basis of subject differences. In other words, students that are better in language than in mathematics are not systematically more likely to select into schools with teachers that are better in language than in mathematics (or vice versa).

While we cannot unambiguously rule out either of these concerns, we present additional evidence in Table 7 and 8 suggesting that neither of these assumptions is likely to be violated. In column (1) of Table 7, we repeat our main specification using first differences across subjects and restricting the sample to students who have the same teacher in grade 3 and 4. In columns (2)-(4), we then examine whether differences between teacher language and mathematics scores might be driven by a common underlying factor that also affects student subject differences. For example, it might be the case that language knowledge of both students and teachers varies systematically across contexts, such as districts, or urban and rural areas, simply because of differences in the prevalence of the official language. To assess this, we include district (column 2) and urban/rural dummies (column 3). As can be seen, compared to the main specification reported in column (1), the estimates change only marginally.[22]

Similarly, other teacher behavior and skills that vary by subject might be correlated with teacher knowledge and affect learning. While we do not have any measure of teacher behavior that varies across subjects for a given teacher, we have already seen that the results remain basically unchanged when we move from a sample where students may have different teachers in each subject to a sample in which they are restricted to have the same teacher in both subjects (Table 6, columns (6) and (7)), suggesting that unobserved teacher behavior and

[22] A Mundlak (1978) test indicates that we cannot reject the null that the additional fixed effects are redundant. Results available upon request.

skill are unlikely to confound the estimates. This also presents us with an opportunity to test directly how teacher subject knowledge correlates with other teacher skills. Specifically, in column (4) of Table 7, we repeat the student fixed effects specification reported in column (6), Table 6, thus including also students taught by different teachers in language and mathematics in the sample, and add a measure of teachers' pedagogy knowledge as an additional explanatory variable.[23] While pedagogy knowledge has a positive effect on student learning, the coefficients of interest are only marginally affected by the inclusion of this variable. Hence, we would argue that unmeasured differences in teacher skills—at least pedagogical skills—across subjects (be they the same or different teachers) are unlikely to confound the coefficient of interest.

To further test for sorting across (and within) schools we report the results of two specifications where we constrain the sample to only include schools in rural areas—where the choice of schools to attend for students are more limited (column 5)—and schools with only one classroom (column 6); thus effectively ruling out sorting into different classes within schools. While the estimate on current teacher knowledge falls somewhat and the estimate on previous teacher slightly increases, the cumulative effect, as discussed in section 5.2, remains largely unchanged.

To further bolster the causal interpretation, Table 8 presents a set of placebo tests in line with Chetty et al. (2014). Column (1) uses the subject differences of test scores of teachers in higher grades as an additional control. The argument is as follows: if there is purely a sorting relationship between subject differences of student and teacher test scores in the school, then the teacher test scores in other grades should also be correlated with student test scores (and the teacher test scores should be correlated across grades because of teacher sorting). Hence, including such test scores should change the coefficient on current and previous teacher test scores if sorting is taking place, but not, if the effect is causal. As seen by comparing columns (1) and (2) of Table 8, both using the same sample of students for whom teacher test scores in higher grades are available, the coefficients on current and previous teacher test scores are unchanged by the inclusion of teacher test scores in higher grades, which themselves have an insignificant and essentially zero effect. Second, if the relationship between teacher and student knowledge is purely due to sorting, then the length of exposure to a given teacher should not matter. We test this in columns (3) and (4), where we compare the coefficient on current teacher knowledge for those who have kept their grade class teachers in grade 4 and those who changed

[23] Pedagogical knowledge is measured as the score on a lesson preparation exercise that was administered to all teachers. The assessment pedagogy knowledge and skills is described in Bold et. al. (2017).

class teacher. The coefficient is almost twice as large in the first case, implying that length of exposure indeed matters.

5.2. Estimates of the cumulative effect of teacher content knowledge on student learning

We now present the range of structural parameters, α, the contemporaneous effect, γ, its persistence and the cumulative effect of teacher knowledge on student learning after four years, i.e. $\alpha \sum_{t=1}^{4} \gamma^{4-t}$, implied by the reduced form coefficients in Table 6, a $\hat{\rho} = 0.51$, and the restrictions imposed on the correlations between observed and unobserved subject differences in test scores as described in section 4.2.

Figure 5 depicts the density function of the estimated α, γ, and the cumulative effect. Table 9 summarizes the finding by reporting the median estimates from the estimated distribution (column 1), as well as the estimates from the two "limiting" data generating processes described in section 4.2.

The contemporaneous effect is estimated to lie between 0.043−0.072 (1st−99th percentile) with a median of 0.061 (Figure 5, Panel A), implying that being taught by a teacher with one more year of effective education would increase student learning, using the median estimate as a reference point, by three-quarters of a month after one year.[24] To put this number in context with other findings, a 1 SD increase in effective years of education for a teacher increases student learning (effective years of schooling) by 0.087 standard deviations, implying that moving a student from the 5th to the 95th percentile of the teacher effective years of education distribution increases students' effective years of schooling by 0.27 SD in one year.

The range for the degree of persistence, 0.39−0.72 (1st–99th percentile) with a median of 0.485 (Figure 5, Panel B), is consistent with what has been reported using data from Pakistan and the US (see Kane and Staiger 2008; Jacob, Lefgren, and Sims 2010; Rothstein 2010; and Andrabi et al., 2011). This implies that approximately 50 percent of the short-run effect persists between grades, again using the median of the estimated persistence parameters as a reference point.

Figure 5, Panel C plots the cumulative effect, which is estimated to lie between 0.103 and 0.129, with a median effect total effect across all parameterizations estimated to be 0.113. That is, being taught, throughout lower primary (grade 1-4), by a teacher with one more year of effective education would increase student learning by almost a month and a half after four years, or by 0.08 SD of effective years of schooling.

---

[24] Note that α and γ that solve (6) and (7) are not independent, and hence $\text{median}(\alpha \sum_{t=1}^{4} \gamma^{4-t}) \neq (\text{median}(\alpha) \sum_{t=1}^{4} \text{median}(\gamma)^{4-t})$. In Table 9, we report the parameter estimates for α and γ at the median cumulative effect.

The minimum for the cumulative effect coincides with the scenario in which students have the same teacher throughout grade 1-3. On the other hand, 97% of values lie below the cumulative effect estimated in the case where subject differences in test scores follow an AR(1) process. Within these, fairly tight, bounds, the total cumulative effect tends to be larger, the smaller the correlation between included and omitted test scores and the less the correlation structure departs from the exogeneity condition in the AR(1) model (i.e. $\mathrm{cov}\left(\Delta x_t, v_{t',t'-1}\right) = 0\ \forall t, t'$).

We also present one additional exercise in Table 9, namely the estimation of the contemporaneous and cumulative effect of teacher knowledge on student learning when constructing the dependent and independent variables using item response analysis (IRT) rather than the curriculum adjustment as in Panel A. The IRT measure of teacher knowledge and student learning (Panel B) yield similar structural estimates, though the actual magnitudes for the cumulative effects are somewhat smaller when expressing both sets of effects in standard deviations. Finally, as an additional robustness exercise, we also estimate the structural parameters and the cumulative effects based on the reduced for estimates reported in columns (5) and (6) in Table 7; i.e., in the sample of schools in rural areas and sample of schools with only one classroom. While the median estimates for the contemporaneous effect is now somewhat larger, and the persistence parameter somewhat smaller compared to the median estimates reported in Table 9, the cumulative effect is essentially the same.[25]

5.3. Counterfactual policy experiments

Given that students lag behind 2.2 years of effective schooling already after four years of primary school and their teachers do not master the primary curriculum, what do these results imply for policy reforms designed to combat the learning crisis?[26]

[25] Test scores are an inherently noisy measure of the underlying teacher knowledge. Failing to account for classical measurement error usually implies that coefficient estimates are conservative because of attenuation bias, but this is not necessarily the case when there are several variables measured with error and the resulting bias therefore warrants further examination. Although it is generally impossible to obtain estimates of the measurement error, when the variable measured with error is a test score, psychometric test theory can be used to estimate its reliability (which is inversely related to the variability of measurement error). We therefore re-estimate the range of coefficients adjusting for each subject test's reliability (either via Crohnbach's alpha or via the information matrix from the partial credit model for test scores, see Section 3). Adjusting for classical measurement error in this way increases the median estimates (and their range) by 15-25%.

[26] As mentioned in footnote 17, students and teachers were assessed approximately half way through year 4. Linearly approximating students' achievement after four years, based on their achievement after 3.6 years, results in 1.78 years of effective schooling after four years and an education gap of 2.22.

De jure all countries in our sample have well-established systems for teacher training, which confer training at or below the post-secondary non-tertiary level and the large majority of teachers holds such a training certificate. The minimum entry requirement for teacher training is lower secondary education, equivalent to ten years of schooling, which 90% of teachers in our sample have completed. De facto, however, we have shown that teachers' effective years of education are far lower.

We now ask how many effective years of schooling students would accumulate after four years if teachers' effective years of education rose to the lower secondary level, the minimum entry requirement for teaching—and thus equaled the number of years most of them spent in school. This policy experiment is equivalent to an increase of 6.5 years of teachers' effective years of education relative to the current average of 3.5 years. Extrapolating the quasi-experimental results, this would increase effective years of schooling by 0.67−0.84 year, with the median of the full range of parameterizations equal to three quarters of a year. This in turn implies that raising teacher content knowledge to the lower bar for primary teachers in Africa would in itself, holding teacher effort and pedagogical skills constant, reduce the observed schooling gap after four years by 28−36 percent (median = 30 percent).[27]

These results apply holding other dimensions of teacher quality, such as effort and skill, constant. As we show in Bold et al. (2017), there are also large shortfalls along these dimensions; for example teachers are absent from classroom roughly half of the scheduled teaching time. Specifically, while the effect sizes reported above suggest that low teacher content knowledge may be the most important factor in accounting for the fact that children learn little from attending school, reforms that focus purely on teacher knowledge and training would require teachers in Sub-Saharan Africa to complete (effective) education exceeding university level in order to completely close the gap in student learning that has opened up by grade 4.[28]

6. Discussion

Recent estimates suggest that differences in (the quality of) human capital can explain a dominant share of world income differences (Caselli and Coleman, 2006; Jones, 2014; Malmberg, 2017). Thus, the fact that many children in low income countries learn little from

[27] This result is arrived at by multiplying the cumulative effect of four years of teaching in the third row of Panel B in Table 9 by the number of effective years of education required to increase from the current average (3.5 years) to the minimum requirement (10 years).
[28] This result is arrived at by dividing students' shortfall in human capital after four years, 2.5, by the amount of learning acquired after four years if teachers increased their human capital by one year.

attending school may be one of the most pressing development challenges. In this paper, we focus on one component of the education production function—teachers' knowledge of the subject they are teaching. While a growing literature has shown that teachers matter, much less is known about the link between specific teacher characteristics and student learning (see Glewwe and Muralidharan, 2015). Here we show that teachers' content knowledge, or lack thereof, is an important explanation for why students in primary schools in Africa already after a few years of schooling are far behind their counterparts in most developed countries. Potential human capital for cohorts of students is consequently lost.

Our results have obvious implications for both policy and research. Regarding the latter, there are few, if any, well-identified studies on how to effectively improve teacher knowledge and skills and the impact thereof (Glewwe and Muralidharan, 2015). Our results strongly suggest that this evidence gap is important to address.

Overall, our findings highlight the importance of improving teacher quality in low income countries. It is also easy to see, given our findings, how a vicious circle is created, with today's teachers having gone through an education system that does not prepare them adequately to be able to teach the next generation of students. Importantly, however, the continued rapid expansion of new teachers (see Bold et al, 2017)—two million new teachers are anticipated to be hired in the next 15 years in Sub-Saharan Africa alone—ought to provide ample opportunities to break this cycle. That is, although it may be costly, and difficult, to systematically and significantly raise the quality of the existing stock of teachers, a focus on how to ensure that the next cohort of teachers is better prepared to teach well, and rewarded for doing so when deployed, can potentially go a long way to improve outcomes. Related interventions, that either supplement current teachers with additional instructors, or automate certain aspects of teaching using computer-aided learning programs or scripted lesson plans, also show promising results (Murnane and Ganimian, 2014; Glewwe and Muralidharan, 2015; Evans and Popova, 2016).

References

Aaronson, Daniel, Lisa Barrow, William Sander. 2007. "Teachers and Student Achievement in the Chicago Public High Schools." *Journal of Labor Economics* 25 (1), pp. 95–135.

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja. 2009. "Do Value-Added Estimates Add Value? Accounting for Learning Dynamics, Harvard Kennedy School Faculty Research Working Papers Series, RWP09-034

ASER. 2013. *Annual Status of Education Report (Rural) 2013*. ASER Center. New Delhi.

Ashenfelter, Orley and David J. Zimmerman. 1997. "Estimates of the returns to schooling from sibling data: fathers, sons, and brothers." *The Review of Economics and Statistics* 79 (1), pp. 1–9.

Bau Natalee and Jishnu Das. 2016. "The Misallocation of Pay and Productivity in the Public Sector: Evidence from the Labor Market for Teachers." Working Paper, The World Bank.

Behrman, Jere R. 2010. "Investment in Education: Inputs and Incentives." In Dani Rodrik and Mark Rosenzweig, eds., *Handbook of Development Economics*, Vol. 5, Elsevier, pp. 4883–4975.

Bietenbeck, Jan, Marc Piopiunik, and Simon Wiederhold. 2017. "Africa's Skill Tragedy: Does Teachers' Lack of Knowledge Lead to Low Student Performance?" *Journal of Human Resources* (forthcoming).

Bold, Tessa, Deon Filmer, Gayle Martin, Ezequiel Molina, Brian Stacy, Christophe Rockmore, Jakob Svensson, and Waly Wane. 2017. "Enrollment Without Learning: Teacher Effort, Knowledge, and Skill in Primary Schools in Africa". *Journal of Economic Perspectives* (forthcoming).

Caselli, Francesco, and Wilbur John Coleman. 2006. "The World Technology Frontier." American Economic Review, 96(3): 499-522.

Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan and F. Halsey Rogers. 2006. "Missing in action: teacher and health worker absence in developing countries." *Journal of Economic Perspectives* 20:1, pp. 91–116.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood". *American Economic Review*, 104(9): 2633–2679.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2010. "Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects." *Journal of Human Resources*, 45, 655–681.

Dee, Thomas S. 2005. "A Teacher like Me: Does Race, Ethnicity, or Gender Matter?" *American Economic Review, Papers and Proceedings*, 95, 158–165.

Dee, Thomas S. 2007. “Teachers and the Gender Gaps in Student Achievement.” *Journal of Human Resources*, 42, 528–554.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2015. “School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from Kenyan primary schools.” *Journal of Public Economics*, Volume 123, pp: 92–110.

Duflo, Esther, Rema Hanna, and Stephen P. Ryan. 2012. “Incentives Work: Getting Teachers to Come to School.” *American Economic Review*, 102(4): 1241–78.

Evans, David K. and Anna Popova. 2016. “What Really Works to Improve Learning in Developing Countries: An Analysis of Divergent Findings in Systematic Reviews”. *World Bank Research Observer*. 31:2, pp. 242-270.

Glewwe, Paul, Michael Kremer. 2006. “Schools, Teachers, and Education Outcomes in Developing Countries”. In Eric A. Hanushek, Finis Welch (eds.), *Handbook of the Economics of Education*, Volume 2, pp. 945–1017. Amsterdam: North-Holland.

Glewwe, Paul and Karthik Muralidharan. 2015. “Improving School Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications.” RISE Working Paper No. 15/001.Glewwe and Muralidharan (2015)

Hanushek, Eric A. and Steven G. Rivkin. 2006. “Teacher Quality.” In Eric A. Hanushek, Finis Welch (eds.), *Handbook of the Economics of Education*, Volume 2, pp. 1051–1078. Amsterdam: North-Holland.

Johnson, David, Andrew Cunningham and Rachel Dowling (2012), Teaching Standards and Curriculum Review, mimeo, The World Bank.

Jones, Benjamin F. 2014. “The Human Capital Stock: A Generalized Approach.” *American Economic Review*, 104(11): 3752-77.

Kremer, Michael, Conner Brannen and Rachel Glennerster. 2013. “The Challenge of Education and Learning in the Developing World”, *Science* 340: 297-300.

Lavy, Victor. 2015. “Do Differences in Schools’ Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries.” *Economic Journal*, 125, F397–F424.

Malmberg, Hannes. 2017 “Human Capital and Development Accounting Revisited”, Working Paper, IIES.

Metzler, Johannes and Ludger Woessmann. 2012. “The Impact of Teacher Subject Knowledge on Student Achievement: Evidence from Within-Teacher Within-Student Variation.” *Journal of Development Economics*, 99, 486–496.

Muralidharan, Karthik and Venkatesh Sundararaman. 2011. “Teacher Performance Pay: Experimental Evidence from India”. *Journal of Political Economy*, 119, No. 1, pp. 39-77.

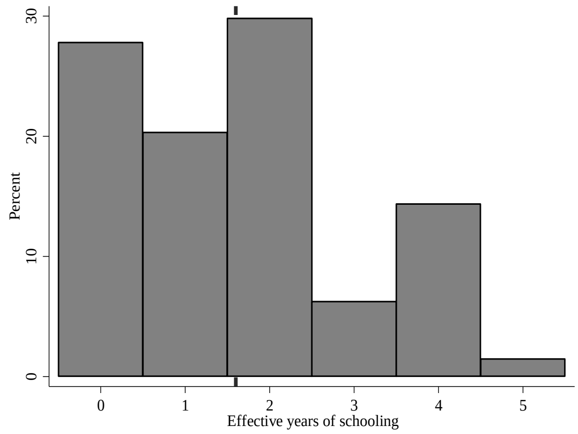
Murnane, Richard J. and Alejandro J. Ganimian. 2014. "Improving Educational Outcomes in Developing Countries: Lessons from Rigorous Evaluations." NBER Working Paper No. 20284.

Rivkin, Steven G., Eric A. Hanushek, John F. Kain. 2005. "Teachers, Schools, and Academic Achievement". *Econometrica* 73 (2), pp. 417–458.

Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data". *American Economic Review* 94 (2), pp. 247–252.

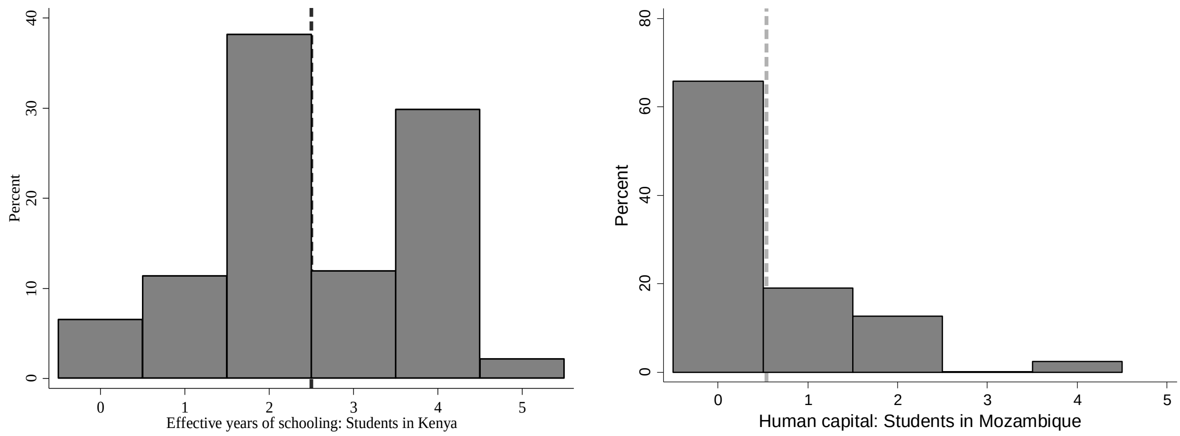
Petra E. Todd and Kenneth I. Wolpin, 2003, "On the Specification and Estimation of the Production Function for Cognitive Achievment", *Economic Journal*, 113, pp.F3-F33.

Figure 1: Effective years of schooling after four years of primary education



Note: Distribution of effective years of schooling for students (pooled data across countries and subjects). Dashed vertical line depicts mean.

Figure 2: Effective years of schooling after four years of primary education: Kenya and Mozambique

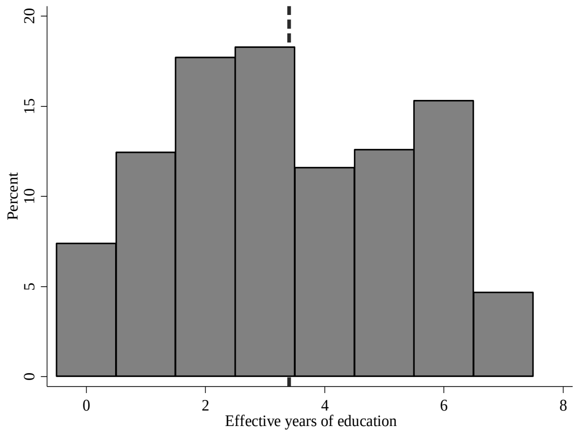


Panel A: Kenya

Panel B: Mozambique

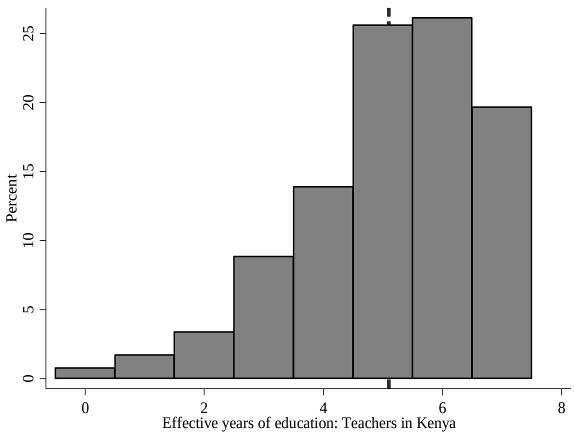
Note: Distribution of effective years of schooling for students in Kenya (Panel A) and Mozambique (Panel B). Pooled data across subjects for each country. Dashed vertical lines depict means.

Figure 3: Effective years of education for primary school teachers

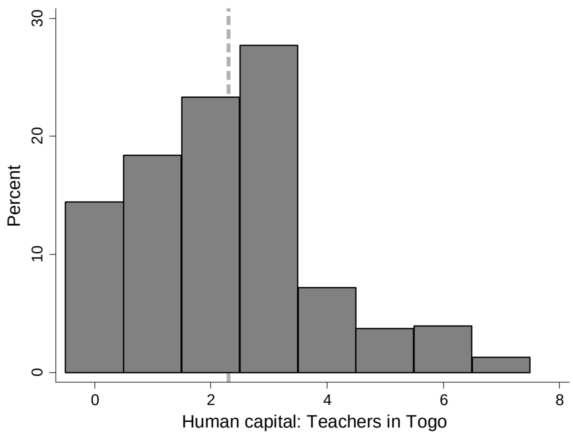


Note: Distribution of effective years of education for primary school teachers (pooled data across countries and subjects). Dashed vertical line depicts mean.

Figure 4: Effective years of education for primary school teachers: Kenya and Togo



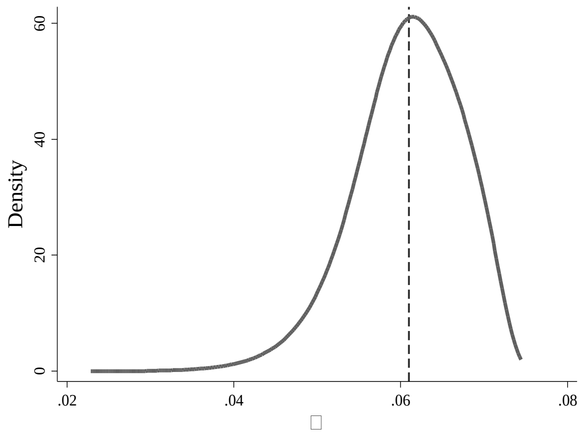
Panel A: Kenya



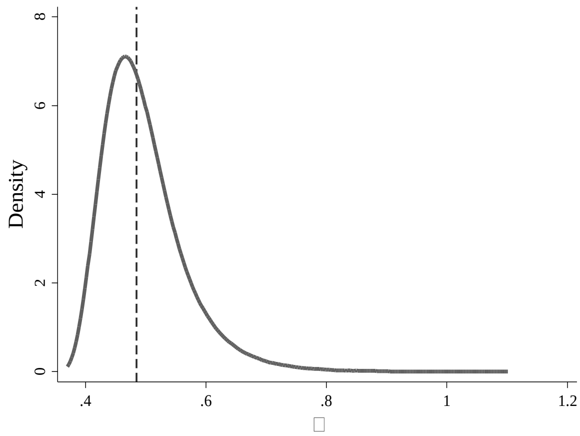
Panel B: Togo

Note: Distribution of effective years of education for primary school teachers in Kenya (Panel A) and Togo (Panel B). Pooled data across subjects for each country. Dashed vertical lines depict means.

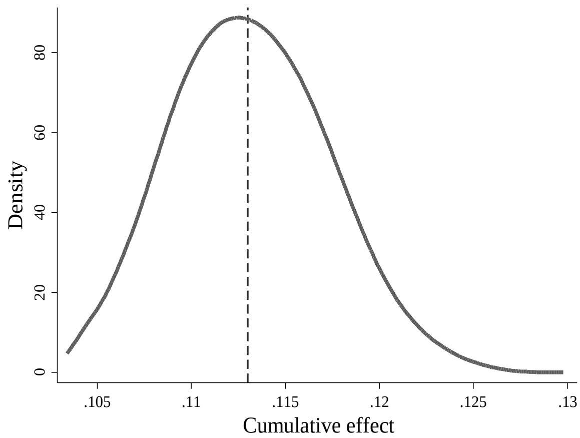
Figure 5: Probability density functions of the estimated α, γ, and the cumulative effects



Panel A: Contemporaneous effect



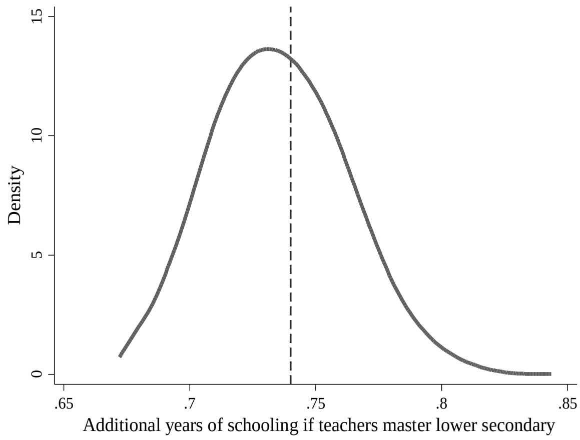
Panel B: Persistence



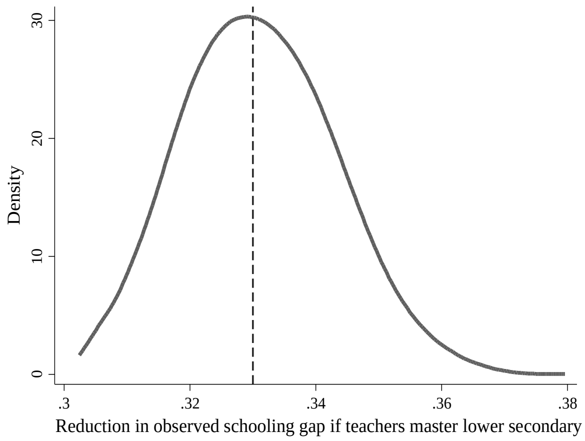
Panel C: Cumulative effect

Note: The figures show the distributions of the estimated contemporaneous effect of teacher knowledge on student learning, its persistence and the total cumulative effect after four years. Dashed vertical lines depict median values.

Figure 6: Probability density functions for cumulative total effect in policy simulation



Panel A: Additional years of schooling



Panel B: Reduction in schooling gap

Note: The figures show the distributions of the estimated total cumulative effect on student learning of increasing effective teacher knowledge from its current level to the level of lower secondary education after four years and the resulting reduction in students' learning gap. Dashed vertical lines depict median values.

Table 1: Summary statistics

	Mean
Absence from class (%)	44
Absence from school (%)	23
Scheduled teaching time (h min)	5h 27mins
Time spent teaching (h min)	2h 46mins
Minimum general pedagogy knowledge (%)	11
Minimum knowledge assessing students (%)	0

Note: See Bold et al. (2017) for details. Pooled data for Kenya, Mozambique, Nigeria, Senegal, Tanzania, Togo, and Uganda on teacher quality. All individual country statistics are calculated using country-specific sampling weights. The average for the pooled sample is taken by averaging over the country averages. Teachers are marked as absent from school if during an unannounced visit they are not found anywhere on the school premises. Otherwise, they are marked as present. Teachers are marked as absent from class if during an unannounced visit, they are absent from school or present at school but absent from the classroom. Otherwise, they are marked as present. The scheduled teaching time is the length of the school day minus break time. Time spent teaching adjusts the length of the school day by the share of teachers who are present in the classroom, on average, and the time the teacher spends teaching while in the classroom. A teacher is defined as having minimum knowledge of general pedagogy if she scores at least 80% on the tasks that relate to general pedagogy (factual text comprehension and being able to formulate learning outcomes and lesson aims). A teacher is defined as having minimum knowledge for assessing students if they score least 80% on the tasks that relate to assessment (comparing students' writing and monitoring progress among a group of students).

Table 2: Distribution of effective years of schooling for students

Effective years of schooling	Language	Mathematics
0	27%	36%
1	15%	21%
2	45%	18%
3	5%	6%
4	8%	17%
5	n/a	3%
N	23,884	23,016

Note: Distribution of effective years of schooling for students, calculated using country-specific sampling weights.

Table 3: Effective years of schooling, Item Response Scores (IRT), and average scores for students

	Language		Mathematics	
Effective years of schooling	Average	IRT	Average	IRT
0	12%	-1.19	24%	-0.99
1	23%	-0.56	43%	-0.06
2	58%	0.15	55%	0.49
3	93%	0.98	65%	0.92
4	98%	1.48	67%	0.94
5	-	-	83%	1.57
N	23,884	23,884	23,016	23,016

Note: Item Response Scores (IRT) and average scores for students, conditional on effective years of schooling, calculated using country-specific sampling weights.

Table 4: Distribution of effective years of education for teachers

Effective years of education	Language	Mathematics
0	8%	8%
1	7%	16%
2	19%	14%
3	22%	14%
4	18%	6%
5	23%	2%
6	2%	28%
7	3%	12%
N	4755	4970

Note: Distribution of effective years of education, calculated using country-specific sampling weights.

Table 5: Effective years of education, average scores and Item Response Scores (IRT) for teachers

	Language		Mathematics	
Effective years of education	Raw	IRT	Raw	IRT
0	20%	-1.57	4%	-1.99
1	33%	-0.94	26%	-0.77
2	42%	-0.38	40%	-0.31
3	49%	-0.02	49%	0.08
4	54%	0.33	68%	0.57
5	65%	0.88	78%	0.92
6	72%	1.10	91%	0.74
7	82%	1.51	89%	1.13
N	4755	4755	4970	4970

Note: Item Response Scores (IRT) and average scores for teachers, conditional on effective years of education, calculated using country-specific sampling weights.

Table 6: Relationship between student and teacher content knowledge

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Variable	Effective years of schooling: Students						
Effective years of education of current teacher	0.073***	0.089***	0.087***	0.067***	0.094***	0.027**	0.031*
	(.012)	(.014)	(.009)	(.013)	(.012)	(.013)	(.018)
Effective years of education of previous year's teacher				0.049***		0.047***	0.046***
				(.012)		(.012)	(.016)
Language			0.148***	0.148***	0.124***	0.167***	0.216***
			(.019)	(.026)	(.027)	(.012)	(.033)
Constant	2.005***	2.136***	2.062***	1.927***	2.033***	1.221***	1.029***
	(.086)	(.065)	(.065)	(.088)	(.083)	(.042)	(.042)
Observations	14,926	15,435	30,361	17,294	17,294	17,294	8,969
Adj. R-squared	0.131	0.146	0.136	0.132	0.129	0.497	0.524
Number of schools	1,432	1,817	1,974	1,503	1,503	1,503	626
Number of students	14,926	15,435	16,922	10,324	10,324	10,324	4,503
Subject	Math	Language	Both	Both	Both	Both	Both
Country FE	X	X	X	X	X		
Student FE						X	X
Same teacher in language and mathemathics							X

Note: Fixed effects specifications with clustered, by school, standard errors in parenthesis. *** 1% , ** 5% , * 10% significance.

Table 7: Specification tests

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable	Effective years of schooling: Students					
Effective years of education of current teacher	0.031* (.018)	0.033* (0.018)	0.030* (0.018)	0.025* (0.013)	0.017 (0.019)	0.021 (0.019)
Effective years of education of previous teacher	0.046*** (.016)	0.039** (0.017)	0.049*** (0.016)	0.048*** (0.012)	0.050*** (0.017)	0.050*** (0.017)
Teacher pedagogy score				0.239 (0.175)		
Constant	-0.216*** (0.033)	-0.216*** (0.030)	-0.216*** (0.033)	-0.165*** (0.026)	-0.181*** (0.035)	-0.186*** (0.036)
Specification/Sample	Main	District FE	Urban FE	Student FE	Rural	One grade 4 class
Observations	4,466	4,466	4,466	6,970	3,626	3,895
Number of schools	619	619	619	1037	501	537

Note: First difference (across subjects) specification, with clustered, by school, standard errors in parenthesis. Specification: (1) main specification with the sample of students with the same teacher in language and mathematics in a given year (grade 3 and 4); (2) main specification with subject variant district fixed effects; (3) main specification with subject variant urban fixed effects; (4) sample of all students with data on current and previous teacher score and current teacher pedagogy score; (5) main specification on sample of rural schools; (6) main specification on sample of schools with one grade 4 classroom. *** 1% , ** 5% , * 10% significance.

Table 8: Placebo tests

	(1)	(2)	(3)	(4)
Dep. Variable	Effective years of schooling: Students			
Effective years of education of current teacher	0.046 (0.034)	0.047 (0.034)	0.072*** (0.027)	0.043*** (0.016)
Effective years of education of previous teacher	0.072** (0.030)	0.074** (0.030)		
Effective years of education of higher grade teachers	0.005 (0.031)			
Constant	-0.277*** (0.066)	-0.275*** (0.066)	-0.117** (0.057)	-0.296*** (0.031)
Specification/Sample	Placebo	Comparison	Grade 3 & 4	Grade 4 only
Observations	1,201	1,201	1,773	3,901
Number of schools	167	167	277	525

Note: First difference (across subjects) specification, with clustered, by school, standard errors in parenthesis. Specification: (1) main specification controlling for teacher content knowledge (effective years of education) of higher grade teachers in school; (2) main specification using the same sample as in column (2); (3) Sample of students with the same teacher in both subjects in both years (grade 3 and grade 4); (4) Sample of students with a new teacher teaching both subjects in grade 4. *** 1% , ** 5% , * 10% significance.

Table 9: Structural parameters

	(1)	(2)	(3)
Panel A: Effect of teacher test score on student test score (effective years of learning)			
Contemporaneous effect (α)	0.058*	0.057*	0.057*
	(0.033)	(0.033)	(0.033)
Persistence (γ)	0.524	0.479	0.588
	(0.608)	(0.430)	(0.585)
Total effect after 4 years	0.113***	0.103***	0.121***
	[0.000]	[0.000]	[0.000]
Panel B: Effect of teacher test score on student test score (IRT scores)			
Contemporaneous effect (α)	0.065	0.063	0.063
	(0.045)	(0.045)	(0.045)
Persistence (γ)	0.502	0.467	0.551
	(0.773)	(0.737)	(0.737)
Total effect after 4 years	0.123***	0.111***	0.126***
	[0.001]	[0.002]	[0.001]
Specification			
Median across all correlation structures	X		
Same teacher (1-3) structures		X	
AR(1) process			X

Note: Estimates of the contemporaneous effect (α) and persistence parameter (γ) under different scenarios varying assumptions about the correlation between teacher knowledge in grade 3 and 4 and previous teachers' knowledge. Panel A reports estimates using effective years of education of teachers (explanatory variables) and effective years of learning of students (dependent variable). Panel B: Estimates using IRT scores. Specifications: Column (1): Estimates for α and γ at the median cumulative effect across all correlation structures; Column (2): Students are assumed to be taught by the same teacher in lower primary (i.e. grade 1, 2 and 3); Column (3): Correlation coefficient between any two grades t and t' is $\rho^{|t-t'|}$. Total (cumulative) effect after 4 years is $\alpha \sum_{t=1}^{4} \gamma^{t-1}$. Clustered, by school, standard errors in parenthesis, with p-value in brackets giving the probability of the null hypothesis that the cumulative effect is zero.

Appendix:

A. Definition of curriculum-adjusted years of human capital

We define a student to have 0 years of human capital in language, if they cannot read three letters. A student is defined as having one year of human capital in language, if they can read three letters, but cannot do more advanced tasks. They are scored as having two years of human capital in language, if they can read three words, but cannot do any more advanced tasks. They are scored as having accumulated three years of human capital if they have basic vocabulary, can read a sentence, half a paragraph and answer a basic comprehension question, but cannot do more advanced tasks. Finally, they are scored as having four years of human capital if they can read the whole paragraph and answer an advanced comprehension question.

In mathematics, we score a student as having zero years of human capital if they cannot recognize numbers or cannot do single digit addition or cannot do single digit subtraction. We score them as having one year of human capital if they can recognize numbers, do single digit addition and single digit subtraction, but not any of the more advanced tasks. We score them as having two years of human capital if they can perform double digit addition, triple digit addition and order numbers between 0 and 999. We class them as having three years of human capital if they can multiply single digits, divide single digits and do double digit subtraction. We class them as having accumulated four years of human capital if they can divide double by single digits and compare fractions and as having five years of human capital if they can multiply double digits.

On the teacher side, we score teachers as having no years of human capital, if they could not answer the simplest grammar question, namely forming a question with "Where is…?" and using 'who' in order to define what person is doing. We scored them as having one year of human capital if they could formulate such a question, but could not do any of the more advanced material. We scored them as having two years of human capital if they could use 'when' as a conjunction, could form a sentence asking 'how much' and used 'which' correctly. We scored them as having three years of human capital if they could use because and so correctly as conjunctions, and we scored them as having four years of human capital if they could form a sentence with a conditional statement, use past passive and use unless correctly. We score them as having 5 years of human capital if they could complete more than 70% of an unprompted Cloze passage, as six years if they could correct more than 70% of the mistakes in a letter written by a fourth grade students and as seven years if they could complete both these tasks.

For mathematics, we score teachers as having 1 year of human capital if they could not add double digits (without borrowing). We score them as having one year of education if they could add double digits (without borrowing), but could not do any of the more advanced tasks. We scored them as having two years of human capital if they could add triple digits and recognize basic geometric shapes, but could not do any of the more advanced tasks. We score them as having three years of human capital if they could subtract double digits (with borrowing), and divide a double digit by a single digit. We scored them as having four years of human capital if they could add decimals, solve a multiplication problem involving monetary unity, subtract decimals. We scored them as having five years of human capital if they could multiply double digits, manipulate fractions and solve a problem involving units of time. We scored them as having six years of human capital if they could solve square roots up to twelve, solve for an unknown in an algebraic equation. We scored them as having seven years of human capital if they could analyze data in a graph, divide fractions, and calculate the perimeter and area of a rectangle.

B. *Expression for the OLS estimator of β_3 and β_4*

The OLS estimators of β_4 and β_3 in equation (4) are

(A1)
$$\hat{\beta}_4 = \frac{1}{1-\frac{\sum \Delta\tilde{x}_3\Delta\tilde{x}_4}{\sum \Delta\tilde{x}_3^2}\frac{\sum \Delta\tilde{x}_3\Delta\tilde{x}_4}{\sum \Delta\tilde{x}_4^2}}\left(\frac{\sum \Delta\tilde{x}_4\Delta\tilde{y}}{\sum \Delta\tilde{x}_4^2} - \frac{\sum \Delta\tilde{x}_3\Delta\tilde{x}_4}{\sum \Delta\tilde{x}_4^2}\frac{\sum \Delta\tilde{x}_3\Delta\tilde{y}}{\sum \Delta\tilde{x}_3^2}\right),$$

and

(A2)
$$\hat{\beta}_3 = \frac{1}{1-\frac{\sum \Delta\tilde{x}_3\Delta\tilde{x}_4}{\sum \Delta\tilde{x}_3^2}\frac{\sum \Delta\tilde{x}_3\Delta\tilde{x}_4}{\sum \Delta\tilde{x}_4^2}}\left(\frac{\sum \Delta\tilde{x}_3\Delta\tilde{y}}{\sum \Delta\tilde{x}_3^2} - \frac{\sum \Delta\tilde{x}_3\Delta\tilde{x}_4}{\sum \Delta\tilde{x}_4^2}\frac{\sum \Delta\tilde{x}_4\Delta\tilde{y}}{\sum \Delta\tilde{x}_4^2}\right),$$

where a tilde above the variable denotes a demeaned variable. To find their probability limits, substitute the true model for $\Delta\tilde{y}$ from equation (5), divide each of the summed terms by N, the number of observations, and let N go to infinity.

C. *Inference*

The asymptotic variance-covariance matrix of the parameters $\hat{\theta} = \{\hat{\alpha},\ \widehat{\alpha\gamma},\ \widehat{\alpha\gamma^2},\ \widehat{\alpha\gamma^3}\}$ in the structural model (5) is given by

(A3)
$$V = \sigma_\epsilon^2 E(\Delta\tilde{x}'\Delta\tilde{x})^{-1}/N = \sigma_\epsilon^2/N \begin{pmatrix} \sigma^2_{\Delta x_4} & \rho_{4,3}\sigma^2_{\Delta x_3} & \rho_{4,2}\sigma^2_{\Delta x_2} & \rho_{4,1}\sigma^2_{\Delta x_1} \\ \rho_{4,3}\sigma^2_{\Delta x_3} & \sigma^2_{\Delta x_3} & \rho_{3,2}\sigma^2_{\Delta x_2} & \rho_{3,1}\sigma^2_{\Delta x_1} \\ \rho_{4,2}\sigma^2_{\Delta x_2} & \rho_{3,2}\sigma^2_{\Delta x_2} & \sigma^2_{\Delta x_2} & \rho_{2,1}\sigma^2_{\Delta x_1} \\ \rho_{4,1}\sigma^2_{\Delta x_1} & \rho_{3,1}\sigma^2_{\Delta x_1} & \rho_{2,1}\sigma^2_{\Delta x_1} & \sigma^2_{\Delta x_1} \end{pmatrix}^{-1},$$

where σ_ϵ^2 is the error variance, $\Delta\tilde{x} = \{\Delta\tilde{x}_4, \Delta\tilde{x}_3, \Delta\tilde{x}_2, \Delta\tilde{x}_1\}$ is the matrix of de-meaned test scores (differenced across subjects) in each year, and $\sigma_{\Delta x_t}^2$ is the population variance of $\Delta x_t\ \forall t = 1,\ldots,4$.

We now show that each of the terms in the product can be written as a function of known population moments. Consider first $E(\Delta\tilde{x}'\Delta\tilde{x})$: After imposing the usual stationarity assumption on the test score differences across subjects, we can write this as

(A4) $$E(\Delta\tilde{x}'\Delta\tilde{x}) = \sigma_{\Delta x_4}^2 \begin{pmatrix} 1 & \rho_{4,3} & \rho_{4,2} & \rho_{4,1} \\ \rho_{4,3} & 1 & \rho_{3,2} & \rho_{3,1} \\ \rho_{4,2} & \rho_{3,2} & 1 & \rho_{2,1} \\ \rho_{4,1} & \rho_{3,1} & \rho_{2,1} & 1 \end{pmatrix},$$

which can be estimated by replacing $\sigma_{\Delta x_4}^2$ and $\rho_{4,3}$ by their sample analogues $\frac{\sum_i \Delta\tilde{x}_{i,4}{}^2}{N}$ and $\hat{\rho}$, the coefficient estimate from equation (8), and setting the remaining $\rho_{t,t'}$ equal to their values in the particular parameterization under consideration.

Next consider the error variance, σ_ϵ^2. If we could observe $\Delta\tilde{x}_2$ and $\Delta\tilde{x}_1$, then we could simply estimate this as $\widehat{\sigma_\epsilon^2} = \frac{1}{N}\sum\left(\Delta\tilde{y}_i - \hat{\alpha}\,\Delta\tilde{x}_{4i} - \widehat{\alpha\gamma}\,\Delta\tilde{x}_{3i} - \widehat{\alpha\gamma^2}\,\Delta\tilde{x}_{2i} - \widehat{\alpha\gamma^3}\,\Delta\tilde{x}_{1i}\right)^2$. However, since we do not observe the test scores in earlier grades, we relate σ_ϵ^2 to σ_μ^2, the error variance in the reduced form model, which we can estimate.

Writing

(A5) $$\hat{\mu} = \Delta\tilde{y} - \hat{\beta}_4\,\Delta\tilde{x}_4 - \hat{\beta}_3\Delta\tilde{x}_3,$$

and adding and subtracting $\epsilon = \Delta\tilde{y} - (\alpha\Delta\tilde{x}_4 + \alpha\gamma\Delta\tilde{x}_3 + \alpha\gamma^2\Delta\tilde{x}_2 + \alpha\gamma^3\Delta\tilde{x}_1)$, gives

(A6) $$\hat{\mu} = \epsilon + (\alpha - \hat{\beta}_4)\,\Delta\tilde{x}_4 + (\alpha\gamma - \hat{\beta}_3)\,\Delta\tilde{x}_3 + \alpha\gamma^2\,\Delta\tilde{x}_2 + \alpha\gamma^3\Delta\tilde{x}_1\,.$$

Hence, we get

(A7) $$\text{plim}\frac{\hat{\mu}'\hat{\mu}}{N} = \sigma_\mu^2 = \sigma_\epsilon^2 + B_4^2\sigma_{\Delta x_4}^2 + B_3^2\sigma_{\Delta x_3}^2 + \alpha^2\gamma^4\sigma_{\Delta x_2}^2 + \alpha^2\gamma^6\sigma_{\Delta x_1}^2 + 2B_4B_3\rho_{4,3}\sigma_{\Delta x_3}^2 - 2B_4\alpha\gamma^2\rho_{4,2}\sigma_{\Delta x_2}^2 - 2B_4\alpha\gamma^3\rho_{4,1}\sigma_{\Delta x_1}^2 - 2B_3\alpha\gamma^2\rho_{3,2}\sigma_{\Delta x_2}^2 - 2B_3\alpha\gamma^3\rho_{3,1}\sigma_{\Delta x_1}^2 + 2\alpha^2\gamma^5\rho_{2,1}\sigma_{\Delta x_1}^2,$$

where we have used the fact that $\text{plim}\frac{1}{N}\sum \epsilon_i\,(\alpha\Delta\tilde{x}_{4i} + \alpha\gamma\Delta\tilde{x}_{3i} + \alpha\gamma^2\Delta\tilde{x}_{2i} + \alpha\gamma^3\Delta\tilde{x}_{1i}) = 0$ by assumption, and $B_4 = \alpha\gamma^2\left(\frac{\rho_{2,4}-\rho_{2,3}\rho}{1-\rho^2}\right) + \alpha\gamma^3\left(\frac{\rho_{1,4}-\rho_{1,3}\rho}{1-\rho^2}\right)$ and $B_3 = \alpha\gamma^2\left(\frac{\rho_{2,3}-\rho_{2,4}\rho}{1-\rho^2}\right) + \alpha\gamma^3\left(\frac{\rho_{1,3}-\rho_{1,4}\rho}{1-\rho^2}\right)$ are the asymptotic biases on the OLS estimates of $\hat{\beta}_4$ and $\hat{\beta}_3$.

Once again imposing the stationarity condition on the test score differences across subjects, the asymptotic error variance is,

(A7) $$\sigma_{\epsilon}^2 = \sigma_{\mu}^2 - \big(B_4^2 + B_3^2 + \alpha^2\gamma^4 + \alpha^2\gamma^6 + 2B_4B_3\rho_{4,3} - 2B_4\alpha\gamma^2\rho_{4,2} - 2B_4\alpha\gamma^3\rho_{4,1} - 2B_3\alpha\gamma^2\rho_{3,2} - 2B_3\alpha\gamma^3\rho_{3,1} + 2\alpha^2\gamma^5\rho_{2,1}\big)\sigma_{\Delta x_4}^2.$$

This can be estimated by replacing σ_{μ}^2 with the mean of the sum of squared residuals, $\frac{\hat{\mu}'\hat{\mu}}{N}$, in the reduced form model, $\sigma_{\Delta x_4}^2$ with $\frac{\sum_i \Delta\tilde{x}_{i,4}{}^2}{N}$, $\rho_{4,3}$ with $\hat{\rho}$, and setting the remaining $\rho_{t,t'}$ equal to their values in the particular parameterization under consideration.

With an estimate of the asymptotic variance-covariance matrix, $\hat{V}$, we can then compute the usual Wald statistic of the hypothesis that the cumulative effect of teacher knowledge on student learning is zero. Second, we can compute the standard error on the persistence parameter by applying the delta method to, for instance, the ratio of $\widehat{\alpha\gamma}$ and $\hat{\alpha}$.

In all calculations, we adjust for the fact that test scores are measured half way through year 4, while the structural coefficient of interest, α, measures the effect of one year of schooling. We also adjust parametrically for clustering of the standard errors at the school level by multiplying the estimate of the asymptotic variance-covariance matrix by the Moulton factor (which equals 2.35 in our data set).