

Primary Education Expansion and Quality of Schooling: Evidence from Tanzania

Christine Valente*

University of Bristol, U.K.

Abstract

I estimate the effect of increased enrollment following the removal of primary school fees in Tanzania on cohorts fully exposed to the new fee regime. I address the potential endogeneity of regional enrollment growth using an instrument based on predetermined fertility decisions. I find that enrollment growth led to large increases in the pupil-teacher ratio, and a reduction in average teacher experience and subject-specific knowledge, but rule out a substantial effect on national average test scores. However, in urban areas, where baseline achievement is higher, I find evidence of a deterioration of test scores where enrollment growth was larger.

JEL codes: I21; I28; O15

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

*Email address: 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 Sami Stouli and Frank Windmeijer, as well as seminar participants at the University of Bristol, Universite Catholique de Louvain, ISER (University of Essex), the 2014 Royal Economic Society Conference and 2014 Econometric Society European Meeting for their useful comments. All errors are my own.

1 Introduction

The net enrollment rate in Tanzanian primary schools went from 53% in 2000 to 73% in 2002, coinciding with the removal of primary school fees announced in 2001 (World Development Indicators (2013), Kattan & Burnett (2004)). Tanzania is no outlier - a sizeable number of Sub-Saharan African countries (17 since 1994 according to Lucas & Mbiti (2012)) have recently implemented Free Primary Education (FPE) policies and many of these reforms resulted in large increases in enrollment (Figure A-1).

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 poor quality of schooling provision in developing countries (Kremer et al. 2013). Increased pressure on education systems due to large rises in enrollment is likely to add to the challenges of providing schooling of a sufficient quality to ensure the satisfactory production of cognitive skills.

Despite considerable concern about this issue in policy circles, there is close to no arguably causal, quantitative evidence of the impact of the large increases in enrollment seen in recent years on the quality of the learning environment. The exception is a recent study by Lucas & Mbiti (2012), which shows that the removal of primary school fees in 2003 in Kenya increased access to primary schooling without substantially reducing the test scores of children who had been in school for at least 3 years prior to FPE and would have been likely to take the end of primary school test even in the absence of the reform. These children will not, however, have experienced the full effect of the reform - since new entrants were enrolled in year one irrespective of age, children in Lucas & Mbiti (2012)'s sample will, for instance, only have experienced larger class sizes through increased retention of students or if resources were reallocated across grades and will, at most, have been affected from grade 4 onwards.

In this paper, I estimate the impact of the enrollment growth coinciding with FPE in Tanzania on the learning outcomes of the first cohort of pupils who started primary school post-reform and were hence fully exposed to its consequences. In addition, I shed light not only on the impact of the

reform on test scores, but also on a rich set of measures of the quantity and quality of educational inputs and teachers.

The removal of primary school fees took place simultaneously in the whole country. However, 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, local schooling systems 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. I use this source of arguably exogenous variation, which is based on past fertility decisions, in order to establish the causal effect of enrollment growth on schooling inputs and cognitive skills acquisition. I sign the potential direction of the bias in my IV estimates and, where this bias might be driving my conclusions, I explore the robustness of my findings to deviations from the perfect instrument exogeneity case using the approach suggested by Conley et al. (2012).

I find that primary enrollment growth has led to large 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. Access to pupil-specific physical inputs was little affected and classroom equipment may have improved somewhat, suggesting that the capitation grant accompanying the reform and aimed at covering *non-salary* costs was effective in maintaining access to physical inputs. Estimates of the effect of enrollment growth on pupil test scores are 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 move of about 5 percentiles down the distribution of test scores. 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 results are consistent with the expectation that pupil-teacher ratios increased in the wake of FPE, with previous literature finding no effect of smaller class sizes on test scores in developing country settings (Banerjee et al. (2007); Duflo et al. (2011); Duflo et al. (2012)) and with the well-established low correlation between pupil test scores and observable teacher

characteristics such as training and qualifications (Rivkin et al. 2005).

However, when investigating the possibility of heterogeneous effects for urban and rural areas, which start from very different achievement levels, 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 this conclusion is robust to substantial departures from the perfectly exogenous instrument case. These findings contrast with Lucas & Mbiti (2012), who find no effect of FPE in Kenya on the scores of cohorts who were partially affected across the distribution of scores, so that their conclusion that FPE was welfare enhancing in the short run in Kenya may not extend to Tanzania when considering cohorts fully exposed to FPE.

The rest of the paper is organised as follows. Section 2 describes the Tanzanian free primary education policy background, Section 3 reviews the relevant literature, Section 4 presents the identification strategy, Section 5 describes the data and provides a preliminary graphical analysis, Section 6 presents the estimation results and Section 7 concludes.

2 Universal Primary Education Policy

Primary education in Tanzania comprises 7 years (Standard I-VII), with a normal entry age of 7 years old. In 2000, only 53% of primary-school age children (i.e., age 7-13) attended school. This was despite early attempts at achieving universal primary education in the late 1970s, which culminated in a net enrollment rate of 70.3 in 1980 (World Development Indicators 2013). 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, 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, although increasing enrollment was the “highest priority” of the Primary Education

Development Plan for 2002-2006 (PEDP I) (Basic Education Development Committee (2001), p.4), the “second primary education policy priority” was to “revitalise and improve the quality of primary education” by “(a) improving teachers’ styles and methods in the classroom, (b) ensuring the availability of good quality learning and teaching materials; and (c) ensuring the necessary support for maintaining educational standards” (Basic Education Development Committee (2001), p.9).

PEDP I was very successful in increasing enrollment rates, with an increase in the net enrollment rate from 53.1% in 2000 to 96.5% in 2006 (World Development Indicators 2013). However, qualitative assessments suggest that very little was done towards achieving the quality improvement objective during 2002-2006 (Wedgwood 2007), and the quantitative objectives set by PEDP I for the pupil-teacher ratio (45:1) and the pupil-textbook ratio (1:1 in 2006) were far from achieved (52:1 and 1:3, respectively, in 2006 according to Basic Education Development Committee (2006)). The foreword to the second Primary Education Development Plan (2007-2011) by the Minister for Education and Vocational Training is clear: “The challenges are many, but the key one is that unequal attention has been paid to enrollment expansion and other objectives, namely quality improvement, capacity building and institutional arrangements or management.”.

Similar to other countries that implemented free primary education, a donor-funded capitation grant of US\$10 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 is expended at the discretion of the school committee for other non-salary costs (Basic Education Development Committee 2001).

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).

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 et al. 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 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.

In the next section, I review the literature on the impact of FPE on the quality of the schooling environment, as well as previous work on the effect of schooling inputs on learning.

3 Literature Review

3.1 Impact of Free Primary Education Reforms on School Quality

Most of the literature on the impact of FPE reforms focuses on their impact on enrollment and 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). The question of whether the quality of schooling was affected is often discussed, but no direct evidence is provided except in Lucas & Mbiti (2012). 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 at least 3 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 due to the program. 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. 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.

In Tanzania, Hoogeveen & Rossi (2013) estimate the impact of FPE on attendance and grade completion. Their household data confirm that enrollment rates 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 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, which are considered “treated”, and older children, which are not, Hoogeveen & Rossi (2013) conclude that the reform had a negative effect on educational attainment, which they hypothesize is due to a deterioration of the quality of schooling.¹

¹More specifically, Hoogeveen & Rossi (2013) estimate the effect of the reform on years of education completed by estimating a difference-in-difference specification in which the authors compare the number of years of education accumulated between 2001 and 2007 *minus the age-specific normal years of education* for a group of “treated” children (i.e., aged 8-10 in 2002) relative to a group of “untreated” children (i.e., children aged 11 in 2002). They find a negative treatment effect, which is statistically significant in rural areas. Children age 11 and older in 2002 were not prioritised by the reform. However, the extent to which the exclusion of children age 11 and above in 2002 was enforced in practice is unclear as final admission enrollment decisions for children older than the normal starting age of 7 were left to the discretion of the school committees (Basic Education Development Committee (2001), p.5). In addition, all new entrants were expected to start school in grade 1, so that treatment effects on the number of school years accumulated *minus the age-specific normal years of education* are difficult

3.2 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 recent 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).

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 pupils². 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 (Banerjee et al. (2007); Duflo et al. (2011); Duflo et al. (2012)).

to interpret. Finally, spillovers are likely across age cohorts both at the household level (between siblings) and at the school level (between school grades), which may lead to contamination of the control group.

²Seven pupils corresponds both to the estimated increase in the pupil-teacher ratio for a one standard deviation increase in enrollment growth found in Section 6.1 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)).

Studies considering the impact of teachers' observable measures of quality such as experience and education generally find little evidence that these characteristics play a role in students' learning (except, for experience, after the first few initial years). Rivkin et al. (2005) show that the variance of learning outcomes across teachers is large but uncorrelated to teacher education, teacher experience, or class-size. A review of the evidence on the effect of teacher subject-specific knowledge on students' test scores, however, finds strong positive effects (Glewwe et al. 2011).

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, there is no evidence that small class sizes, more educated and more experienced teachers, or more textbooks and other physical inputs have a noticeable effect on test scores in developing countries, except for teacher subject-specific knowledge.³ Therefore, it is unclear whether a reform such as FPE, 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.

4 Identification Strategy

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):

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

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, quantity of schooling, and teacher and non-teacher school inputs (e.g., pupil-teacher ratio,

³Instead, 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.

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. 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. 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 average student enrolled under FPE is likely to have lower SES, and may thus have a worse cognitive and non-cognitive endowment than previous students - a worsening of average test scores after FPE may therefore occur due to a negative composition effect, even when the test scores of pupils who would have been enrolled in the absence of FPE do not deteriorate.

I first estimate the effect of primary enrollment growth on an observable set of schooling inputs Z_{ij} , 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 quality between 2000 and 2007 across regions that experienced different rates of growth in primary enrollment. More precisely, I estimate the following equation, in which some measure of quality y_{irt} is regressed on a post-reform (survey) dummy ($\mathbf{1}(t = 2007)_t$), region dummies (β_r), individual and regional (time-varying) controls (X_{irt}), and the interaction between the post-survey dummy and the relative size of post-reform enrollment (cumulated over 2002-2007) relative to baseline enrollment:

$$y_{irt} = \beta_0 + \beta_1 \left(\frac{\text{post_enrol}}{\text{baseline_enrol}} \right)_r \times \mathbf{1}(t = 2007)_t \quad (2)$$

$$+ \mathbf{1}(t = 2007)_t + \beta_r + X'_{irt} \beta_X + \epsilon_{irt}$$

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 $\text{post_enrol}_r = \sum_{j=2002}^{2007} E_{rj}$ is the sum of the number of pupils enrolled in primary schools in region r during 2002-2007, the 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 . 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 confidence intervals and 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 important 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 would result in β_1 being an underestimate of the worsening of quality (i.e., would bias β_1 towards less negative values). I address this concern in a robustness check in which I control for discretionary government transfers. 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 never spoken at home, a household item ownership score (based on 14 items), and for maternal and paternal education levels.

There may remain unobservable sources of endogeneity. 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. In order to address these remaining issues, I use *potential* growth in enrollment

based on predetermined fertility decisions 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:

$$\begin{aligned} \left(\frac{post_enrol}{baseline_enrol}\right)_r \times \mathbf{1}(t = 2007)_t &= \gamma_0 + \gamma_1 \left(\frac{post_age7_13}{baseline_age7_13}\right)_r \times \mathbf{1}(t = 2007)_t \\ &+ \mathbf{1}(t = 2007)_t + \gamma_r + X'_{irt} \gamma_X + \nu_{irt} \end{aligned} \quad (3)$$

where $baseline_age7_13_r$ 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 2002-2007.

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

$$\begin{aligned} y_{irt} &= \lambda_0 + \lambda_1 \left(\frac{post_enrol}{baseline_enrol}\right)_r \times \mathbf{1}(t = 2007)_t \\ &+ \mathbf{1}(t = 2007)_t + \lambda_r + X'_{irt} \lambda_X + mu_{irt} \end{aligned} \quad (4)$$

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

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. Therefore, this issue would tend to lead to an overestimation of the worsening of quality coinciding with higher enrollment growth.

To see this, consider the following system:

$$Y = \beta X + \gamma Z + \epsilon \quad (5)$$

$$X = \nu Z + \mu \quad (6)$$

and the corresponding reduced-form:

$$Y = \beta\nu Z + \gamma Z + \phi \quad (7)$$

Consider the case in which $\beta \leq 0$ (higher enrollment growth may worsen test scores), $\nu > 0$ (e.g., 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). If $\gamma = 0$, then 2SLS identifies $\frac{\beta\nu}{\nu} = \beta$. If $\gamma \neq 0$, then 2SLS identifies $\frac{\beta\nu + \gamma}{\nu}$, 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 effect of enrollment growth on test scores, an omitted variable would have to be positively correlated both with fertility (between 1988 and 2000) and with investments in educational quality or child human capital. 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 Section 6.2 I provide empirical evidence in support of the argument that the most likely direction of the bias, if any, is indeed to overestimate any worsening in quality. In the urban sample, in which I find a significant negative effect of enrollment growth on achievement, I explore the robustness of my findings to deviations from the perfect instrument exogeneity case ($\gamma = 0$) using the approach suggested by Conley et al. (2012).

5 Data and Preliminary Analysis

5.1 Pupil 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 col-

lected 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 ensures that the survey is representative of all 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 (TIMMS) 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.

The final sample is obtained as follows. Five observations 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.⁴

⁴Some 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.

5.2 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.

5.3 Summary Statistics

Table 1 reports summary statistics separately for pupils observed in the 2000 and 2007 SACMEQ surveys.

The first three rows 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 would have 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 covered by the data.

The third row shows that the total amount received by regions in education block 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, 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 (by close to three years). Consistent with the overall improvement in the education of teachers, the performance of teachers in subject-specific tests have 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 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 1 paint a striking pattern: despite the large 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).

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 children coming from poorer backgrounds has been more than compensated by the positive trend in

standards of living between 2000 and 2007.⁵

5.4 Graphical Analysis

All in all, the summary statistics reported above seem to confirm the expectation that the sharp increase in enrollment had a large effect on pupil-teacher ratios, that new teachers with lower subject-specific knowledge had to be recruited at lower levels of teacher training to meet the demand, but that physical inputs have tended to improve over time while pupil achievement progressed substantially. At the aggregate level, it is not possible to disentangle the role of enrollment growth from that of national trends that are independent of enrollment growth (e.g., the increase in the proportion of teachers with O-level qualifications is likely to be due to a positive trend in education rather than enrollment growth). In this section I present a graphical analysis comparing changes in schooling quality between 2000 and 2007 across regions with different enrollment growth rates to shed more light on the particular role played by the latter.

In Figures 1 to 3, I sort regions, from left to right, by enrollment growth, which is represented by the unlabelled dots. In Figure 1, I then superimpose first-differences (i.e., the change between the 2000 and 2007 surveys) in the mean pupil-teacher ratio in the region (the labelled dots), which shows the positive correlation between regional increases in the pupil-teacher ratio between 2000 and 2007 and regional enrollment growth. In Figures 2 and 3, I instead superimpose the first-differences in reading (Figure 2) and math (Figure 3), which shows no correlation between changes in regional average pupil test scores and regional enrollment growth.

Of course, enrollment growth may be endogenous, and therefore in what follows I test the robustness of my results to instrumenting enrollment growth. Figure 4 shows the strong correlation between the growth rates of the primary-school age cohorts and that of actual enrollment, thus illustrating the relevance of the instrument.

One may expect regions with lower scores in 2000 to experience larger improvements in test scores between 2000 and 2007 due to there being more

⁵The 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.

low-hanging fruits to be picked or due to noise in the measure of achievement based on test scores. This is indeed the case, as illustrated by Figures A-2 and A-4. 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-2 and A-4 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-3 and A-5 show no systematic relationship between baseline test scores and enrollment growth. In order to confirm that mean reversion is not driving my results, in the next section 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.

6 Results

In this section, I first present the main results, starting with the effect of enrollment growth on observable schooling inputs and then turning to its total effect on average test scores. Second, I discuss the robustness of my findings to a number of changes in the main specification. Finally, I investigate the effect of enrollment growth across the distribution of test scores and for different population groups.

6.1 Main Results

Table 2 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.⁶ The wild cluster bootstrap-t p-value is 0.008, thus confirming the conclusions based on analytical standard er-

⁶Here I refer to the standard deviation of the distribution of enrollment growth in the 2007 sample (0.693, see Table 1)

rors. 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 not weak. 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 actual enrollment (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, although confidence intervals include large increases in both (namely, a 4%-points increase in the probability of operating multiple shifts compared to the sample mean of 10% and 3.4 additional weekly teaching hours, each for a one s.d. increase in enrollment growth).

Table 3 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. The lower bounds of the estimate for O-level qualifications correspond to small

effects relative to the sample mean at -0.035 and -0.040 for Kiswahili and Mathematics, respectively (for a 1 standard deviation increase in actual enrollment). I cannot rule out larger decreases in the share of teachers with at least two years of teacher training (-0.17 for Kiswahili and -0.08 for Mathematics teachers, respectively).

Table 4 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 have less experienced teachers (by 1.4 years for Kiswahili teachers and 2.8 years for Math teachers), 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 5 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 earmarked 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 5, 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. The estimates of the effect of enrollment growth on the average pupil's number of exercise books and pupil equipment score are both statistically insignificant, and the confidence intervals imply that I can rule out positive or negative effects larger than 3% of the sample mean, except for the upper bound of the effect of (a 1 s.d.) increase in enrollment growth on the number of exercise books, which corresponds to a 9% increase relative to the sample mean. The OLS point estimate for the effect of enrollment growth on Kiswahili class equipment is 0.155, which corresponds to a negligible effect of 0.11 for an increase in enrollment growth by 1 standard deviation (just 2.5% of the sample mean equipment index) and is statistically insignificant. The IV estimate is much larger, however, significant at the 5% level and corresponds to an 11% increase.⁷ Equally large improvements in Mathematics class equipment cannot be ruled out, although the IV point estimates are smaller and not statistically significant.

Results so far indicate that enrollment growth has led to large increases in pupil-teacher ratios, a substantial 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. I now turn to estimating the total effect of these observable and other unobservable changes in the arguments entering the test scores production function on pupil test scores.

In Table 6, 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 (0.15 using the wild cluster bootstrap-t method) of a standard deviation for an increase in enrollment growth by 1 standard deviation. In Column (2), I add controls

⁷The difference between the OLS and IV estimates suggest that some omitted variable positively correlated with actual enrollment growth is negatively correlated with changes in class equipment. One possible such omitted variable may be that head teachers whose schools were not benefiting as much from the increasing trend in classroom equipment were more likely to over-report enrollment figures in order to receive more capitation grants.

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 5 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. 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.

In Column (3), I report IV estimates showing a small, positive point estimate with a 95% CI lower bound which is less negative than that obtained with OLS for the same set of regressors. An exogeneity test that is robust to heteroskedasticity fails to reject the null hypothesis that enrollment growth is exogenous in Equation 2. Therefore, OLS is the preferred estimator as it is more efficient.

Table 7 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., and when controlling for all pupils' characteristics and physical inputs (as in Column 2 of Table 7), pupils in rural areas perform on average 0.31 s.d. below their urban counterparts, girls obtain 0.37 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 0.38 s.d..

6.2 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 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 experience faster (potential) enrollment growth and been increasingly targeted by government education funding, which could bias my estimates if government transfers also increased achievement. To assess the validity of this concern, in Columns (4) of Tables 6 and 7, I add 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 an interaction term between a post-reform survey dummy and the baseline regional average pupil score (in reading in Table 6 and in math in Table 7). 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 neither on Kiswahili nor on 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 Section 4, I showed that, if regions where fertility decreased more slowly during the period 1988-2000 (i.e., regions with higher potential enrollment growth) also saw smaller increases in investments in the human capital of children, then my IV estimates would tend to be biased downwards - i.e., they would tend to overestimate any worsening in test scores due to enrollment growth. If regions where fertility decreased more slowly during this period somehow saw a relative improvement of investments in human capital or of living conditions, then a concern would be that my IV estimates would in fact tend to underestimate the worsening of test scores due to enrollment growth. Given that I find no evidence of a substantial worsening of test scores, finding that regions with higher potential enrollment growth also saw a relative improvement in child outcomes observed before the reform would cast doubt on my results. On the contrary, finding that regions with higher potential enrollment growth also saw a relative worsening in child outcomes observed before the reform would reinforce my conclusion that enrollment growth did not substantially worsen achievement on average.

In order to empirically sign the most likely direction of the bias, 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:

$$health_{irt} = \xi_0 + \xi_1 \left(\frac{post_enrol}{baseline_enrol} \right)_r \times \mathbb{1}(t \geq 1995)_t + \mathbb{1}(t \geq 1995)_t + \xi_r + \phi_{irt} \quad (8)$$

where $health_{irt}$ 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.

Results are reported in Table A-1. All the point estimates go in the direction of smaller improvements in child health inputs and outcomes in regions in which fertility was decreasing more slowly during the period 1988-2000, statistically significantly so in the case of full immunization and delivery by a health professional. 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 average achievement.

6.3 Looking Beyond the Mean

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. In order to explore potential heterogeneous effects across the distribution of test scores, I first use Athey & Imbens (2006)'s Changes-in-Changes estimator. 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 Change-in-Changes estimate of the treatment effect at quantile q can be written:

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

where $F_{Y,gt}(y)$ is outcome Y 's cumulative distribution function in group g in period t and Y_i^I denotes the outcome for individual i when treated. I define the treated group as those regions with above-median potential enrollment growth.⁸ Similar to Lucas & Mbiti (2012), I exclude observations at the bottom of the distribution of test-scores (here, the bottom 25%) in order to satisfy the requirement that the underlying ability distribution within group does not vary over time, as FPE may have led children with

⁸I present here reduced-form estimates based on potential rather than actual enrollment growth because I cannot test for the endogeneity of enrollment growth at each quantile.

lower ability to enroll. Also in order to maintain comparability between characteristics of pupils within group over time, I control for the set of covariates included in Column (2) of Tables 6 and 7. Results are displayed in Figures 5 and 6. Looking first at Figure 5, we can see that there is no evidence of a worsening of test scores, except perhaps some moderate worsening at the very top of the distribution where the lower bound of the 95% CI translates into a 0.10 of a standard deviation decrease in test scores for a one standard deviation increase in potential enrollment growth (which is equal to 0.30, as shown in Table 1). There is more heterogeneity in the effect of potential enrollment growth on mathematics test scores, as point estimates become negative from the 70th percentile (of the truncated distribution) onwards, with lower bounds of the 95% CI corresponding to the effect of a one standard deviation increase in potential enrollment growth ranging from -0.09 to -0.22 (Figure 6).

One limitation of this Changes-in-Changes exercise is that I cannot exploit all the variation in enrollment growth as it is reduced to a binary treatment variable. I therefore now look beyond treatment effects at the mean in two different ways. First, I repeat my initial Difference-in-Differences analysis using indicators of achievement at high or low levels of competency, instead of test scores. Second, I split the sample between rural and urban areas, as urban areas have a 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).

Table A-2 reports results of regressions in which the dependent variables are dummy variables that switch on, in turn, when a pupil: demonstrates analytical or critical reading (Panel A), displays competent numeracy to abstract problem solving ability (Panel B), has no more than “basic reading” skills (Panel C), has no more than “emergent numeracy” skills (Panel D). These indicators were created based on two variables (one for reading, and one for math) comprising eight competency levels each. The point estimates in the two top panels are very close to zero, and tend to be negative in the bottom two panels (i.e., going in the direction of a *decrease* in the incidence of low competency), although the 95% CI do not allow me to rule out non-negligible changes in the proportion of pupils with high or low reading or mathematics competency with respect to the sample means.

Finally, I repeat the analysis of Section 6.1 separately on the rural and urban sample. 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 8 reports the results. Starting with the rural sample, results are similar to the average effects reported in Section 6.1, 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. 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 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. Given the large baseline differences in average math teachers' z-scores in rural (-0.448) and urban (0.187) areas, the contrasting effect of the recruitment of new teachers is perhaps not so surprising: with 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. Similarly, 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 last rows of Table 8 Panels B and C). It would be tempting to attribute the change in pupils' test scores to that in teacher subject-knowledge. 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.⁹ On the other hand, these controls are based on characteristics of the pupils 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.

For the urban sample, IV and OLS estimates of the effect of enrollment growth on achievement are broadly similar. As argued in Sections 4 and 6.2, the sign of the bias which my instrument may suffer from is negative - i.e., it may overestimate the worsening of achievement due to enrollment growth. In order to shed light on the robustness of my conclusions for the urban sample to this possible source of bias, I implemented the Local-to-Zero procedure suggested by Conley et al. (2012), which 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 and clustering the standard errors at the region level, as in the main analysis. As shown in Figure 7, 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 effect of potential growth on achievement after conditioning on actual enrollment growth does not exceed 30% of its reduced-form unconditional effect on achievement, 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 8 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

⁹Full results are available on request.

these point estimates to achieve statistical significance.

7 Conclusion

The past two decades have seen many poor countries remove user fees in primary education in order to increase enrollment. These policies have largely succeeded in doing so, sometimes within a very short period of time. One such example is Tanzania. Despite considerable concern about this issue in policy circles, there is no previous direct, arguably causal evidence on the impact of large increases in enrollment on the quality of learning for cohorts fully exposed to a Free Primary Education reform.¹⁰

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 large 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. Access to pupil-specific physical inputs was little affected and classroom equipment (such as writing boards) may have improved somewhat, suggesting the the capitation grant for non-salary costs accompanying the reform was effective in maintaining access to physical inputs. When analyzing the effect of enrollment growth in rural and urban areas separately, findings for schooling inputs are qualitatively similar except for math teachers' subject-specific knowledge, which worsened dramatically in urban areas but improved in rural areas with the influx of new teachers. These opposite effects may be explained by the inability of urban schools to "cream-skin" new teachers when faced with large increases in enrollment, and the very low baseline levels of subject-specific knowledge of rural math teachers.

Estimates of the effect of enrollment growth on learning outcomes, as measured by average pupil test scores for the country as a whole, are 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

¹⁰In Kenya, Lucas & Mbiti (2012) estimate the impact of FPE on test scores of students who had been in school for at least 3 years before FPE.

reading (math) scores by 0.15 (0.16) of a standard deviation. How “large” are these effects? As a point of comparison, 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. These conclusions are robust to a range of robustness checks, including to instrumenting enrollment growth using predetermined fertility decisions, and I show that any plausible source of instrument endogeneity is unlikely to be driving these findings.

However, when investigating the possibility of heterogeneous effects for urban and rural areas, which start from very different achievement levels, 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). 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. The estimated effect of enrollment growth on urban test scores is not reduced when including controls for 6th-grade teacher experience, 6th-grade teacher z-score and the pupil-teacher ratio in the school during 6th grade, thus giving no indication that these are important mechanisms in the transmission of enrollment growth to worse test scores, although the same variables are not observed for previous grades.

This study shows that large increases in enrollment in primary schooling do not have to come at the cost of a substantial worsening of the quality of the learning environment across the board, but that quality losses may be concentrated within specific schooling environments. My results are consistent with the expectation that pupil-teacher ratios increased because of FPE, with previous literature finding no effect of smaller class sizes on test scores in developing country settings (Banerjee et al. (2007); Duflo et al. (2011); Duflo et al. (2012)) and with the well-established low correlation between pupil test scores and observable teacher characteristics such as

training and qualifications (Rivkin et al. 2005). Importantly, these findings contrast with Lucas & Mbiti (2012)'s conclusion that FPE in Kenya was welfare-enhancing for cohorts affected by FPE half-way through their primary schooling, as I find that, in Tanzania, when estimating the effect of enrollment growth in the wake of FPE for cohorts fully exposed to FPE, wider access came at the cost of quality in urban areas. 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.

References

- Al-Samarrai, S. & Zaman, H. (2007), ‘Abolishing school fees in malawi: The impact on education access and equity 1’, *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.
- 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.
- Basic Education Development Committee (2006), *Primary Education Development Plan II (2007-2011)*, Dar Es Salaam: Government of the United Republic of Tanzania.
- Bennell, P. & Mukyanuzi, F. (2005), ‘Is there a teacher motivation crisis in tanzania?’.
- 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. (2011), ‘Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya’, *American Economic Review* **101**(5), 1739–74.
- Duflo, E., Dupas, P. & Kremer, M. (2012), School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from kenyan primary schools, Technical report, National Bureau of Economic Research.
- 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, Technical report, National Bureau of Economic Research WP17554.
- 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.

- 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.
- 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.
- Ministry of Education [Tanzania] (1999-2007), Basic education statistics in tanzania, Technical report, Ministry of Education [Tanzania]. Dar Es Salaam.
- MoEVT, Office, U. D. & de Dakar, P. (2012), *Tanzania Education Sector Analysis: Beyond primary education, the quest for balanced and efficient policy choices for human development and economic growth*, United Republic of Tanzania, Ministry of Education and Vocational Training.
- 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.
- Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (2005), ‘Teachers, schools, and academic achievement’, *Econometrica* **73**(2), 417–458.
- 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.
- Wedgwood, R. (2007), 'Education and poverty reduction in tanzania', *International Journal of Educational Development* **27**(4), 383 – 396. Beyond Basic Education and Towards an Expanded Vision of Education - for Poverty Reduction and Growth?
- World Development Indicators (2013), *World Development Indicators*, The World Bank.

Table 1: 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 \geq 2 y (K)	0.94		0.81	
=1 if Training \geq 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

	SACMEQ 2000		SACMEQ 2007	
	mean	sd	mean	sd
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	
=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	
<i>N</i>	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. Class equipment score items: writing board, chalk, wall chart, cupboard, bookshelves, library, teacher table, teacher chair. 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 2: Effect on Quantity of Teachers

	(1)	(2)	(3)	(4)
	Pupil/Teacher	=1 if Multiple Shifts	Teaching Hours (K)	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 95% CI	[2.207,18.047]	[-.128,.058]	[-3.131,5.921]	[-2.867,2.689]
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 endogeneity test	0.486	0.181	0.182	0.022

(1)	(2)	(3)	(4)
Pupil/Teacher	=1 if Multiple Shifts	Teaching Hours (K)	Teaching Hours (M)

Sample size: 6933. Region-correlated robust standard errors in parentheses. 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 95% confidence intervals were obtained using the method proposed in Cameron et al. (2008).

Table 3: 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 95% CI	[-.048,.083]	[-.064,.131]	[-.187,.167]	[-.118,.061]
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 endogeneity test	0.069	0.165	0.067	0.476

Sample size: 6933. Region-correlated robust standard errors in parentheses. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * p<0.10, ** p<0.05, *** p<0.01. Further details as per notes under Table 2.

Table 4: 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
Panel A: OLS Estimates				
(=1 if 2007)*Enrol. 2002-2007	-2.042**	-4.018***	-0.207**	0.049
/Enrol. in 2001	[-3.796,-0.288]	[-5.508,-2.528]	[-0.367,-0.047]	[-0.177,0.274]
Wild cluster bootstrap-t 95% CI	[-4.134,-.140]	[-6.324,-1.789]	[-.398,-.018]	[-.237,.326]
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]	[-0.439,-0.013]	[-0.410,0.353]
Panel C: Reduced-Form Estimates				
(=1 if 2007)*Age 7-13 in 2002-2007	-4.634**	-7.324***	-0.374*	-0.048
/Age 7-13 in 2001	[-8.133,-1.134]	[-10.887,-3.760]	[-0.761,0.012]	[-0.733,0.638]
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 endogeneity test	0.297	0.597	0.801	0.473

Sample size: 6933. Region-correlated robust standard errors in parentheses. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * p<0.10, ** p<0.05, *** p<0.01. Further details as per notes under Table 2.

Table 5: Effect on Physical Inputs

	(1)	(2)	(3)	(4)	(5)	(6)
	=1 if Own K. Bk	=1 if Own M. Bk	Class Equip.	Class Equip.	# Ex. Books	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 95% CI	[-.026,.026]	[-.025,.019]	[-.848,1.215]	[-.454,.347]	[-.700,.484]	[-.216,.262]
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 endogeneity test	0.662	0.469	0.093	0.048	0.089	0.157

Sample size: 6933. Region-correlated robust standard errors in parentheses. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * p<0.10, ** p<0.05, *** p<0.01. Further details as per notes under Table 2.

Table 6: Effect on Pupil Reading Test Scores

	Pupil Reading Test Scores					
	(1)	(2)	(3)	(4)	(5)	(6)
(=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 95% CI	[-.211,.227]	[-.211,.201]		[-.237,.166]	[-.251,.115]	
(=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 95% CI						[-.304,.471]
Pupil Characteristics	No	Yes	No	No	No	No
Physical Inputs	No	Yes	No	No	No	No
Observations	6933	6933	6933	6933	6933	6933
No. of clusters	19	19	19	19	19	19
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			

	(1)	(2)	(3)	(4)	(5)	(6)
	Pupil Reading Test Scores					
KP underid stat H0=underid			4.658			
KP underid p-val			0.031			
p-val robust endogeneity test			0.437			

Sample size: 6933. Region-correlated robust standard errors in parentheses. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Further details as per notes under Table 2.

Table 7: Effect on Pupil Math Test Scores

	Pupil Math Test Scores					
	(1)	(2)	(3)	(4)	(5)	(6)
(=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 95% CI	[-.227,.311]	[-.198,.265]		[-.250, .193]	[-.330, .242]	
(=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 95% CI						[-.642,.669]
Pupil Characteristics	No	Yes	No	No	No	No
Physical Inputs	No	Yes	No	No	No	No
Observations	6933	6933	6933	6933	6933	6933
No. of clusters	19	19	19	19	19	19
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			

	(1)	(2)	(3)	(4)	(5)	(6)
	Pupil Math Test Scores					
KP underid stat H0=underid			4.658			
KP underid p-val			0.031			
p-val robust endogeneity test			0.648			

Sample size: 6933. Region-correlated robust standard errors in parentheses. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Further details as per notes under Table 2.

Table 8: Heterogeneity between Rural and Urban Areas

	(1)	(2)	(3)	(4)
	Rural Sample		Urban Sample	
	OLS	IV	OLS	IV
Panel A: Teacher Variables				
Pupil-Teacher Ratio	10.063** [2.069,18.058]	9.854* [-1.326,21.035]	7.260 [-4.664,19.183]	16.660* [-0.612,33.933]
Years Experience (K)	-1.526 [-3.873,0.822]	-1.417 [-5.005,2.171]	-0.722 [-5.016,3.571]	-3.194 [-7.918,1.529]
Years Experience (M)	-4.373*** [-6.461,-2.285]	-3.337** [-6.029,-0.646]	-1.298 [-3.888,1.292]	-4.039** [-7.861,-0.217]
Teacher Z-Score (K)	-0.260** [-0.497,-0.022]	-0.275* [-0.563,0.013]	-0.065 [-0.536,0.407]	-0.062 [-0.471,0.348]
Teacher Z-Score (M)	0.253** [0.043,0.463]	0.272** [0.023,0.520]	-0.649*** [-1.109,-0.190]	-0.994** [-1.756,-0.232]
Panel B: Reading Scores				
OLS, No Controls	0.038 [-0.150,0.226]		-0.193 [-0.466,0.079]	
OLS, Controls	0.027 [-0.093,0.148]		-0.150 [-0.378,0.078]	
IV		0.100 [-0.105,0.304]		-0.227 [-0.517,0.062]
KP F-Stat		14.315		23.392
Endog. P-Value		0.429		0.677
OLS, Gov. Transfers	0.018 [-0.169,0.206]		-0.247* [-0.530,0.036]	
OLS, Mean Rev.	-0.026 [-0.189,0.136]		-0.314** [-0.574,-0.054]	
OLS, Full Set	-0.018 [-0.130,0.093]		-0.261* [-0.530,0.008]	

	(1)	(2)	(3)	(4)
	Rural Sample		Urban Sample	
	OLS	IV	OLS	IV
Panel C: Math Scores				
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]	

Sample size: 6933. Region-correlated robust standard errors in parentheses. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Further details as per notes under Table 2.

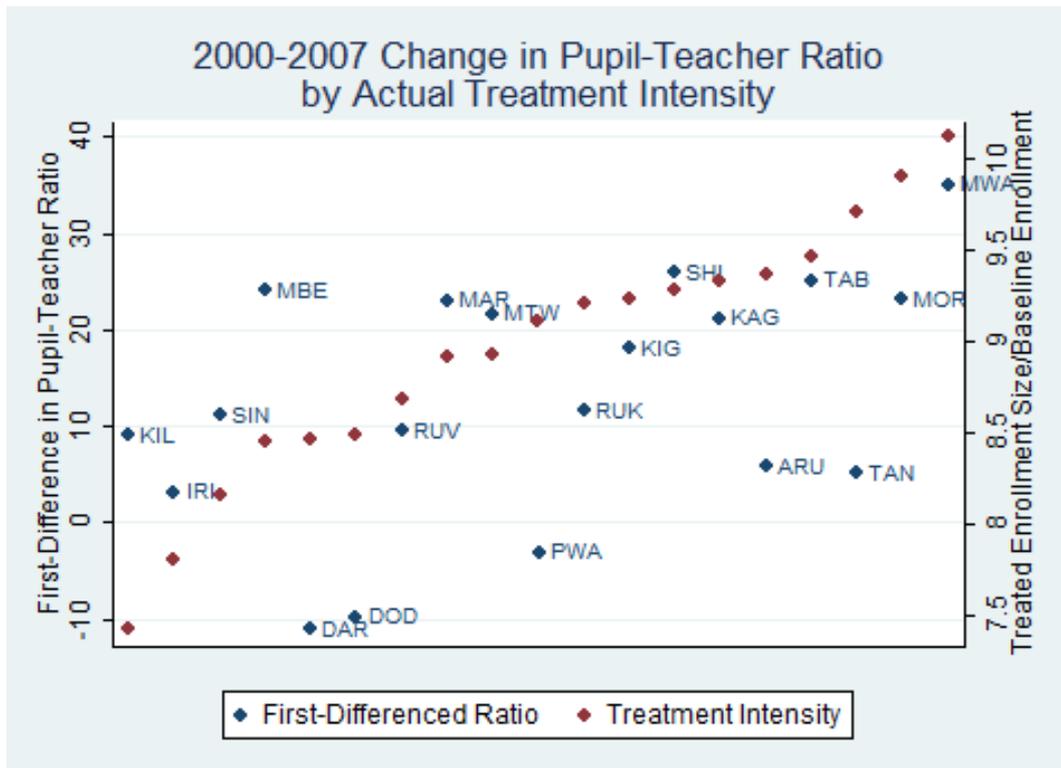


Figure 1

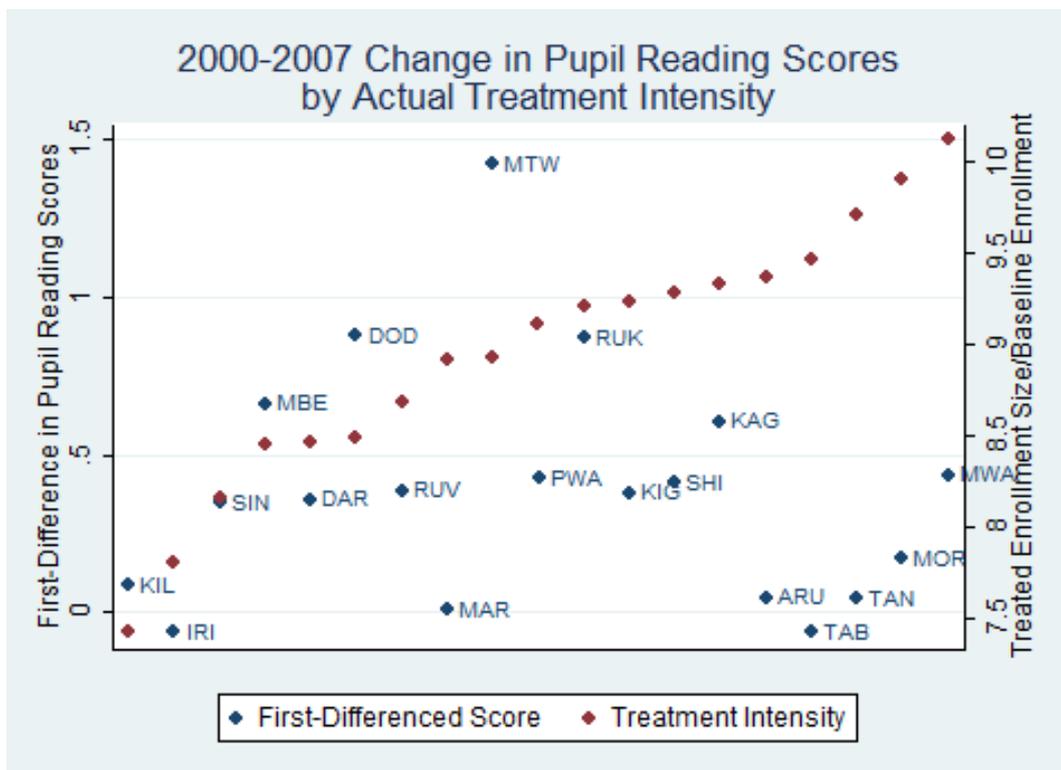


Figure 2

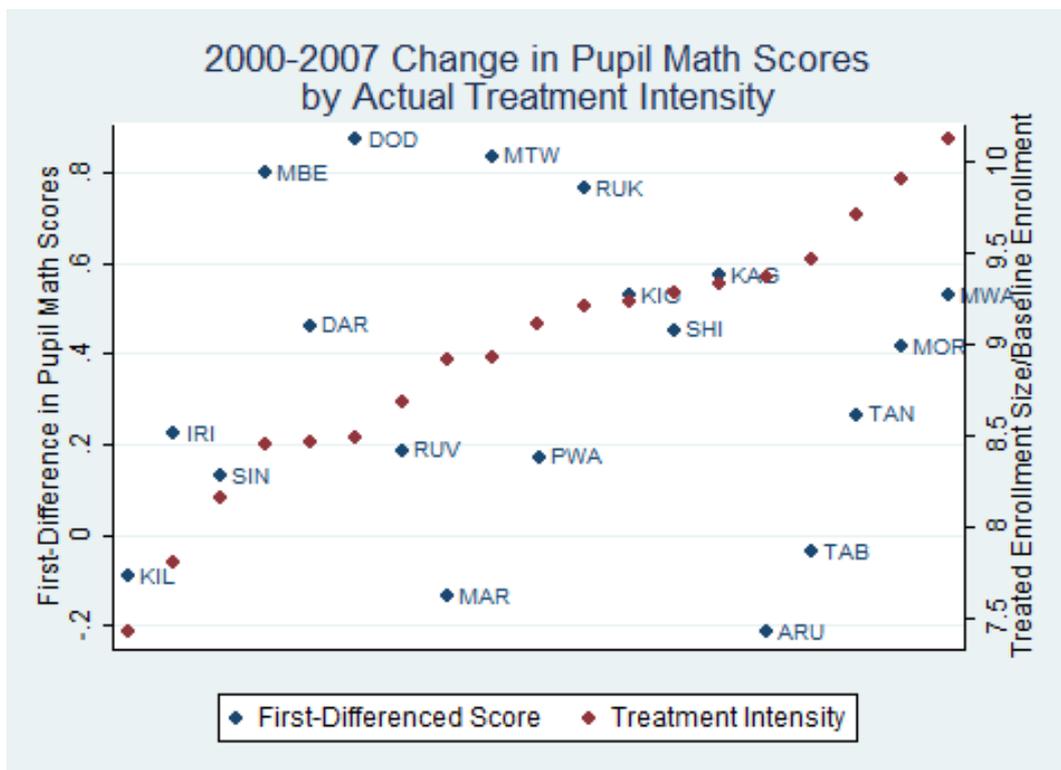


Figure 3

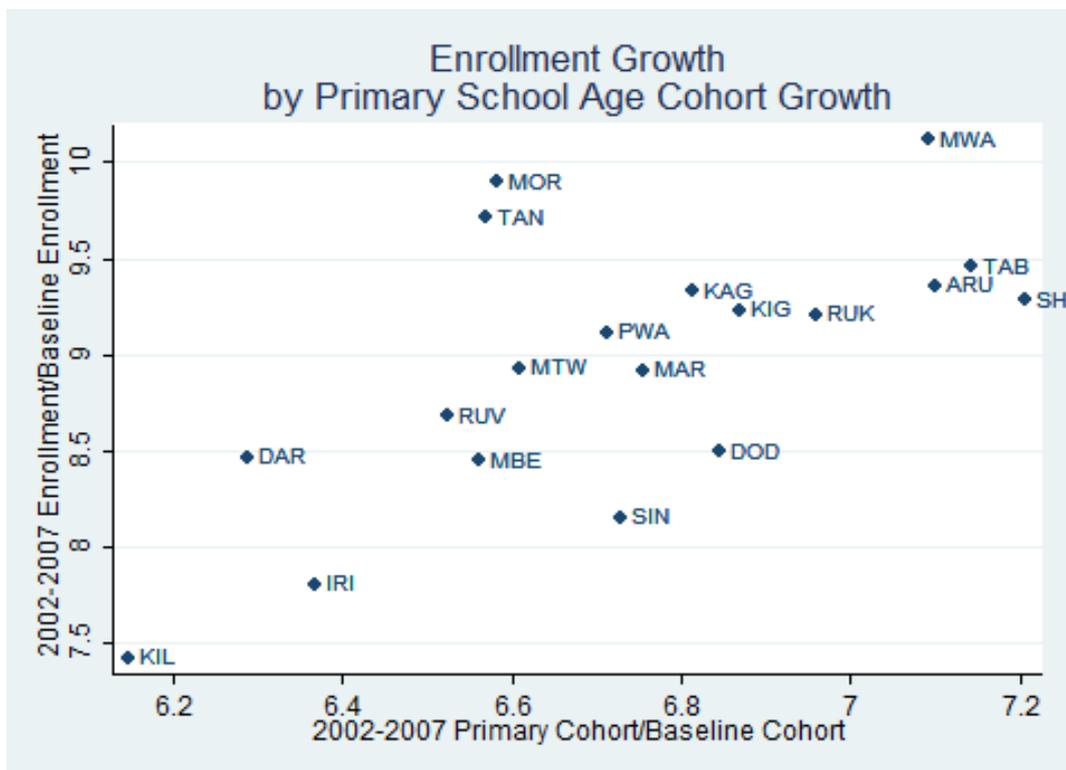


Figure 4

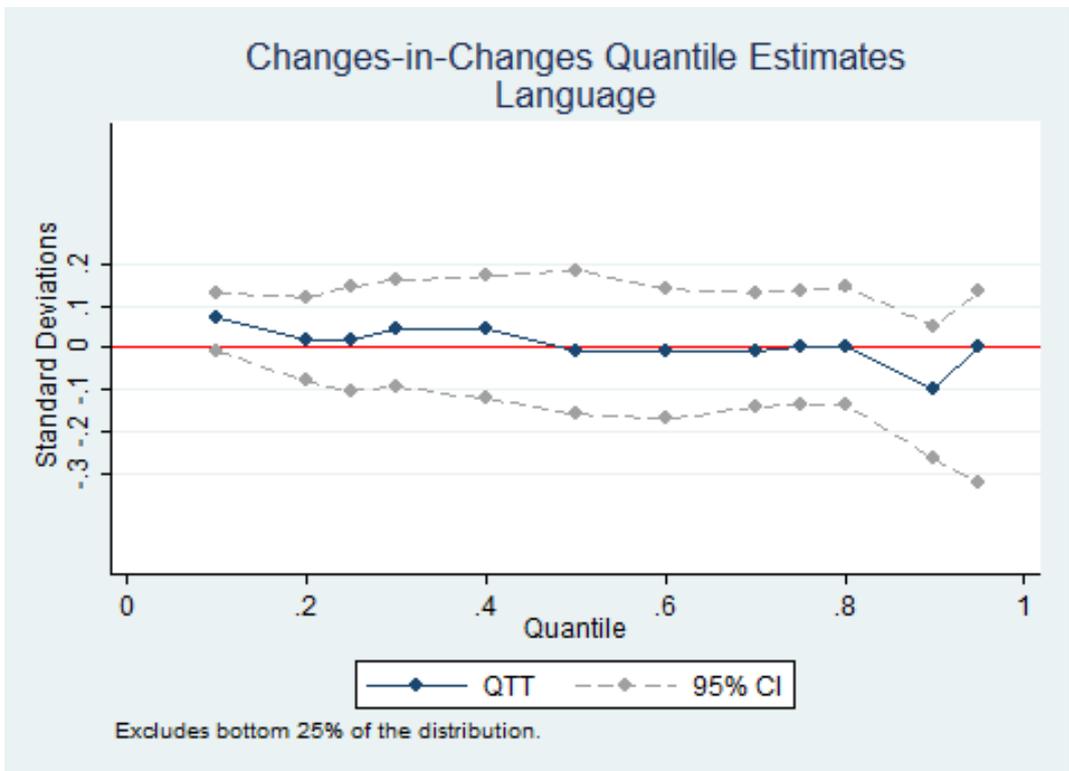


Figure 5

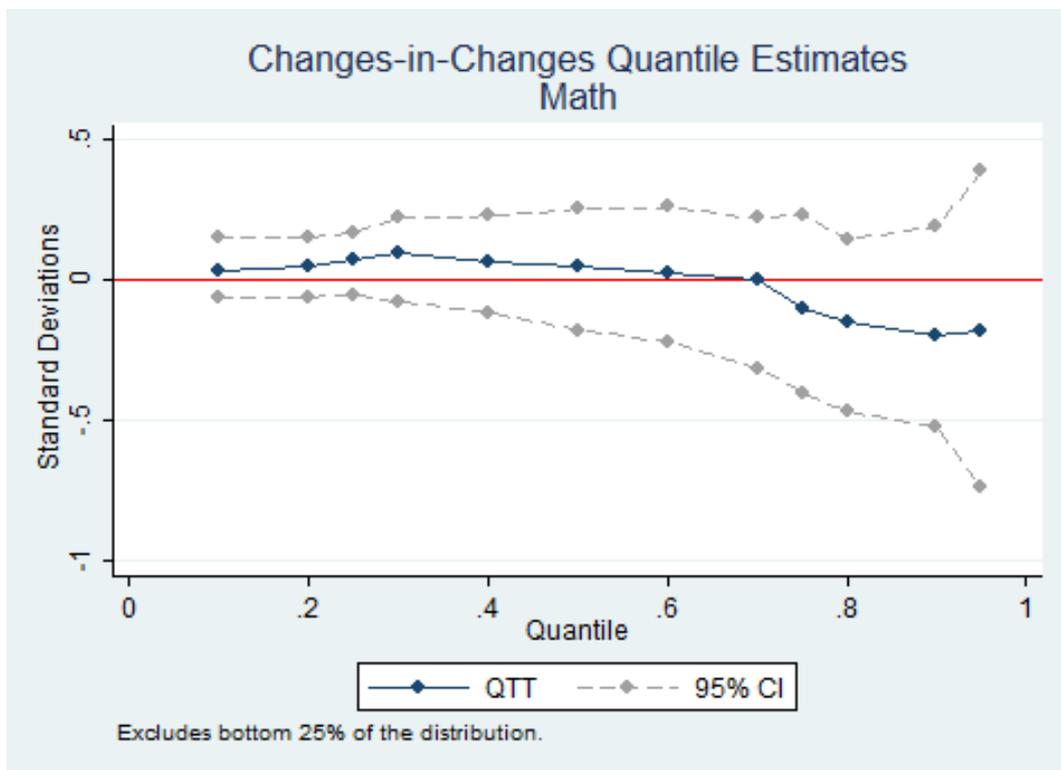


Figure 6

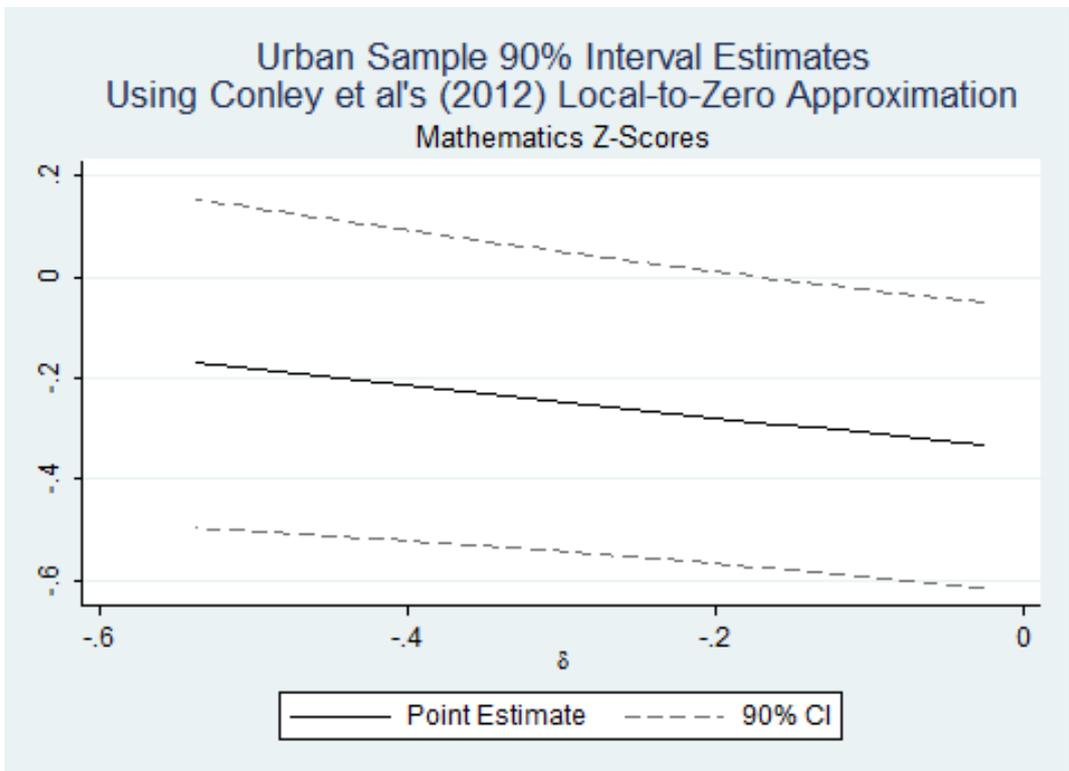


Figure 7

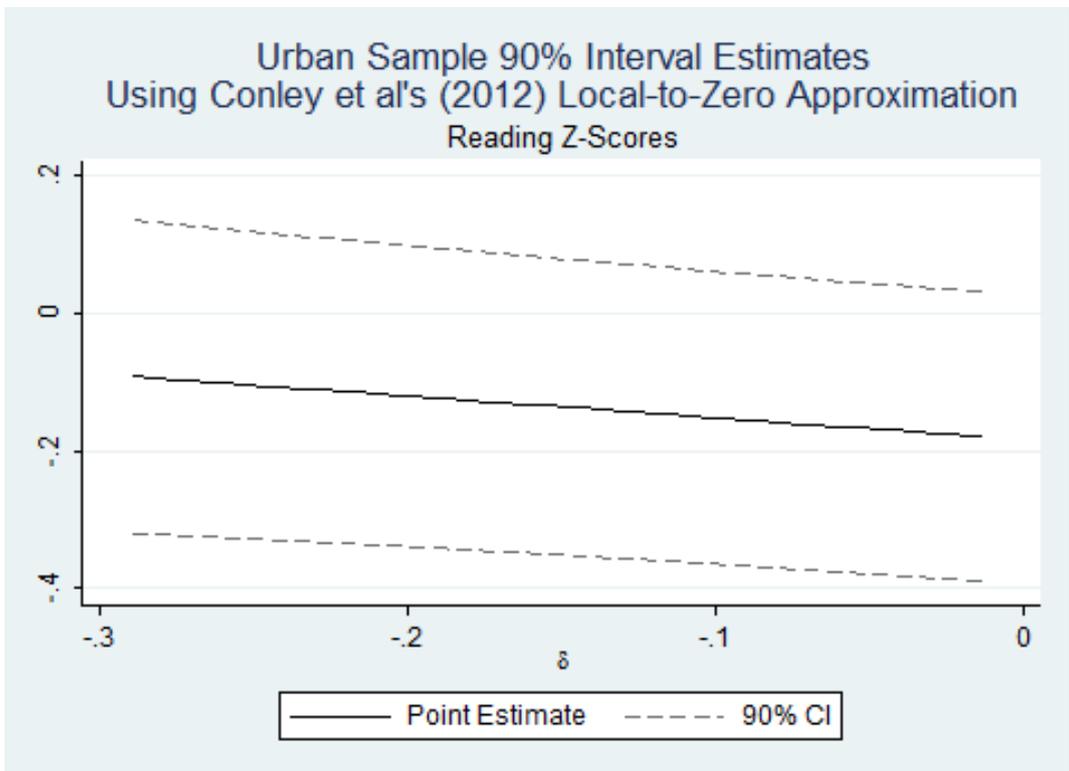


Figure 8

Appendix

Table A-1: 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/ Age 7-13 in 2001	-0.161** [-0.299,-0.022]	-0.248*** [-0.420,-0.076]	0.048 [-0.030,0.127]	0.023 [-0.030,0.076]	0.068 [-0.020,0.156]
=1 if born 1995-2000	1.066** [0.146,1.986]	1.631*** [0.463,2.799]	-0.343 [-0.879,0.192]	-0.152 [-0.512,0.208]	-0.547* [-1.149,0.056]
Region Dummies	Yes	Yes	Yes	Yes	Yes
Observations	11093	16233	16233	12492	12675
No. of clusters	20	20	20	20	20
R-squared	0.03	0.06	0.06	0.00	0.04
Mean Y	0.71	0.49	0.06	0.09	0.43

Region-correlated robust standard errors in parentheses. 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 whose mother was ever married. Samples vary due to immunization sample including only children age 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., age 12 months old or older), as well as 1378, 75 and 3633 truly missing values for immunizations, help with delivery and height measurement, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel D: =1 if low mathematics competency						
(=1 if 2007)*Enrol. 2002-2007	-0.030	-0.029	-0.035	-0.012	0.002	
/Enrol. in 2001	[-0.070,0.010]	[-0.067,0.010]	[-0.101,0.030]	[-0.055,0.030]	[-0.031,0.035]	
(=1 if 2007)*Age 7-13 in 2002-2007						-0.058
/Age 7-13 in 2001						[-0.180,0.063]

Sample size: 6933. Region-correlated robust standard errors in parentheses. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * p<0.10, ** p<0.05, *** p<0.01. ^aThe p-values of tests of the endogeneity of enrollment growth are above 0.10 for all outcome variables except low reading competency, so that results presented in Columns (4) and (5) are 2SLS estimates for low reading competency and OLS estimates for the other three dependent variables. Further details as per notes under Table 2.

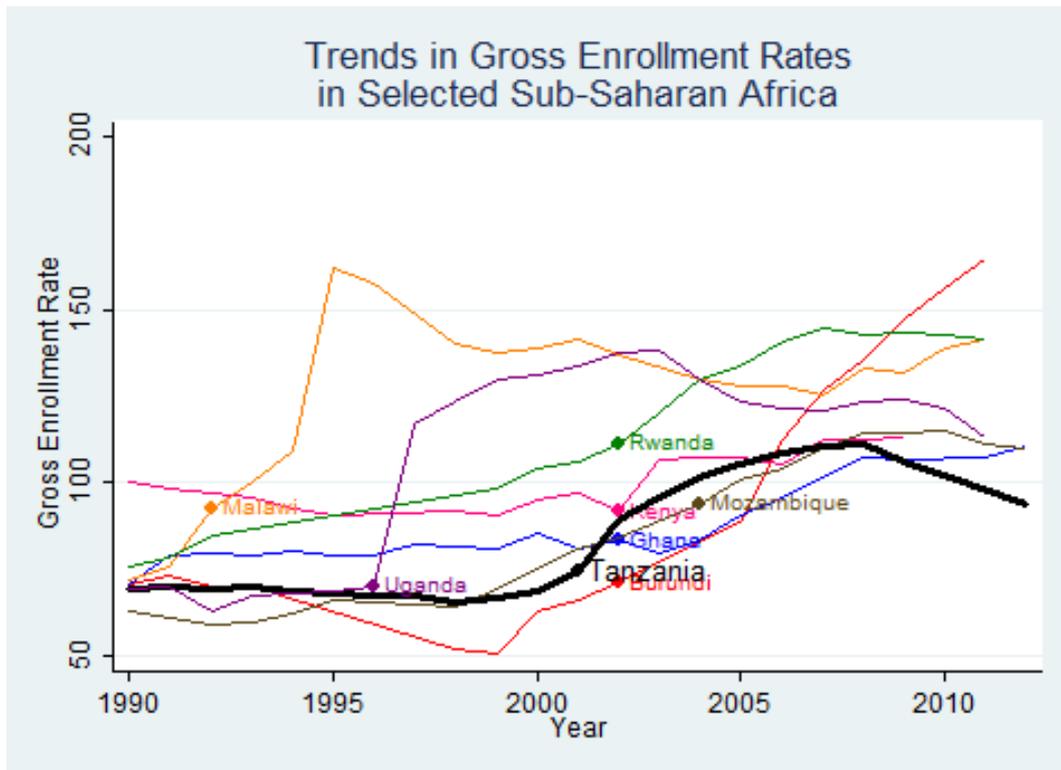


Figure A-1

Countries represented here removed primary school fees the year following that at which their name appears on the figure.

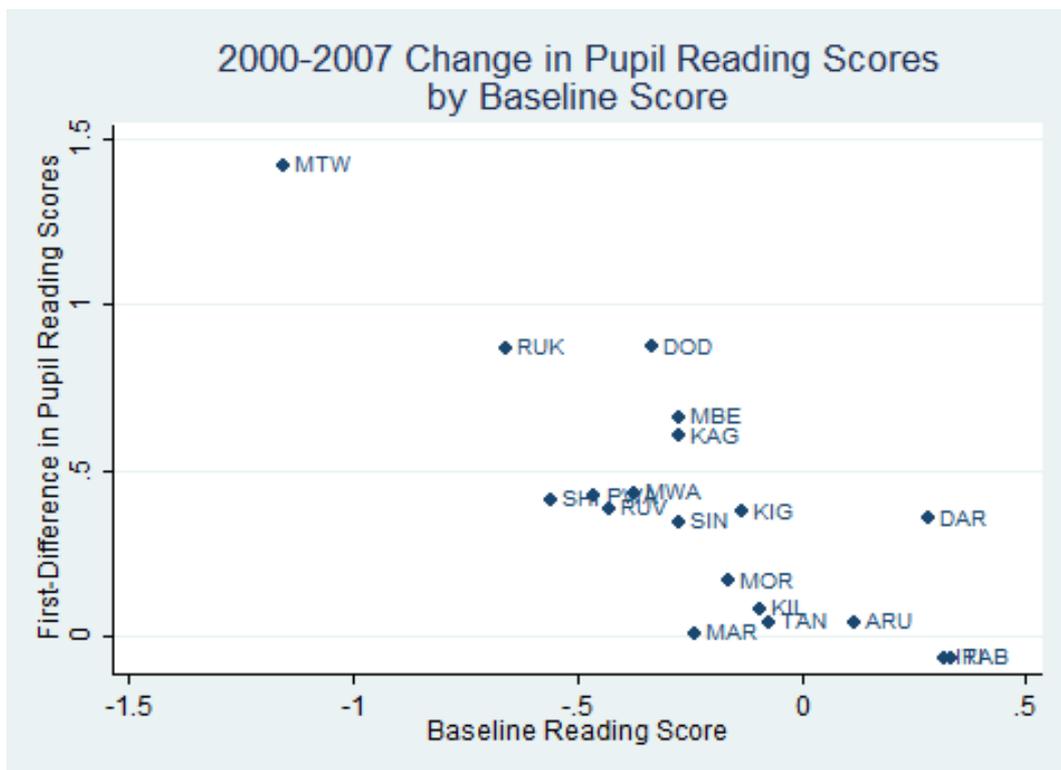


Figure A-2

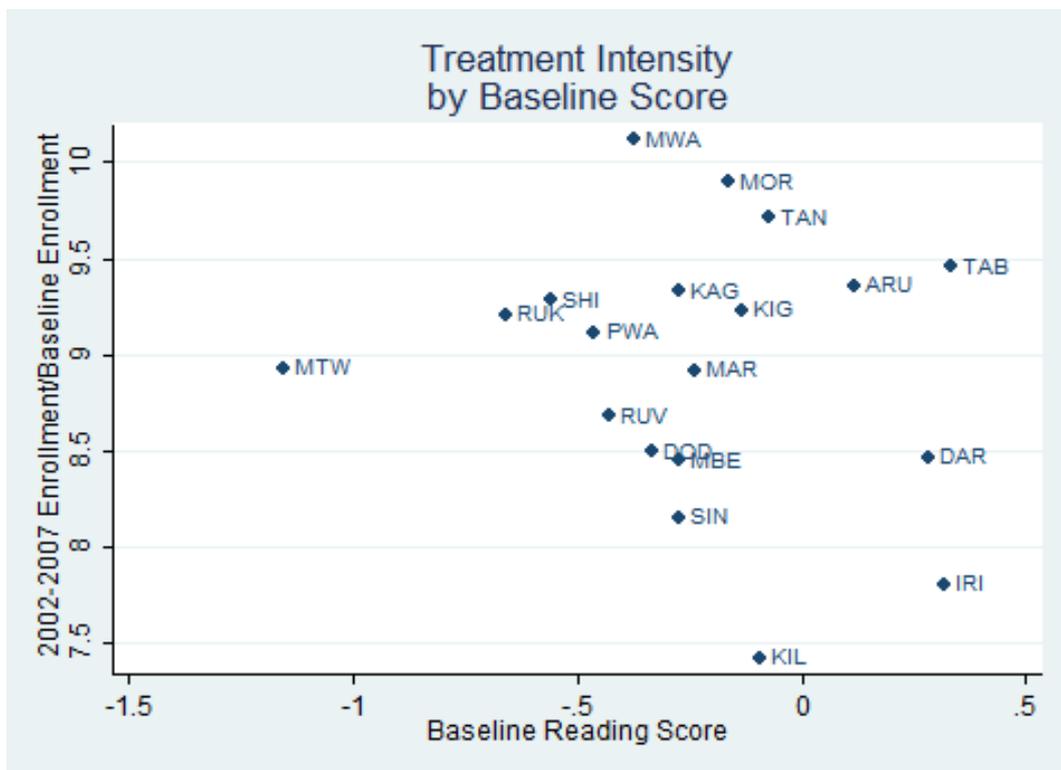


Figure A-3

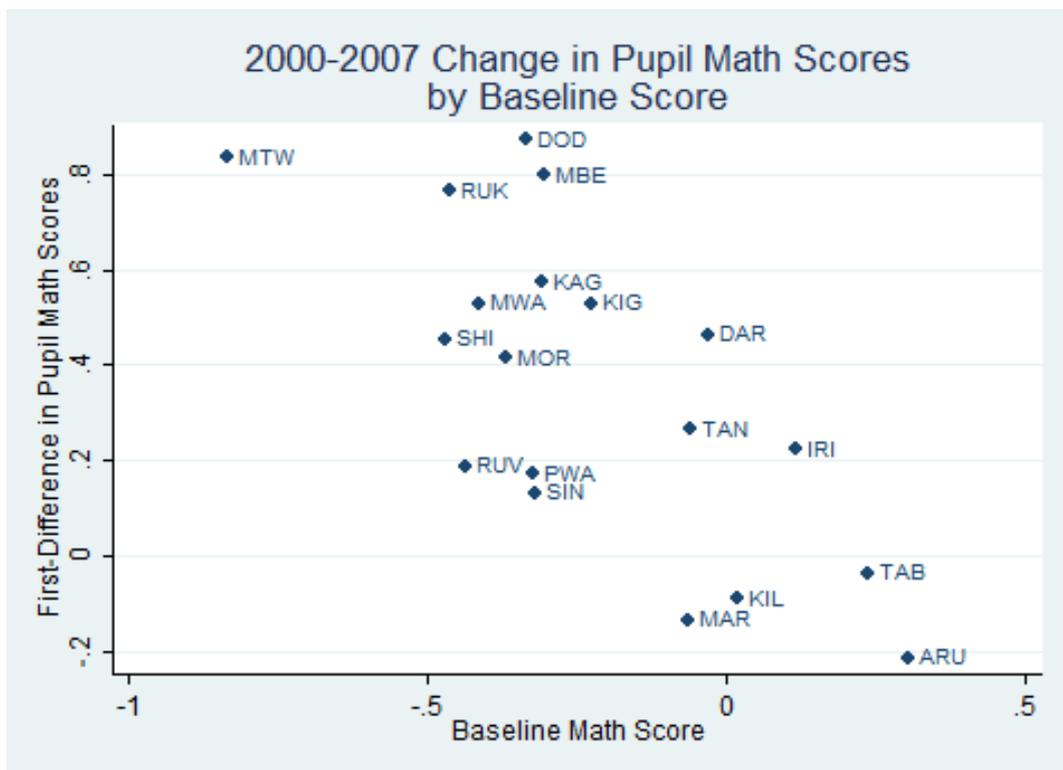


Figure A-4

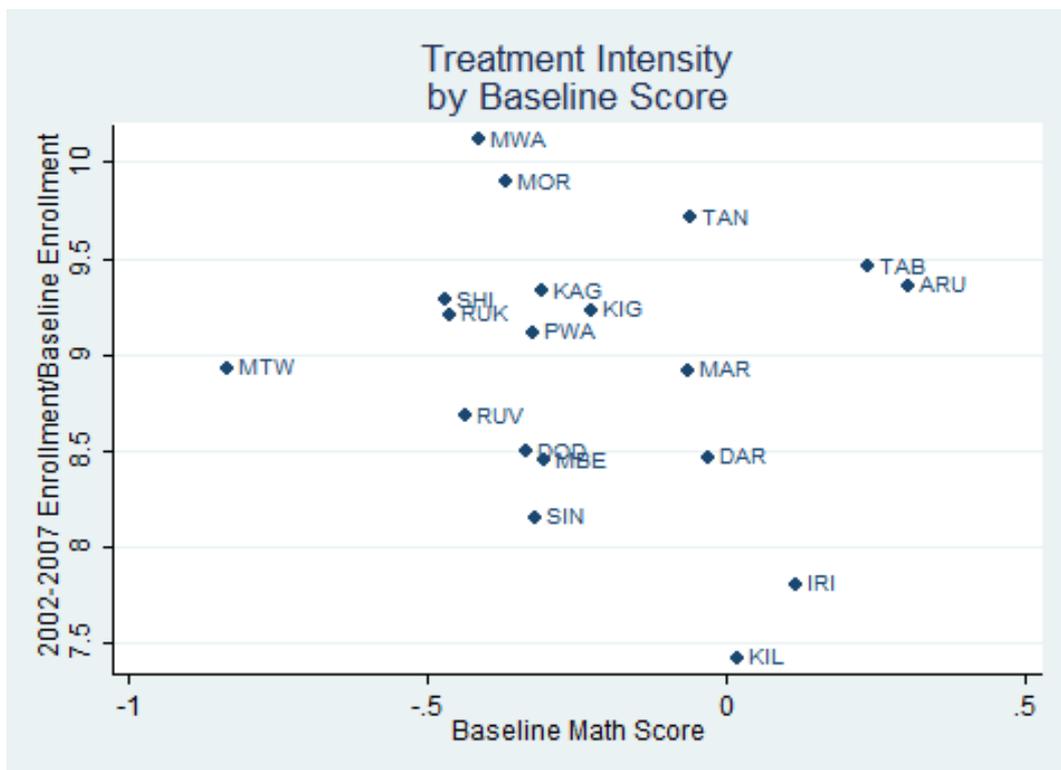


Figure A-5