ECONSTOR
*Make Your Publications Visible.*

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Valente, Christine

**Working Paper**

# Primary Education Expansion and Quality of Schooling: Evidence from Tanzania

IZA Discussion Papers, No. 9208

**Provided in Cooperation with:**
IZA – Institute of Labor Economics

*Suggested Citation:* Valente, Christine (2015) : Primary Education Expansion and Quality of Schooling: Evidence from Tanzania, IZA Discussion Papers, No. 9208, Institute for the Study of Labor (IZA), Bonn

This Version is available at:
https://hdl.handle.net/10419/114072

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

DISCUSSION PAPER SERIES

IZA DP No. 9208

# Primary Education Expansion and Quality of Schooling: Evidence from Tanzania

Christine Valente

July 2015

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

# Primary Education Expansion and Quality of Schooling: Evidence from Tanzania

**Christine Valente**
*University of Bristol*
*and IZA*

Discussion Paper No. 9208
July 2015

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 9208
July 2015

# ABSTRACT

# Primary Education Expansion and Quality of Schooling: Evidence from Tanzania*

The rapid increase in primary enrollment seen in many developing countries might worsen schooling quality. I estimate the effect of enrollment growth following the removal of primary school fees in Tanzania and find that it led to large increases in the pupil-teacher ratio and a reduction in observable teacher quality, but rule out a substantial effect on test scores overall. These results are robust to instrumenting enrollment growth using predetermined fertility and migration decisions, and not driven by compositional changes. In urban areas, however, where baseline achievement was higher, test scores deteriorated where enrollment growth was larger.

JEL Classification: I21, I28, O15

Keywords: universal primary education, pupil-teacher ratio, test scores, Tanzania

Corresponding author:

Christine Valente
Department of Economics
University of Bristol
8 Woodland Road
BS8 1TN, Bristol
United Kingdom
E-mail: 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 Simon Burgess, Damon Clark, Robert Garlick, Ellen Greaves, Sami Stouli, Frank Windmeijer, and Yanos Zylberberg, as well as seminar participants at the University of Bristol, Universite Catholique de Louvain, ISER (University of Essex), Institute for Fiscal Studies (EDePo), the University of Manchester, Universidade NOVA de Lisboa, the 2014 Royal Economic Society Conference and 2014 Econometric Society European Meeting. All errors are my own.

# 1 Introduction

After over a decade of stagnation, the net enrollment rate in Tanzanian primary schools went from 53% in 2000 to 73% in 2002, coinciding with the removal of primary school fees announced in 2001. Tanzania is no outlier - a sizeable number of Sub-Saharan African countries have recently implemented Free Primary Education (FPE) policies and many of these reforms were followed by large increases in enrollment.[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). It is therefore important to understand the consequences of accelerated enrollment growth on the quality of the learning environment.

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), in which the authors estimate the effect of the removal of primary school fees in Kenya on the test scores of children who had been in school for three to seven years prior to FPE. They find no effect on test scores at the end of Grade 8, which they hypothesize may be due to the expansion of private education. Their study is, however, silent on the effect of FPE for new entrants into the schooling system, who are exposed to dramatically larger cohorts in their own grades, from the onset of their schooling careers. The main contribution of the present paper is to provide a more stringent test of the effect of rapid enrollment growth on the quality of schooling by focusing on its effect on these new entrants. I find that, in Tanzania, larger increases in enrollment than those seen in Kenya were not accompanied by a substantial degradation of test scores for pupils who were affected by rapid enrollment growth from Grade 1, while the private sector accounted

[1] Seventeen countries have removed primary school fees since 1994 according to Lucas & Mbiti (2012). Figure A-1 illustrates the magnitude of the increases in enrollment that followed in a selected sample of countries.

for less than 1% of pupils throughout the period of analysis and could therefore only have played a negligible mitigating role. In addition, I shed light on the impact of primary schooling expansion 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 across Tanzania. But there is variation in the subsequent rate of enrollment growth across regions, which I first exploit using a difference-in-differences approach. In order to address concerns regarding the endogeneity of enrollment growth, I then note that regions whose post-reform primary-school age population was larger relative to the pre-reform school-age population experienced larger primary enrollment growth rates. Therefore, schools in these areas experienced larger increases in the demand for primary education, independently of the potentially endogenous response of the regional enrollment rate to the school fee reform. I use this source of arguably exogenous variation, which is based on past fertility and past migration decisions, in order to establish the causal effect of enrollment growth on schooling inputs and cognitive skills acquisition. The main threat to the exogeneity of this instrument is that regions experiencing slower decreases in fertility (and thus larger enrollment growth) may also have seen slower improvements over time in parental or governmental investments in human capital. I sign the potential direction of the bias in my IV estimates and show that, if anything, this bias would lead me to overestimate the worsening of quality due to rapid enrollment growth. My main conclusion is that there was no substantial decrease in test scores overall, which can therefore not be driven by this bias.

Another concern is that the composition of pupils may have changed differently over time across regions with differing rates of enrollment growth. There is no reason to think that the marginal student should be better than the average student, which would be required for changes in composition to account for my main finding. I show that my results are robust to controlling for a rich set of observable characteristics, and in additional results using a changes-in-changes approach, I exclude students in the bottom 25% of the distribution and still reach the same conclusion.

I also 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 en-

rollment growth by one standard deviation) and a worsening of average teacher experience and subject-specific knowledge. 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).

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

The rest of the paper is organised as follows. Section 2 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, Section 6 presents the main results, Section 7 explores the robustness of my findings, and Section 8 investigates the effect of enrollment growth across the distribution of test scores and by rural or urban location. Section 9 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. Throughout the 1990s, only about half of primary-school age children (i.e., age 7-13) attended school (with annual

net enrollment rates varying between 49 and 51%). This was despite early attempts at achieving universal primary education in the late 1970s, which culminated in a net enrollment rate of 70.3 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 seemingly 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 three to seven years before FPE. Using a changes-in-changes estimation approach exploiting the variation in FPE treatment intensity due to differential drop-out rates across Kenyan districts before the country-wide reform, they find that students who would have taken the exam in the absence of FPE lost no more than 0.05 of a standard deviation 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 considered a control group because their enrollment was not prioritized by the reform (see Section 2), Hoogeveen & Rossi (2013) find a statistically significant decrease in attainment in rural areas, which they hypothesize to be due to a deterioration of the quality of schooling.

## 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).[2]

One of the most researched aspects of the achievement production function is the effect of class size on test scores. In developed country settings, the range of estimates is generally between 0.07 and 0.27 of a standard deviation increase in test scores for a decrease of 7 pupils[3]. 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" treat-

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

[3]Seven pupils is a convenient point of reference as it corresponds both to the estimated increase in the pupil-teacher ratio for a one standard deviation increase in enrollment growth found in Section 6 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)).

ment group and the "normal class" control group was about 7 pupils (see Table 3 in Krueger (1999)). Perhaps more illustrative than these effect sizes, the effects found in Krueger (1999) translate into 64% (82%) of the white-black gap in kindergarten (third grade). In developing country settings, two randomized experiments nearly halving class size, one in Kenya and one in India, did not find any statistically significant effects on test scores (Banerjee et al. (2007); Duflo et al. (2011); Duflo et al. (2012)).

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, during the two or three 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 (exclusively non-experimental) evidence on the effect of teacher subject-specific knowledge on students' test scores, however, reports consistently positive effects (Glewwe et al. 2011).

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

# 4 Identification Strategy

## 4.1 Conceptual Framework

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

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

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

We can distinguish two mechanisms, one working through changes in the test scores of inframarginal students (i.e., students who would have been enrolled in school even if enrollment had stayed constant), and another one working through changes in the composition of students but leaving the achievement of inframarginal students unaffected. Only decreases in test scores resulting from the first of these mechanisms would denote a worsening of the quality of the schooling environment. Such worsening of quality could come about for several reasons. First, enrollment growth is likely to increase pupil-teacher ratios and decrease the quality of the average teacher in terms of teacher training and experience. A change in average teacher quality may arise for a number of reasons, including: the mechanical decrease in teacher experience due to the need to hire more teachers, decreased selection in the recruitment of teacher trainees and shorter teacher training to meet the increased demand, and possibly increased turnover. The effect on non-teacher school inputs is less clear *a priori* because the increase in enrollment was accompanied by a capitation grant targeted at non-salary expenditure. Finally, the marginal student is likely to have lower SES, and may thus have a worse cognitive and non-cognitive endowment than previous students, which may lead to negative peer effects on the performance

of the inframarginal students. Even when the performance of inframarginal students is not significantly affected by the increase in enrollment, we may observe a worsening of average test scores among enrolled students through a composition effect, if the marginal student has lower ability than the average inframarginal student.

## 4.2 Difference-in-Differences Approach

In order to answer the question of whether rapid enrollment growth worsened the quality of schooling, I first estimate the effect of primary enrollment growth on an observable set of schooling inputs $Z_{ij}$ observed in Grade 6, and then estimate its overall effect on achievement in Grade 6, as captured by test scores in Kiswahili and mathematics. The baseline identification strategy relies on a comparison of changes in schooling inputs or test scores between 2000 and 2007 across regions that experienced different rates of growth in primary enrollment. More precisely, I estimate the following equation using the 2000 and 2007 SACMEQ surveys described in Section 5, in which a measure of quality of inputs or outcomes $y_{irt}$ is regressed on a survey dummy ($\mathbb{1}(t = 2007)_t$), region dummies ($\beta_r$), individual and regional (time-varying) controls ($X_{irt}$), and the interaction between the 2007 survey dummy and the relative size of post-reform enrollment (cumulated over 2002-2007) relative to baseline enrollment:

$$\begin{aligned} y_{irt} =& \beta_0 + \beta_1 (\frac{post_enrol}{baseline_enrol})_r \times \mathbb{1}(t = 2007)_t \\ &+ \mathbb{1}(t = 2007)_t + \beta_r + X'_{irt}\beta_X + \epsilon_{irt} \end{aligned} \tag{2}$$

where $baseline_enrol_r$ is the number of pupils enrolled in primary schools in region $r$ in 2001 (as enrollment statistics broken down by region are not available for 2000) and $post_enrol_r = \sum_{j=2002}^{2007} E_{rj}$ is the sum of the number of pupils enrolled in primary schools in region $r$ during 2002-2007, the years during which the Grade 6 students of 2007 should have been in primary school and which coincides with the post-FPE period. I focus on the cumulative effect of exposure to larger school cohorts from Grade 1 to Grade 6 in order to reflect the cumulative nature of learning illustrated in Equation

1. From here onwards, I refer to $(\frac{post_enrol}{baseline_enrol})_r$ as "enrollment growth" in region $r$, which can be thought of as a continuous measure of treatment intensity in a difference-in-difference setting. 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 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 $\beta_1$ 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 $\beta_1$ being an underestimate of the worsening of quality (i.e., would bias $\beta_1$ towards less negative values). I address this concern in a robustness check in which I control for the growth in discretionary government transfers to the region. As previously mentioned, FPE-led enrollment growth may lead to the recruitment of less able students. This compositional effect would result in $\beta_1$ 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.

## 4.3 Instrumental Variable Approach

There may remain unobservable sources of endogeneity even after controlling for growth in discretionary government transfers and observable characteristics of students. For instance, if expected returns to education increased faster between 2000 and 2007 in some regions than others, then one might expect both increases in enrollment and in study effort, so that

enrollment growth would be endogenous when $y_{irt}$ is a pupil's test score. Or it could be the case that regions where local administrations became more committed to education, higher increases in both enrollment and education quality were achieved. Or one may worry about measurement error in enrollment figures, since there is an incentive to over-report enrollment rates in order to increase the number of capitation grants. In order to address these remaining issues, I use *potential* growth in enrollment based on predetermined fertility decisions and migration decisions up to 2002 as an instrumental variable for *actual* growth in enrollment. More specifically, I exploit the fact that actual enrollment depends not only on contemporaneous decisions of policy makers, parents and children, but also on the size of the primary-school age population, which is predetermined, and use the growth in the size of the primary-school age population as an instrument for the actual enrollment growth. The first stage of my two-stage least squares system is as follows:

$$(\frac{post_enrol}{baseline_enrol})_r \times \mathbb{1}(t=2007)_t = \gamma_0 + \gamma_1(\frac{post_age7_13}{baseline_age7_13})_r \times \mathbb{1}(t=2007)_t + \mathbb{1}(t=2007)_t + \gamma_r + X'_{irt}\gamma_X + \nu_{irt} \tag{3}$$

where $\gamma_r$ are region fixed effects, $baseline_age7_13_r$ is the number of children aged 7-13 in region $r$ in 2001 and $post_age7_13_r = \sum_{j=2002}^{2007} Age7_13_{rj}$ is the sum of the number of children aged 7-13 in region $r$ in each year from 2002 to 2007. The size of the relevant cohorts is calculated using a single population census carried out in 2002 and based on the individual's region of residence at the time of the census. Differences in $(\frac{post_age7_13}{baseline_age7_13})_r$ across regions can therefore be interpreted as differences in fertility trends (between 1988 and 2000) and migration patterns up to 2002.[4] In a robustness check, I instead construct the instrument based on the individual's region of residence in 2001 (hence based on migration decisions before FPE) using migration data, and show that this does not affect the results.

[4]The relevant fertility period is 1988-2000 because $post_age7_13$ includes children born between 1989 and 2000 and $baseline_age7_13$ corresponds to children born between 1988 and 1994.

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

$$\begin{aligned} y_{irt} =& \lambda_0 + \lambda_1 (\frac{post_age7_13}{baseline_age7_13})_r \times \mathbb{1}(t = 2007)_t \\ &+ \mathbb{1}(t = 2007)_t + \lambda_r + X'_{irt}\lambda_X + \mu_{irt} \end{aligned} \tag{4}$$

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

## 4.4 Signing the Direction of Any Remaining Bias

The main concern regarding the exclusion restriction required for the instrument to be valid is that regions having experienced faster fertility declines in the pre-reform period (1988-2000) may also have experienced faster increases in investments in the human capital of children by parents or policy makers. If this were the case, then this would lead quality measures that can be influenced by such investments in human capital (e.g., children test scores) to increase more in regions experiencing slower growth in potential enrollment. Similarly, if some older children migrate to regions with more positive school quality trends, then regions with lower values of $(\frac{post_age7_13}{baseline_age7_13})_r$ may experience more positive changes in learning outcomes over time, since larger older children cohorts increase the denominator of this ratio relative to the numerator. All these potential issues would tend to lead to an overestimation of the worsening of quality coinciding with higher enrollment growth.

To see this, consider the following system:

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

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

and the corresponding reduced-form:

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

Consider the case in which $\beta \leq 0$ (higher enrollment growth may worsen test scores), $\nu > 0$ (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 or worse school quality trends). 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 improvements in educational quality or child human capital. Or there would have to be pre-reform migration patterns such that younger children are more likely than older children to be observed in areas with more favorable trends in schooling quality. It is hard to think of such omitted variable other than government transfers targeting less developed areas, which I control for in a robustness check.

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

$$health_{irt} = \xi_0 + \xi_1 (\frac{post_age7_13}{baseline_age7_13})_r \times \mathbb{1}(t \geq 1995)_t + \mathbb{1}(t \geq 1995)_t + \xi_r + \phi_{irt} \tag{8}$$

where $health_{irt}$ refers to child $i$ in region $r$ born in year $t$ and is, in turn, an indicator for whether the child has received a full course of immunization, a dummy for whether delivery was assisted by a health professional, a dummy for whether the mother received no help at all during delivery, an infant mortality indicator equal to one if the child died within 12 months of birth, and zero otherwise, and a stunting indicator which is equal to one if the child's height-for-age z-score is below 2 standard deviations of the reference median, and equal to zero otherwise. $\xi_r$ are region fixed

effects. The coefficient of interest is $\xi_1$, as a non-zero coefficient indicates a differential trend in the outcome variable in areas with slower fertility decline (using 1995, the mid-point of the relevant period, as threshold).

Results are reported in Table A-1. All the point estimates go in the direction of smaller improvements in child health inputs and outcomes in regions with higher potential enrollment growth, statistically significantly so in the case of full immunization and delivery by a health professional. This confirms that, if anything, my IV estimates of the effect of enrollment growth on test scores are likely to over- rather than understate any worsening in achievement.

# 5 Data and Summary Statistics

## 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 collected in 2007. SACMEQ II surveyed 2,854 pupils Grade 6 in 181 schools, and SACMEQ III surveyed 4,194 Grade 6 pupils in 196 schools and stratified sampling 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. Six pupils in the 2007 survey are dropped due to missing math scores, I also drop observations from two schools (and their 26 pupils) with a pupil-teacher ratio above 250, as well as 18 pupils with no information about father education. Finally, I drop 65 pupils from the Lindi region in the 2007 survey because this small region was not surveyed in 2000, resulting in a sample of 6,933 pupils.[5]

## 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

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

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 discretionary education block grants (i.e., excluding capitation grants) between 2002-2007 is on average 11.4 times larger than that received in 2001 (in real terms).

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

Turning now to measures of the "quality" of the teachers teaching the average pupil, there is evidence of an improvement in the level of academic qualifications of teachers over time, with a particularly steep increase in the proportion of Kiswahili teachers with O-level qualifications (which is the exam taken after four years of secondary education). There has however been a decline in the proportion of teachers with at least two years of teacher training, which is expected since initial training in teachers colleges went from two- to one-year in order to speed up the supply of qualified teachers. As expected, with the increased demand for new teachers, the average experience of teachers also decreased between 2000 and 2007. Consistent with the overall improvement in the education of teachers, the performance of teachers in subject-specific tests 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.[6]

All in all, the summary statistics reported above seem to suggest that the rapid expansion of the primary schooling system did not lead to a drop in test scores. Results reported in the next section show that what is suggested by these raw data is confirmed in the regression analysis.[7]

# 6 Main Results

## 6.1 Effect of Enrollment Growth on Schooling Inputs

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

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

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

deviation increases the pupil-teacher ratio by 6.9, and this effect is statistically significant at the 1% level.[8]. The wild cluster bootstrap-t p-value is 0.008, thus confirming the conclusions based on analytical standard errors. Panel B reports 2SLS results obtained when instrumenting enrollment growth with potential enrollment growth. The point estimate is somewhat larger but qualitatively similar, and significant at the 5% significance level. The Kleibergen-Paap F-statistic - which allows for intra-regional correlation of standard errors - for the first stage is 17.296, which suggests that the instrument is not weak.[9] 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 since the net enrollment rate was increasing during the period under scrutiny (i.e., $\gamma_1 = 1.66 > 1$ in the first-stage Equation 3).

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

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

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

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 growth). 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 gained less experienced teachers (by 1.4 years for Kiswahili teachers and 2.8 years for Math teachers for 1 s.d. increase in actual enrollment growth), on average, which is consistent with the expectation that local governments in these regions had to recruit a larger number of new teachers. Similarly, there is a statistically significant worsening of the reading scores of language teachers in regions experiencing larger enrollment growth (a worsening by 0.14 of a standard deviation for one standard deviation larger enrollment growth). These findings are robust to the different estimation methods employed.

Table 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 acquisi-

tion 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.[10]. 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

[10]The 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.

arguments entering the test scores production function on pupil test scores.

## 6.2 Effect of Enrollment Growth on 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 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..

# 7 Robustness Checks

Having found no evidence of a sizeable effect of enrollment growth on pupils' test scores, which is robust to instrumenting enrollment growth with potential enrollment growth based on past fertility and migration decisions, I now turn to testing the robustness of these findings to controlling for growth in primary education funding across regions. The main concern here is that less-developed regions may have experience faster (potential and actual) 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 reversion to the mean. More specifically, I include an interaction term between a post-reform survey dummy and the baseline regional average pupil score (in reading in Table 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 statisti-

cally insignificant, and the lower bounds of the 95% CI for reading (math) scores correspond to a -0.15 s.d. effect on both reading and math test scores for an increase in enrollment growth by 1 standard deviation. Finally, Column (6) reports estimates of the reduced-form equation. Potential enrollment growth has no statistically significant effect on either Kiswahili or mathematics test scores, with the lower bound of the 95% CI indicating a maximum worsening of test scores by 0.08 (Kiswahili) and 0.16 (math) of a s.d. for a 1 s.d. increase in potential enrollment growth.

In Table 8, I test the robustness of my instrumental variable. First, I add, in the reduced-form equation (Eq. 4), the 7-year "lag" of the instrumental variable, namely $\sum_{j=1995}^{2000} Age7_13_{rj}$ divided by the number of children aged 7-13 in region $r$ in 1994. The idea of this test is to check whether differential changes in test scores between 2000 and 2007 associated with demographic trends that affected potential enrollment growth before FPE could be confounding my estimates. Second, I show the robustness of my findings to defining the instrument based on region of residence as of 2001, using information on region of residence one year before the 2002 census for those respondents who said they had migrated in the past year.[11]

In the first two columns, I present the reduced-form regressions for Kiswahili z-scores (Column (1)) and mathematics z-scores (Column (2)) including a control for primary-age cohort growth between 1994 and 2000. The correlation coefficient between potential enrollment growth between 1994 and 2000 and between 2001 and 2007 is weak (-0.17), so that including $\frac{\sum_{j=1995}^{2000} Age7_13_{rj}}{Age7_13_{r1994}}$ as a regressor brings little change to the point estimates (comparing with the last column of Tables 6 and 7: the reduced-form estimate becomes 0.068 (-0.006) for reading (math) compared to 0.083 (0.012)). In the next two columns, I present 2SLS estimates obtained when the instrument is constructed by assigning individuals to the region where they lived in August 2001, hence before primary fees were removed, instead of

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

the region of residence at the time of the census in August 2002. The point estimates are almost identical to those obtained with my instrument based on region of residence in 2002 (0.046 (0.008) instead of 0.05 (0.007) for reading (math) scores), which suggests that migration patterns correlated with school quality trends are not driving my findings.

# 8 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. On the other hand, less able students may be more reliant on schooling inputs in order to learn, and may thus suffer more. In order to explore potential heterogeneous effects across the distribution of test scores, I 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:

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

where $F_{Y,gt}(y)$ is outcome $Y$'s cumulative distribution function in group $g$ in period $t$, and region 1 only is treated in period 1. I define the treated group as those regions with above-median potential enrollment growth.[12]. 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 lower ability to enroll. Also in order to maintain comparability between characteristics of

[12]I 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.

pupils within group over time, I control for the set of covariates included in Column (2) of Tables 6 and 7.[13] Results are displayed in Figures 1 and 2. Looking first at Figure 1, we can see that there is no evidence of a worsening of reading test scores. Even at the very top of the distribution, the lower bound of the 95% CI translates into a 0.09 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 -0.18 (Figure 2).

I also investigate heterogeneity in treatment effects by splitting 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). 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 9 reports the results. Starting with the rural sample, results are similar to the average effects reported in Section 6, except for the finding that subject-specific knowledge *increases* among mathematics teachers. The baseline OLS specification picks up a statistically significant, positive, correlation between pupils math test scores and enrollment growth, but the point estimates become statistically insignificant when instrumenting for enrollment growth or when adding controls for government transfers or mean reversion (Panel C). On the contrary, in the urban sample, there is a large worsening of subject-specific knowledge among mathematics teachers, as well as a substantial, robust, worsening of reading and mathematics scores with enrollment growth.

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

[13]Covariates are adjusted for using the parametric approach proposed by Athey & Imbens (2006).

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 9 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.[14] On the other hand, these controls are based on characteristics of the pupils and their schools during 6th Grade, not throughout their primary schooling experience, so that it is not possible to completely rule out a worsening in teachers' subject-specific knowledge or increases in the pupil-teacher ratio as relevant pathways to the worsening in pupils' math test scores in urban areas.

For the urban sample, IV and OLS estimates of the effect of enrollment growth on achievement are broadly similar. As argued in Sections 4.4, 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 $\beta$ for a certain prior distribution for $\gamma$ in Equation 5. The reduced-form estimate of the effect of potential enrollment growth on Z-scores gives a natural lower bound (maximum negative

[14]Full results are available on request.

magnitude) of $\gamma$, and therefore I obtained Local-to-Zero estimates for a range of possible distributions of $\gamma$ of the form $\gamma \sim \mathcal{U}(\delta, 0)$, with values of $\delta$ ranging from the reduced-form point estimate (-0.538 for math scores) to -0.027 ($\frac{1}{20th}$ of -0.538) and clustering the standard errors at the region level, as in the main analysis. As shown in Figure 3, the point estimate for math scores is equal to -0.2 or less except for very large possible values of $\gamma$, and it is significantly negative at 10% for maximum prior values of $\gamma$ above -0.17. In other words, we can be confident that the effect of enrollment growth on math test scores is negative in urban areas provided the direct effect of potential growth on achievement after conditioning on actual enrollment growth does not exceed about 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 4 shows that the point estimate for $\beta$ is between -0.10 and -0.20 for the range of possible $\delta$ values, but that the clustered standard errors associated with these estimates are too large for these point estimates to achieve statistical significance.

# 9 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.[15]

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 ex-

[15]In Kenya, Lucas & Mbiti (2012) estimate the impact of FPE on test scores of students who had been in school for three to seven years before FPE.

perience and subject-specific knowledge in the country taken as a whole. 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-skim" 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 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 and migration decisions, and I show that plausible sources of instrument endogeneity are unlikely to be driving these findings.

However, when investigating the possibility of heterogeneous effects for urban and rural areas, I find evidence of a deterioration of test scores in urban areas in mathematics (0.27 s.d. for one s.d. increase in enrollment growth), and, to a lesser extent, in reading (0.18 s.d. for one s.d. increase in enrollment growth). One plausible explanation for this differential effect on achievement in urban relative to rural areas is the much higher baseline achievement in urban areas, and hence the larger potential for a worsening of the learning environment due to the pressures of rapid enrollment growth. 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 with rapid enrollment growth, 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).

How do these results compare to those reported by Lucas & Mbiti (2012), who estimate the impact of FPE in Kenya on test scores of students who had been in school for three to seven years before FPE? Tanzania and Kenya had similar baseline reading levels, although Kenya enjoyed higher levels of mathematics proficiency in 2000 according to the SACMEQ evaluation exercise (Makuwa 2010). Compared to the case of Kenya, however, enrollment growth following FPE was more pronounced in Tanzania, with a gross enrollment rate going, in the five years following FPE, from 91% to 112% in Kenya, and from 74% to 111% in Tanzania (Figure A-10). My findings therefore show that even larger increases in enrollment than previously known can be achieved without substantial deterioration of the learning environment for most pupils, even when considering children exposed to larger cohorts from the start of their primary schooling experience rather than halfway through. However, in Tanzania, wider access appears to have come at the cost of quality in urban areas, so that whether or not enrollment growth was welfare-enhancing depends on how the gains of the many winners are weighted against the losses of the (fewer) losers.

# 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?*, Unpublished report.

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', *NBER Working Paper No. 17939* .

Glewwe, P., Kremer, M. & Moulin, S. (2009), 'Many children left behind? textbooks and test scores in kenya', *American Economic Journal: Applied Economics* **1**(1), 112–135.

Glewwe, P., Kremer, M., Moulin, S. & Zitzewitz, E. (2004), 'Retrospective vs. prospective analyses of school inputs: the case of flip charts in kenya', *Journal of Development Economics* **74**(1), 251–268.

Glewwe, P. W., Hanushek, E. A., Humpage, S. D. & Ravina, R. (2011), 'School resources and educational outcomes in developing countries: a review of the literature from 1990 to 2010', *NBER Working Paper No.17554* .

Grogan, L. (2009), 'Universal primary education and school entry in uganda', *Journal of African Economies* **18**(2), 183–211.

Hanushek, E. A. & Woessmann, L. (2008), 'The role of cognitive skills in economic development', *Journal of economic Literature* pp. 607–668.

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

Hoxby, C. M. (2000), 'The effects of class size on student achievement: New evidence from population variation', *Quarterly Journal of economics* pp. 1239–1285.

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.

Makuwa, D. (2010), 'What are the levels and trends in reading and mathematics achievement?', *SACMEQ Policy Issues Series* (2).

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.

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

| | 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 | |
| *N* | 2849 | | 4084 | |

Source: Author's calculations using SACMEQ II, SACMEQ III, IPUMS (2011), Ministry of Education "Basic Education Statistics in Tanzania" and Budget Plans for various years. Class equipment score items: writing board, chalk, wall chart, cupboard, bookshelves, library, teacher table, teacher chair. Pupil equipment score items: exercise book, notebook, pencil, sharpener, eraser, ruler, pen, folder. Household items ownership items: newspaper, magazine, radio, TV set, VCR, cassette player, telephone, refrigerator/freezer, car, motorcycle, bicycle, piped water, electricity, table to write on.

Table 2: Effect on Quantity of Teachers

| | (1) Pupil/Teacher | (2) =1 if Multiple Shifts | (3) Teaching Hours (K) | (4) Teaching Hours (M) |
|---|---|---|---|---|
| **Panel A: OLS Estimates** | | | | |
| (=1 if 2007)×Enrol. 2002-2007 | 9.989*** | -0.031 | 1.354 | -0.073 |
| /Enrol. in 2001 | [3.172,16.806] | [-0.121,0.058] | [-2.210,4.918] | [-2.216,2.070] |
| Wild cluster bootstrap-t 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 |

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

Table 3: Effect on Quality of Teachers - Initial Training

| | (1) =1 if O-level (K) | (2) =1 if O-level (M) | (3) =1 if Training≥2 y (K) | (4) =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 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * p<0.10, ** p<0.05, *** p<0.01. Further details as per notes under Table 2.

Table 4: Effect on Quality of Teachers - Experience and Subject-Specific Knowledge

| | (1) Years Experience (K) | (2) Years Experience (M) | (3) Teacher Reading Score | (4) Teacher Math Score |
|---|---|---|---|---|
| **Panel A: OLS Estimates** | | | | |
| (=1 if 2007)×Enrol. 2002-2007 | -2.042** | -4.018*** | -0.207** | 0.048 |
| /Enrol. in 2001 | [-3.796,-0.288] | [-5.508,-2.528] | [-0.367,-0.047] | [-0.177,0.274] |
| Wild cluster bootstrap-t 95% CI | [-4.134,-.140] | [-6.324,-1.789] | [-.398,-.018] | [-.237,.325] |
| **Panel B: 2SLS Estimates** | | | | |
| (=1 if 2007)×Enrol. 2002-2007 | -2.796** | -4.419*** | -0.226** | -0.029 |
| /Enrol. in 2001 | [-5.161,-0.431] | [-6.654,-2.183] | [-0.440,-0.013] | [-0.410,0.352] |
| **Panel C: Reduced-Form Estimates** | | | | |
| (=1 if 2007)×Age 7-13 in 2002-2007 | -4.634** | -7.324*** | -0.375* | -0.048 |
| /Age 7-13 in 2001 | [-8.133,-1.134] | [-10.887,-3.760] | [-0.761,0.011] | [-0.732,0.637] |
| Mean Y | 13.537 | 12.049 | 0.008 | -0.001 |
| Kleibergen-Paap F-Stat | 17.296 | 17.296 | 17.296 | 17.296 |
| KP underid stat H0=underid | 4.658 | 4.658 | 4.658 | 4.658 |
| KP underid p-val | 0.031 | 0.031 | 0.031 | 0.031 |
| p-val robust endogeneity test | 0.297 | 0.597 | 0.796 | 0.475 |

Sample size: 6933. Region-correlated robust 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * p<0.10, ** p<0.05, *** p<0.01. Further details as per notes under Table 2.

Table 5: Effect on Physical Inputs

| | (1) =1 if Own Bk (K) | (2) =1 if Own Bk (M) | (3) Class Equip. (K) | (4) Class Equip. (M) | (5) # Ex. Books | (6) Pupil Equip. |
|---|---|---|---|---|---|---|
| **Panel A: OLS Estimates** | | | | | | |
| (=1 if 2007)×Enrol. 2002-2007 | 0.001 | -0.006 | 0.155 | -0.051 | -0.095 | 0.014 |
| /Enrol. in 2001 | [-0.021,0.022] | [-0.022,0.010] | [-0.444,0.755] | [-0.405,0.303] | [-0.619,0.428] | [-0.214,0.242] |
| Wild cluster bootstrap-t 95% CI | [-.026,.026] | [-.025,.019] | [-.848,1.215] | [-.454,.347] | [-.700,.484] | [-.199,.246] |
| **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 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * p<0.10, ** p<0.05, *** p<0.01. Class equipment and pupil equipment defined under Table 1. 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 |
| R-squared | 0.1400 | 0.2133 | 0.0965 | 0.1408 | 0.1484 | 0.1401 |
| Mean Y | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |
| Kleibergen-Paap F-Stat | | | 17.296 | | | |
| KP underid stat H0=underid | | | 4.658 | | | |
| KP underid p-val | | | 0.031 | | | |
| p-val robust endogeneity test | | | 0.437 | | | |

Sample size: 6933. Region-correlated robust 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * p<0.10, ** p<0.05, *** p<0.01. Pupil characteristics and physical inputs variables as listed in Table 1. Other 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 |
| R-squared | 0.1026 | 0.1899 | 0.0700 | 0.1054 | 0.1110 | 0.1024 |
| Mean Y | -0.004 | -0.004 | -0.004 | -0.004 | -0.004 | -0.004 |
| Kleibergen-Paap F-Stat | | | 17.296 | | | |
| KP underid stat H0=underid | | | 4.658 | | | |
| KP underid p-val | | | 0.031 | | | |
| p-val robust endogeneity test | | | 0.648 | | | |

Sample size: 6933. Region-correlated robust 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * p<0.10, ** p<0.05, *** p<0.01. Pupil characteristics and physical inputs variables as listed in Table 1. Other details as per notes under Table 2.

Table 8: Instrument Robustness Checks

| | Reduced-Form, Control for $\frac{\sum_{j=1995}^{2000} Age7_13_{rj}}{Age7_13_{r1994}}$ | | IV Based on Region in 2001 | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Reading Score | Math Score | Reading Score | Math Score |
| (=1 if 2007)×Age 7-13 in 2002-2007 | 0.068 | -0.006 | | |
| /Age 7-13 in 2000 | [-0.267,0.403] | [-0.508,0.496] | | |
| (=1 if 2007)×Enrol. 2002-2007 | | | 0.046 | 0.008 |
| /Enrol. in 2001 | | | [-0.148,0.240] | [-0.302,0.318] |
| R-squared | 0.1412 | 0.1040 | 0.0965 | 0.0700 |
| Mean Y | 0.002 | -0.004 | 0.002 | -0.004 |

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

Table 9: Heterogeneity between Rural and Urban Areas

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

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

Sample size: 5054 (rural sample) and 1879 (urban sample). Region-correlated robust 95% confidence intervals in brackets. Source: Author's calculations using SACMEQ II and III and Tanzania Census Extract (2002). * p<0.10, ** p<0.05, *** p<0.01. "Controls" refers to the inclusion of the same pupil characteristics and physical inputs variables as in the second column of Tables 6 and 7, while "Full set" of controls refers to the inclusion of these pupil characteristics and physical inputs variables as well as growth in government transfers and ((=1 if 2007)×Baseline Average Score) to control for mean reversion.

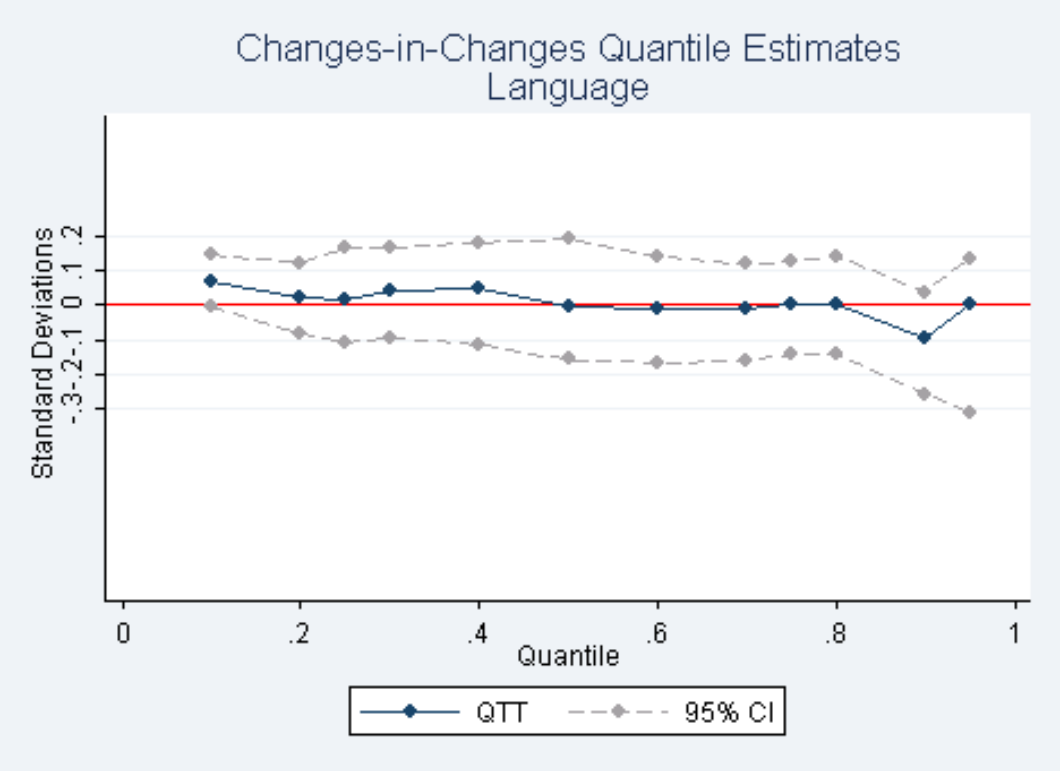


Figure 1

Note: the treated group is defined as those pupils in regions with above-median potential enrollment growth. Sample excludes bottom 25% of the distribution. Results obtained when controlling for the same set of covariates as in Column (2) of Tables 6. Bootstrapped standard errors clustered at the region level.

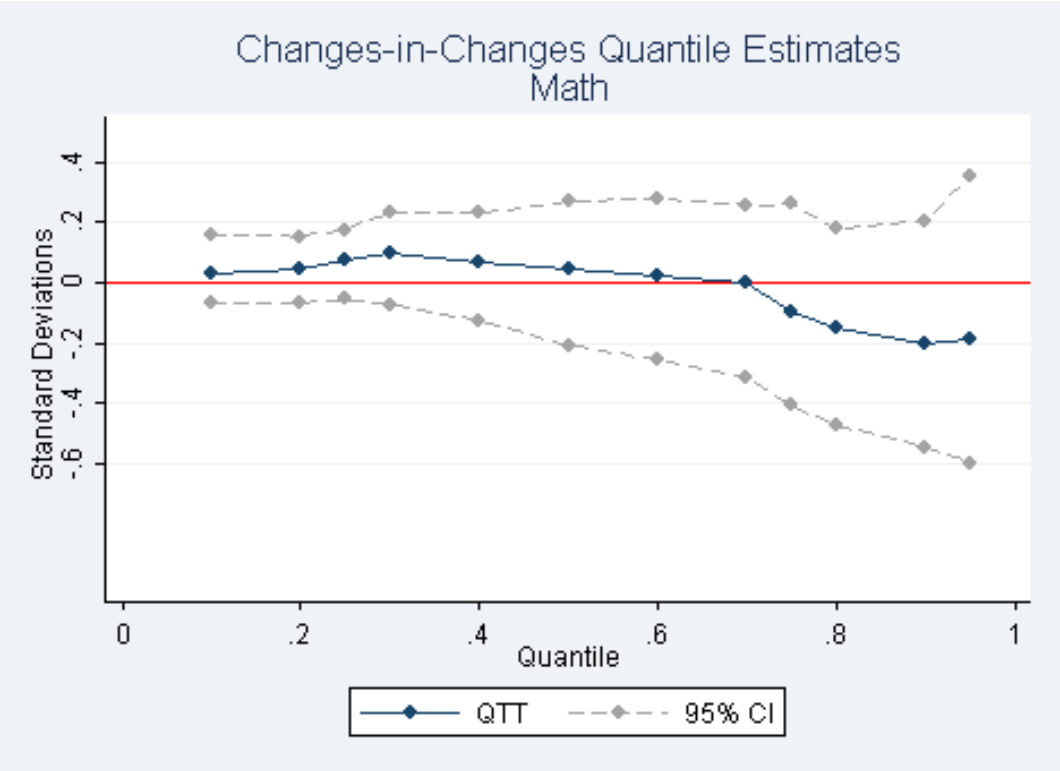


Figure 2
Note: the treated group is defined as those pupils in regions with above-median potential enrollment growth. Sample excludes bottom 25% of the distribution. Results obtained when controlling for the same set of covariates as in Column (2) of Tables 7. Bootstrapped standard errors clustered at the region level.

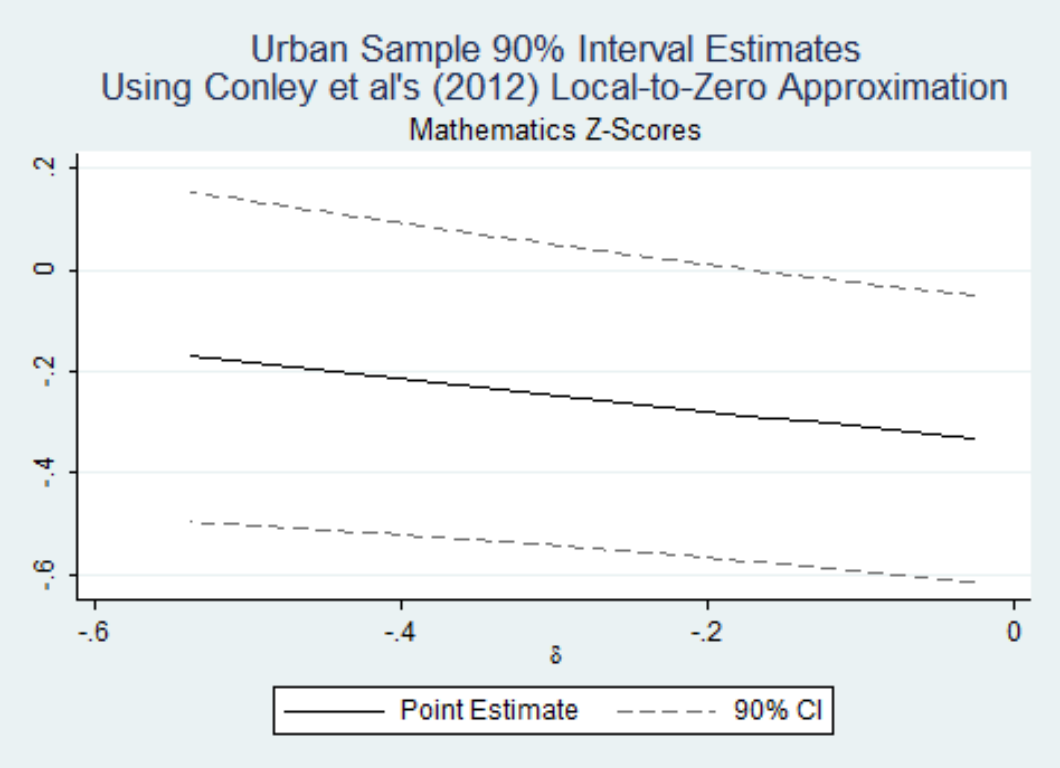


Figure 3
Note: Effect of enrollment growth allowing for departures from the perfectly exogenous instrument case ($\gamma = 0$) of the form $\gamma \sim \mathcal{U}(\delta, 0)$.

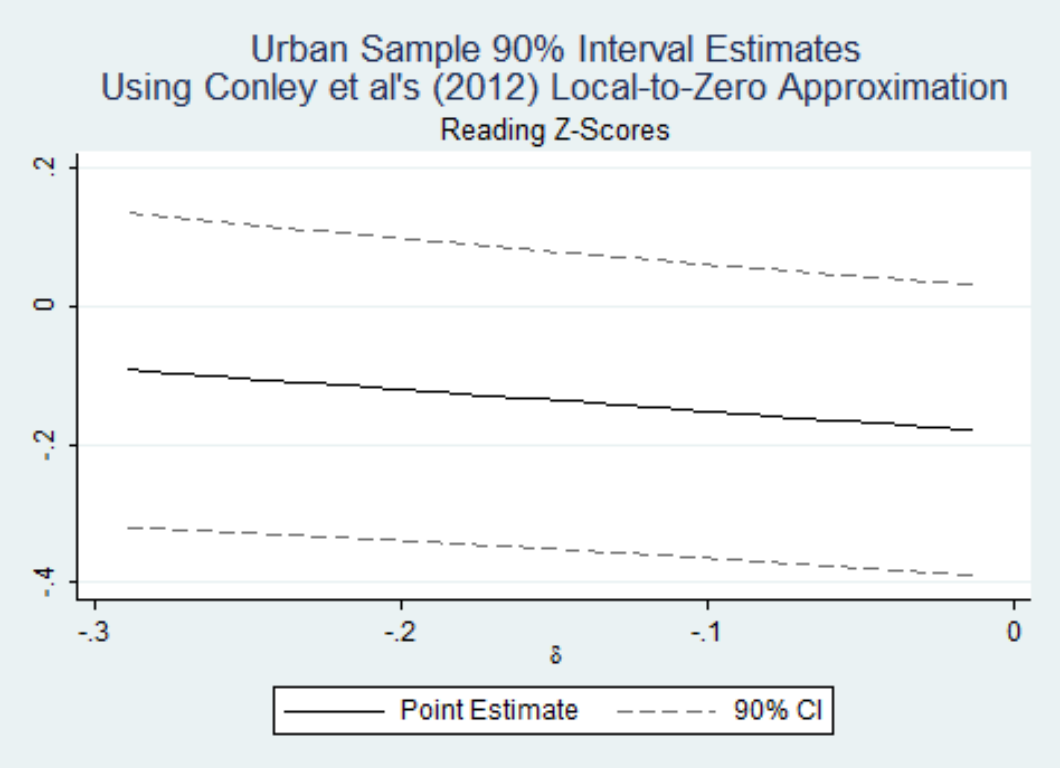


Figure 4
Note: Effect of enrollment growth allowing for departures from the perfectly exogenous instrument case ($\gamma = 0$) of the form $\gamma \sim \mathcal{U}(\delta, 0)$.

## Appendix - For Online Publication Only

### Graphical Analysis

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

One may expect regions with lower scores in 2000 to experience larger improvements in test scores between 2000 and 2007 (e.g., due to there being more low-hanging fruits to be picked). This is indeed the case, as illustrated by Figures A-5 and A-7. One concern could be that less developed regions started off with lower average test scores and also experienced faster enrollment growth under FPE, and/or experienced slower fertility declines in the past. In this case, the "mean reversion" observed in Figures A-5 and A-7 could bias my (OLS and/or IV) estimates and lead to an underestimation of the worsening of test scores due to enrollment growth. However, Figures A-6 and A-8 show no systematic relationship between baseline test scores and enrollment growth. In order to confirm that mean reversion is not driving my results, in Section 7 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.

**Appendix Tables**

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 | -0.159** | -0.257*** | 0.051 | 0.021 | 0.068 |
| /Age 7-13 in 2001 | [-0.292,-0.027] | [-0.428,-0.085] | [-0.026,0.129] | [-0.031,0.074] | [-0.019,0.155] |
| =1 if born 1995-2000 | 1.047** | 1.685*** | -0.362 | -0.141 | -0.549* |
| | [0.166,1.928] | [0.523,2.847] | [-0.890,0.166] | [-0.498,0.216] | [-1.143,0.045] |
| Region Dummies | Yes | Yes | Yes | Yes | Yes |
| Observations | 10944 | 16294 | 16294 | 12556 | 12730 |
| R-squared | 0.03 | 0.06 | 0.06 | 0.00 | 0.04 |
| Mean Y | 0.72 | 0.49 | 0.06 | 0.09 | 0.43 |

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

## Appendix Figures

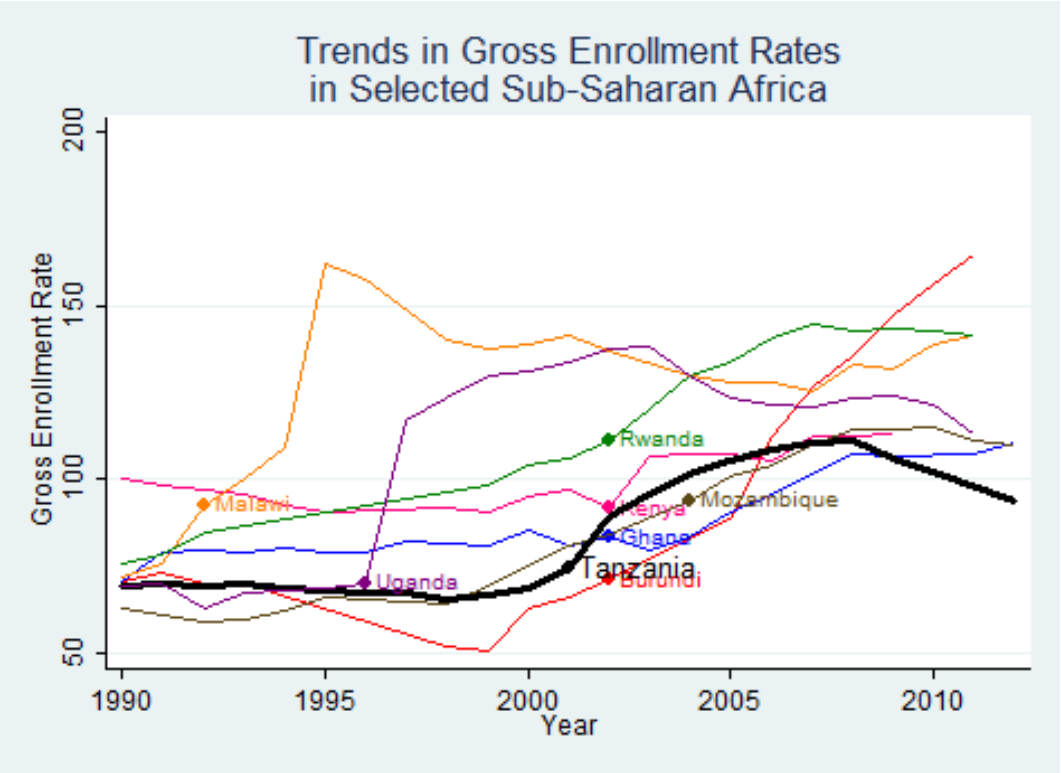


Figure A-1
Note: Primary school fees removed in the year following that at which the country's name appears on the figure.

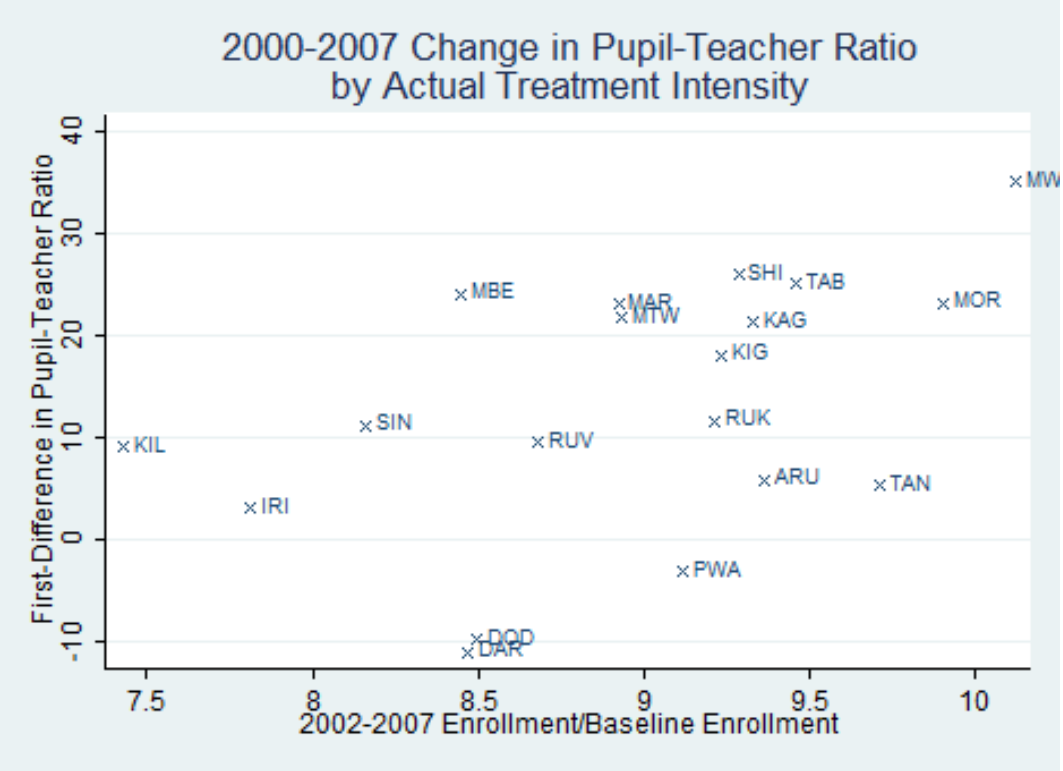
2000-2007 Change in Pupil-Teacher Ratio
by Actual Treatment Intensity
First-Difference in Pupil-Teacher Ratio
-10
0
10
20
30
40
7.5
8
8.5
9
9.5
10
2002-2007 Enrollment/Baseline Enrollment
MW
SHI
TAB
MOR
MBE
MAR
MTW
KAG
KIG
RUK
SIN
KIL
RUV
ARU
TAN
IRI
PWA
DOD
DAR

Figure A-2

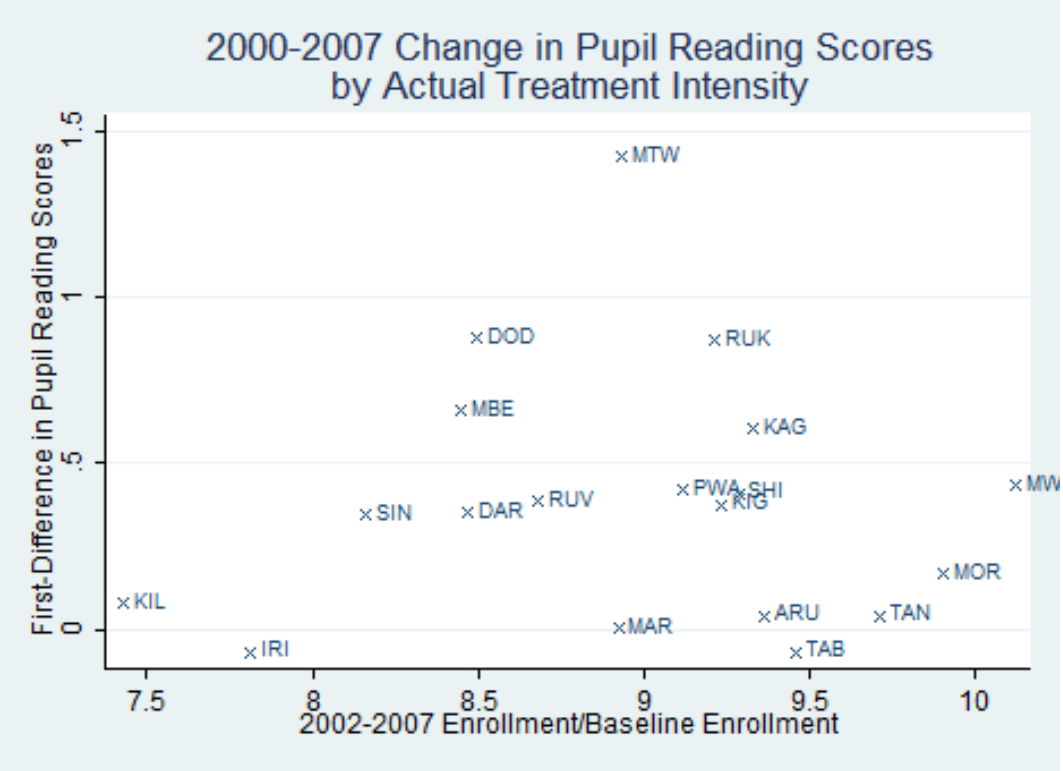
2000-2007 Change in Pupil Reading Scores
by Actual Treatment Intensity
First-Difference in Pupil Reading Scores
0
.5
1
1.5
MTW
DOD
RUK
MBE
KAG
SIN
DAR
RUV
PWA
SHI
KIG
MW
MOR
KIL
MAR
ARU
TAN
IRI
TAB
7.5
8
8.5
9
9.5
10
2002-2007 Enrollment/Baseline Enrollment

Figure A-3

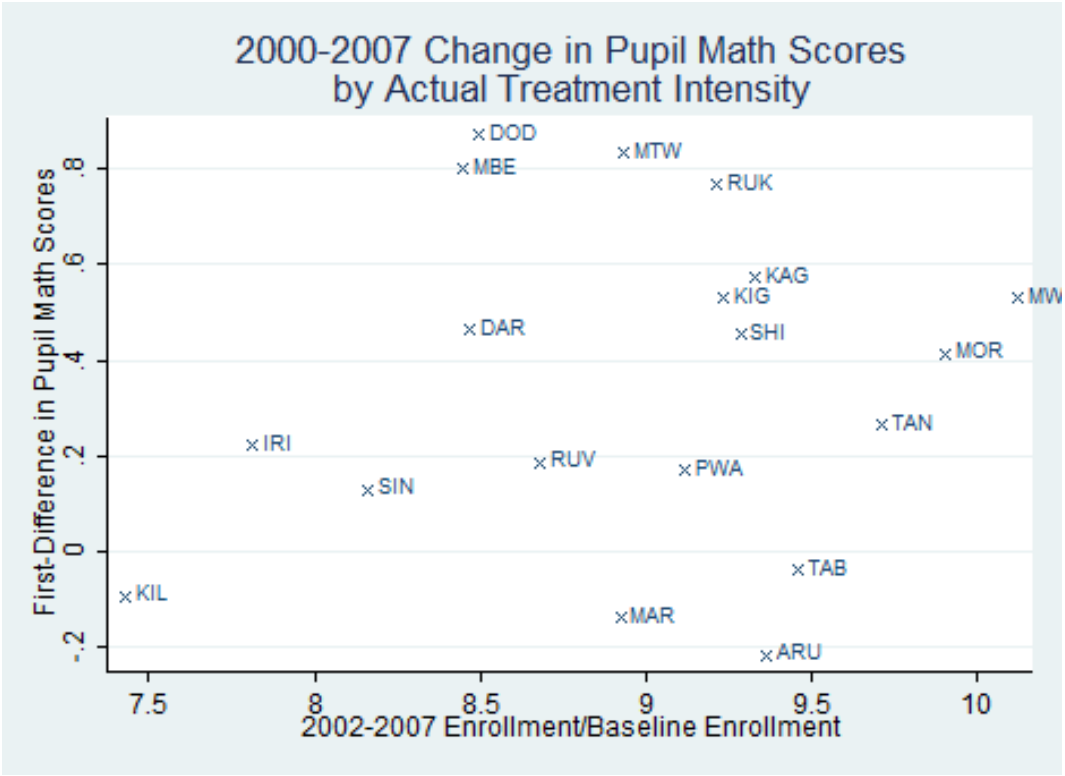
2000-2007 Change in Pupil Math Scores
by Actual Treatment Intensity
First-Difference in Pupil Math Scores
-.2
0
.2
.4
.6
.8
7.5
8
8.5
9
9.5
10
2002-2007 Enrollment/Baseline Enrollment
DOD
MBE
MTW
RUK
KAG
KIG
MW
DAR
SHI
MOR
TAN
IRI
RUV
PWA
SIN
TAB
KIL
MAR
ARU

Figure A-4

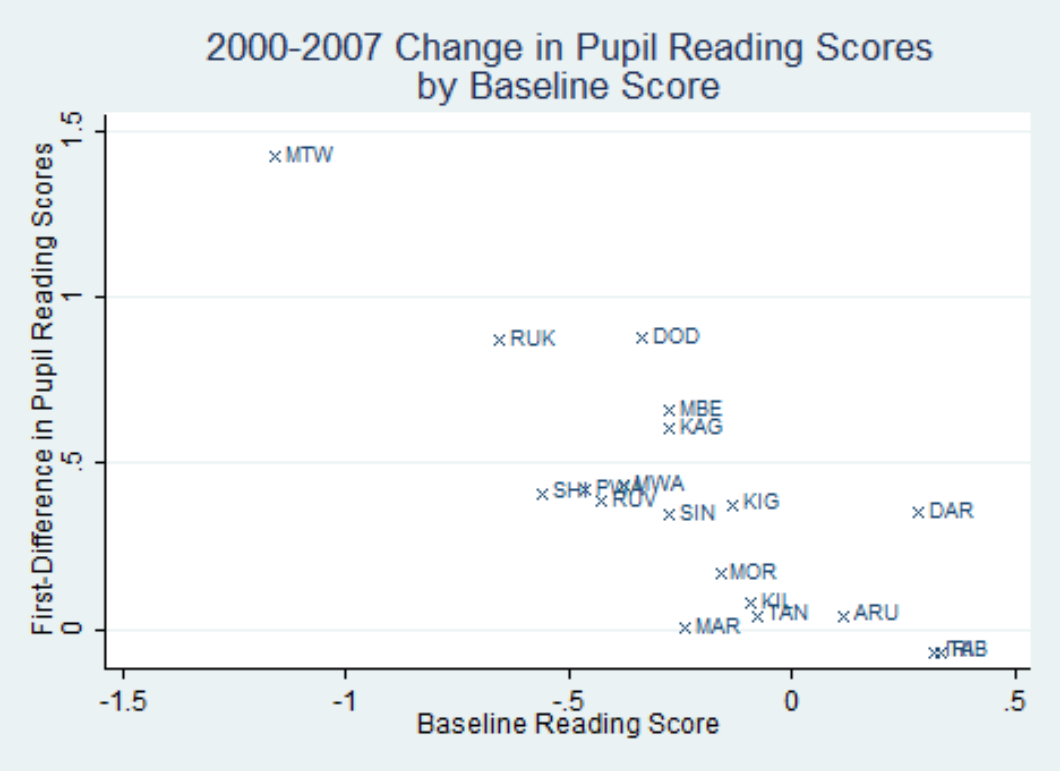
2000-2007 Change in Pupil Reading Scores
by Baseline Score
First-Difference in Pupil Reading Scores
0
.5
1
1.5
MTW
RUK
DOD
MBE
KAG
SH
RUV
SIN
KIG
DAR
MOR
KIL
TAN
ARU
MAR
-1.5
-1
-.5
0
.5
Baseline Reading Score

Figure A-5

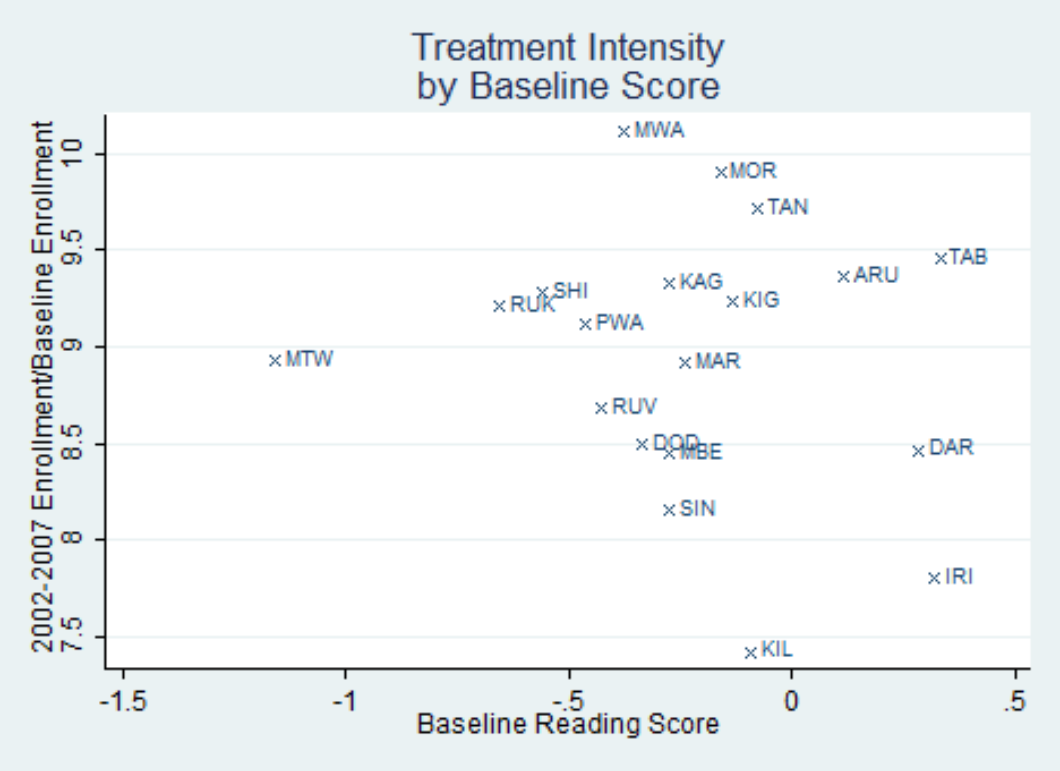
Treatment Intensity
by Baseline Score
2002-2007 Enrollment/Baseline Enrollment
7.5
8
8.5
9
9.5
10
MWA
MOR
TAN
TAB
ARU
KAG
SHI
KIG
RUK
PWA
MTW
MAR
RUV
DOD
MBE
DAR
SIN
IRI
KIL
-1.5
-1
-.5
0
.5
Baseline Reading Score

Figure A-6

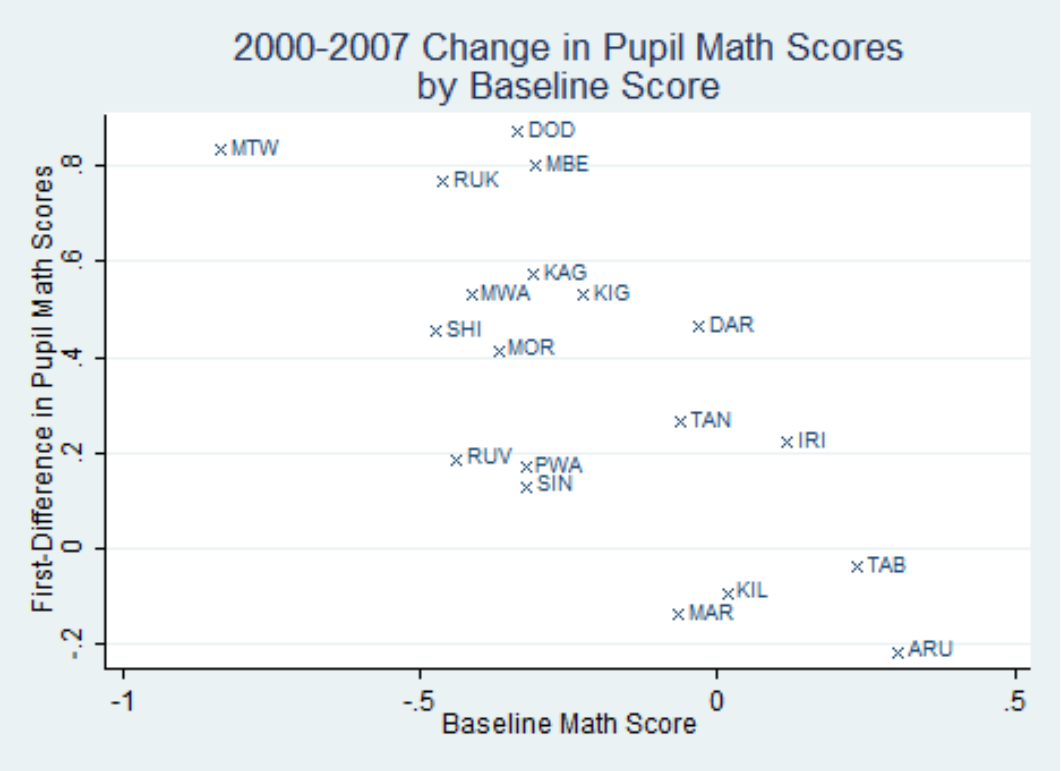
2000-2007 Change in Pupil Math Scores
by Baseline Score
First-Difference in Pupil Math Scores
-.2
0
.2
.4
.6
.8
-1
-.5
0
.5
Baseline Math Score
DOD
MTW
MBE
RUK
KAG
MWA
KIG
SHI
DAR
MOR
TAN
IRI
RUV
PWA
SIN
TAB
KIL
MAR
ARU

Figure A-7

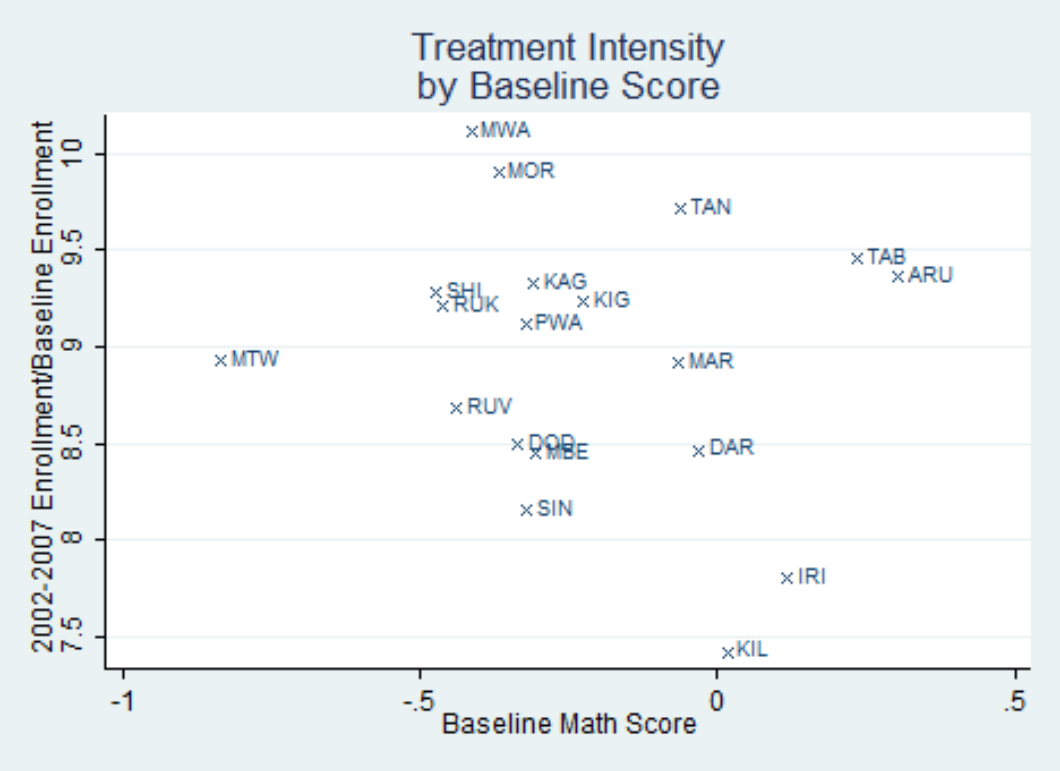
Treatment Intensity
by Baseline Score
2002-2007 Enrollment/Baseline Enrollment
7.5
8
8.5
9
9.5
10
MWA
MOR
TAN
TAB
ARU
SHI
KAG
RUK
KIG
PWA
MTW
MAR
RUV
DOD
MBE
DAR
SIN
IRI
KIL
-1
-.5
0
.5
Baseline Math Score

Figure A-8

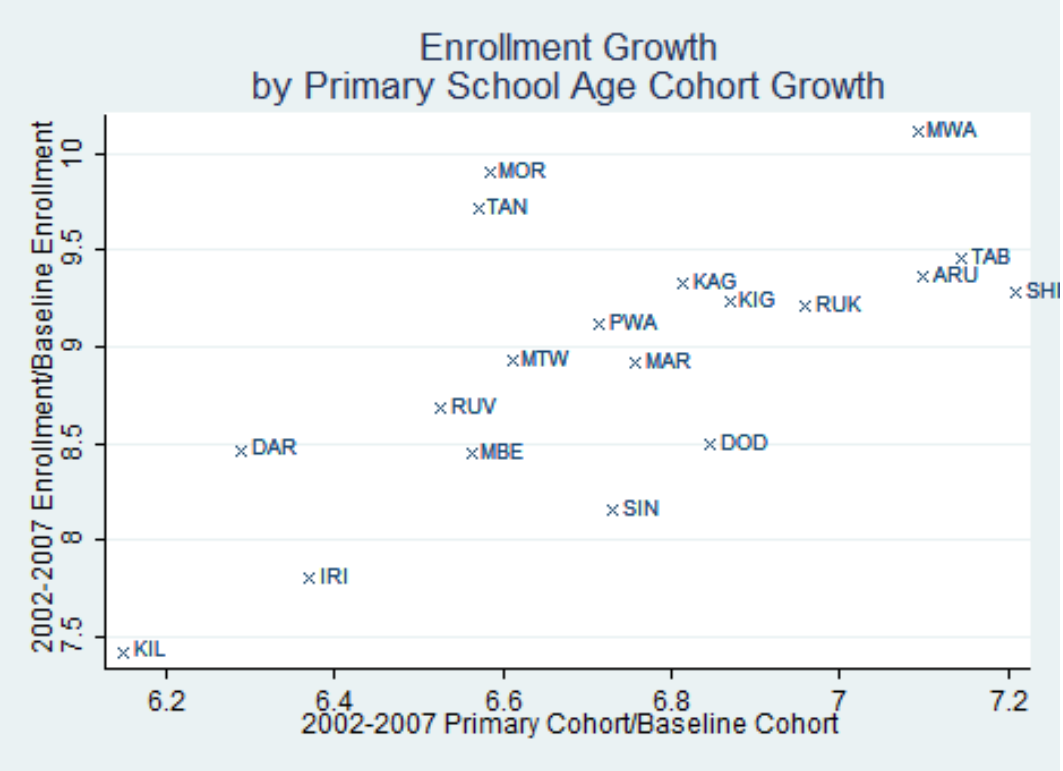
Enrollment Growth
by Primary School Age Cohort Growth
2002-2007 Enrollment/Baseline Enrollment
7.5
8
8.5
9
9.5
10
6.2
6.4
6.6
6.8
7
7.2
2002-2007 Primary Cohort/Baseline Cohort
MWA
MOR
TAN
TAB
ARU
SHI
KAG
KIG
RUK
PWA
MTW
MAR
RUV
DAR
MBE
DOD
SIN
IRI
KIL

Figure A-9

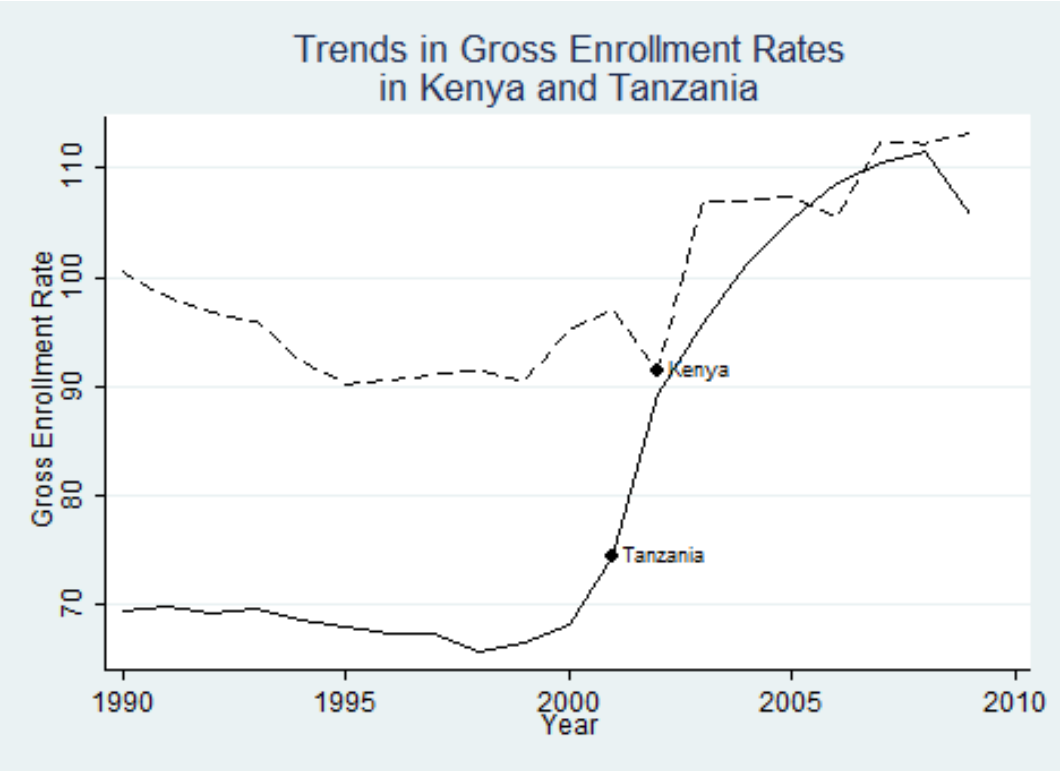


Figure A-10

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