Primary Education Expansion and Quality of Schooling

Christine Valente*

University of Bristol and IZA

Abstract

The rapid increases in enrollment seen in many developing countries might further worsen the poor schooling quality found in these countries. I estimate the effect of enrollment growth following the removal of primary school fees in Tanzania and find evidence of a sizeable increase in pupil-teacher ratios and a reduction in observable teacher quality, but rule out a substantial effect on test scores overall. These results are robust to instrumenting enrollment growth using predetermined fertility and migration decisions, and to a number of checks including the use of baseline enrollment rates as an alternative source of variation in enrollment growth. However, when investigating the possibility of heterogeneous effects for urban and rural areas, I find evidence of a deterioration of test scores in urban areas.

JEL codes: I21; I28; O15

Keywords: Universal Primary Education; Pupil-Teacher Ratio; Test Scores; Tanzania

---

*Department of Economics, University of Bristol, Priory Road Complex, Priory Road, Bristol, BS8 1TU, U.K., christine.valente@bristol.ac.uk. Thanks go to Ben Kett for excellent research assistance. I also thank the Bureau of Statistics, Tanzania for providing the census data making this research possible. For their useful comments and suggestions, I thank three anonymous reviewers, Doug Almond, Erlend Berg, Simon Burgess, Damon Clark, Robert Garlick, Ellen Greaves, Sami Stouli, Frank Windmeijer, Yanos Zylberberg, as well as seminar participants at the Universite Catholique de Louvain, ISER (University of Essex), Institute for Fiscal Studies (EDePo), the University of Manchester, Universidade NOVA de Lisboa, the 2014 Royal Economic Society Conference and 2014 Econometric Society European Meeting. All errors are my own. Declarations of interest: none.

1 Introduction

After over a decade of stagnation, the net enrollment rate in Tanzanian primary schools went from 53% in 2000 to 73% in 2002 (World Development Indicators 2015), coinciding with the removal of primary school fees announced in 2001. This increase was driven mainly by an 82% increase in the number of children enrolled in Grade 1.[1] Tanzania is no outlier–a sizeable number of Sub-Saharan African countries have recently experienced large, sudden increases in enrollment, which often followed Free Primary Education (FPE) policies (World Bank 2009).[2] And with 63 Million children of primary school age still out of school in the world, over half of whom in sub-Saharan Africa (UNESCO Institute for Statistics n.d.), more countries may follow suit.

However, there is evidence that, more than years of education accumulated, it is the cognitive skills acquired during schooling that matter for both individual outcomes on the labor market and for macroeconomic growth (Hanushek & Woessmann 2008). In addition, there is mounting evidence of the "often abysmal" quality of schooling provision in developing countries (Kremer et al. (2013), p.297). The levels of learning in many developing countries are so low, that there has recently been calls for a complete rethink of education systems across the developing world amidst a "learning crisis" caused not least by the international donor community's narrow focus on education as school enrollment (Pritchett 2013). It is therefore important to understand the consequences of accelerated enrollment growth, through FPE or other measures such as school construction programmes, on the quality of the learning environment.

Despite considerable concern about this issue in policy circles, there is no arguably causal evidence on the effect of the very large increases in enrollment seen in a number of countries on the test scores of the cohorts entering school at the time of expansion, whose learning would be most affected by the likely decrease in educational inputs per capita.[3] In addition to filling this gap, I shed light on the impact of this rapid primary schooling expansion on a rich set of measures of educational inputs, thus contributing to our existing knowledge on the effect on

[1] Author calculations based on figures reported in Ministry of Education [Tanzania] (1999-2007). In absolute terms, the number of pupils enrolled in the first grade of primary school went from 894,894 in 2000 (before FPE was announced) to 1,628,195 in 2002, the first academic year after it was announced.

[2] Figure A-1 illustrates the magnitude of the increases in enrollment that followed in a selected sample of countries.

[3] In the case of Kenya, Lucas & Mbiti (2012) estimate the effect of FPE on test scores of pupils who were in Grades 4 to 8 at the time of FPE, while enrollment growth beyond Grade 1 was limited (13% growth in grades 2 to 8). See Sections 2.2 and A-1 for further details.

test scores of schooling inputs such as pupil-teacher ratios and teacher characteristics in a poor country setting.

The removal of primary school fees took place simultaneously across Tanzania. But there is variation in the subsequent rate of enrollment growth across regions, which I first exploit using a difference-in-differences approach. Test scores are only available at two points in time (2000 and 2007), which prevents me from testing for differences in pre-existing trends. Instead, in order to address concerns regarding pre-existing trends as well as other potential sources of endogeneity of enrollment growth, I note that regions whose post-reform primary-school age population was larger relative to the pre-reform school-age population experienced larger primary enrollment growth rates. Therefore, schools in these areas experienced larger increases in the demand for primary education, independently of the potentially endogenous response of the regional enrollment rate to the school fee reform. Instrumental variable estimation exploiting this source of arguably exogenous variation, which is based on past fertility and past migration decisions, bolsters the causal interpretation of the effect of enrollment growth on schooling inputs and cognitive skills acquisition.

My main conclusion is that there was no substantial decrease in test scores overall. I also find that primary enrollment growth has led to sizeable increases in the pupil-teacher ratio (an increase by 6.9 pupils for an increase in enrollment growth by one standard deviation) and a worsening of average teacher experience and subject-specific knowledge. Point estimates of the effect of enrollment growth on pupil test scores are small in magnitude and statistically insignificant. The lower bounds of the 95% confidence intervals imply that an increase in enrollment growth by 1 standard deviation led at most to a decrease in test scores of 0.15-0.16 standard deviation, which corresponds to about a fourth (third) of the rural-urban gap in language (math), or roughly a fifth (fourth) of the gap between children of fathers with more than primary schooling and the children of fathers who did not complete primary schooling. In other words, I cannot rule out some deterioration in the quality of the learning environment for the average pupil at the national level, but I can rule out a substantial worsening of quality overall.

The main message is therefore one of cautious optimism with respect to the possibility of broadening rapidly and comprehensively access to primary education without worsening schooling quality. However, when investigating the possibility of heterogeneous effects for urban and rural areas, I find evidence of a deterioration of test scores in urban areas, so that whether or not enrollment growth was welfare-enhancing depends on how the gains of the many winners are weighted against the losses of the (fewer) losers.

The rest of the paper is organised as follows. Section 2 gives an overview of schooling expansion in Tanzania and summarizes the existing international evidence on the effect of rapid enrollment growth, Section 3 presents the identification strategy, Section 4 describes the data, Section 5 presents the main results, Section 6 explores the robustness of my findings, and Section 7 investigates the effect of enrollment growth across the distribution of test scores and by rural or urban location. Section 8 concludes.

2 Schooling Expansion in Tanzania and Previous Literature

2.1 Schooling Expansion in Tanzania

Primary education in Tanzania comprises 7 years (Standard I-VII), with a normal entry age of 7 years old. Throughout the 1990s, only about half of primary-school age children (i.e., aged 7-13) attended school (with annual net enrollment rates varying between 49 and 51%). This was despite early attempts at achieving universal primary education in the late 1970s, which culminated in a net enrollment rate of 70% in 1980 (World Development Indicators 2015). A decline in the quality of education induced by this first attempt at universal primary education has been blamed for part of the subsequent decline in enrollment, not least due to the need to recruit less qualified teachers (Wedgwood 2007). When the government decided to remove primary school fees starting in January 2002, some lessons from the past seemed to have been learnt. In order to help manage the absorption of new entrants into the school system, the government plan stipulated that "admission priority" should be given to children who are seven years old, with older children being admitted at the discretion of the school committee (Basic Education Development Committee (2001), p.5). In addition, similar to other countries that implemented free primary education, a donor-funded capitation grant of US$10 (9,000 Tanzanian Shillings) per pupil enrolled was introduced to cover *non-salary costs* in order to compensate for the loss of revenue from user fees. While US$4 of the grant were ring-fenced for the purchase of textbooks and other teaching and learning materials, the remaining of the grant was expended at the discretion of the school committee for other non-salary costs (Basic Education Development Committee 2001). In practice, the average per capita grant received by schools is believed to have been US$6 in 2002-2003, and US$4.7 in 2007/2008 (Twaweza 2010). As a point of comparison, expenditure data from National Bureau of Statistics (2001) suggests mean (median) primary school fees of US$4.6 (US$3.3).[4] The replacement of fees by the capitation grant may therefore have, at first, slightly increased the average school's ability to cover *non-salary* costs.

Contrary to the physical inputs covered by the capitation grant, however, the number of teachers per pupil and the average quality of their training should have decreased with the steep growth in student numbers. In order to help accommodate for the expected increases in enrollment, teacher training programs were shortened from two years of academic training to one year academic training plus one year of practice with supervised on-the-job training. In a study produced for the UK Department For International Development, the authors express concerns that "the previous two-year curriculum has been crammed into one year, (which means that there is insufficient subject content)" (Bennell & Mukyanuzi (2005), p.19).[5]

Other relevant institutional features of the Tanzanian primary school system are as follows. Teachers are recruited by the Local Government Authorities (LGA), which are responsible for providing primary education. Most of the LGAs' budget is made up of central government transfers, and salary payments are made directly to teachers by the Ministry of Finance and Economic Affairs (MoEVT and UNESCO 2012). Despite an increase in the share of primary school pupils attending private schools, the share of the private sector is negligible at primary level with 1% in 2007, up from 0.12% in 2001 (own calculations based on 2003 and 2007 editions of Ministry of Education [Tanzania] (1999-2007)).

All in all, the Tanzanian primary school system had to absorb a near-doubling of the number

[4]These figures were obtained by restricting the 2001 Household Budget Survey sample to households with individuals enrolled in primary school only, dividing annual household expenditure on school fees by the number of children enrolled, and removing outliers (defined as the top 1% expenditure on school fees).

[5]Contrary to other Sub-Saharan African countries facing large increases in enrollment (Bourdon et al. 2010), new teachers were not enrolled in fixed-term contracts in Tanzania. There may well have been motivational changes related to the increase in enrollment, for which there is no "hard" data, but qualitative work by Bennell & Mukyanuzi (2005) suggest that, while low, levels of teacher motivation have not systematically deteriorated with the expansion of primary enrollment (p.11).

of pupils within a few years–more precisely, there was an 81% increase between 2000 and 2006 (author's calculations based on figures reported in the 2003 and 2007 editions of Ministry of Education [Tanzania] (1999-2007)). Resources were made available to schools in order to cover non-salary costs, but an expansion on this scale was unlikely to be met with a commensurate increase in the supply of equally qualified teachers.

2.2 Summary of the Previous Literature

An abundant literature has documented the substantial effect of FPE on *enrollment* (Deininger (2003), Grogan (2009), Nishimura et al. (2008) for Uganda; Al-Samarrai & Zaman (2007) for Malawi; Lucas & Mbiti (2012) and Bold et al. (2015) for Kenya; and Hoogeveen & Rossi (2013) for Tanzania). The study by Lucas & Mbiti (2012) goes further by providing a rich picture of the consequences of FPE in Kenya, shedding light on its effects on sorting between public and private schools, and providing estimates of the effect of FPE in Kenya on the end-of-Grade 8 test scores of pupils who were in Grades 4 to 8 at the time of the removal of the school fees (in the absence of data for pupils who were in earlier grades at that time). Enrollment growth in Grades 2-8 was limited at 13% (Oketch & Somerset 2010), and pupils in Grades 4 to 8 at the time of the reform were only exposed to any deterioration of quality post-FPE during a few years before being tested. Still, Lucas & Mbiti (2012) interestingly find that above-average predicted intensity in exposure to FPE had, at most, small negative effects on scores at the end of primary school.

In previous work focussing on Tanzania, Hoogeveen & Rossi (2013) estimate the impact of FPE on school attendance and grade completion. Their household data confirm that enrollment at age 7 is more likely and less predicted by socioeconomic status in 2007 than in 2001. However, comparing years of education accumulated between 2001 and 2007 between children aged 8 to 10 in 2002, who are considered "treated", and children aged 10, whose enrollment was not prioritized by the reform, Hoogeveen & Rossi (2013) find a statistically significant decrease in grade attainment, which they hypothesize to be due to a deterioration of the quality of schooling.

A detailed literature review, covering studies of the effect of enrollment growth on education outcomes as well as the evidence on the effect of schooling inputs on test scores, is provided in

Section A-1.

3 Identification Strategy

3.1 Conceptual Framework

To fix ideas, consider the general achievement production function relating test scores at age a to all prior investments in child i in household j at age a (Todd & Wolpin 2007):

$$A_{ija} = A_a(Z_{ij}(a), \mu_{ij0}) \tag{1}$$

where $Z_{ij}(a)$ is the vector of all inputs having entered the achievement production function of individual i at any time until age a. This includes parental investments, environmental factors (including peers), quantity of schooling, and teacher and non-teacher school inputs (e.g., pupil-teacher ratio, teacher quality, physical inputs). μ_{ij0} denotes the child's cognitive and non-cognitive endowment. Large, sudden increases in enrollment following FPE may affect a number of inputs, and thus affect test scores.

We can distinguish two mechanisms, one working through changes in the test scores of inframarginal students (i.e., students who would have been enrolled in school even if enrollment had stayed constant), and another one working through changes in the composition of students but leaving the achievement of inframarginal students unaffected. Only decreases in test scores resulting from the first of these mechanisms would denote a worsening of the quality of the schooling environment. Such worsening of quality could come about for several reasons. First, enrollment growth is likely to increase pupil-teacher ratios and decrease the quality of the average teacher in terms of teacher training and experience. A change in average teacher quality may arise for a number of reasons, including: the mechanical decrease in teacher experience due to the need to hire more teachers, decreased selection in the recruitment of teacher trainees and shorter teacher training to meet the increased demand, and possibly increased turnover. The effect on non-teacher school inputs is less clear *a priori* because the increase in enrollment was accompanied by a capitation grant targeted at non-salary expenditure. Finally, the marginal student is likely to have lower socioeconomic status (SES), and may thus have a worse cognitive and non-cognitive endowment than previous students, which may lead to negative peer effects on the performance of the inframarginal students. Even when the performance of inframarginal students is not significantly affected by the increase in enrollment (and hence there is no decrease in the quality of schooling), we may observe a worsening of average test scores among enrolled students through a composition effect, if the marginal student has lower ability μ_{ij0} than the average inframarginal student.

3.2 Difference-in-Differences Approach

In order to answer the question of whether rapid enrollment growth worsened the quality of schooling, I first estimate the effect of primary enrollment growth on an observable set of schooling inputs Z_{ij} observed while the pupil is in Grade 6, and then estimate its overall effect on achievement in Grade 6, as captured by test scores in Kiswahili and mathematics. The baseline identification strategy relies on a comparison of changes in schooling inputs or test scores between 2000 and 2007 across regions that experienced different rates of growth in primary enrollment. More precisely, I estimate the following equation using the 2000 and 2007 SACMEQ surveys described in Section 4, in which a measure of quality of inputs or outcomes y_{irt} is regressed on a survey dummy ($\mathbb{1}(t = 2007)_t$), a set of region dummies (R_r), individual and regional (time-varying) controls (X_{irt}), and the interaction between the 2007 survey dummy and the size of post-reform enrollment (cumulated over 2002-2007) relative to baseline enrollment:

$$\begin{aligned} y_{irt} =& \beta_0 + \beta_1 \left(\frac{post\_enrol}{baseline\_enrol}\right)_r \times \mathbb{1}(t = 2007)_t \\ &+ \mathbb{1}(t = 2007)_t + R_r' \beta_r + X_{irt}' \beta_X + \epsilon_{irt} \end{aligned} \tag{2}$$

where $baseline\_enrol_r$ is the number of pupils enrolled in primary schools in region r in 2001 (as enrollment statistics broken down by region are not available for 2000 and most previous years) and $post\_enrol_r = \sum_{j=2002}^{2007} E_{rj}$ is the sum of the number of pupils enrolled in primary schools in region r during 2002-2007, the years during which the Grade 6 students of 2007 should have been in primary school and which coincides with the post-FPE period. I focus on the cumulative effect of exposure to larger school cohorts from Grade 1 to Grade 6 in order to reflect the cumulative nature of learning illustrated in Equation 1. From here onwards, I refer to $(\frac{post\_enrol}{baseline\_enrol})_r$ as "enrollment growth" in region r, which can be thought of as a continuous measure of treatment intensity in a difference-in-differences setting.[6]

The ratio $(\frac{post\_enrol}{baseline\_enrol})_r$ does not vary over time within region, and is therefore subsumed in the regional dummies R_r, which capture any baseline difference in the outcome variable that is specific to each region and constant over the two survey years.

Standard errors are clustered at the regional level to allow for an intra-region error correlation structure of an arbitrary nature. All regressions are weighted using the pupil weights provided in the dataset. Given the comparatively small number of regions (19), I also report p-values based on the wild cluster bootstrap-t procedure recommended by Cameron et al. (2008).

If the enrollment growth rate is not correlated with omitted variables that also affected changes in schooling quality or other inputs in the achievement production function, then a simple OLS estimation of β_1 in Equation 2 will yield the causal effect of an increase in the growth rate of primary school enrollment on schooling quality. There are a number of reasons why one may expect the enrollment growth rate not to be exogenous, however. Some potential sources of bias can be controlled for directly. Less developed regions may have experienced larger enrollment growth and been increasingly targeted over time by government transfers, which could bias β_1 towards less negative values if "more of the same" resources increased test scores (which evidence suggest they do not, Kremer et al. (2013)). I address this concern in a robustness check in which I control for the growth in discretionary government transfers to the region. As previously mentioned, FPE-led enrollment growth may lead to the recruitment of less able students. This compositional effect would result in β_1 being an *overestimate* of the worsening of quality. In order to address this concern, I check the robustness of my findings to controlling for the following observable pupil characteristics: age, gender, whether English is

[6] An alternative would have been to use as denominator the number of pupils enrolled in primary schools in region r during 1995-2000 instead of the number of pupils enrolled in 2001. This is impossible since enrollment data broken down by region is not available for the 1995-2000 period. Another possibility would have been to replace the numerator ($post\_enrol$) with the number of pupils enrolled in primary schools in region r during 2007 only, but this goes against the idea presented in the conceptual framework of Section 3.1 that achievement at age a depends on all inputs having entered the achievement production function of individual i at any time until age a. Different dynamics such as different drop-out rates across regions may have led to potentially large differences in the variation captured by the treatment variable depending on the choice of numerator between $\sum_{j=2002}^{2007} E_{rj}$ and E_{r2007}. In practice, however, there is a 99.7% correlation between this alternative treatment variable and that used in the paper, so that my conclusions are not sensitive to this choice. Full results available on request.

never spoken at home, a household item ownership score (based on 14 items), and for maternal and paternal education levels.

Test scores are only available at two points in time (2000 and 2007), which prevents me from testing for differences in pre-existing trends. Instead, in order to address concerns regarding pre-existing trends as well as other potential sources of endogeneity of enrollment growth, I test the robustness of my findings to instrumenting enrollment growth, as described in the next section.

3.3 Instrumental Variable Approach

There may remain unobservable sources of endogeneity even after controlling for growth in discretionary government transfers and observable characteristics of students. For instance, if expected returns to education increased faster between 2000 and 2007 in some regions than others, then one might expect both increases in enrollment and in study effort, so that enrollment growth would be endogenous when y_{irt} is a pupil's test score. Or it could be the case that regions where local administrations became more committed to education, higher increases in both enrollment and education quality were achieved. Or one may worry about measurement error in enrollment figures, since there is an incentive to over-report enrollment rates in order to increase the number of capitation grants.[7] In order to address these remaining issues, I use *potential* growth in enrollment based on predetermined fertility decisions and migration decisions up to 2002 as an instrumental variable for *actual* growth in enrollment. More specifically, I exploit the fact that actual enrollment depends not only on contemporaneous decisions of policy makers, parents and children, but also on the size of the primary-school age population, which is predetermined, and use the growth in the size of the primary-school age population as an instrument for the actual enrollment growth. The first stage of my two-stage least squares system is as follows:

$$\left(\frac{post\_enrol}{baseline\_enrol}\right)_r \times \mathbb{1}(t=2007)_t = \gamma_0 + \gamma_1\left(\frac{post\_age7\_13}{baseline\_age7\_13}\right)_r \times \mathbb{1}(t=2007)_t \qquad (3)$$
$$+ \mathbb{1}(t=2007)_t + R_r'\gamma_r + X_{irt}'\gamma_X + \nu_{irt}$$

[7] Joshi & Gaddis (2015) however find no evidence of over-reporting of total enrollments by schools.

where γ_r are region fixed effects, $baseline\_age7\_13_r$ is the number of children aged 7-13 in region r in 2001 and $post\_age7\_13_r = \sum_{j=2002}^{2007} Age7\_13_{rj}$ is the sum of the number of children aged 7-13 in region r in each year from 2002 to 2007. The size of the relevant cohorts is calculated using a single population census carried out in 2002 and based on the individual's age and region of residence at the time of the census.[8] Differences in $(\frac{post\_age7\_13}{baseline\_age7\_13})_r$ across regions can therefore be interpreted as differences in fertility trends (between 1988 and 2000) and migration patterns up to 2002.[9] In a robustness check, I instead construct the instrument based on the individual's region of residence in 2001 (hence based on migration decisions before FPE) using migration data, and show that this does not affect the results.

The figures in Table 1 shed light on the nature of the variation captured by the instrument. The first (last) three columns report the total number of births (average number of births per woman) in 1988 and in 2000 and the change in annual births between the two years, by region. There are large differences across regions during this period, from an 86% increase in the number of births in 2000 relative to 1988 in Tabora to a 1% decrease in the Kilimanjaro. Increases in the total number of annual births are observed in some regions despite decreases in the number of births *per woman* due to high levels of past fertility (which translate into an increase in the number of fertile women between 1988 and 2000). In all but three regions, however, a woman of fertile age was less likely to give birth in 2000 than in 1988, as shown in the last column. Although changes in the total number of births depend both on changes in fertility per woman and differences in the number of women of fertile age found in the region, overall, regions with smaller increases in cohort size also experienced larger decreases in fertility rates, except for the capital Dar-es-Salaam, where the 40% cohort growth is driven by immigration since it took place despite a 38% decrease in fertility.

The reduced-form equation corresponding to the two-stage least squares system is:

$$\begin{aligned} y_{irt} =& \lambda_0 + \lambda_1 (\frac{post\_age7\_13}{baseline\_age7\_13})_r \times \mathbb{1}(t = 2007)_t \\ &+ \mathbb{1}(t = 2007)_t + R_r' \lambda_r + X_{irt}' \lambda_X + \mu_{irt} \end{aligned} \tag{4}$$

[8]For instance, the number of students age 13 in 2001 (2005) is inferred from the number of individuals age 14 (10) in the 2002 Census.

[9]The relevant fertility period is 1988-2000 because $post\_age7\_13$ includes children born between 1989 and 2000 and $baseline\_age7\_13$ corresponds to children born between 1988 and 1994.

And the effect of enrollment growth obtained using the instrumentation procedure is $\frac{\lambda_1}{\gamma_1}$.

Rapid enrollment growth raises two main concerns in terms of schooling quality, the first related to the increase in the number of pupils to be accommodated in the system and the second to a worsening of the quality of peers. While any instrument with power in explaining variation in the number enrolled will yield estimates that speak to the first concern, different instruments may lead to different local average treatment effects (LATE) depending on the way the first stage influences peer composition. The LATE estimated based on my IV may differ from one based on, e.g., the pre-FPE net enrollment rate in the region if the quality of additional pupils due to demographic growth differs from that due to lower initial enrolment rates. The direction of the difference is unclear *a priori*: in both cases, enrollees induced by an increase in the value of the instrument should be of lower SES,[10] but with an IV using variation in initial enrolment rates, these additional enrollees may also come from more motivated households, or from regions with more proactive local governments. Similarly, the LATE obtained using the pre-FPE net enrollment rate in the region as an IV may differ from one obtained using another education-related pre-FPE regional variation such as differences in school fees–e.g., in some regions, fees may be low pre-FPE due to low demand (low net enrolment rate), or on the contrary, pre-FPE enrolment may be already high because the fees were low. Here I chose to focus on a powerful instrument for which I have good data and for which the direction of the remaining potential bias is easiest to sign, as discussed below.[11]

3.4 Signing the Direction of Any Remaining Bias

The main concern regarding the exclusion restriction required for the instrument to be valid is that regions having experienced faster fertility declines in the pre-reform period (1988-2000) may also have experienced faster increases in investments in the human capital of children by parents or policy makers. If this were the case, then this would lead quality measures that can be influenced by such investments in human capital (e.g., children test scores) to increase more in regions experiencing slower growth in potential enrollment. Similarly, if some older children migrate to regions with more positive school quality trends, then regions with lower values of

[10] In the 2002 Tanzanian census, uneducated mothers of fertile age had had on average 4.12 children compared to 2.65 for mothers with at least some primary schooling.

[11] See Figure A-9 for a graphical illustration of the strong correlation between the growth rates of the primary-school age cohorts and that of actual enrollment.

$(\frac{post\_age7\_13}{baseline\_age7\_13})_r$ may experience more positive changes in learning outcomes over time, since larger older children cohorts increase the denominator of this ratio relative to the numerator. All these potential issues would tend to lead to an overestimation of the worsening of quality coinciding with higher enrollment growth.

To see this, consider the following system:

$$Y = \beta X + \gamma Z + \epsilon \tag{5}$$

$$X = \nu Z + \mu \tag{6}$$

and the corresponding reduced-form:

$$Y = \beta \nu Z + \gamma Z + \phi \tag{7}$$

Consider the case in which $\beta \leq 0$ (higher enrollment growth may worsen test scores), $\nu > 0$ (higher potential enrollment growth leads to higher actual enrollment growth), and $\gamma \leq 0$ (higher potential enrollment growth may be correlated with slower increases in parental or public investment in child quality or worse school quality trends). If $\gamma = 0$, then 2SLS identifies $\frac{\beta\nu}{\nu} = \beta$. If $\gamma \neq 0$, then 2SLS will provide an estimate of $\frac{\beta\nu+\gamma}{\nu}$, and the magnitude of the negative effective of higher enrollment growth on test scores is overestimated as $\frac{\gamma}{\nu} \leq 0$.

In order to account for my IV findings on the whole sample, for which I find no statistically significant effect of enrollment growth on test scores, an omitted variable would have to be positively correlated both with fertility trends (between 1988 and 2000) and with *improvements* over time in educational quality or child human capital. Or there would have to be pre-reform migration patterns such that younger children are more likely than older children to be observed in areas with more favorable trends in schooling quality. It is hard to think of such omitted variable other than government transfers targeting less developed areas, which I control for in a robustness check.

In order to confirm empirically the most likely direction of the bias, if any, I use data from the Tanzanian Demographic and Health Surveys of 1991-92, 1996, 1999, and 2004-2005, which collected data on a range of under-5 children's health inputs and outcomes, and test for differential trends in these inputs and outcomes between regions with different potential enrollment growth. More precisely, I run the following regressions on the sample of children born between 1988 and 2000 (i.e., during the period that is relevant to the construction of the instrumental variable, which spans children aged 13 in 2001 to children aged 7 in 2007):

$$health_{irt} = \xi_0 + \xi_1 (\frac{post\_age7\_13}{baseline\_age7\_13})_r \times \mathbb{1}(t \geq 1995)_t + \mathbb{1}(t \geq 1995)_t + R_r' \xi_r + \phi_{irt} \quad (8)$$

where $health_{irt}$ refers to child i in region r born in year t and is, in turn, an indicator for whether the child has received a full course of immunization, a dummy for whether delivery was assisted by a health professional, a dummy for whether the mother received no help at all during delivery, an infant mortality indicator equal to one if the child died within 12 months of birth, and zero otherwise, and a stunting indicator which is equal to one if the child's height-for-age z-score is below 2 standard deviations of the reference median, and equal to zero otherwise. ξ_r are region fixed effects. The coefficient of interest is ξ_1, and a non-zero coefficient indicates a differential trend in the outcome variable in areas with slower fertility decline (using 1995, the mid-point of the relevant period, as threshold).

Results are reported in Table 2. All the point estimates go in the direction of smaller improvements in child health inputs and outcomes in regions with higher potential enrollment growth, statistically significantly so in the case of full immunization and delivery by a health professional.[12] This confirms that, if anything, my IV estimates of the effect of enrollment growth on test scores are likely to over- rather than understate any worsening in achievement.

4 Data and Summary Statistics

4.1 Pupil, Teacher, and School Data

SACMEQ is a consortium of 15 Ministries of Education in Southern and Eastern Africa. I use data from the two surveys available for Tanzania, namely SACMEQ II, collected in 2000, and SACMEQ III, which was collected in 2007. SACMEQ II surveyed 2,854 pupils Grade 6 in 181 schools, and SACMEQ III surveyed 4,194 Grade 6 pupils in 196 schools and stratified sampling

[12] The exercise is repeated separately for rural and urban areas in Tables A-1 and A-2. There is some variation between rural and urban areas in terms of which indicators show significant changes, but the overall pattern is the same in both sectors.

ensures that the survey is representative of all Grade 6 pupils in government schools.

In addition to testing the numeracy and literacy skills of Grade 6 pupils and their teachers, the survey collected data from pupils, teachers and the school headteacher, thus providing an exceptionally rich level of detail on schooling inputs and learning outcomes. The pupil mathematics test was based partly on Trends in Mathematics and Science Study (TIMSS) items and partly on other items newly written by the SACMEQ National Research Coordinators. The tests carried out in 2000 and 2007 differ in order to reflect changes in curricula between the two periods. However, there was an overlap in questions in order to create scores that are comparable over time using item response theory–the same approach as that taken, e.g., in the well-known Trends in Mathematics and Science Study and Progress in Reading Literacy Study.

The timing of the surveys is ideal to evaluate the effect of the large increases in enrollment following the removal of primary school fees on the quality of the learning environment since Grade 6 students in 2007 will have started school in 2002 and therefore been fully exposed to the larger cohorts that entered school after primary school fees were removed. On the contrary, students in Grade 6 in 2000 will not have been affected since the policy was only announced in 2001.

SACMEQ surveys first selected schools within each sampling stratum (defined as between one and two Tanzanian regions in the present sample) by probability proportional to size sampling. A sample of 20 (in 2000) or 25 (in 2007) pupils within each school was then selected based on a random draw from all Grade 6 class registers in the school. If a selected pupil was absent during the survey, he or she was *not* replaced by another pupil. All the Kiswahili and math teachers teaching the randomly drawn students were eligible to be surveyed (Mrutu et al. 2015). Local authorities and schools were informed several weeks in advance of the enumerators' visit, and the survey took place over two consecutive days, with language tests administered on the first day and math tests on the second day (ACER 2015). Although the sample of pupils (and hence teachers eligible for interview and testing) was drawn randomly, as in any survey without a perfect response rate, non-response can lead to selection bias. There is, however, no correlation between the number of pupils successfully interviewed/tested and regional enrollment growth, which suggests that non-response is unlikely to bias my estimates of the effect of enrollment growth on pupil test scores.[13]

[13]More specifically, when estimating Equation 2 with the number of pupils successfully interviewed/tested

The final sample is obtained as follows. Six pupils in the 2007 survey are dropped due to missing math scores, I also drop observations from two schools (and their 26 pupils) with a pupil-teacher ratio above 250, as well as 18 pupils with no information about father education. Finally, I drop 65 pupils from the Lindi region in the 2007 survey because this small region was not surveyed in 2000, resulting in a sample of 6,933 pupils.[14]

4.2 Other Data Sources

Regional enrollment data are taken from statistical yearbooks produced by the Ministry of Education and Vocational Training ("Basic Education Statistics"). Primary-school age cohort sizes are based on the 2002 Population Census microdata extract provided by IPUMS. Education grants data come from district-level budget plan data for the period 2000-2007.

4.3 Summary Statistics

Table 3 reports summary statistics separately for pupils observed in the 2000 and 2007 SACMEQ surveys. Here I describe changes over time in the country as a whole before analysing differences in changes over time across regions with different rates of enrollment growth in Section 5.[15]

The first three rows of Table 3 show regional demographic and education statistics based on calculations from government statistics (rows 1 and 3) and census data (row 2). On average, the size of the cumulated enrollment in the 2002-2007 period is nearly 9 times larger than that in 2001 (it would have been 6 times larger if enrollment had been stable over the 2001-2007 period), while the total primary school-age population for the period 2002-2007 is 6.8 times larger than that in 2001. Two reasons why growth in actual enrollment is larger than that in the total primary school-age population are that (i) the net enrollment rate also increased over time and (ii) more underage and overage children may have been enrolled in the later years

as dependent variable, the coefficient associated with enrollment growth is -0.20 and its associated p-value is 0.764.

[14] Some of the school- or class-level variables are missing for 513 pupils. Given that the main interest of this study is to analyze the effect on achievement rather than schooling inputs, instead of dropping these observations, I impute the value of these missing variables to be equal to the school sample mean (mode) for continuous (categorical) variables. I repeated the analysis excluding these 513 pupils instead and found nearly identical results, which are available on request.

[15] See Appendix A-2 for a graphical analysis comparing changes in schooling quality between 2000 and 2007 across regions with different rates of enrollment growth.

covered by the data due to the FPE reform.

Note that a one-unit increase in $(\frac{post\_enrol}{baseline\_enrol})_r$ corresponds to an increase of 33% in the size of the cohorts entering Grade 1 from 2002 onwards, under the simplifying assumptions that there is no population growth and that no pupil drops out.[16]

The third row shows that the total amount received by regions in discretionary education block grants (i.e., excluding capitation grants) between 2002-2007 is on average 11.4 times larger than that received in 2001 (in real terms).

The remaining statistics are based on SACMEQ data. The average pupil-teacher ratio increased by 32%, going from 47 to 62.2 between 2000 and 2007. However, the incidence of multiple shifts teaching reported by head teachers decreased, and the number of teaching hours (self-reported by teachers) did not change much despite statistically significant decreases.

Turning now to measures of the "quality" of the teachers teaching the average pupil, there is evidence of an improvement in the level of academic qualifications of teachers over time, with a particularly steep increase in the proportion of Kiswahili teachers with O-level qualifications (which is the exam taken after four years of secondary education). There has however been a decline in the proportion of teachers with at least two years of teacher training, which is expected since initial training in teachers colleges went from two- to one-year in order to speed up the supply of qualified teachers. As expected, with the increased demand for new teachers, the average experience of teachers also decreased between 2000 and 2007. Consistent with the overall improvement in the education of teachers, the performance of teachers in subject-specific tests has improved by 0.32 (0.41) standard deviations in Kiswahili (Math).

Most indicators suggest that access to physical educational inputs has improved despite the increase in enrollment, which would suggest that the capitation grant aimed at covering non-salarial costs was effective in maintaining expenditure on teaching and learning materials. For instance, the proportion of pupils having no access at all to a reading textbook has gone down from 36% in 2000 to 23% in 2007, although some ground has been lost on the government's

[16] See Table A-3 for an illustration of the structure of grade enrollment under these assumptions. A one-unit increase in $(\frac{post\_enrol}{baseline\_enrol})_r$ corresponds to a $\delta_r = \frac{1}{3}$ increase in intake since in this case $(\frac{post\_enrol}{baseline\_enrol})_r = \frac{21N+21(1+\delta_r)N}{7N}$. The ratio $(\frac{post\_enrol}{baseline\_enrol})_r$ being equal to 9, on average, thus implies a 100% increase in the number of children enrolled in Grade 1 from 2002 onwards under these assumptions. This is to be compared to a figure of 82% computed from official statistics and mentioned in the introduction, which may be slightly smaller, for instance, because drop out also decreased from 2002 onwards, or because of positive population growth.

target of achieving a one-to-one pupil-textbook ratio. Similarly, classroom equipment has improved, on average, by more than one of the following items: writing board, chalk, wallchart, cupboard, bookshelves, library, teacher table, teacher chair.

The learning outcomes statistics reported in Table 3 paint a striking pattern: despite the sizeable increase in the pupil-teacher ratio, and the decrease in teacher experience, there has been sizeable progress in the performance of pupils both in reading and mathematics tests of around 0.35 standard deviations, translating in the halving of the proportion of pupils with low reading or math competency (i.e., no more than "basic reading" or "emergent numeracy" competency) and a substantial increase in the proportion of pupils with high reading or math competency (i.e., demonstrating analytical and critical reading or competent numeracy to abstract problem solving).[17]

Finally, turning to socioeconomic and demographic pupil characteristics, the average Grade 6 pupil in 2007 compared to that in 2000 is younger, slightly more likely to be male, and enjoys a more favorable socioeconomic background, which indicates that despite the extension in the schooling "franchise", any increase in the proportion of Grade 6 children coming from poorer backgrounds has been more than compensated by the positive trend in standards of living between 2000 and 2007.[18] Recall that the cohort of Grade 6 pupils observed here in 2007 was the first affected by FPE from Grade 1. Despite the initial jump in enrollment (from 58% in 2001 to 73% in 2002), the change in socioeconomic composition among this first post-reform cohort of new entrants is likely to be less than in subsequent years. Put differently, the marginal child in the first post-reform cohort studied here is likely to have higher SES than the marginal child enrolled in later post-reform cohorts, if anything because of the priority given to 7-year old new entrants and the tendency for poorer households to delay entry. In addition, even if enrollment in the lower primary grades increased more among poorer households (as suggested by Hoogeveen & Rossi (2013)), the change in SES composition is likely to be weaker by Grade 6 due to higher dropout rates among the poor.

All in all, the national trends discussed above suggest that the rapid expansion of the pri-

[17] Between 2000 and 2007, learning outcomes improved by at least 0.10 s.d. in 6 out of the 14 countries included in the SACMEQ exercise, while they decreased by 0.10 s.d. or more only in Mozambique (in reading and math) and Uganda (in math)(Makuwa 2010).

[18] The slight increase in the proportion of Grade VI students who are male is consistent with official figures from BEST 2003 and BEST 2007. Own calculations based on these figures indicate an increase from 49.3% in 2000 to 51.2% in 2007.

mary schooling system did not lead to a drop in test scores. Results reported in the next section show that what is suggested by these raw data aggregated at the country-level is confirmed in the regression analysis exploiting regional variation in enrollment growth.

5 Main Results

5.1 Effect of Enrollment Growth on Schooling Inputs

Table 4 reports estimates of the effect of enrollment growth on the "quantity" of teachers. The first column reports estimated effects on the pupil-teacher ratio. Results in Panel A were obtained from an OLS regression of Equation 2. I find that an increase in enrollment growth by one standard deviation increases the pupil-teacher ratio by 6.9, and this effect is statistically significant at the 1% level.[19]. The wild cluster bootstrap-t p-value is 0.008, thus confirming the conclusions based on analytical standard errors. Panel B reports 2SLS results obtained when instrumenting enrollment growth with potential enrollment growth. The point estimate is somewhat larger but qualitatively similar, and significant at the 5% significance level. The Kleibergen-Paap F-statistic–which allows for intra-regional correlation of standard errors–for the first stage is 17.296, which suggests that the instrument is reasonably strong. The p-value of the Kleibergen-Paap underidentification test is 0.031, and thus I can reject the null hypothesis that the model is underidentified. Interestingly, except for maths teaching hours (last column), I cannot reject that enrollment growth is exogenous with respect to the measures of teaching quantity used as dependent variables here (see last row). Whenever this is the case, from here onwards I focus on the estimated effects based on OLS rather than 2SLS, as OLS is more efficient than 2SLS.

Panel C reports reduced-form estimates of the impact of potential enrollment growth on the dependent variable (Equation 4). These results do not require imposing the restriction that the only way in which potential enrollment growth affects y_{irt} is through actual enrollment growth. The estimated effect of potential enrollment growth is larger than that for actual enrollment growth in Panel A because an increase in the size of the cohort of primary school age post-reform relative to the pre-reform primary school-age cohort translates into a larger increase in

[19]Here I refer to the standard deviation of the distribution of enrollment growth in the 2007 sample (0.693, see Table 3).

actual enrollment since the net enrollment rate was increasing during the period under scrutiny (i.e., $\gamma_1 = 1.66 > 1$ in the first-stage Equation 3).

Looking now at Columns (2) to (4), we see that there is no statistically significant effect of enrollment growth on whether schools operate multiple shifts, and on the (self-reported) number of hours taught by teachers.

Table 5 reports OLS, IV, and reduced-form estimates of the impact of enrollment growth on measures of teacher quality, namely: whether they have O-levels (Columns 1 and 2), and whether they have completed at least two years of teacher training (Columns 3 and 4). Irrespective of the estimation approach, no statistically significant effect is observed, although I cannot rule out non-negligible decreases in the share of teachers with at least two years of teacher training (-0.17 or a decrease by 20% of the sample mean for Kiswahili teachers).

Table 6 considers two other measures of teacher quality: years of experience (Columns 1 and 2) and subject-specific standardized test-scores (Columns 3 and 4). Regions that experience larger increases in enrollment gained less experienced teachers (by 1.4 years for Kiswahili teachers and 2.8 years for Math teachers for 1 s.d. increase in actual enrollment growth), on average, which is consistent with the expectation that local governments in these regions had to recruit a larger number of new teachers. Similarly, there is a statistically significant worsening of the reading scores of language teachers in regions experiencing larger enrollment growth (a worsening by 0.14 of a standard deviation for one standard deviation larger enrollment growth). These findings are robust to the different estimation methods employed.

Table 7 considers the effect of enrollment growth on access to physical inputs, namely whether pupils have access to a textbook without having to share it with any other pupil (Columns 1 and 2), the number of equipment items available in the classroom (Columns 3 and 4), the pupil's number of exercise books (Column 5), and the number of equipment items available to the pupil (Column 6). Access to physical inputs, and textbooks in particular, was emphasized as a priority area in the Primary Education Development Plan which accompanied the removal of primary school fees (Basic Education Development Committee 2001). In particular, US$4 out of the US$10 donor-funded capitation grant received for each enrolled pupil to cover non-salary costs was explicitly ring-fenced for the acquisition of textbooks and other teaching and learning materials. Contrary to teacher-related inputs, the expected effect of enrollment growth on these physical inputs is therefore unclear *a priori* since the extra source of funding coming from the capitation grant may have more than compensated the increase in needs and the loss of school fee revenues experienced by schools. Looking at the estimates in Table 7, one can see that the total effect of enrollment growth on pupil equipment appears to have been small and statistically insignificant, while some improvement is observed in terms of classroom equipment in the two-stage least squares estimates.

Results so far indicate that enrollment growth led to sizeable increases in pupil-teacher ratios, a non-negligible decrease in the level of experience of the average teacher and some worsening of average teacher subject-specific knowledge (in Kiswahili, at least), but that access to pupil-specific physical inputs was little affected. The availability of classroom equipment (such as writing boards) may have improved somewhat thanks to the targeting of the capitation grant for non-salary expenditures.

Based on the existing body of knowledge on the impact of schooling inputs on test scores, however, it is not clear that larger class sizes should lead to lower test scores, especially in a developing country context (Banerjee et al. 2007, Duflo et al. 2012) (see detailed review in Section A-1.2). Pupils of teachers at the start of their careers have been found to perform less well in developed countries, but the estimated effect of an additional year of experience is small (between 0.014-0.018 standard deviation in Rockoff (2004)) and in India, Azam & Kingdon (2015) do not find that teacher experience explains any of the between-teacher test score variation. Metzler & Woessmann (2012) find that teacher's subject-specific knowledge matters in mathematics, but given the size of this effect (an increase by 0.087 SD for one SD increase in teacher's knowledge) and the size of my estimates on the effect of enrollment growth on teacher knowledge (-0.14 in Kiswahili teacher knowledge for one SD increase in enrollment growth), the implied effect on test scores is small. Finally, turning to the effect of non-teacher inputs, the most reliable evidence available, obtained through randomized controlled trials in Kenya, suggests no effect of flipcharts (Glewwe et al. 2004) or textbooks (Glewwe et al. 2009) on test scores, except for the best students in the case of textbooks.

Therefore, the impact on test scores of the changes in inputs per pupil observed in the data is expected to be small. In the next section, I estimate the effect of enrollment growth on test scores overall (i.e., through changes in both observed and unobserved inputs).

5.2 Effect of Enrollment Growth on Test Scores

In Table 8, I estimate the impact of enrollment growth on reading test scores. The first column reports OLS estimates controlling only for region fixed effects, survey round, and the rural or urban location of the school. The point estimate is essentially zero (-0.002), and the lower bound of the 95% CI is -0.168, implying a maximum decrease of 0.12 of a standard deviation for an increase in enrollment growth by 1 standard deviation. In Column (2), I add controls for pupil characteristics to control for differential changes in composition across regions with different enrollment growth rates, and for each of the physical inputs used as dependent variables in Table 7 in order to control for possible improvements in access to these inputs due to the capitation grant accompanying enrollment growth. The inclusion of these controls barely changes the estimates.[20] To put the lower bound of the 95% CI into perspective, the difference between the 50th and 55th percentiles in the distribution of reading (math) scores is 0.15 (0.11) standard deviations and the difference between the 55th and the 60th percentiles is 0.11 (0.14) standard deviations. I can therefore not rule out a worsening of test scores of the order of a 5-percentile drop in the distribution, but I can rule out a larger effect for the average pupil. The IV estimates reported in Column (3) lead to similar conclusions, and an exogeneity test that is robust to heteroskedasticity fails to reject the null hypothesis that enrollment growth is exogenous in Equation 2.

Table 9 reports estimates of the effect of enrollment growth on pupil math test scores. Results are very similar to those obtained for reading test scores. The OLS (with or without controls) and IV estimates are statistically insignificant, I cannot reject the exogeneity of enrollment growth in Equation 2, and the lower bound of the OLS 95% CI is -0.12 of a standard deviation for an increase in enrollment growth by 1 standard deviation. In comparison, the overall improvement in test scores in math between the two SACMEQ surveys was 0.35 s.d., pupils in rural areas perform on average 0.47 s.d. below their urban counterparts, girls obtain 0.32 s.d. lower scores than boys, and children whose fathers have completed more than primary education outperform children of fathers who did not complete primary schooling by 0.59 s.d..

[20] Although pupil characteristics are correlated with test scores, estimating Equation 2 using pupil characteristics as dependent variables shows that there is no robust statistically significant effect of enrollment growth on pupil SES for this first post-FPE cohort when observed in Grade 6 (see Table A-11). As discussed in Section 4.3, this is not unexpected given the priority initially given to correct-age entrants and larger dropout rates among the poor.

6 Robustness Checks

Having found no evidence of a sizeable effect of enrollment growth on pupils' test scores, which is robust to instrumenting enrollment growth with potential enrollment growth based on past fertility and migration decisions, I now turn to testing the robustness of these findings to controlling for growth in primary education funding across regions. The main concern here is that less-developed regions may have experienced faster (potential and actual) enrollment growth and been increasingly targeted by government education funding. This could bias my estimates if government transfers also increased achievement. Evidence suggests that test scores are largely unresponsive to "more of the same" inputs and flexible grants (Kremer et al. 2013). I nonetheless assess the validity of this concern in Columns (4) of Tables 8 and 9 by adding to the OLS regression a control for the growth in government funding for primary education (at the regional level). The estimated effect of enrollment growth on pupil reading and math z-scores remains statistically insignificant, and the lower bound of the 95% CI for reading (math) scores now translates into a -0.14 (-0.16) s.d. effect for an increase in enrollment growth by 1 standard deviation, a small increase in magnitude relative to the baseline specification.

Column (5) then reports OLS estimates obtained when controlling for reversion to the mean. More specifically, I include an interaction term between a post-reform survey dummy and the baseline regional average pupil score (in reading in Table 8 and in math in Table 9). Consistent with the expectation that progress was larger in areas where achievement was lower at baseline, I find that an additional standard deviation in mean reading (math) scores at baseline is correlated with a 0.61 (0.74) s.d. smaller increase in reading test scores between 2000 and 2007. The estimated effect of enrollment growth on pupil reading and math z-scores are still statistically insignificant, and the lower bounds of the 95% CI for reading (math) scores correspond to a -0.15 s.d. effect on both reading and math test scores for an increase in enrollment growth by 1 standard deviation. Finally, Column (6) reports estimates of the reduced-form equation. Potential enrollment growth has no statistically significant effect on either Kiswahili or mathematics test scores, with the lower bound of the 95% CI indicating a maximum worsening of test scores by 0.08 (Kiswahili) and 0.16 (math) of a s.d. for a 1 s.d. increase in potential enrollment growth.

In Table 10, I test the robustness of my instrumental variable. First, I add, in the reduced-form equation (Eq. 4), the 7-year "lag" of the instrumental variable, namely $\sum_{j=1995}^{2000} Age7\_13_{rj}$ divided by the number of children aged 7-13 in region r in 1994. The idea of this test is to check whether differential changes in test scores between 2000 and 2007 associated with demographic trends that affected potential enrollment growth before FPE could be confounding my estimates. Second, I show the robustness of my findings to defining the instrument based on region of residence as of 2001, using information on region of residence one year before the 2002 census for those respondents who said they had migrated in the past year.[21]

In the first two columns, I present the reduced-form regressions for Kiswahili z-scores (Column (1)) and mathematics z-scores (Column (2)) including a control for primary-age cohort growth between 1994 and 2000. The correlation coefficient between potential enrollment growth between 1994 and 2000 and between 2001 and 2007 is weak (-0.17), so that including $\frac{\sum_{j=1995}^{2000} Age7\_13_{rj}}{Age7\_13_{r1994}}$ as a regressor brings little change to the point estimates (comparing with the last column of Tables 8 and 9: the reduced-form estimate becomes 0.068 (-0.006) for reading (math) compared to 0.083 (0.012)). In the next two columns, I present 2SLS estimates obtained when the instrument is constructed by assigning individuals to the region where they lived in August 2001, hence before primary fees were removed, instead of the region of residence at the time of the census in August 2002. The point estimates are almost identical to those obtained with my instrument based on region of residence in 2002 (0.046 (0.008) instead of 0.05 (0.007) for reading (math) scores), which suggests that migration patterns correlated with school quality trends are not driving my findings.

Finally, I consider an alternative identification strategy to that adopted in this paper, namely one exploiting differences in baseline enrollment rates instead of predetermined fertility as a source of plausibly exogenous variation in enrollment growth. Contrary to the instrumental variable used in the paper, it is difficult to sign *a priori* the direction of the bias in IV (or reduced-form) estimates using pre-FPE enrollment rates to create an instrumental variable for actual enrollment growth. Lower enrollment levels at baseline could indeed have been correlated with either higher or lower subsequent growth in test scores irrespective of the increase in enrollment. For instance, lower baseline enrollment could be correlated with higher future

[21] The population census was carried out in August 2002, while the abolition of primary school fees was announced by the president of Tanzania in April 2001 (Kattan & Burnett 2004). Region of location in August 2001 is therefore unlikely to have been affected by the school fee regime change, as inter-regional migration decisions are likely to take more than a few months to plan and act upon, and since the abolition of school fees was only effective from the start of the 2002 school year.

growth in test scores due to mean reversion in investments both in education quantity and quality, or, on the opposite, with lower subsequent growth in test scores if, for example, baseline enrollment proxies for tastes for education and more education-oriented parents increasingly care about the quality of education. Therefore I do not follow this strategy in the main analysis. For completeness, I however repeated the main analysis exploiting differences in baseline enrollment rates as a source of variation in exposure to FPE, and find remarkably consistent results (see Appendix A-3).

7 Treatment Effect Heterogeneity

Although I find no evidence of substantial worsening of average test scores, the effect of enrollment growth may vary across pupils. For instance, in schools where very little learning took place before 2002, then only small achievement losses should be expected from enrollment growth. Therefore, students at the top of the distribution of test scores may suffer more. On the other hand, less able students may be more reliant on schooling inputs in order to learn, and may thus suffer more. In order to explore potential heterogeneous effects of rapid enrollment growth, I first use Athey & Imbens (2006)'s Changes-in-Changes estimator.

7.1 Treatment Effects Across the Distribution of Test Scores

The main appeal of this approach is that it estimates the impact of a binary treatment at any point in the distribution of test scores while relaxing the standard Difference-in-Difference assumption that the unobserved component of the outcome variable depends additively on the treatment group. Formally, the Changes-in-Changes estimate of the treatment effect at quantile q can be written:

$$\tau_q^{CiC} = F^{-1}_{Y^I,11}(q) - F^{-1}_{Y,01}(F_{Y,00}(F^{-1}_{Y,10}(q))) \tag{9}$$

where $F_{Y,gt}(y)$ is outcome Y's cumulative distribution function in group g in period t, and region 1 only is treated in period 1, Y^I denotes the potential treated outcome and Y denotes the realized outcome. I define the treated group as those regions with above-median enrollment growth. This method can be summarized in three steps. First, in the pre-treatment period, find the quantile q' in the control group's outcome (Y) distribution corresponding to the same value of the outcome as quantile q in the treatment group's distribution. Second, use data from the control group to compute the change Δ, between the before- and after treatment periods, in the value of Y at quantile q'. Third, compare the value of Y at quantile q in the treated group's post-treatment distribution ($F^{-1}_{Y^I,11}(q)$) to that which would be predicted from adding Δ to the value of Y observed at quantile q in the treated group's pre-treatment distribution ($F^{-1}_{Y,01}(F_{Y,00}(F^{-1}_{Y,10}(q)))$). The difference is the Changes-in-Changes estimate.

One limitation when applying the Changes-in-Changes approach here is that estimates rely on the assumption that the distribution of ability does not vary within the treatment group over time. If areas with larger enrollment growth drew more pupils with lower ability into schools relative to the control areas, for instance, then this assumption would be violated. A common way of addressing the issue that the probability of being observed in the sample may be systematically different for treated and control observations is to compute bounds on the treatment effect by making some assumptions about the direction of the potential bias arising from sample selection and trimming the sample accordingly. Following Blanco et al. (2013)'s insights, a lower bound (i.e., here, the largest negative effect) is obtained without trimming, while the upper bound is obtained by trimming the bottom of the distribution of test scores in the treated group to reflect the influx of marginal students in the extreme-case scenario that they might all have lower scores than inframarginal students, as explained in detail in Appendix A-4.

Results are displayed in Figures 1 and 2. Looking first at Figure 1, we can see that, in most cases, the lower bound point estimates suggest no more than a decrease in KiSwahili test scores by 0.10-0.15 standard deviations. There is more heterogeneity in the effect of enrollment growth on mathematics test scores, as point estimates for the lower bounds of the quantile treatment effects become negative from the 75th percentile onwards, and take some large absolute values from the 80% percentile onwards (Figure 2), although the estimates are too imprecise to achieve statistical significance.[22] For both language and mathematics, the upper bound estimates are

[22]Exact zero effects arise because of bunching in the distribution of test scores. For instance, looking at the lower-bound estimate for the 10th percentile in Math, $F_{Y,00}(F^{-1}_{Y,10}(0.1))$ is equal to 0.0959, meaning that the test score $F^{-1}_{Y,10}(0.1)$ corresponding to the 10th percentile of the distribution in the high-growth areas in 2000 was found at the 9.59th percentile of the 2000 distribution in low-growth areas. Due to some bunching in the distribution of test scores, the test score corresponding to the 9.59th percentile of the distribution of test scores in the low-growth areas in 2007 ($F^{-1}_{Y,01}(0.0959)$) is equal to exactly the same test score as that

positive and large in magnitude for at least the bottom half of the trimmed distribution, which suggests that the assumption that all marginal students are less able than inframarginal students in treated areas unduly removes poorly performing inframarginal students from the distribution of treated individuals in the 2007 distribution.

7.2 Treatment Effects in Rural vs. Urban Schools

I also investigate heterogeneity in treatment effects by splitting the sample between rural and urban areas, as urban areas have a very large advantage in test scores at baseline (of 0.82 and 0.56 of a standard deviation in reading and math, respectively, in the raw data).[23] Table A-10 reports summary statistics broken down by rural and urban areas illustrating the differences between the two sectors, and in particular that urban pupils have higher SES, higher quality and quantity of teachers, and test scores. Given the much higher standard of achievement in urban areas, there may be more to lose in terms of the quality of the learning environment in these areas. Table 11 reports the results obtained when analyzing the effect of enrollment growth separately for rural and urban areas. Starting with the rural sample (consisting of 5054 out of the 6933 pupils included in the main analysis), results are similar to the average effects reported in Section 5, except for the finding that subject-specific knowledge *increases* among mathematics teachers. The baseline OLS specification picks up a statistically significant, positive, correlation between pupils math test scores and enrollment growth, but the point estimates become statistically insignificant when instrumenting for enrollment growth or when adding controls for government transfers or mean reversion (Panel C). On the contrary, in the urban sample, there is a large worsening of subject-specific knowledge among mathematics teachers, as well as a substantial, robust, worsening of reading and mathematics scores with enrollment growth.

The most striking contrast between urban and rural areas is found in the response to

corresponding to the 10th percentile of the high-growth group in 2007 ($F^{-1}_{Y^I,11}(0.1) = -0.967$) and hence $\tau^{CiC}_{0.1} = F^{-1}_{Y^I,11}(0.1) - F^{-1}_{Y,01}(F_{Y,00}(F^{-1}_{Y,10}(0.1))) = 0$. The same happens until the 70th percentile. At the 75th percentile, the test score corresponding to the 75th percentile of the 2000 high-growth distribution is found at the 78.8th percentile of the 2000 low-growth distribution, and the test score at the 78.8th percentile of the 2007 low-growth distribution is 0.891, compared to the 0.740 found at the 75th percentile of the 2007 high-growth distribution, resulting in a quantile treatment effect estimate of -0.151.

[23] Head teachers were asked to indicate whether their school was located in an isolated area, rural area, small town or large city. Schools in isolated or rural areas are classified here as 'rural', and the others as 'urban'. Each region has both rural and urban schools.

enrollment growth of the math test scores of math teachers and pupils.

While pupils in rural areas saw the new math teachers recruited to face enrollment growth improve average teacher subject-specific knowledge, the reverse happened in urban areas. In light of the large baseline differences in average math teachers' z-scores in rural (-0.45) and urban (0.19) areas, the contrasting effect of the recruitment of new teachers is perhaps not so surprising: given the initial sorting of high-scores teachers in urban areas and low-scores teachers in rural areas, the need to recruit many new teachers rapidly is likely to have led to an influx of lower-scores teachers in urban, but not in rural areas, where new recruits even appear to have improved the initial low average subject-specific teacher knowledge. At least until 2005, the Ministry of Education and Culture still deployed centrally newly qualified teachers in order to redress inequalities in teacher deployment between rural and urban schools, but many teachers refused assignments far from home–in 2003, as many as 2000 out of 9000 new teachers refused the posts they were assigned (Bennell & Mukyanuzi 2005). Given that most teachers find rural posts less desirable than urban posts (Bennell & Mukyanuzi 2005), the process through which teachers turn down job postings is likely to differ in the rural and urban sectors. In particular, this process should lead to negative sorting on teacher quality in rural- relative to urban areas, since only teachers with high outside options will refuse posts in urban schools, but teachers with both high and modest outside options will refuse posts in rural schools. Provided that the quality of outside options and the human capital of teachers are positively correlated, this alone could account for the lower average subject-specific knowledge in rural- compared to urban schools (Table A-10).[24]

Turning now to pupils' scores, there is no evidence of a change in pupil test scores in rural areas (after controlling for the full set of covariates), whereas there is a clear worsening of pupil test scores in urban areas, which is robust to simultaneously including controls for pupil characteristics, growth in government education transfers to the region, access to physical inputs (which may have been boosted by the capitation grant), and allowing for reversion to the mean (see the one but last rows of Table 11 Panels B and C).

[24]The Basic Education Development Committee (2001) pledged "To provide teacher housing as a deployment incentive, with priority given to female teachers in remote and rural areas" (p.7), raising the possibility that the living conditions of rural teachers might have improved relative to those of their urban counterparts. The SACMEQ surveys asked teachers to rate the condition of their housing on a four-point scale from poor to good. I estimated the effect of enrollment growth on their responses to this question, but found no significant effects or differences between rural and urban areas. Full results are available on request.

It would be tempting to attribute the change in pupils' test scores to that in teacher subject-knowledge, as these two variables are correlated (Fehrler et al. 2009). However, including controls for teacher experience, teacher z-score and the pupil-teacher ratio does not reduce the estimated worsening of test scores in urban areas, as can be seen in the last row of Panels B and C of Table 11. On the other hand, these controls are based on characteristics of the pupils and their schools during 6th Grade, not throughout their primary schooling experience, so that it is not possible to completely rule out a worsening in teachers' subject-specific knowledge or increases in the pupil-teacher ratio as relevant pathways to the worsening in pupils' math test scores in urban areas.

Another possible explanation for the contrasting effects on test scores observed in rural and urban areas is that the composition of students may have changed more in urban- than in rural areas. However, the results reported in Table A-11 suggest that this is unlikely to be the case. When estimating Equation 2 using pupil SES characteristics as dependent variables, the pattern of results for rural and urban areas is similar, despite the stronger decrease in the number of books in the home of the average Grade 6 pupil in urban- than in rural areas, and the larger increase in the proportion of Grade 6 pupils who do not have a father or male legal guardian in rural- compared to urban areas.

Finally, I investigate the possibility that enrollment growth may have been more marked in urban areas. To the best of my knowledge, the annual enrollment data used to construct the treatment variable are not available separately for rural and urban areas. However, the SACMEQ dataset contains the total number of pupils per school as well as the number of Tanzanian pupils represented by each sampled pupil, which can be used to obtain the total number of Grade 6 pupils in the region and rural/urban sector in 2000 and 2007. Results reported in Table A-12 indicate that the average increase in the number of pupils *per school* was larger in a region's urban- than rural areas. On the other hand, the increase in the total number of Grade 6 pupils was larger in a region's *rural* areas. For the rural school sector to accommodate a larger increase in the total number of pupils while experiencing a smaller increase in the number of pupils per individual school than in the urban sector, there must have been a larger ratio of new schools to new entrants in the rural sector than in the urban sector. The more negative effect of regional enrollment growth on learning in urban areas may therefore stem from more intense pressure on existing schools in unobserved ways such as the disruption of school management and organization systems.

As discussed in Section 3, unobserved changes in pupil composition or other omitted variables might bias my OLS estimates downwards. My IV strategy addresses several potential sources of endogeneity such as changes in pupil composition not due to changes in predetermined fertility and migration trends, but there may remain a downward bias so that my IV results may overestimate the worsening of achievement due to enrollment growth (Section 3.4). This raises the question of whether my results for the urban sample are driven by omitted variable bias. The IV and OLS estimates of the effect of enrollment growth on achievement are broadly similar, which is reassuring. In addition, in order to shed light on the robustness of my conclusions for the urban sample to the possible remaining downward bias, I obtained treatment effect estimates under a range of possible departures from the assumption that my instrument is exogenous using the Local-to-Zero procedure suggested by Conley et al. (2012). Conley et al. (2012) suggest several methods to account for the imperfect nature of most instrumental variables. The main intuition is to construct confidence intervals for IV estimates, which span the range of confidence intervals which would be obtained under different degrees of deviation from perfect exogeneity (i.e., when $\gamma \neq 0$ in Equation 5). These confidence intervals can then be used to discuss the informativeness of the IV estimates.

More specifically, the Local-to-Zero procedure used here allows estimating point estimates and confidence intervals for β for a certain prior distribution for γ in Equation 5. The reduced-form estimate of the effect of potential enrollment growth on z-scores gives a natural lower bound (maximum negative magnitude) of γ, and therefore I obtained Local-to-Zero estimates for a range of possible distributions of γ of the form $\gamma \sim \mathcal{U}(\delta, 0)$, with values of δ ranging from the reduced-form point estimate (-0.538 for math scores) to -0.027 ($\frac{1}{20th}$ of -0.538) and clustering the standard errors at the region level, as in the main analysis. As shown in Figure A-10, the point estimate for math scores is equal to -0.2 or less except for very large possible values of γ, and it is significantly negative at 10% for maximum prior values of γ above -0.17. In other words, we can be confident that the effect of enrollment growth on math test scores is negative in urban areas provided the direct effect of potential growth on achievement after conditioning on actual enrollment growth does not exceed about a third of its reduced-form unconditional effect on achievement (since $\frac{-0.17}{-0.538} = 0.32$), which seems like a reasonable assumption. A similar analysis for reading test scores is less conclusive, as could be expected from the smaller magnitude of the negative effect of enrollment growth on reading compared to mathematics. More precisely, Figure A-11 shows that the point estimate for β is between -0.10 and -0.20 for the range of possible δ values, but that the clustered standard errors associated with these estimates are too large for these point estimates to achieve statistical significance.

8 Conclusion

The past two decades have seen large and sudden increases in primary school enrollment in many poor countries, often in the wake of the scrapping of user fees. One such country is Tanzania. Despite considerable concern about this issue in policy circles, there is a dearth of direct, arguably causal evidence on the impact of large, sudden increases in enrollment on the quality of the learning environment.

Comparing changes over time across regions of Tanzania which experienced different rates of growth in the number of pupils enrolled, I find that enrollment growth following FPE has led to sizeable increases in the pupil-teacher ratio (an increase by 6.9 pupils for an increase in enrollment growth by one standard deviation) and a worsening of average teacher experience and subject-specific knowledge in the country taken as a whole.

Estimates of the effect of enrollment growth on learning outcomes, as measured by average pupil test scores for the country as a whole, are small in magnitude and statistically insignificant for both reading and math. More specifically, I find that the lower bounds of the 95% confidence intervals imply that an increase in enrollment growth by 1 standard deviation led at most to a decrease in the reading (math) scores by 0.15 (0.16) of a standard deviation. This corresponds to about a fourth (third) of the rural-urban gap in language (math), or roughly a fifth (fourth) of the language (math) gap between children of fathers with more than primary schooling and the children of fathers who did not complete primary schooling. In other words, I cannot rule out some deterioration in the quality of the learning environment for the average pupil at the national level, but I can rule out a substantial worsening of quality overall. These conclusions are robust to a range of robustness checks, including to instrumenting enrollment growth using predetermined fertility and migration decisions, and I show that plausible sources of instrument endogeneity are unlikely to be driving these findings.

However, when investigating the possibility of heterogeneous effects for urban and rural areas, I find evidence of a deterioration of test scores in urban areas in mathematics (0.27 s.d. for one s.d. increase in enrollment growth), and, to a lesser extent, in reading (0.18 s.d. for one s.d. increase in enrollment growth). One plausible explanation for this differential effect on achievement in urban relative to rural areas is the much higher baseline achievement in urban areas, and hence the larger potential for a worsening of the learning environment due to the pressures of rapid enrollment growth. I also find some evidence that existing urban schools in regions with higher enrollment growth may have had to accommodate larger increases in student numbers than their rural counterparts, which could contribute to the different effect of regional enrollment growth in rural and urban areas. An exploration of the robustness of my findings for the urban sample to departures from the perfectly exogenous instrument case indicates that the conclusion that math test scores worsened due to enrollment growth in urban areas is robust to substantial departures from the perfectly exogenous instrument case.

This study shows that larger increases in enrollment in primary schooling than previously known can be achieved without substantial deterioration of the learning environment for most pupils. Quality losses may however be concentrated within specific environments–here, better-performing urban schools. A fruitful area for future research would be to shed further light on the sources of heterogeneous effects across schools, both in Tanzania and elsewhere.

References

ACER (2015), 'The Southern and Eastern Africa Consortium for Monitoring Educational Quality', *Australian Council for Educational Research, Assessment GEMS no. 8* .

Al-Samarrai, S. & Zaman, H. (2007), 'Abolishing School Fees in Malawi: The Impact on Education Access and Equity', *Education Economics* **15**(3), 359–375.

Angrist, J. D. & Lavy, V. (1999), 'Using Maimonides' Rule To Estimate The Effect Of Class Size On Scholastic Achievement', *The Quarterly Journal of Economics, MIT Press* **114**(2), 533–575.

Athey, S. & Imbens, G. W. (2006), 'Identification and Inference in Nonlinear Difference-in-Differences Models', *Econometrica* **74**(2), 431–497.

Azam, M. & Kingdon, G. G. (2015), 'Assessing Teacher Quality in India', *Journal of Development Economics* **117**, 74–83.

Banerjee, A. V., Cole, S., Duflo, E. & Linden, L. (2007), 'Remedying Education: Evidence from two randomized experiments in India', *The Quarterly Journal of Economics* **122**(3), 1235–1264.

Basic Education Development Committee (2001), *Primary Education Development Plan (2002-2006)*, Dar Es Salaam: Government of the United Republic of Tanzania.

Bennell, P. & Mukyanuzi, F. (2005), *Is There a Teacher Motivation Crisis in Tanzania?*, Unpublished report.

Blanco, G., Flores, C. A. & Flores-Lagunes, A. (2013), 'Bounds on Average and Quantile Treatment Effects of Job Corps Training on Wages', *Journal of Human Resources* **48**(3), 659–701.

Blimpo, M., Gajigo, O. & Pugatch, T. (2016), 'Financial Constraints and Girls' Secondary Education: Evidence from School Fee Elimination in The Gambia', *The World Bank Economic Review* . Forthcoming.

Bold, T., Kimenyi, M., Mwabu, G. & Sandefur, J. (2015), 'Can Free Provision Reduce Demand for Public Services? Evidence from Kenyan education', *The World Bank Economic Review* **29**(2), 293–326.

Bourdon, J., Frölich, M. & Michaelowa, K. (2010), 'Teacher shortages, teacher contracts and their effect on education in africa', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **173**(1), 93–116.

Cameron, A. C., Gelbach, J. B. & Miller, D. L. (2008), 'Bootstrap-Based Improvements for Inference with Clustered Errors', *The Review of Economics and Statistics* **90**(3), 414–427.

Chay, K. Y., McEwan, P. J. & Urquiola, M. (2005), 'The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools', *The American Economic Review* **95**(4), 1237–1258.

Conley, T. G., Hansen, C. B. & Rossi, P. E. (2012), 'Plausibly Exogenous', *Review of Economics and Statistics* **94**(1), 260–272.

Deininger, K. (2003), 'Does Cost of Schooling Affect Enrollment by the Poor? Universal primary education in Uganda', *Economics of Education Review* **22**(3), 291–305.

Duflo, E., Dupas, P. & Kremer, M. (2012), 'School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental evidence from Kenyan primary schools', *NBER Working Paper N^o 17939* .

Fehrler, S., Michaelowa, K. & Wechtler, A. (2009), 'The effectiveness of inputs in primary education: Insights from recent student surveys for sub-saharan africa', *The Journal of Development Studies* **45**(9), 1545–1578.

Glewwe, P., Kremer, M. & Moulin, S. (2009), 'Many Children Left Behind? Textbooks and test scores in Kenya', *American Economic Journal: Applied Economics* **1**(1), 112–135.

Glewwe, P., Kremer, M., Moulin, S. & Zitzewitz, E. (2004), 'Retrospective vs. Prospective Analyses of School Inputs: The case of flip charts in Kenya', *Journal of Development Economics* **74**(1), 251–268.

Glewwe, P. W., Hanushek, E. A., Humpage, S. D. & Ravina, R. (2011), 'School Resources and Educational Outcomes in Developing Countries: A review of the literature from 1990 to 2010', *NBER Working Paper Nº 17554* .

Grogan, L. (2009), 'Universal Primary Education and School Entry in Uganda', *Journal of African Economies* **18**(2), 183–211.

Hanushek, E. A. & Woessmann, L. (2008), 'The Role of Cognitive Skills in Economic Development', *Journal of economic Literature* pp. 607–668.

Hoogeveen, J. & Rossi, M. (2013), 'Enrolment and Grade Attainment following the Introduction of Free Primary Education in Tanzania', *Journal of African Economies* **22**(3), 375–393.

Hoxby, C. M. (2000), 'The Effects of Class Size on Student Achievement: New evidence from population variation', *Quarterly Journal of economics* pp. 1239–1285.

Joshi, A. R. & Gaddis, I. (2015), *Preparing the Next Generation in Tanzania: Challenges and Opportunities in Education*, World Bank Publications.

Kattan, R. B. & Burnett, N. (2004), 'User Fees in Primary Education', *World Bank Education Advisory Service* .

Kremer, M., Brannen, C. & Glennerster, R. (2013), 'The Challenge of Education and Learning in the Developing World', *Science* **340**(6130), 297–300.

Krueger, A. B. (1999), 'Experimental Estimates of Education Production Functions', *The Quarterly Journal of Economics* **114**(2), 497–532.

Lee, D. (2009), 'Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects', *Review of Economic Studies* **76**(3), 1071–1102.

Leuven, E., Oosterbeek, H. & Rnning, M. (2008), 'Quasi-experimental Estimates of the Effect of Class Size on Achievement in Norway', *Scandinavian Journal of Economics* **110**(4), 663–693.

Lucas, A. M. & Mbiti, I. M. (2012), 'Access, Sorting, and Achievement: The short-run effects of free primary education in Kenya', *American Economic Journal: Applied Economics* **4**(4), 226–253.

Makuwa, D. (2010), 'What Are the Levels and Trends in Reading and Mathematics Achievement?', *SACMEQ Policy Issues Series* (2).

Masino, S. & Niño-Zarazúa, M. (2016), 'What works to improve the quality of student learning in developing countries?', *International Journal of Educational Development* **48**, 53–65.

Metzler, J. & Woessmann, L. (2012), 'The Impact of Teacher Subject Knowledge on Student Achievement: Evidence from within-teacher within-student variation', *Journal of Development Economics* **99**(2), 486–496.

Ministry of Education [Tanzania] (1999-2007), Basic Education Statistics in Tanzania, Technical report, Ministry of Education [Tanzania]. Dar Es Salaam.

MoEVT and UNESCO (2012), *Tanzania Education Sector Analysis: Beyond primary education*, United Republic of Tanzania, Ministry of Education and Vocational Training.

Mrutu, A., Ponera, G. & Nkumbi, E. (2015), 'The SACMEQ II Project in Tanzania: A study of the conditions of schooling and the quality of education', *SACMEQ Educational Policy Research Series* .

National Bureau of Statistics (2001), 'Tanzania Household Budget Survey 2001 [database]'.

Nishimura, M., Yamano, T. & Sasaoka, Y. (2008), 'Impacts of the Universal Primary Education Policy on Educational Attainment and Private Costs in Rural Uganda', *international Journal of Educational development* **28**(2), 161–175.

Oketch, M. & Somerset, A. (2010), 'Free Primary Education and After in Kenya: Enrolment impact, quality effects, and the transition to secondary school', *CREATE Research Monograph N^o 37* .

Pritchett, L. (2013), *The Rebirth of Education: Schooling ain't learning*, Center for Global Development (CDG) Books.

Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (2005), 'Teachers, Schools, and Academic Achievement', *Econometrica* **73**(2), 417–458.

Rockoff, J. E. (2004), 'The Impact of Individual Teachers on Student Achievement: Evidence from panel data', *American Economic Review* pp. 247–252.

Schaffer, M. (2010), ‘xtivreg2: Stata module to perform extended iv/2sls, gmm and ac/hac, liml and k-class regression for panel data models’.

Todd, P. E. & Wolpin, K. I. (2007), ‘The Production of Cognitive Achievement in Children: Home, school, and racial test score gaps’, *Journal of Human Capital* **1**(1), 91–136.

Twaweza (2010), ‘Capitation grant for education: When will it make a difference?, Twaweza Policy brief TZ.08/2010E’.

UNESCO Institute for Statistics (n.d.), ‘One in five children, adolescents and youth is out of school’, *Fact Sheet No. 48, February 2018* .

Wedgwood, R. (2007), ‘Education and Poverty Reduction in Tanzania’, *International Journal of Educational Development* **27**(4), 383 – 396.

World Bank (2009), ‘Abolishing school fees in africa lessons from ethiopia, ghana, kenya, malawi, and mozambique’.

World Development Indicators (2015), *World Development Indicators*, The World Bank.

Table 1: Regional Demographic Variation

	Total Births			Births per Woman Age 15-49		
Region	1988	2000	% Change	1988	2000	% Change
Tabora	35816	66672	86	0.18	0.21	14
Rukwa	27649	47506	72	0.21	0.22	1
Shinyanga	68825	112640	64	0.2	0.22	10
Arusha	50559	82379	63	0.2	0.18	-10
Mwanza	75502	111272	47	0.21	0.21	-1
Kigoma	35253	50333	43	0.22	0.2	-9
Dar es Salaam	42038	58767	40	0.16	0.1	-38
Ruvumba	27526	37377	36	0.17	0.16	-7
Morogoro	40685	54714	34	0.17	0.16	-8
Kagera	51507	69257	34	0.22	0.2	-10
Mbeya	50580	67626	34	0.17	0.16	-6
Mara	38309	51075	33	0.21	0.2	-7
Dodoma	44391	58227	31	0.18	0.17	-5
Singida	30545	37431	23	0.19	0.18	-4
Pwani	21785	26504	22	0.19	0.16	-16
Tanga	44417	53530	21	0.18	0.16	-10
Mtwara	28190	33348	18	0.16	0.13	-14
Iringa	41428	45823	11	0.18	0.15	-12
Kilimanjaro	38787	38378	-1	0.19	0.15	-23

Source: Author's calculations using Tanzania Census Extract (1988) and (2002). Total Births in 1988 (2000) are the total number of children less than one year old (age 2) in the 1988 (2002) Census. Births per Woman Age 15-49 is the average number of children born to women of ages 15-49 in 1988 and 2000 based on household composition in the 1988 and 2002 Census, respectively.

Table 2: Differential Trends in Health Inputs and Outcomes

	(1) =1 if Full Immunization	(2) =1 if Delivered by Health Professional	(3) =1 if Delivered Without Any Help	(4) =1 if Infant Death	(5) =1 if Stunted
(=1 if born 1995-2000)×Age 7-13 in 2002-2007	-0.159**	-0.257***	0.051	0.021	0.068
/Age 7-13 in 2001	[-0.292,-0.027]	[-0.428,-0.085]	[-0.026,0.129]	[-0.031,0.074]	[-0.019,0.155]
=1 if born 1995-2000	1.047**	1.685***	-0.362	-0.141	-0.549*
	[0.166,1.928]	[0.523,2.847]	[-0.890,0.166]	[-0.498,0.216]	[-1.143,0.045]
Region Dummies	Yes	Yes	Yes	Yes	Yes
Observations	10944	16294	16294	12556	12730
R-squared	0.03	0.06	0.06	0.00	0.04
Mean Y	0.72	0.49	0.06	0.09	0.43

Region-correlated robust 95% confidence intervals in brackets. Source: Author's calculations using Demographic and Health Surveys of Tanzania children recode (1991-2, 1996, 1999, and 2004-2005). Sample of children born between 1988 and 2000. Samples vary due to immunization sample including only children aged at least one year old, infant death being only defined for children who have been fully exposed to the risk of infant death (i.e., born at least 12 months before the survey), height being only measured for children alive at the time of the survey, as well as 433, 78 and 1791 truly missing or unreasonable values for immunizations, help with delivery and height measurement, respectively.

Table 3: Summary Statistics

	SACMEQ 2000		SACMEQ 2007	
	mean	sd	mean	sd
Regional Education Characteristics				
(=1 if 2007)×Enrol. 2002-2007 /Enrol. in 2001	0.00	0.000	9.00	0.693
(=1 if 2007)×Age 7-13 in 2002-2007 /Age 7-13 in 2001	0.00	0.000	6.75	0.304
(=1 if 2007)×Ed. grants 2002-2007 /Ed. grant 2001, deflated	0.00	0.000	11.36	1.366
Teacher Quantity Variables				
Pupil-Teacher Ratio	47.05	19.758	62.24	31.842
=1 if Multiple Shifts	0.18		0.06	
Kiswahili Teaching Hours	16.93	7.287	16.05	6.176
Math Teaching Hours	17.15	7.124	16.40	5.798
Teacher Quality Variables				
=1 if O-level (K)	0.75		0.93	
=1 if O-level (M)	0.92		0.96	
=1 if Training≥2 y (K)	0.94		0.81	
=1 if Training≥2 y (M)	0.97		0.77	
Teacher Experience (K)	14.10	7.800	12.36	10.672
Teacher Experience (M)	12.47	7.249	10.98	9.933
Teacher Reading Z-score	-0.17	0.909	0.15	1.038
Teacher Math Z-score	-0.27	1.011	0.14	0.961
Physical Inputs Variables				
Pupil-Specific Variables				
=1 if Pupil Has Own Kiswahili Book	0.06		0.03	
=1 if Pupil Has Own Math Book	0.07		0.03	
=1 if No Reading Textbook	0.36		0.23	
=1 if No Math Textbook	0.33		0.23	
# Exercises Books	8.88	2.997	7.16	3.140
Total Pupil Equipment Score (0 to 8)	5.45	1.784	6.15	1.512
Class-Specific Variables				
Total Kiswahili Equipment Score (0 to 8)	3.59	1.788	4.83	1.649
Total Math Equipment Score (0 to 8)	3.33	1.660	4.73	1.686
Learning Outcomes				
Pupil Reading Z-score	-0.21	0.996	0.15	0.975
Pupil Math Z-score	-0.21	1.005	0.14	0.973
=1 if Low Competency (M)	0.25		0.13	

	SACMEQ 2000		SACMEQ 2007	
	mean	sd	mean	sd
=1 if Low Competency (K)	0.18		0.10	
=1 if High Competency (M)	0.18		0.31	
=1 if High Competency (K)	0.22		0.33	
Pupil Characteristics				
=1 if Rural	0.71		0.69	
=1 if Male Pupil	0.48		0.49	
Pupil's Age	14.44	1.537	13.94	1.596
=1 if English is Never Spoken at Home	0.10		0.08	
Household Items Ownership (0 to 14)	3.42	2.671	5.07	2.195
Parental Education Variables				
=if if Father < Completed Primary	0.24		0.17	
=if if Mother < Completed Primary	0.23		0.25	
=if if Father = Completed Primary	0.39		0.53	
=if if Mother = Completed Primary	0.51		0.60	
=if if Father > Completed Primary	0.30		0.23	
=if if Mother > Completed Primary	0.19		0.13	
=1 if Does Not Know Dad's Educ. Level	0.06		0.04	
=1 if Does Not Know Mum's Educ. Level	0.07		0.02	
=1 if No Father or Male Guardian	0.02		0.02	
=1 if No Mother or Female Guardian	0.01		0.00	
N	2849		4084	

Source: Author's calculations using SACMEQ II, SACMEQ III, IPUMS (2011), Ministry of Education "Basic Education Statistics in Tanzania" and Budget Plans for various years. Statistics weighted by the same SACMEQ pupil weights as in the regressions. The pupil-teacher ratio is calculated as the ratio of the total number of pupils to the total number of teachers in the school based on the information collected during interviews with head teachers. Class equipment score items: writing board, chalk, wall chart, cupboard, bookshelves, library, teacher table, teacher chair. Pupil equipment score items: exercise book, notebook, pencil, sharpener, eraser, ruler, pen, folder. Household items ownership items: newspaper, magazine, radio, TV set, VCR, cassette player, telephone, refrigerator/freezer, car, motorcycle, bicycle, piped water, electricity, table to write on.

Table 4: Effect on Quantity of Teachers

	(1) Pupil/Teacher	(2) =1 if Multiple Shifts	(3) Teaching Hours (K)	(4) Teaching Hours (M)
Panel A: OLS Estimates				
(=1 if 2007)×Enrol. 2002-2007	9.989***	-0.031	1.354	-0.073
/Enrol. in 2001	[3.172,16.806]	[-0.121,0.058]	[-2.210,4.918]	[-2.216,2.070]
Wild cluster bootstrap-t p-values	0.008	0.462	0.436	0.882
Panel B: 2SLS Estimates				
(=1 if 2007)×Enrol. 2002-2007	12.670**	0.068	-0.204	-1.726
/Enrol. in 2001	[0.941,24.400]	[-0.146,0.282]	[-3.916,3.508]	[-3.821,0.369]
Panel C: Reduced-Form Estimates				
(=1 if 2007)×Age 7-13 in 2002-2007	21.000**	0.113	-0.338	-2.861
/Age 7-13 in 2001	[1.240,40.760]	[-0.241,0.467]	[-7.164,6.488]	[-6.641,0.919]
Mean Y	53.179	0.096	16.660	16.723
Kleibergen-Paap F-Stat	17.296	17.296	17.296	17.296
KP underid stat H0=underid	4.658	4.658	4.658	4.658
KP underid p-val	0.031	0.031	0.031	0.031
p-val robust test of regressor endogeneity	0.486	0.181	0.182	0.022

Sample size: 6933. Region-correlated robust 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * $p<0.10$, ** $p<0.05$, *** $p<0.01$. 2SLS estimates were computed using the command XTIVREG2 (Schaffer 2010). Wild cluster bootstrap-t p-values were obtained using the method proposed in Cameron et al. (2008).

Table 5: Effect on Quality of Teachers: Initial Training

	(1)	(2)	(3)	(4)
	=1 if O-level (K)	=1 if O-level (M)	=1 if Training≥2 y (K)	=1 if Training≥2 y (M)
Panel A: OLS Estimates				
(=1 if 2007)×Enrol. 2002-2007	0.019	0.034	-0.008	-0.030
/Enrol. in 2001	[-0.056,0.095]	[-0.058,0.125]	[-0.138,0.122]	[-0.114,0.053]
Wild cluster bootstrap-t p-values	0.540	0.496	0.904	0.458
Panel B: 2SLS Estimates				
(=1 if 2007)×Enrol. 2002-2007	0.113	-0.021	-0.087	-0.077
/Enrol. in 2001	[-0.050,0.277]	[-0.095,0.054]	[-0.248,0.074]	[-0.226,0.071]
Panel C: Reduced-Form Estimates				
(=1 if 2007)×Age 7-13 in 2002-2007	0.188	-0.034	-0.144	-0.128
/Age 7-13 in 2001	[-0.060,0.436]	[-0.180,0.111]	[-0.427,0.138]	[-0.374,0.118]
Mean Y	0.845	0.942	0.865	0.852
Kleibergen-Paap F-Stat	17.296	17.296	17.296	17.296
KP underid stat H0=underid	4.658	4.658	4.658	4.658
KP underid p-val	0.031	0.031	0.031	0.031
p-val robust test of regressor endogeneity	0.069	0.165	0.067	0.476

Sample size: 6933. Region-correlated robust 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 6: Effect on Quality of Teachers: Experience and Subject-Specific Knowledge

	(1) Years Experience (K)	(2) Years Experience (M)	(3) Teacher Reading Score	(4) Teacher Math Score
Panel A: OLS Estimates				
(=1 if 2007)×Enrol. 2002-2007	-2.042**	-4.018***	-0.207**	0.048
/Enrol. in 2001	[-3.796,-0.288]	[-5.508,-2.528]	[-0.367,-0.047]	[-0.177,0.274]
Wild cluster bootstrap-t p-values	0.058	0.002	0.036	0.655
Panel B: 2SLS Estimates				
(=1 if 2007)×Enrol. 2002-2007	-2.796**	-4.419***	-0.226**	-0.029
/Enrol. in 2001	[-5.161,-0.431]	[-6.654,-2.183]	[-0.440,-0.013]	[-0.410,0.352]
Panel C: Reduced-Form Estimates				
(=1 if 2007)×Age 7-13 in 2002-2007	-4.634**	-7.324***	-0.375*	-0.048
/Age 7-13 in 2001	[-8.133,-1.134]	[-10.887,-3.760]	[-0.761,0.011]	[-0.732,0.637]
Mean Y	13.537	12.049	0.008	-0.001
Kleibergen-Paap F-Stat	17.296	17.296	17.296	17.296
KP underid stat H0=underid	4.658	4.658	4.658	4.658
KP underid p-val	0.031	0.031	0.031	0.031
p-val robust test of regressor endogeneity	0.297	0.597	0.796	0.475

Sample size: 6933. Region-correlated robust 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 7: Effect on Physical Inputs

	(1) =1 if Own Bk (K)	(2) =1 if Own Bk (M)	(3) Class Equip. (K)	(4) Class Equip. (M)	(5) # Ex. Books	(6) Pupil Equip.
Panel A: OLS Estimates						
(=1 if 2007)×Enrol. 2002-2007	0.001	-0.006	0.155	-0.051	-0.095	0.014
/Enrol. in 2001	[-0.021,0.022]	[-0.022,0.010]	[-0.444,0.755]	[-0.405,0.303]	[-0.619,0.428]	[-0.214,0.242]
Wild cluster bootstrap-t p-values	0.940	0.446	0.600	0.804	0.706	0.918
Panel B: 2SLS Estimates						
(=1 if 2007)×Enrol. 2002-2007	0.006	-0.014	0.679**	0.346	0.391	0.201
/Enrol. in 2001	[-0.030,0.043]	[-0.039,0.012]	[0.032,1.325]	[-0.220,0.911]	[-0.304,1.087]	[-0.142,0.545]
Panel A: Reduced-Form Estimates						
(=1 if 2007)×Age 7-13 in 2002-2007	0.011	-0.023	1.125**	0.573	0.649	0.334
/Age 7-13 in 2001	[-0.055,0.076]	[-0.069,0.024]	[0.163,2.086]	[-0.323,1.469]	[-0.481,1.779]	[-0.191,0.859]
Mean Y	0.042	0.041	4.311	4.175	7.977	5.831
Kleibergen-Paap F-Stat	17.296	17.296	17.296	17.296	17.296	17.296
KP underid stat H0=underid	4.658	4.658	4.658	4.658	4.658	4.658
KP underid p-val	0.031	0.031	0.031	0.031	0.031	0.031
p-val robust test of regressor endogeneity	0.662	0.469	0.093	0.048	0.089	0.157

Sample size: 6933. Region-correlated robust 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Class equipment and pupil equipment defined under Table 3.

Table 8: Effect on Pupil Reading Test Scores

	Pupil Reading Test Scores					
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	2SLS	OLS	OLS	OLS
(=1 if 2007)×Enrol. 2002-2007	-0.002	-0.004	0.050	-0.035	-0.070	
/Enrol. in 2001	[-0.168,0.165]	[-0.134,0.126]	[-0.139,0.239]	[-0.206,0.135]	[-0.221,0.081]	
Wild cluster bootstrap-t p-values	0.940	0.936		0.702	0.418	
(=1 if 2007)×Ed. grants 2002-2007				0.046		
/Ed. grant 2001, deflated				[-0.024,0.117]		
(=1 if 2007)×Baseline Average					-0.606***	
					[-0.883,-0.328]	
(=1 if 2007)×Age 7-13 in 2002-2007						0.083
/Age 7-13 in 2001						[-0.276,0.442]
Wild cluster bootstrap-t p-values						0.642
Pupil Characteristics	No	Yes	No	No	No	No
Physical Inputs	No	Yes	No	No	No	No
R-squared	0.1400	0.2133	0.0965	0.1408	0.1484	0.1401
Mean Y	0.002	0.002	0.002	0.002	0.002	0.002
Kleibergen-Paap F-Stat			17.296			
KP underid stat H0=underid			4.658			
KP underid p-val			0.031			
p-val robust test of regressor endogeneity			0.437			

Sample size: 6933. Region-correlated robust 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * p<0.10, ** p<0.05, *** p<0.01. Pupil characteristics and physical inputs variables as listed in Table 3.

Table 9: Effect on Pupil Math Test Scores

	Pupil Math Test Scores					
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	2SLS	OLS	OLS	OLS
(=1 if 2007)×Enrol. 2002-2007	0.041	0.039	0.007	-0.022	-0.030	
/Enrol. in 2001	[-0.167,0.250]	[-0.143,0.221]	[-0.298,0.312]	[-0.236,0.192]	[-0.213,0.152]	
Wild cluster bootstrap-t p-values	0.666	0.728		0.872	0.772	
(=1 if 2007)×Ed. grants 2002-2007				0.087		
/Ed. grant 2001, deflated				[-0.019,0.193]		
(=1 if 2007)×Baseline Average					-0.744***	
					[-1.113,-0.376]	
(=1 if 2007)×Age 7-13 in 2002-2007						0.012
/Age 7-13 in 2001						[-0.548,0.572]
Wild cluster bootstrap-t p-values						0.944
Pupil Characteristics	No	Yes	No	No	No	No
Physical Inputs	No	Yes	No	No	No	No
R-squared	0.1026	0.1899	0.0700	0.1054	0.1110	0.1024
Mean Y	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
Kleibergen-Paap F-Stat			17.296			
KP underid stat H0=underid			4.658			
KP underid p-val			0.031			
p-val robust test of regressor endogeneity			0.648			

Sample size: 6933. Region-correlated robust 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * p<0.10, ** p<0.05, *** p<0.01. Pupil characteristics and physical inputs variables as listed in Table 3.

Table 10: Instrument Robustness Checks

	Reduced-Form, Control for $\frac{\sum_{j=1995}^{2000} Age7\_13_{rj}}{Age7\_13_{r1994}}$		IV Based on Region in 2001	
	(1)	(2)	(3)	(4)
	Reading Score	Math Score	Reading Score	Math Score
(=1 if 2007)×Age 7-13 in 2002-2007	0.068	-0.006		
/Age 7-13 in 2000	[-0.267,0.403]	[-0.508,0.496]		
(=1 if 2007)×Enrol. 2002-2007			0.046	0.008
/Enrol. in 2001			[-0.148,0.240]	[-0.302,0.318]
R-squared	0.1412	0.1040	0.0965	0.0700
Mean Y	0.002	-0.004	0.002	-0.004

Sample size: 6933. Region-correlated robust 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * p<0.10, ** p<0.05, *** p<0.01.

Table 11: Heterogeneity between Rural and Urban Areas

	(1)	(2)	(3)	(4)
	Rural Sample		Urban Sample	
	OLS	IV	OLS	IV
Panel A: Effect on Teacher Variables				
Dependent Variable				
Pupil-Teacher Ratio	10.063**	9.854*	7.260	16.660*
	[2.069,18.058]	[-1.326,21.035]	[-4.664,19.183]	[-0.612,33.933]
Years Experience (K)	-1.526	-1.417	-0.722	-3.194
	[-3.873,0.822]	[-5.005,2.171]	[-5.016,3.571]	[-7.918,1.529]
Years Experience (M)	-4.373***	-3.337**	-1.298	-4.039**
	[-6.461,-2.285]	[-6.029,-0.646]	[-3.888,1.292]	[-7.861,-0.217]
Teacher Z-Score (K)	-0.259**	-0.275*	-0.064	-0.062
	[-0.497,-0.022]	[-0.563,0.013]	[-0.537,0.408]	[-0.473,0.349]
Teacher Z-Score (M)	0.252**	0.271**	-0.650***	-0.994**
	[0.043,0.462]	[0.023,0.520]	[-1.110,-0.191]	[-1.756,-0.232]
Panel B: Effect on Reading Scores				
Specification				
OLS, No Controls	0.038		-0.193	
	[-0.150,0.226]		[-0.466,0.079]	
OLS, Controls	0.027		-0.150	
	[-0.093,0.148]		[-0.378,0.078]	
IV		0.100		-0.227
		[-0.105,0.304]		[-0.517,0.062]
KP F-Stat		14.315		23.392
Endog. P-Value		0.429		0.677
OLS, Government	0.018		-0.247*	
Transfers	[-0.169,0.206]		[-0.530,0.036]	
OLS, Mean Rev.	-0.026		-0.314**	
	[-0.189,0.136]		[-0.574,-0.054]	
OLS, Full Set	-0.018		-0.261*	
	[-0.130,0.093]		[-0.530,0.008]	
OLS, Teacher Controls	0.017		-0.250*	
	[-0.086,0.120]		[-0.514,0.013]	

	(1)	(2)	(3)	(4)
	Rural Sample		Urban Sample	
	OLS	IV	OLS	IV
Panel C: Effect on Math Scores				
Specification				
OLS, No Controls	0.144*		-0.309	
	[-0.029,0.316]		[-0.684,0.066]	
OLS, Controls	0.127*		-0.267	
	[-0.013,0.268]		[-0.591,0.057]	
IV		0.118		-0.401**
		[-0.168,0.404]		[-0.773,-0.029]
KP F-Stat		14.315		23.392
Endog. P-Value		0.726		0.335
OLS, Gov. Transfers	0.087		-0.351*	
	[-0.089,0.263]		[-0.736,0.034]	
OLS, Mean Rev.	0.061		-0.442***	
	[-0.070,0.193]		[-0.748,-0.136]	
OLS, Full Set	0.029		-0.395**	
	[-0.099,0.156]		[-0.724,-0.067]	
OLS, Teacher Controls	0.033		-0.444**	
	[-0.112,0.177]		[-0.799,-0.089]	

Sample size: 5054 (rural sample) and 1879 (urban sample). Region-correlated robust 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * p<0.10, ** p<0.05, *** p<0.01. "Controls" refers to the inclusion of the same pupil characteristics and physical inputs variables as in the second column of Tables 8 and 9, while "Full set" of controls refers to the inclusion of these pupil characteristics and physical inputs variables as well as growth in government transfers and ((=1 if 2007)×Baseline Average Score) to control for mean reversion. "Teacher Controls" refers to the inclusion of all the regressors included in the "Full set" of regressors as well as the relevant (i.e., Kiswahili or Math) teacher's years of experience, teacher z-score, and pupil-teacher ratio.

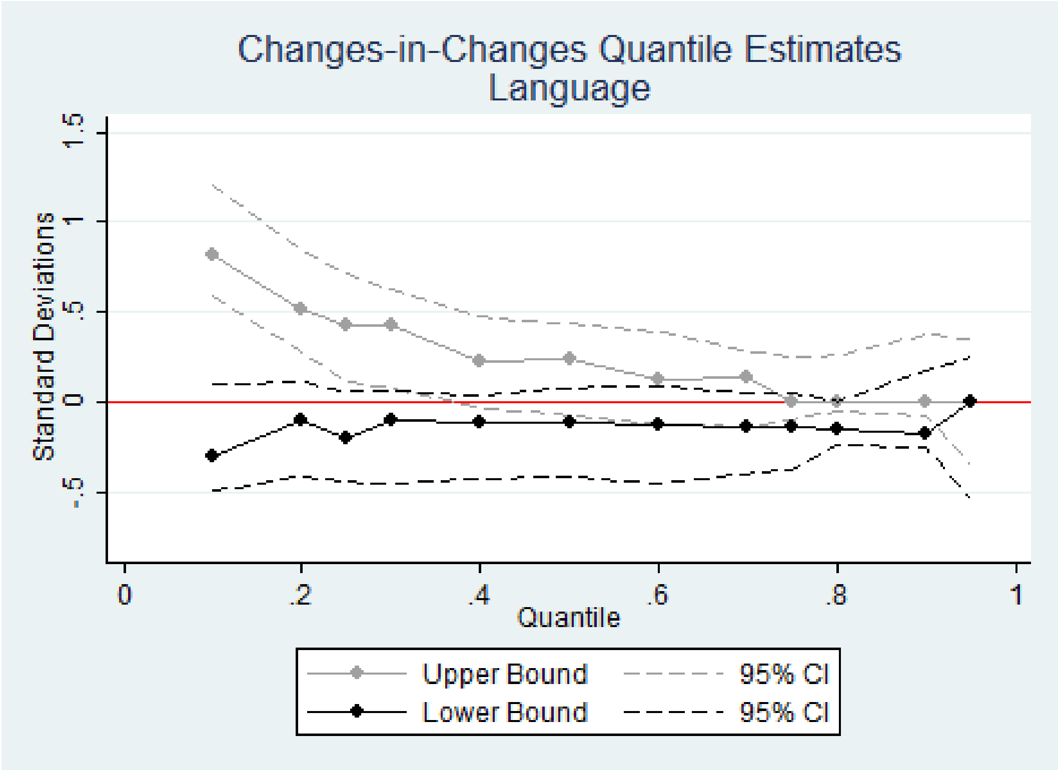

Figure 1

Notes: the treated group is defined as those pupils in regions with above-median enrollment growth. Lower bound obtained by estimating Changes-in-Changes quantile treatment effects on the full sample. Upper bound obtained by trimming the bottom of the 2007 distribution for the treated group by $[\frac{Pr(S_i=1|T_i=1)-Pr(S_i=1|T_i=0)}{Pr(S_i=1|T_i=1)}]_{2007} - [\frac{Pr(S_i=1|T_i=1)-Pr(S_i=1|T_i=0)}{Pr(S_i=1|T_i=1)}]_{2000}$, where $S_i = 1$ if individual i is observed in the data, and $T_i = 1$ if individual i is treated, and zero otherwise (see Appendix A-4 for further detail). The confidence intervals correspond to the 5th and 95th percentiles of the cluster-bootstrapped distributions of quantile effects and are therefore not centered around the point estimates. Please refer to Footnote 22 for a detailed explanation of how estimates are obtained.

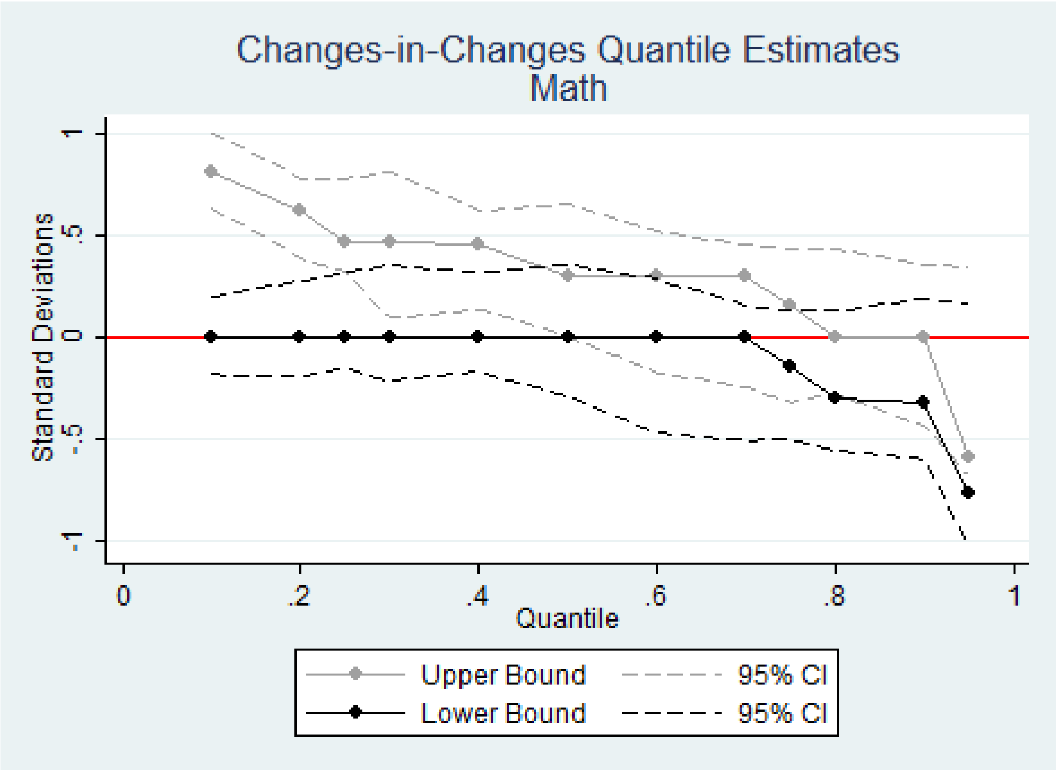

Figure 2

Notes: the treated group is defined as those pupils in regions with above-median enrollment growth. Lower bound obtained by estimating Changes-in-Changes quantile treatment effects on the full sample. Upper bound obtained by trimming the bottom of the 2007 distribution for the treated group by $[\frac{Pr(S_i=1|T_i=1)-Pr(S_i=1|T_i=0)}{Pr(S_i=1|T_i=1)}]_{2007} - [\frac{Pr(S_i=1|T_i=1)-Pr(S_i=1|T_i=0)}{Pr(S_i=1|T_i=1)}]_{2000}$, where $S_i = 1$ if individual i is observed in the data, and $T_i = 1$ if individual i is treated, and zero otherwise (see Appendix A-4 for further detail). The confidence intervals correspond to the 5th and 95th percentiles of the cluster-bootstrapped distributions of quantile effects and are therefore not centered around the point estimates. Please refer to Footnote 22 for a detailed explanation of how (exact zero and other) estimates are obtained.

Appendix (For Online Publication Only)

A-1 Literature Review

A-1.1 Impact of Rapid Enrollment Growth on School Quality

The body of existing evidence on the effect of rapid enrollment growth on schooling outcomes comes nearly exclusively from indirect evidence based on the effect of the removal of user fees. And most of the literature on the impact of FPE reforms focuses on their impact on enrollment and in general conclude that FPE increased access to primary schooling, especially for poorer children (see Deininger (2003), Grogan (2009), Nishimura et al. (2008) for Uganda; Al-Samarrai & Zaman (2007) for Malawi; Lucas & Mbiti (2012) for Kenya; and Hoogeveen & Rossi (2013) for Tanzania). All of these papers explicitly raise the question of whether the quality of schooling was affected, but no direct evidence of the effect of FPE on test scores is provided except for Kenya. In this country, Lucas & Mbiti (2012) estimate the impact of FPE on test scores obtained at the end of primary school exam by students who had been in school for three to seven years before FPE. Using a changes-in-changes estimation approach exploiting the variation in FPE treatment intensity due to differential drop-out rates across Kenyan districts before the country-wide reform, they find that students who would have taken the exam in the absence of FPE lost no more than 0.05 of a standard deviation in districts with above-average predicted FPE intensity relative to students in districts with below-average predicted FPE intensity. However, as acknowledged by the authors, the students in their sample were not fully exposed to the impact of the reform, since they had been in school long before FPE took place. In addition, the increase in enrollment following FPE in Kenya was only 13% in Grades 2 to 8, so that the cohorts considered treated in Lucas & Mbiti (2012), who were in Grades 4 to 8 at the time of the removal of the school fees, did not experience much of the initial enrollment growth. Other findings in Lucas & Mbiti (2012) are that FPE in Kenya increased the number of students who completed primary school, led to a growth in the private schooling sector, and increased the share of primary school students whose parents are illiterate. On the other hand, Bold et al. (2015) estimate that the net enrollment rate in public primary schools increased for poor households but fell for wealthier households in favor of private schools, thus resulting in the stagnation of the enrollment rate in public schools after FPE and suggesting a decrease in the perceived benefits of public primary schools for those most able to choose between free and fee-paying schools.

In Tanzania, Hoogeveen & Rossi (2013) estimate the impact of FPE on attendance and grade completion. Their household data confirm that enrollment rates at age 7 are higher in 2007 than in 2001, and in a multivariate analysis in which the dependent variable is a school enrollment indicator, they find that variables capturing the socio-economic status (SES) of the household are less strongly correlated with attendance at age 7 in 2007 than in 2001, thus suggesting that the reform was effective in increasing enrollment among lower SES children. However, comparing years of education accumulated between 2001 and 2007 between children aged less than 11 in 2002, who are considered "treated", and older children, who are considered a control group because their enrollment was not prioritized by the reform, Hoogeveen & Rossi (2013) find a statistically significant decrease in grade attainment–especially in rural areas, which they hypothesize to be due to a deterioration of the quality of schooling.

Finally, a recent study sheds light on the effect of *secondary* school expansion on test scores in the Gambia. Blimpo et al. (2016) find evidence of small test scores gains in response to a program which pays directly to schools the secondary school fees for girls. Although interesting in itself, this study does not speak to the effect of large increases in enrollment on school quality since the authors estimate that the program only increased the number of test takers by 26 girls and 43 boys per district, and had no effect on pupil-teacher ratios, thus suggesting that the modest increases in enrollment resulting from the program could be effectively managed from the onset of the expansion, contrary to the typical experience of countries pursuing rapid widening of *primary* school access.

A-1.2 Impact of Schooling Inputs on Learning

The body of literature concerned with estimating causal effects of class-size, access to physical inputs, and teacher quality, on learning is vast and a full literature review is beyond the scope of this analysis. A review by Kremer et al. (2013) of randomized controlled trials carried out in developing countries concludes that "test scores are remarkably low and unresponsive to more-of-the-same inputs, such as hiring additional teachers, buying more textbooks, or providing flexible grants" (p. 297).[25] A recent systematic review by Masino & Niño-Zarazúa (2016) emphasizes the importance, for their effectiveness in improving education quality, of combining additional physical inputs and human resources with incentives influencing the behavior of teachers, households, and students and/or with community management interventions.

One of the most researched aspects of the achievement production function is the effect of class size on test scores. In developed country settings, the range of estimates is generally between 0.07 and 0.27 of a standard deviation increase in test scores for a decrease of 7 pupils[26]. At the lower end, Hoxby (2000) can rule out effects of 2 to 4 percent of a standard deviation in scores for a 10% increase in class size in Connecticut, where the average class size is 21 pupils, which we can roughly translate as ruling out effects of 0.07-0.14 s.d. for a 7-pupil decrease. Similarly, Leuven et al. (2008) can rule out effects of 0.11 s.d. for a 7-pupil decrease in Norway. Angrist & Lavy (1999) report that their estimates probably translate into an improvement of 0.18 of a standard deviation in the pupil distribution of test scores for an 8-pupil class size reduction, while Krueger (1999) reports effect sizes of 0.19 to 0.28 standard deviations for the STAR experiment (where the difference in average class size between the "small class" treatment group and the "normal class" control group was about 7 pupils (see Table 3 in Krueger (1999)). Perhaps more illustrative than these effect sizes, the effects found in Krueger (1999) translate into 64% (82%) of the white-black gap in kindergarten (third grade). In developing country settings, two randomized experiments nearly halving class size, one in Kenya and one in India, did not find any statistically significant effects on test scores (Banerjee et al. 2007, Duflo et al. 2012).

Studies considering the impact of teachers' observable measures of quality such as education and training generally find little evidence that these characteristics play a role in students' learning except for teacher experience in developed countries. Rivkin et al. (2005) show that the variance of learning outcomes across teachers is large but uncorrelated to teacher education, teacher experience beyond the two or three initial years, or class-size in Texas. Rockoff (2004)

[25]Instead, Kremer et al. (2013) emphasize the positive role of pedagogical reforms that make teaching better suited to students' learning levels and reforms that improve accountability and incentives.

[26]Seven pupils is a convenient point of reference as it corresponds both to the estimated increase in the pupil-teacher ratio for a one standard deviation increase in enrollment growth found in Section 5 and to the difference in average class size between the "small class" treatment group and the "normal class" control group in the well-know Tennessee STAR experiment (see Table 3 in Krueger (1999)).

also finds substantial variance in learning outcomes across teachers in a New Jersey county, and that an additional year of teacher experience increases reading scores by 0.018 standard deviations, while the effect of experience is statistically insignificant and non-monotonic in math. Applying a pupil fixed-effects approach to data from one district of Uttar Pradesh in India, Azam & Kingdon (2015) similarly find that an additional standard deviation in teacher quality increases exam scores of secondary school pupils by 0.37 standard deviations, but that observable teacher characteristics such as age, experience, and qualifications account for very little of this variation. A review of the (exclusively non-experimental) evidence on the effect of teacher subject-specific knowledge on students' test scores, however, reports consistently positive effects (Glewwe et al. 2011). For instance, in a study controlling for student, teacher, and subject fixed effects (and thus for a wide range of potential omitted variables), Metzler & Woessmann (2012) estimate that a one standard deviation increase in teacher subject-specific knowledge increases math scores of 6th-graders in Peru by 0.087 standard deviations, while the effect on reading test scores (0.022 s.d.) is statistically insignificant. Furthermore, analyzing SACMEQ data for several Anglophone countries, Fehrler et al. (2009) find that teachers' academic achievement, duration of teacher training, and subject-specific knowledge are all positively correlated with student test scores. In particular, they find a correlation coefficient of 0.21 (0.32) between student reading (math) scores and their teacher's score at the same test.

Turning now to the effect of non-teacher inputs, the most reliable evidence available, obtained through randomized controlled trials in Kenya, suggests no effect of flipcharts (Glewwe et al. 2004) or textbooks (Glewwe et al. 2009) on test scores, except for the best students in the case of textbooks.

All in all, it is not clear that larger class sizes, less educated teachers (except perhaps if it translates into lower teacher subject-specific knowledge), less experienced teachers (beyond their initial two to three teaching years), or fewer textbooks and other physical inputs should have a large effect on test scores in a developing country setting such as Tanzania. Therefore, it is unclear whether rapid enrollment growth, which in the short run is bound to increase class sizes, and reduce the education and experience of the average teacher, may or not lead to a substantial deterioration of test scores.

A-2 Graphical Analysis

Figure A-2 shows the positive correlation between the change in the mean regional pupil-teacher ratio between the 2000 and 2007 surveys and regional enrollment growth. In Figures A-3 and A-4, changes in average standardized reading scores (Figure A-3) and math scores (Figure A-4) are plotted against regional enrollment growth, and there appears to be no correlation between changes in pupil test scores and enrollment growth.

One may expect regions with lower scores in 2000 to experience larger improvements in test scores between 2000 and 2007 (e.g., due to there being more low-hanging fruits to be picked). This is indeed the case, as illustrated by Figures A-5 and A-7. One concern could be that less developed regions started off with lower average test scores and also experienced faster enrollment growth under FPE, and/or experienced slower fertility declines in the past. In this case, the "mean reversion" observed in Figures A-5 and A-7 could bias my (OLS and/or IV) estimates and lead to an underestimation of the worsening of test scores due to enrollment growth. However, Figures A-6 and A-8 show no systematic relationship between baseline test scores and enrollment growth. In order to confirm that mean reversion is not driving my results, in Section 6 I check the robustness of my findings to allowing for changes in test scores over time to depend on baseline scores, as suggested by Chay et al. (2005) in an application in which a school treatment is allocated on the basis of the school's initial score.

A-3 Alternative Identification Strategy Relying on Regional Differences in Pre-FPE Enrollment Rates

An alternative identification strategy to that adopted in this paper would have been to exploit differences in baseline enrollment rates. The last pre-FPE year for which official statistics document net primary school enrollment rates (NER) by region is 1998, which is what I use here. More specifically, I estimate a difference-in-differences regression similar to Equation 2 replacing $(\frac{post\_enrol}{baseline\_enrol})_r$ with $(1 - NER_{1998})_r$ in the interaction term. The interaction term can then be interpreted as a proxy for "intensity of exposure to FPE". As explained in the main text, I cannot test for pre-existing differences in trends in test scores between regions with different net enrolment rates pre-FPE since there are only two rounds of test scores data (2000 and 2007). But contrary to the instrumental variable used in the paper, it would be difficult to sign the direction of the bias in IV (or reduced-form) estimates using pre-FPE enrollment rates to create an instrumental variable for actual enrollment growth. Lower enrollment levels at baseline could indeed have been correlated with either higher or lower subsequent growth in test scores irrespective of the increase in enrollment. For instance, lower baseline enrollment could be correlated with higher future growth in test scores due to mean reversion in investments both in education quantity and quality, or, on the opposite, with lower subsequent growth in test scores if, for example, baseline enrollment proxies for tastes for education and parents in more education-oriented regions increasingly care about the quality of education. Therefore I do not follow this strategy in the main analysis.

Reassuringly, however, results obtained with this alternative strategy are remarkably consistent with the results reported in the paper, as shown by Tables A-4, A-5, A-6, A-7, A-8 and A-9 below. Qualitatively, all the average treatment effects that are statistically significant in the main analysis are of the same sign and, except in one case (language teacher subject-specific knowledge), are also significant when using this alternative strategy. The magnitudes of the effects are also quite similar: a one standard deviation decrease in pre-FPE enrollment (i.e., an increase of $(1 - NER_{1998})_r$ by 0.083) increases the pupil-teacher ratio by 7.1 pupils between 2000 and 2007, and decreases the average language (math) teacher's experience by 1.54 (1.82) years. The point estimates of the effect of intensity of exposure to FPE on pupil test scores are statistically insignificant, and the magnitudes implied by the lower bounds of the 95% CIs are also consistent with those reported in the paper, with a maximum decrease of 0.13 (0.15) standard deviations in language (math) test scores for a one standard deviation decrease in pre-FPE net enrollment rates. The only small differences in results between the two identification strategies are that: (i) there is a statistically significant, negative effect on language teacher's subject-specific knowledge of 0.14 SD for a one SD increase in enrollment growth, while a one SD decrease in pre-FPE net enrollment rate (i.e., a likely increase in enrollment growth) is associated with a statistically insignificant decrease in language teacher's subject-specific knowledge by 0.06 SD, (ii) while in the main analysis I do not obtain any statistically significant effect of enrollment growth on the probability that teachers hold O-level qualifications, a one SD decrease in pre-FPE net enrollment rate is associated with a marginally significant 4.6%-points increase in the probability that the Kiswahili teacher has O-levels, which represents 5.4% of the sample mean and (iii) while in the main analysis I do not obtain any statistically significant effect of enrollment growth on physical inputs, a one SD decrease in pre-FPE net enrollment rate is associated with a statistically significant increase of 0.19 in the number of items at the disposal of pupils at school (a 3% increase relative to the sample mean).

A-4 Obtaining Bounds for the Changes-in-Changes Estimates

Blanco et al. (2013) show that, in a sample of treated and control observations with random treatment assignment, in which the probability of observing the individual in the data increases when they receive the treatment, quantile treatment effects for the inframarginal individuals can be bounded by estimates obtained with and without trimming the distribution of outcomes among the treated group. More specifically, the lower bound (i.e., here, the largest negative effect) is obtained without trimming, while the upper bound is obtained by trimming the distribution of outcomes in the treated group by removing the lower $\frac{Pr(S_i=1|T_i=1)-Pr(S_i=1|T_i=0)}{Pr(S_i=1|T_i=1)}$ share of the distribution, where $S_i = 1$ if individual i is observed in the data, and $T_i = 1$ if individual i is treated, and zero otherwise.

Lee (2009) shows that, under random treatment assignment and individual-level positive weak monotonicity of the probability of being observed in treatment status, a lower bound for the average treatment effect can be obtained by trimming the treated observations from above. Instead, here I follow Blanco et al. (2013) who extend Lee (2009)'s logic to quantile treatment effects and show that, under the additional assumption that the distribution of outcomes of the inframarginal group stochastic dominates that of the marginal group, i.e., under the assumption that $F_{Y,11|Inframarginal}(y) \leq F_{Y,11|Marginal}(y)$, *for all* y, a lower bound is provided by the untrimmed distribution of outcomes in the treated group, which results in narrower bounds.

Given the Changes-in-Changes set up, I trim the distribution of the treated group, in the 2007 data only, by the *change* between 2000 and 2007 in the relative probabilities of being observed in the data in the treated and control groups $[\frac{Pr(S_i=1|T_i=1)-Pr(S_i=1|T_i=0)}{Pr(S_i=1|T_i=1)}]_{2007} - [\frac{Pr(S_i=1|T_i=1)-Pr(S_i=1|T_i=0)}{Pr(S_i=1|T_i=1)}]_{2000}$ to obtain an upper bound of the effect of enrollment growth. More specifically, I compute $Pr(S_i = 1|T_i = t), t = 0, 1$ as follows. Step 1: define the relevant age group based on the range of Grade 6 pupils' ages observed in SACMEQ 2000. Step 2: compute the total number of individuals in that age group in 2000 and in 2007, in both the treated and control groups, using the 2002 population census. Step 3: using SACMEQ "raising factors", which give the number of pupils in the Grade 6 population that were represented by a single pupil in the SACMEQ sample, compute the number of Grade 6 pupils represented by the SACMEQ dataset, in 2000 and 2007, in the treated and control groups. Each of the $Pr(S_i = 1|T_i = t)$ are then obtained as the ratio of the number of pupils represented by the SACMEQ dataset from step 3 for group $T_i = t$ divided by the size of the relevant population obtained in step 2.

Appendix Tables

Table A-1: Differential Trends in Health Inputs and Outcomes in Rural Areas

	(1) =1 if Full Immunization	(2) =1 if Delivered by Health Professional	(3) =1 if Delivered Without Any Help	(4) =1 if Infant Death	(5) =1 if Stunted
(=1 if born 1995-2000)*Age 7-13 in 2002-2007	-0.108	-0.281**	0.061	0.014	0.066
/Age 7-13 in 2001	[-0.248,0.032]	[-0.492,-0.070]	[-0.025,0.147]	[-0.056,0.084]	[-0.030,0.161]
=1 if born 1995-2000	0.688	1.844**	-0.435	-0.089	-0.526
	[-0.244,1.621]	[0.409,3.280]	[-1.023,0.152]	[-0.568,0.389]	[-1.185,0.134]
Region Dummies	Yes	Yes	Yes	Yes	Yes
Observations	8954	13369	13369	10280	10470
No. of clusters	20	20	20	20	20
R-squared	0.04	0.04	0.06	0.00	0.04
Mean Y	0.70	0.41	0.07	0.09	0.46

See notes under Table 2. Rural sample only.

Table A-2: Differential Trends in Health Inputs and Outcomes in Urban Areas

	(1) =1 if Full Immunization	(2) =1 if Delivered by Health Professional	(3) =1 if Delivered Without Any Help	(4) =1 if Infant Death	(5) =1 if Stunted
(=1 if born 1995-2000)*Age 7-13 in 2002-2007	-0.167**	-0.029	0.037**	0.037	-0.008
/Age 7-13 in 2001	[-0.319,-0.014]	[-0.131,0.072]	[0.001,0.072]	[-0.016,0.090]	[-0.176,0.159]
=1 if born 1995-2000	1.133**	0.170	-0.242**	-0.248	-0.061
	[0.114,2.151]	[-0.487,0.828]	[-0.474,-0.010]	[-0.595,0.100]	[-1.199,1.076]
Region Dummies	Yes	Yes	Yes	Yes	Yes
Observations	1990	2925	2925	2276	2260
No. of clusters	20	20	20	20	20
R-squared	0.04	0.05	0.03	0.01	0.03
Mean Y	0.79	0.83	0.01	0.09	0.32

See notes under Table 2. Urban sample only.

Table A-3: Grade Enrollment Under Grade 1 Growth

Year	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	N in Year	$N(1+\delta)$ in Year
2001	N	N	N	N	N	N	N	7	0
2002	$N(1+\delta)$	N	N	N	N	N	N	6	1
2003	$N(1+\delta)$	$N(1+\delta)$	N	N	N	N	N	5	2
2004	$N(1+\delta)$	$N(1+\delta)$	$N(1+\delta)$	N	N	N	N	4	3
2005	$N(1+\delta)$	$N(1+\delta)$	$N(1+\delta)$	$N(1+\delta)$	N	N	N	3	4
2006	$N(1+\delta)$	$N(1+\delta)$	$N(1+\delta)$	$N(1+\delta)$	$N(1+\delta)$	N	N	2	5
2007	$N(1+\delta)$	$N(1+\delta)$	$N(1+\delta)$	$N(1+\delta)$	$N(1+\delta)$	$N(1+\delta)$	N	1	6
							2002-2007 total:	21	21

Table A-4: Effect on Quantity of Teachers

	(1)	(2)	(3)	(4)
	Pupil/Teacher	=1 if Multiple Shifts	Teaching Hours (K)	Teaching Hours (M)
(=1 if 2007)*(1-1998 Primary NER)	85.545***	0.177	10.012	-5.024
	[23.652,147.437]	[-1.540,1.893]	[-18.371,38.395]	[-18.585,8.537]
Observations	6933	6933	6933	6933
No. of clusters	19	19	19	19
R-squared	0.3314	0.1549	0.1840	0.1848
Mean Y	53.179	0.096	16.660	16.723

Sample size: 6933. Region-correlated robust 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * p<0.10, ** p<0.05, *** p<0.01.

Table A-5: Effect on Quality of Teachers: Initial Training

	(1)	(2)	(3)	(4)
	=1 if O-level (K)	=1 if O-level (M)	=1 if Training≥2 y (K)	=1 if Training≥2 y (M)
(=1 if 2007)*(1-1998 Primary NER)	0.551*	0.286	-0.340	-0.418
	[-0.058,1.161]	[-0.552,1.125]	[-1.230,0.550]	[-1.222,0.385]
Observations	6933	6933	6933	6933
No. of clusters	19	19	19	19
R-squared	0.1717	0.0678	0.1272	0.1660
Mean Y	0.845	0.942	0.865	0.852

Sample size: 6933. Region-correlated robust 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A-6: Effect on Quality of Teachers: Experience and Subject-Specific Knowledge

	(1)	(2)	(3)	(4)
	Years Experience (K)	Years Experience (M)	Teacher Reading Score	Teacher Math Score
(=1 if 2007)*(1-1998 Primary NER)	-18.578***	-33.972***	-0.693	-0.601
	[-31.757,-5.399]	[-54.204,-13.740]	[-3.135,1.748]	[-3.029,1.828]
Observations	6933	6933	6933	6933
No. of clusters	19	19	19	19
R-squared	0.0805	0.1608	0.1421	0.1425
Mean Y	13.537	12.049	0.008	-0.001

Sample size: 6933. Region-correlated robust 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A-7: Effect on Physical Inputs

	(1) =1 if Own Bk (K)	(2) =1 if Own Bk (M)	(3) Class Equip. (K)	(4) Class Equip. (M)	(5) # Ex. Books	(6) Pupil Equip.
(=1 if 2007)*(1-1998 Primary NER)	0.171	0.074	2.009	2.188	3.259	2.301**
	[-0.054,0.396]	[-0.111,0.258]	[-2.130,6.148]	[-1.018,5.395]	[-1.835,8.352]	[0.371,4.232]
Observations	6933	6933	6933	6933	6933	6933
No. of Regions						
No. of clusters	19	19	19	19	19	19
R-squared	0.0096	0.0160	0.1958	0.2087	0.1429	0.0878
Mean Y	0.042	0.041	4.311	4.175	7.977	5.831

Sample size: 6933. Region-correlated robust 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Class equipment and pupil equipment defined under Table 3.

Table A-8: Effect on Pupil Reading Test Scores

	Pupil Reading Test Scores			
	(1)	(2)	(3)	(4)
	OLS	OLS	OLS	OLS
(=1 if 2007)*(1-1998 Primary NER)	0.895	0.911	0.772	-0.086
	[-0.317,2.107]	[-0.216,2.037]	[-0.631,2.174]	[-1.614,1.441]
(=1 if 2007)*Ed. grants 2002-2007/Ed. grant 2001, deflated			0.018	
			[-0.052,0.088]	
(=1 if 2007)×Baseline Average				-0.575***
				[-0.892,-0.258]
Pupil Characteristics	No	Yes	No	No
Physical Inputs	No	Yes	No	No
Observations	6933	6933	6933	6933
No. of clusters	19	19	19	19
R-squared	0.1414	0.2147	0.1415	0.1479
Mean Y	0.002	0.002	0.002	0.002

Sample size: 6933. Region-correlated robust 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Pupil characteristics and physical inputs variables as listed in Table 3.

Table A-9: Effect on Pupil Math Test Scores

	Pupil Math Test Scores			
	(1)	(2)	(3)	(4)
	OLS	OLS	OLS	OLS
(=1 if 2007)*(1-1998 Primary NER)	0.738	0.771	0.216	-0.168
	[-0.971,2.447]	[-0.804,2.345]	[-1.586,2.018]	[-1.805,1.468]
(=1 if 2007)*Ed. grants 2002-2007/Ed. grant 2001, deflated			0.077	
			[-0.028,0.181]	
(=1 if 2007)×Baseline Average				-0.743***
				[-1.119,-0.367]
Pupil Characteristics	No	Yes	No	No
Physical Inputs	No	Yes	No	No
Observations	6933	6933	6933	6933
No. of clusters	19	19	19	19
R-squared	0.1033	0.1907	0.1054	0.1109
Mean Y	-0.004	-0.004	-0.004	-0.004

Sample size: 6933. Region-correlated robust 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * p<0.10, ** p<0.05, *** p<0.01. Pupil characteristics and physical inputs variables as listed in Table 3.

Table A-10: Summary Statistics by Rural/Urban School Location

	SACMEQ 2000				SACMEQ 2007			
	Urban		Rural		Urban		Rural	
	(1)		(2)		(3)		(4)	
	mean	sd	mean	sd	mean	sd	mean	sd
Regional Education Characteristics								
(=1 if 2007)*Enrol. 2002-2007/Enrol. in 2001	0.00	0.000	0.00	0.000	8.98	0.586	9.00	0.736
(=1 if 2007)*Age 7-13 in 2002-2007/Age 7-13 in 2001	0.00	0.000	0.00	0.000	6.71	0.313	6.77	0.297
(=1 if 2007)*Ed. grants 2002-2007/Ed. grant 2001, deflated	0.00	0.000	0.00	0.000	11.26	1.215	11.41	1.427
Teacher Quantity Variables								
Pupil-Teacher Ratio	39.74	19.117	49.96	19.254	45.59	13.910	69.87	34.714
=1 if Multiple Shifts	0.31		0.12		0.10		0.03	
Kiswahili Teaching Hours	13.53	6.071	18.29	7.291	12.91	4.874	17.50	6.176
Math Teaching Hours	12.73	5.315	18.92	6.989	13.50	4.020	17.72	6.002
Teacher Quality Variables								
=1 if O-level (K)	0.94		0.67		0.98		0.91	
=1 if O-level (M)	0.99		0.90		0.99		0.95	
=1 if Training>=2 y (K)	0.95		0.93		0.78		0.82	
=1 if Training>=2 y (M)	0.96		0.97		0.73		0.79	
Teacher Experience (K)	12.20	6.694	14.86	8.078	11.52	9.633	12.75	11.096
Teacher Experience (M)	10.91	6.940	13.09	7.277	9.68	9.349	11.57	10.137
Teacher Reading Z-score	-0.03	0.781	-0.23	0.949	0.35	0.797	0.06	1.120
Teacher Math Z-score	0.19	0.936	-0.45	0.983	0.18	0.785	0.11	1.031
Physical Inputs Variables								
Pupil-Specific Variables								
=1 if Pupil Has Own Kiswahili Book	0.06		0.06		0.03		0.03	

=1 if Pupil Has Own Math Book	0.07		0.07		0.03		0.03	
=1 if No Reading Textbook	0.29		0.38		0.20		0.25	
=1 if No Math Textbook	0.30		0.34		0.22		0.24	
# Exercises Books	9.45	2.837	8.66	3.031	7.61	3.359	6.95	3.012
Total Pupil Equipment Score (0 to 8)	5.83	1.630	5.30	1.821	6.38	1.465	6.05	1.522
Class-Specific Variables								
Total Kiswahili Equipment Score (0 to 8)	3.32	1.950	3.70	1.708	4.95	1.820	4.78	1.562
Total Math Equipment Score (0 to 8)	3.25	1.520	3.36	1.712	4.87	1.864	4.66	1.594
Learning Outcomes								
Pupil Reading Z-score	0.38	0.835	-0.44	0.959	0.49	0.896	-0.00	0.971
Pupil Math Z-score	0.19	1.002	-0.37	0.961	0.41	1.003	0.02	0.933
=1 if Low Competency (M)	0.14		0.30		0.07		0.16	
=1 if Low Competency (K)	0.03		0.23		0.04		0.13	
=1 if High Competency (M)	0.30		0.13		0.40		0.27	
=1 if High Competency (K)	0.38		0.15		0.46		0.27	
Pupil Characteristics								
=1 if Male Pupil	0.48		0.48		0.48		0.49	
Pupil's Age	13.73	1.430	14.72	1.487	13.53	1.527	14.13	1.591
=1 if English is Never Spoken at Home	0.02		0.13		0.03		0.10	
Household Items Ownership (0 to 14)	5.31	3.511	2.66	1.752	5.54	2.361	4.85	2.079
Parental Education Variables								
=if if Father < Completed Primary	0.10		0.29		0.10		0.20	
=if if Mother < Completed Primary	0.09		0.29		0.14		0.29	
=if if Father = Completed Primary	0.28		0.44		0.44		0.58	
=if if Mother = Completed Primary	0.45		0.53		0.57		0.61	
=if if Father > Completed Primary	0.53		0.21		0.37		0.17	
=if if Mother > Completed Primary	0.39		0.11		0.24		0.07	

=1 if Does Not Know Dad's Educ. Level	0.07	0.05	0.06	0.03
=1 if Does Not Know Mum's Educ. Level	0.08	0.06	0.04	0.02
=1 if No Father or Male Guardian	0.02	0.02	0.03	0.02
=1 if No Mother or Female Guardian	0.00	0.01	0.01	0.00
N	673	2176	1206	2878

Source: Author's calculations using SACMEQ II, SACMEQ III, IPUMS (2011), Ministry of Education "Basic Education Statistics in Tanzania" and Budget Plans for various years. Statistics weighted by the same SACMEQ pupil weights as in the regressions. The pupil-teacher ratio is calculated as the ratio of the total number of pupils to the total number of teachers in the school based on the information collected during interviews with head teachers. Class equipment score items: writing board, chalk, wall chart, cupboard, bookshelves, library, teacher table, teacher chair. Pupil equipment score items: exercise book, notebook, pencil, sharpener, eraser, ruler, pen, folder. Household items ownership items: newspaper, magazine, radio, TV set, VCR, cassette player, telephone, refrigerator/freezer, car, motorcycle, bicycle, piped water, electricity, table to write on.

Table A-11: Effect on Grade 6 Pupil Socioeconomic Status

	(1) SES Tercile	(2) #Books at home	(3) Male Pupil	(4) Ownership count	(5) Pupil has no father	(6) Father education
Entire Sample						
(=1 if 2007)*Enrol. 2002-2007/Enrol. in 2001	-0.020	-5.517	-0.009	-0.052	0.006*	-0.024
	[-0.204,0.163]	[-13.559,2.526]	[-0.038,0.021]	[-0.578,0.475]	[-0.001,0.013]	[-0.166,0.118]
Observations	6933	6933	6933	6933	6933	6460
Mean Y	2.002	13.887	0.487	4.318	0.021	2.053
Rural Sample						
(=1 if 2007)*Enrol. 2002-2007/Enrol. in 2001	0.008	-4.219	-0.014	0.024	0.007*	-0.008
	[-0.178,0.194]	[-14.345,5.907]	[-0.045,0.018]	[-0.182,0.230]	[-0.000,0.015]	[-0.170,0.154]
Observations	5054	5054	5054	5054	5054	4736
No. of clusters	19	19	19	19	19	19
Mean Y	1.827	15.061	0.486	3.960	0.021	1.944
Urban Sample						
(=1 if 2007)*Enrol. 2002-2007/Enrol. in 2001	-0.241	-7.831*	-0.009	-0.883	-0.001	-0.136
	[-0.581,0.100]	[-17.013,1.351]	[-0.050,0.032]	[-2.539,0.774]	[-0.016,0.014]	[-0.329,0.057]
Observations	1879	1879	1879	1879	1879	1724
Mean Y	2.472	10.729	0.490	5.281	0.022	2.354

Region-correlated robust 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * $p<0.10$, ** $p<0.05$, *** $p<0.01$. SES tercile: SES tercile based on a principal component analysis using data on household ownership of a table, radio, tv, vcr, cassette player, telephone, fridge, car, motorcycle, bike, access to water and electricity and dwelling materials. Ownership count: count from 0 to 14 of the number of the following equipment owned by the pupil's household: table, radio, tv, vcr, cassette player, telephone, fridge, car, motorcycle, bike, newspaper, magazine, access to water and electricity. Father education: variable equal to 1 if father has none or some primary education, 2 if father completed primary and 3 if father completed more than primary education.

Table A-12: Effect on Intake in Rural and Urban Areas

	(1)	(2)	(3)	(4)
	# Pupils	# Pupils	# Pupils	# Pupils
	in the School	in the School	in the Region	in the Region
	Rural	Urban	Rural	Urban
(=1 if 2007)*Enrol. 2002-2007	98.632**	154.181	4071.059	2817.135
/Enrol. in 2001	[5.077,192.187]	[-396.892,705.254]	[-5508.206,13650.325]	[-4056.121,9690.390]
Observations	5054	1879	38	34
No. of clusters	19	19	19	17
Mean Y in 2000	437	1072	18429	8736

Region-correlated robust 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * $p<0.10$, ** $p<0.05$, *** $p<0.01$. The unit of observation in columns (3) and (4) is one region observed either in 2000 or 2007.

Appendix Figures

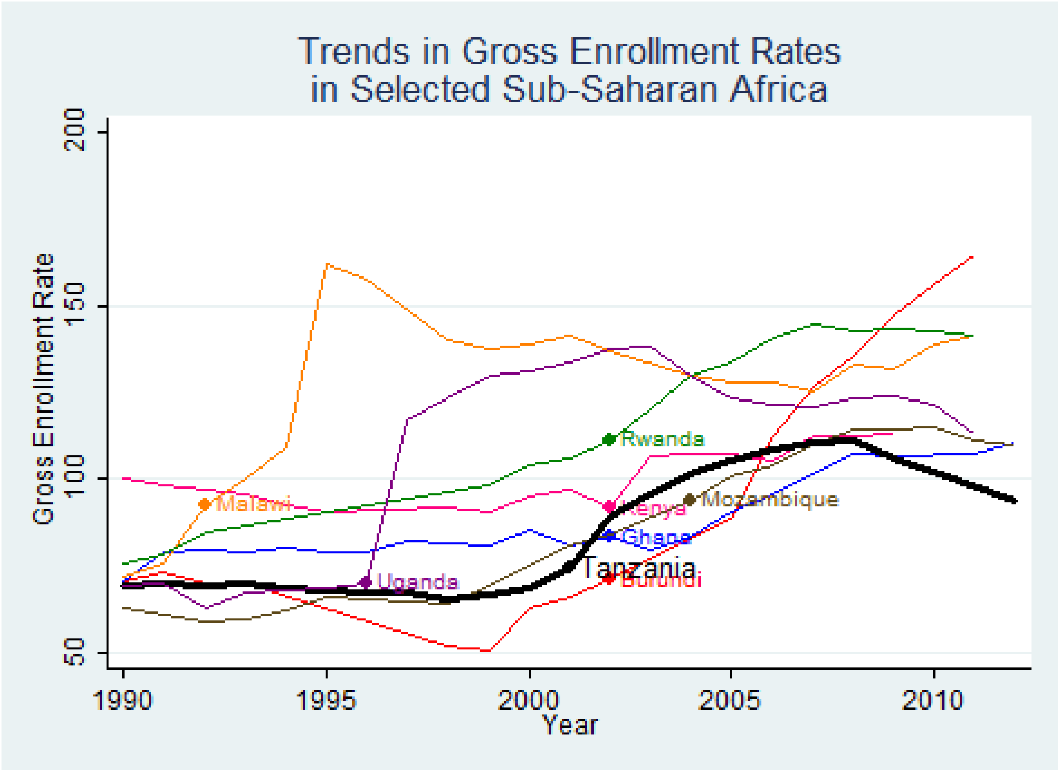

Figure A-1

Note: Primary school fees removed in the year following that at which the country's name appears on the figure.

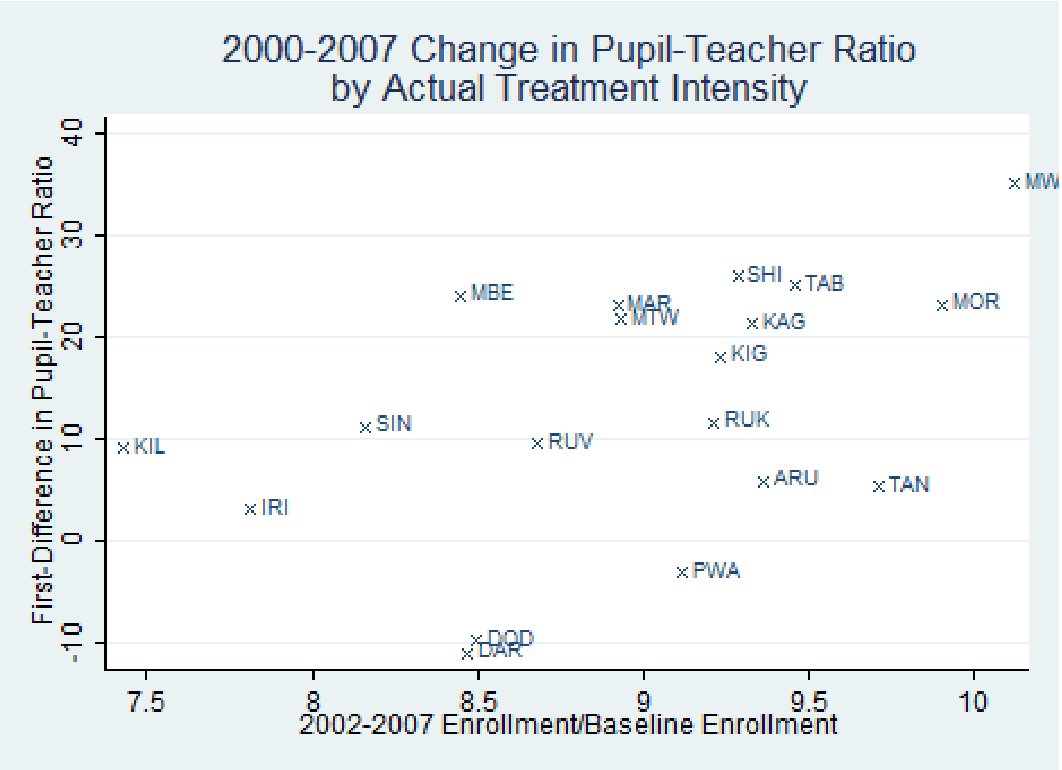

Figure A-2

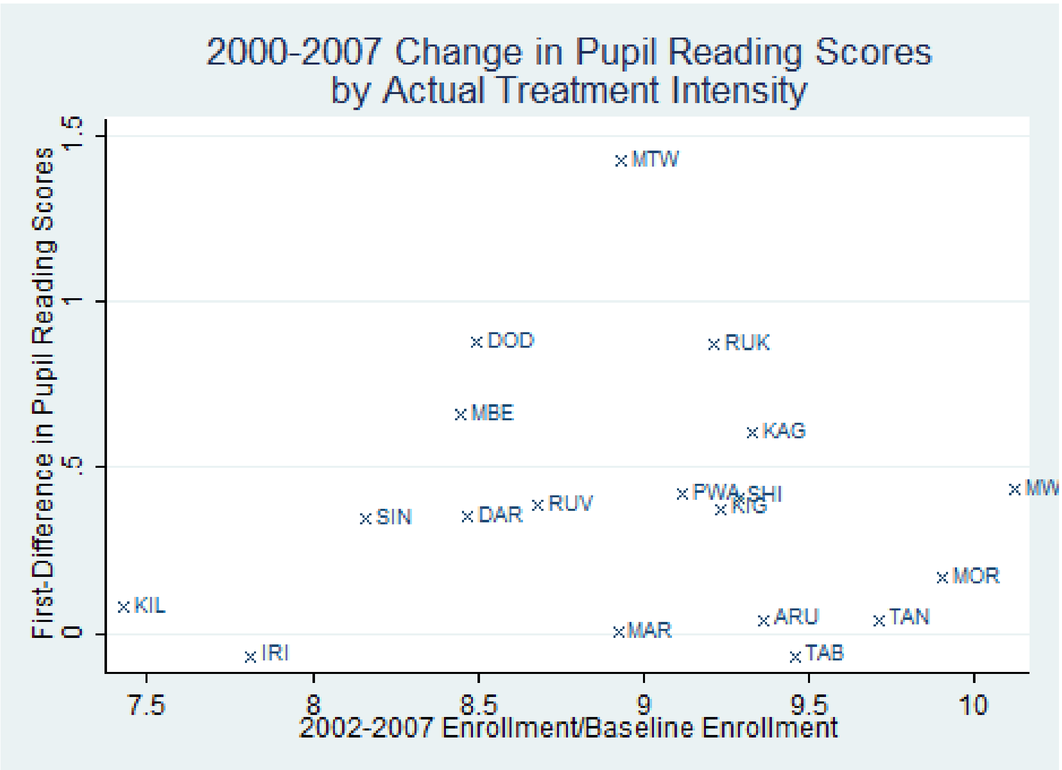

Figure A-3

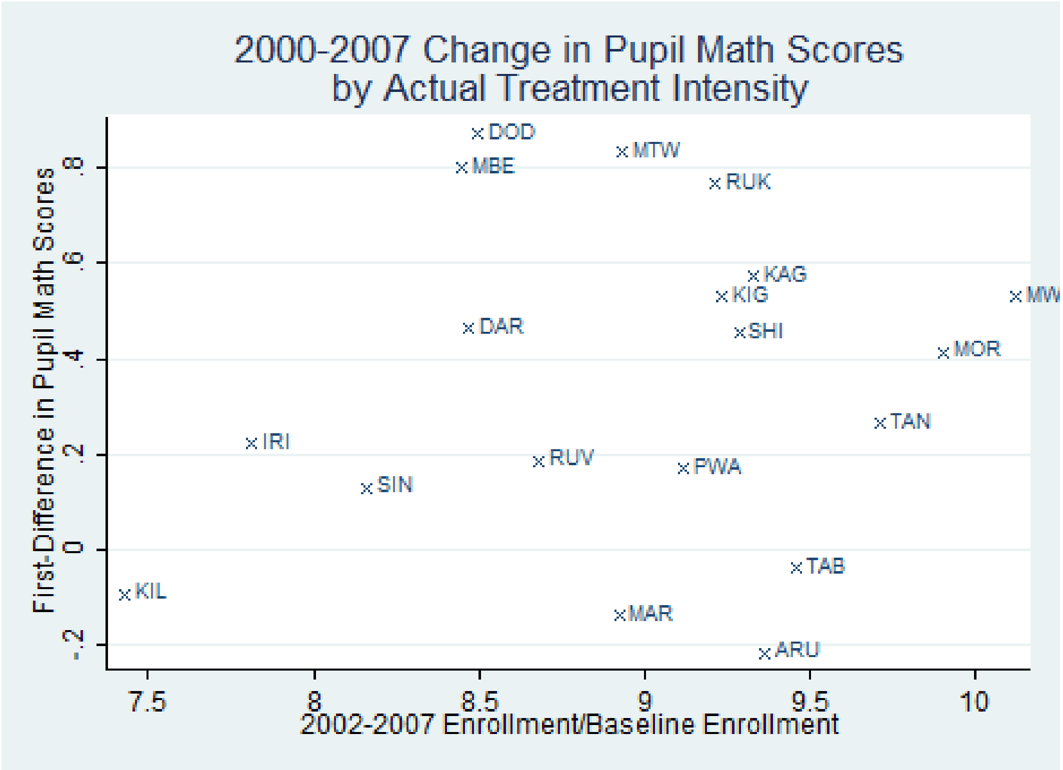

Figure A-4

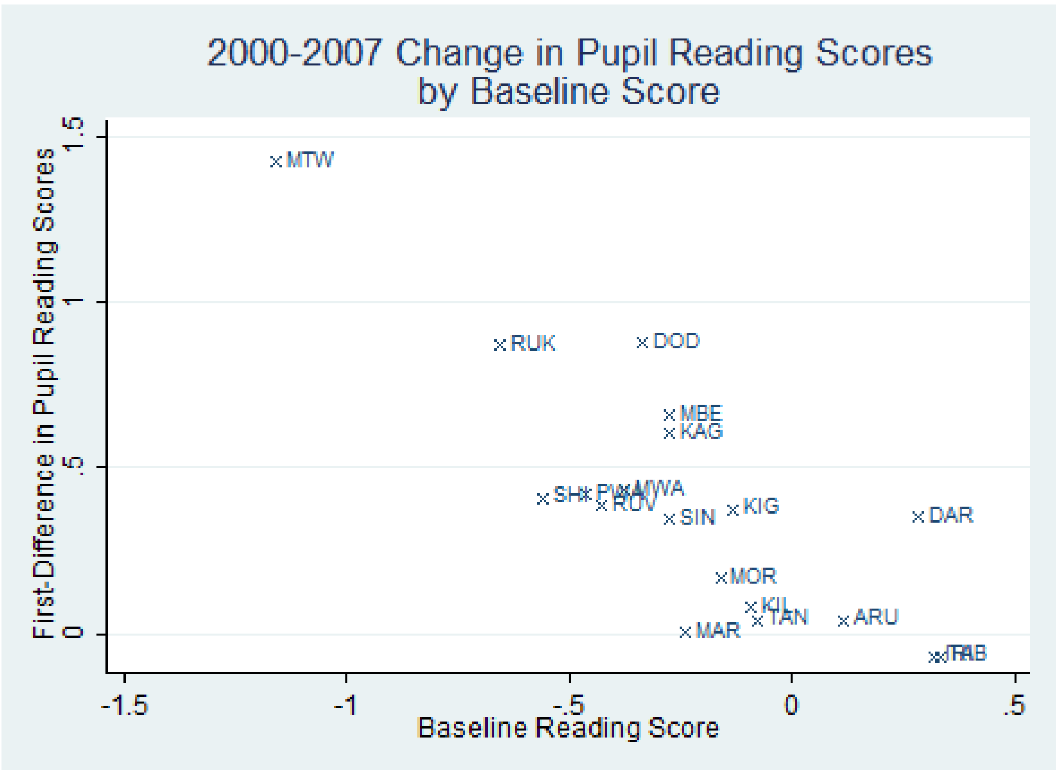

Figure A-5

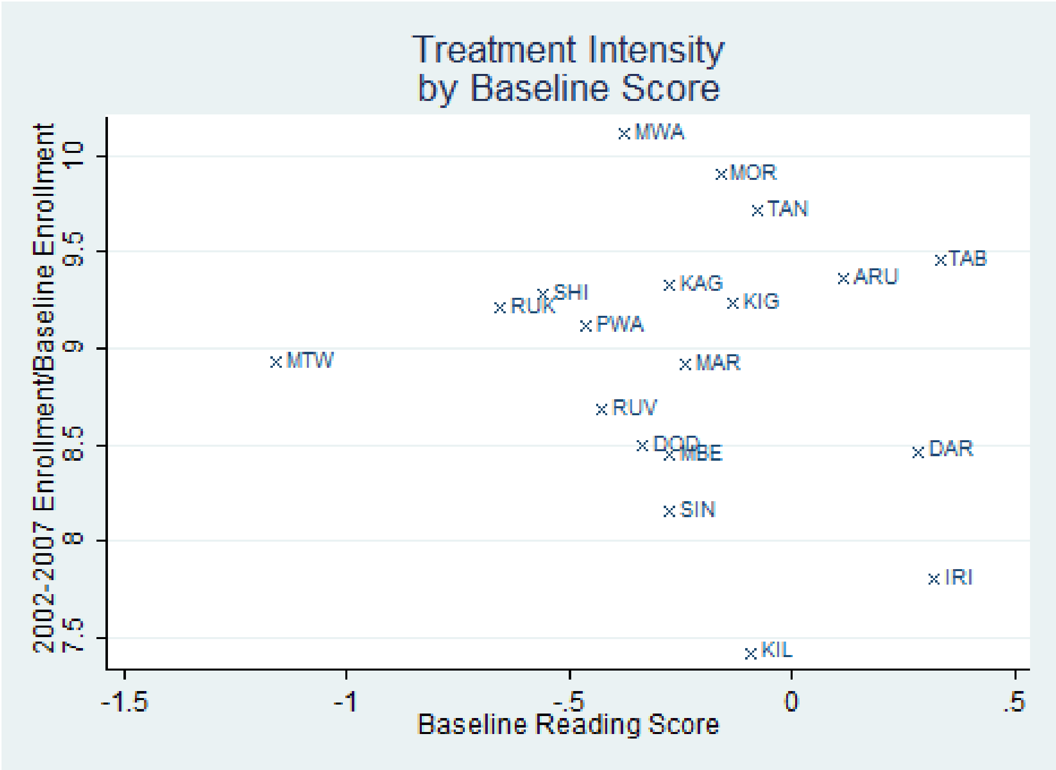

Figure A-6

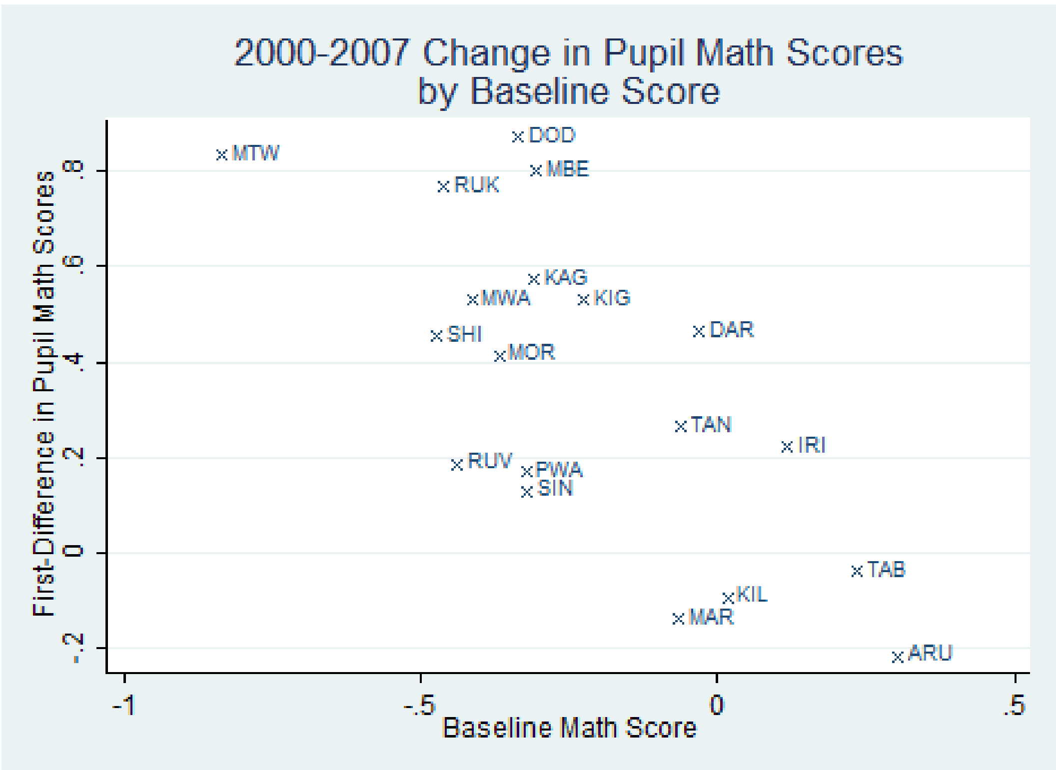

Figure A-7

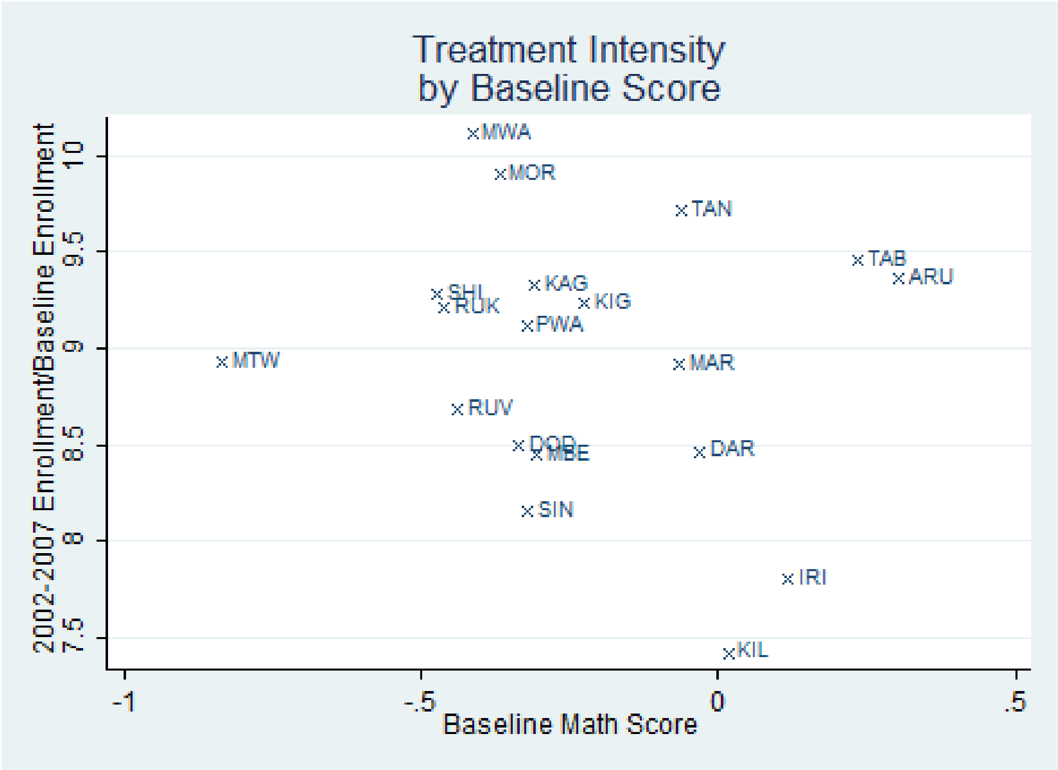

Figure A-8

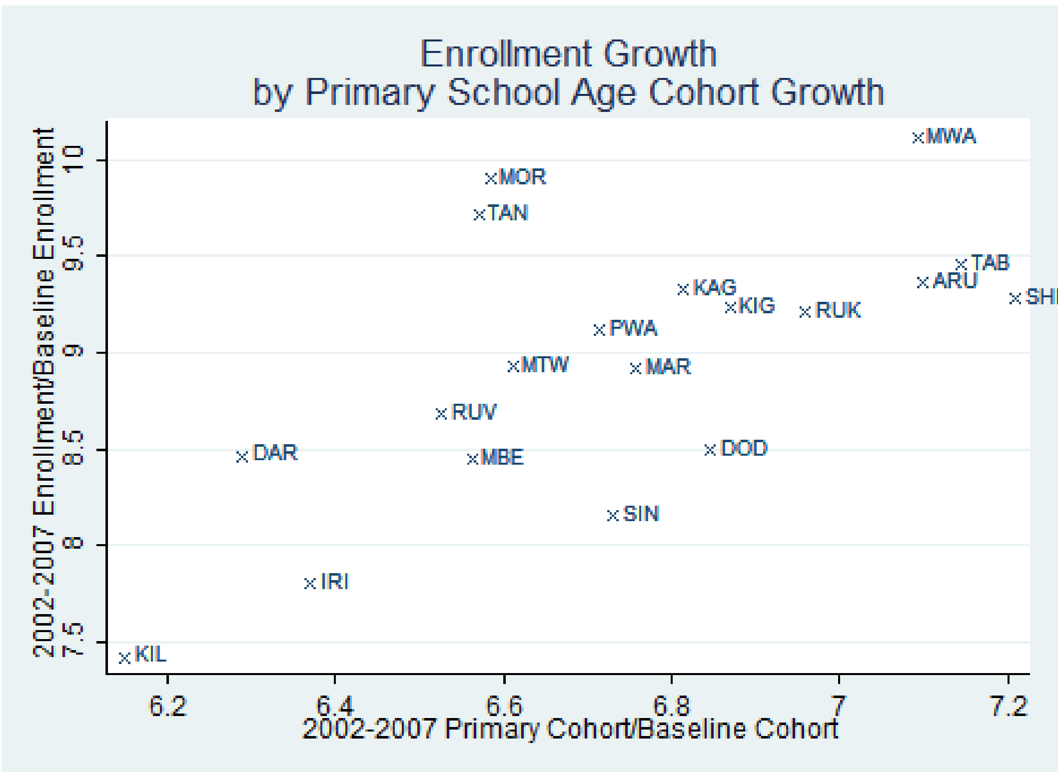

Figure A-9

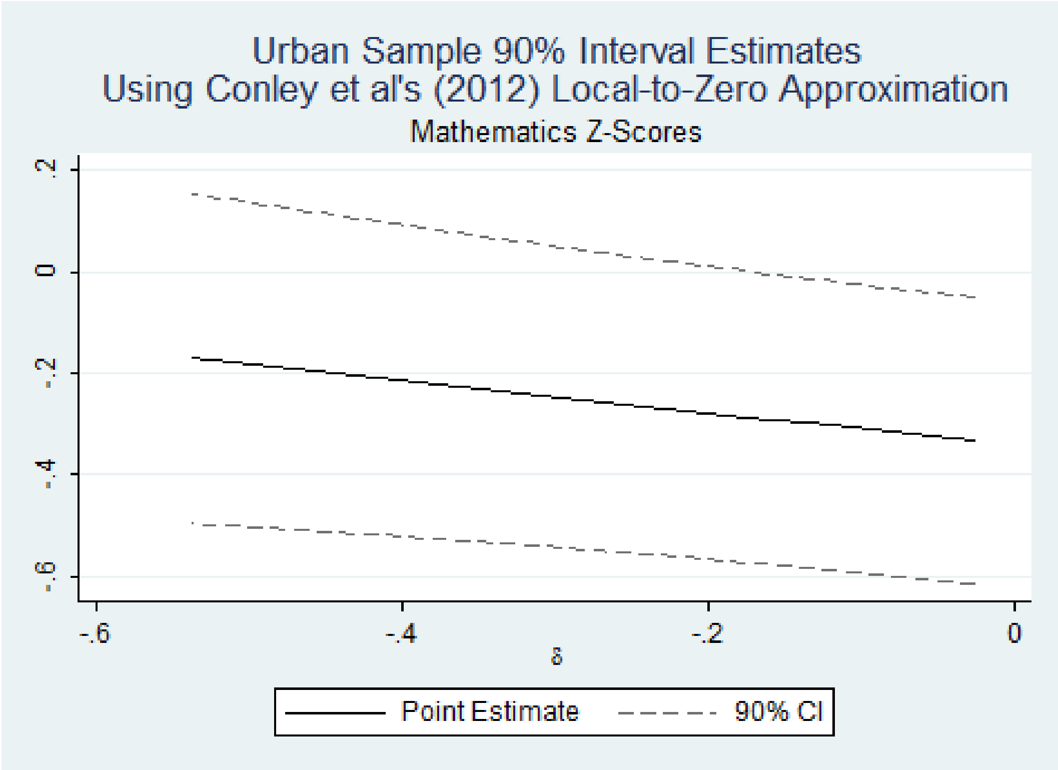

Figure A-10

Note: Effect of enrollment growth allowing for departures from the perfectly exogenous instrument case ($\gamma = 0$) of the form $\gamma \sim \mathcal{U}(\delta, 0)$.

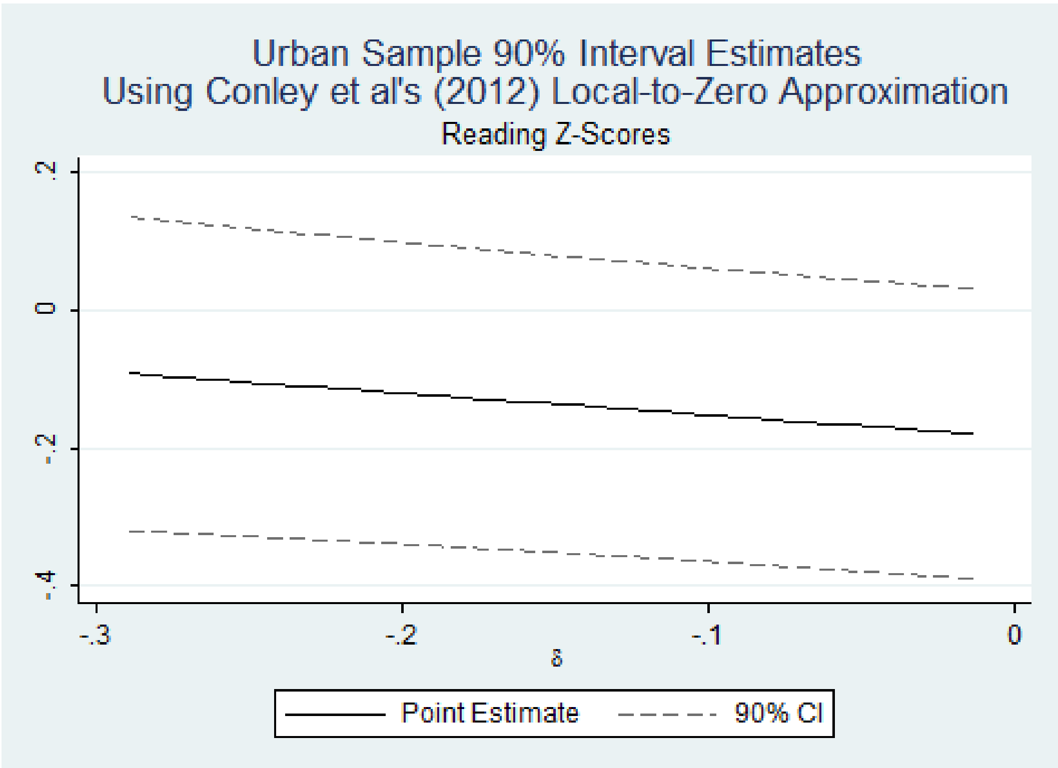

Figure A-11

Note: Effect of enrollment growth allowing for departures from the perfectly exogenous instrument case ($\gamma = 0$) of the form $\gamma \sim \mathcal{U}(\delta, 0)$.