

IZA DP No. 8621

**Teacher Pay and Student Performance:
Evidence from the Gambian Hardship Allowance**

Todd Pugatch
Elizabeth Schroeder

November 2014

Teacher Pay and Student Performance: Evidence from the Gambian Hardship Allowance

Todd Pugatch

*Oregon State University
and IZA*

Elizabeth Schroeder

Oregon State University

Discussion Paper No. 8621

November 2014

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.

ABSTRACT

Teacher Pay and Student Performance: Evidence from the Gambian Hardship Allowance¹

We evaluate the impact of the Gambian hardship allowance, which provides a salary premium of 30-40% to primary school teachers in remote locations, on student performance. A geographic discontinuity in the policy's implementation provides identifying variation. We find no effects of the hardship allowance on average student performance. These null average effects hide important heterogeneity, with learning gains for students at the top of the distribution and losses for those at the bottom. With over two dozen developing countries implementing similar policies to increase teacher compensation in rural schools, this study offers important evidence on their effectiveness.

JEL Classification: I25, I28, J38, J45, O12, O15

Keywords: teacher compensation, rural schools, Gambia, program evaluation, regression discontinuity

Corresponding author:

Todd Pugatch
School of Public Policy
303 Ballard Extension Hall
Oregon State University
Corvallis, OR 97331
USA
E-mail: todd.pugatch@oregonstate.edu

¹ We thank officials at the Gambia Ministry of Basic and Secondary Education, Gambian office of the West African Examinations Council, and the World Bank, particularly Alpha Bah, Momodou Cham, Jenny Hsieh, Sherif Yunus Hydera, Nathalie Lahire, Palamin Mbowe, Aidan Mulkeen, Momodou Sanneh, Ryoko Tomita, and Bassirou Touré. We thank Paul Ferraro, Laura Kawano and numerous seminar participants for helpful comments. We are especially grateful to Candice-Michelle Weems for GIS assistance. Emily Edwards and Andrew Spaeth provided excellent research assistance.

I. Introduction

Disparities in education quality between urban and rural areas are a preoccupation of policymakers throughout the developing world. In Guinea-Bissau, for example, only 27% of rural children can add two single-digit numbers, and only 19% can read a single word (Boone et al. 2013). In response, more than two dozen developing countries have implemented policies to recruit and retain teachers in rural areas. Yet despite the popularity of such policies, little is known about their effectiveness. In the Gambia, teachers in primary schools designated “hardship” due to their remote location earn a salary premium of 30%, 35%, or 40%, depending on the school’s distance from the capital. Teachers earn this hardship allowance unconditionally, i.e., regardless of qualifications or performance. This paper evaluates the effect of the program on student achievement.

An arbitrarily assigned 3-kilometer cutoff determining hardship status provides exogenous variation to identify the program’s impact. We use the universe of administrative records from an achievement test administered to all Gambian students in grades 3 and 5. Despite the sizeable salary increase and resulting increases in teacher qualifications (Pugatch and Schroeder 2014), we find no effects of the hardship allowance on average student performance. These null results persist when pooling the sample across genders, grades, or all available years of data. However, we find suggestive evidence that student performance improved at the top of the test score distribution and fell at the bottom.

We also find no effects of the policy on student enrollment, characteristics, effort, or school quality (other than the teacher characteristics mentioned above), helping to rule out several potential confounding factors to explain these results. Further exploration suggests that test score changes were driven by gains for the most socioeconomically advantaged students regardless of school attended, rather than concentrated within the highest quality schools. Overall, our findings indicate that the substantial, unconditional salary increases earned by Gambian teachers had little to no effect on average student performance, with gains limited to the best students.

Although much recent attention has focused on the effectiveness of teacher compensation that is tied to student performance (Lavy 2002; Lavy 2009; Glewwe, Ilias, and Kremer 2010; Muralidharan and Sundararaman 2011; Woessmann 2011; Fryer 2013), Gambian teachers receive the hardship allowance regardless of performance, as in most programs intended to recruit teachers to remote locations. Estimating the causal effect of unconditional increases in teacher salary is difficult for several

reasons. First, in most settings teacher salaries are determined by inflexible schedules based on formal qualifications and seniority, constraining variation in pay conditional on these characteristics. Second, where salary variation does exist among observationally similar teachers, it is likely correlated with school quality, teacher ability, bargaining power, or other unobservable characteristics, making it difficult to disentangle the effect of pay from these attributes. Accordingly, much of the literature on the relationship between teachers and student performance focuses on teacher characteristics other than salary. An important exception is de Ree et al. (2014), who find that an unconditional, randomly assigned doubling of teacher salary in Indonesia failed to produce gains in student learning.²

In a companion paper (Pugatch and Schroeder 2014), we find that the hardship allowance increases the proportion of qualified (certified) teachers in remote schools, making the present paper closely related to the literature on teacher quality and student performance. Robust evidence is emerging from the United States that the most effective teachers generate substantial gains in student learning and adult earnings (Hanushek 2011; Chetty, Friedman, and Rockoff 2014). But equally robust evidence suggests that a teacher's qualifications, in the form of a master's degree in education, bear no relation to student outcomes after controlling for other teacher and school characteristics (Darling-Hammond, Berry, and Thoreson 2001; Hanushek 2003; Rivkin, Hanushek, and Kain 2005; Jackson, Rockoff, and Staiger 2014). In other words, in developed countries teachers strongly influence student performance, but formal qualifications do not make better teachers.

The relevance of these studies for developing countries is potentially limited, however. Teacher certification in developing countries usually does not require an advanced degree, but instead distinguishes high school graduates from those with any post-secondary schooling, as is the case in the Gambia. Along this margin, increases in teacher qualifications might have a stronger effect than the bachelor's/master's degree margin observed in developed countries. In a comprehensive literature review of the effect of school inputs on educational outcomes in the developing world, Glewwe et al. (2011) find that observable markers of teacher quality, including education, in-service training, and salary tend to exert a positive influence on student learning. Nonetheless, the studies reviewed are not uniform in the direction of their findings, and restricting attention to those with the most credible

² Also related to our study is Kingdon and Teal (2010), who find that teacher unionization, which provides increased pay and other benefits, does not lead to gains in student performance in India.

identification strategies leaves no clear picture on the direction of these effects.³ Results from these studies in developing countries are therefore broadly consistent with the developed country literature.

Another prominent thread in the developing country literature focuses on the role of “contract teachers,” i.e., teachers recruited from the local population who are not subject to the pay and employment regulations of their civil service counterparts, and who typically lack formal teacher training. Galiani and Perez-Truglia (2011), the aforementioned Glewwe et al. (2011), McEwan (2013), and Murnane and Ganimian (2014) survey the literature, and find that the highest quality studies, which rely on experimental variation in the presence of contract teachers from randomized control trials, show positive effects of contract teachers on student outcomes. The most relevant of these for our work is Muralidharan and Sundararaman (2013), who report results from random placement of contract teachers in Indian rural primary schools. They find that students in schools with an extra contract teacher gained 0.16 standard deviations in math performance and 0.15 standard deviations in language, with effects greatest in remotely-located schools. Although not directly comparable to the policy variation used in this paper, the results nonetheless give reason to question the conventional wisdom that increases in qualified teachers should improve school outcomes (the United Nations calls for increases in qualified teacher supply in its proposed Sustainable Development Goals [United Nations 2014, Target 4.a], for instance). By virtue of their ties to the local community or perceived need to distinguish themselves, unqualified teachers may be better positioned to help students learn.⁴

This paper contributes to the literature in several ways. First, we provide evidence on a policy that has been adopted in similar form throughout the developing world, but about which little is known. We found documentation of 40 policies to recruit teachers to rural areas in 29 developing countries, with increased salary the modal incentive (full list available upon request). To our knowledge, we are the first to report credibly identified evidence on the effect of such a policy on student performance. Second, we contribute to the growing literature cited above on teacher salary increases in developing

³ McEwan (2013) also reviews the literature, focusing only on evidence from randomized control trials, and finds that teacher training is among the most effective interventions to improve student learning in developing country primary schools. However, most of the training interventions he reviews are specialized in-service training programs, not the general pre-service training that distinguishes qualified and unqualified teachers in the Gambia.

⁴ Experimental evidence from other studies (Banerjee et al. 2007 for India; Duflo, Dupas, and Kremer 2011 for Kenya) also find positive effects of contract teachers on student learning, but the nature of the interventions (for a remedial education program in India, and using contract teachers to halve class sizes in Kenya) make it difficult to draw connections with our setting. Other evidence on the effect of contract teachers is mixed (Vegas and De Laat 2003 for Togo; Bourdon, Frolich, and Michaelowa 2010 for Niger, Togo and Mali; Goyal and Pandey 2013 for India), possibly because all of these studies use selection-on-observables identification strategies, making it difficult to determine if unobservable teacher differences drive the results.

countries. Despite the excitement around salary increases tied to student performance, unconditional salary increases like the one we analyze are easier to implement politically and therefore arguably more reflective of the policy environment in most countries. Finally, because the Gambian hardship allowance was designed to upgrade the formal qualifications of the rural teaching corps, this study contributes to the broader literature on the relationship between observable markers of teacher quality and student performance. Our results bolster the existing literature suggesting that neither unconditional salary increases nor increases in formal teacher qualifications lead to average test score gains.

In the next section, we describe the hardship allowance and standardized test that we use to gauge its effect. In Sections 3-4, we describe our methodology and data. Section 5 presents results, while Section 6 concludes.

II. Program Description

In the Gambia, the Ministry of Basic and Secondary Education (MoBSE) manages primary schools. Primary schools are of two types: Lower Basic Schools (LBS) include grades 1-6 only, while Basic Cycle Schools (BCS) include grades 1-9. The Ministry divides the country into 6 numbered administrative regions, in increasing distance from the capital Banjul. The hardship allowance applies to teachers of grades 1-6 in public schools in Regions 3-6. The Ministry designated schools as “hardship” in these regions if they were located at least 3 kilometers from the main highways running east-west on either bank of the Gambia River dividing the country. Figure 1 depicts the policy.

The hardship allowance, which has been in place since 2005, is 30% of salary in Regions 3-4, 35% in Regions 5, and 40% in Region 6. Teachers receive the allowance regardless of whether they have completed the 3-year training program offered by Gambia College, though qualified (certified) teachers earn a base salary 2.5 times greater than unqualified teachers. The program is externally financed by the World Bank, and costs an average of US\$23 per teacher, or US\$350,000 per year. The allowance is large in proportion to teacher salaries or to Gambian per capita GDP of US\$43 per month.

The Ministry’s regional offices review teacher allocations to schools each year. Although teachers are nominally required to accept their placement, in practice regional offices honor many teacher requests for specific posts. In a companion paper (Pugatch and Schroeder 2014), we find that the hardship allowance increased the proportion of qualified teachers in hardship schools by 10

percentage points and lowered the pupil-qualified teacher ratio by 61% of the mean. This paper addresses whether these increases in observable school quality boosted student performance.

We measure student performance using results from the National Assessment Test (NAT), which has been administered annually to all third and fifth graders in the Gambia since 2008. Each exam includes separately graded modules for English, Mathematics, and Science, with a module on Social and Environmental Studies added in Grade 5. We follow the literature by focusing on language and mathematics results. MoBSE, selected primary school teachers, and the West African Examination Council (WAEC, an independent body that administers standardized tests in the region) develop the tests based on subject-specific learning outcomes (Gambia Ministry of Basic and Secondary Education 2013). WAEC conducts all test grading. The tests are low stakes, in that they are not tied to student advancement, school budget allocations, or teacher pay. However, school-level results are made publicly available, and schools are required to post a comparison of their scores with national averages for parents and others to view.

III. Methodology

The impact of the allowance is identified by the policy granting hardship status to schools that were at least three kilometers from a main road. Using a regression discontinuity design, the identifying assumption is that school characteristics were distributed continuously across this threshold before the policy was implemented. In this case, there should have been no unobservable differences, on average, between schools that were just above and just below the threshold. We use two methodologies that exploit the regression discontinuity framework, first examining the average impact of the hardship allowance on student test scores, and then looking into the distribution of scores.

A. Mean treatment effects

To examine the average impact of the hardship allowance on student test scores, we use a fuzzy regression discontinuity design, as hardship status is not perfectly predicted by the three-kilometer cutoff. The first stage, which we specify as a linear probability model, tests whether crossing the threshold predicts treatment status for student i in school s :

$$D_{is} = \alpha + \beta T_s + f(d_s) + \gamma X_{is} + \delta_r + \varepsilon_{is} \quad (1)$$

We condition on a flexible function of the road distance, $f(d)$, which we model as a polynomial, as well as the student's age, age-squared, and gender, an indicator of whether the school is a Basic Cycle School (i.e., includes grades 7-9 in addition to grades 1-6), and a set of region fixed effects, δ . The impact of crossing the threshold on the probability of treatment is given by β .

The second-stage outcome of interest is the student's standardized test score in English or Math, denoted y . The reduced-form specification measures the effect of crossing the distance threshold on the outcome

$$y_{is} = \alpha + \rho T_s + f(d_s) + \gamma X_{is} + \delta_r + \varepsilon_{is} \quad (2)$$

Here, T_s is an indicator of crossing the threshold, and ρ measures the intent to treat (ITT).

We also estimate the local average treatment effect (LATE) using an IV strategy.

$$y_{is} = \alpha + \theta D_s + f(d_s) + \gamma X_{is} + \delta_r + \varepsilon_{is} \quad (3)$$

In equation (3), D_s is an indicator of treatment status, which we instrument with T_s . The IV estimate of θ has the interpretation of the LATE for schools induced into treatment by being just across the threshold. If the identifying assumption is met and there are no other policies that rely on the same three-kilometer cutoff, T_s meets the exclusion restriction required to instrument for hardship status. Note that the LATE coefficient is equal to the reduced-form estimate scaled by the first stage (i.e., $\theta = \rho/\beta$). Standard errors are clustered by school throughout.

B. Quantile treatment effects

A focus on mean effects could mask a change in the distribution of test scores, if the hardship allowance affects learning outcomes differently at different points the distribution. To address this possibility, we also analyze the reduced-form outcomes at a set of quantiles. We estimate quantile treatment effects using the nonparametric estimator of Frandsen, Frolich, and Melly (2012), which relies on the discontinuity in program assignment for identification.

First, to maintain consistency with the above analysis, we regress test scores onto a set of region dummies, a dummy variable indicating a Basic Cycle School, and the student's age and age-squared. The residuals from these regressions are used to estimate the quantile treatment effects.⁵

The Frandsen, Frolich and Melly estimator uses local linear regressions to estimate the distribution of potential outcomes for compliers at the threshold. The estimated cumulative density of test scores for treated students is given by:

$$\hat{F}_{Y^1}(y) = \frac{\lim_{r \rightarrow r_0^+} E[1(Y \leq y)D|R = r] - \lim_{r \rightarrow r_0^-} E[1(Y \leq y)D|R = r]}{\lim_{r \rightarrow r_0^+} E[D|R = r] - \lim_{r \rightarrow r_0^-} E[D|R = r]}$$

where R is the running variable (in our case, distance). This distribution can be estimated by local linear two-stage least squares, as can \hat{F}_{Y^0} , the distribution for untreated students. For a sharp discontinuity, or a reduced-form estimation of the intent-to-treat effect (as we estimate below), the denominator is equal to one and these densities are estimated by simple local linear regressions using kernel weights. The estimated densities are then inverted to find the quantiles, where $Q_{Y^1|C}(\tau)$ denotes the τ -quantile. The estimated treatment effect, identified at the threshold, is given by:

$$\hat{\delta}_{LQTE}(\tau) = \hat{Q}_{Y^1|C}(\tau) - \hat{Q}_{Y^0|C}(\tau)$$

where $\hat{\delta}$ is the local quantile treatment effect (LQTE). All regressions use a uniform kernel and a set of bandwidths that minimize asymptotic mean-squared error; monotonicity of the cdfs is guaranteed by rearrangements that preserve the asymptotic properties of the estimator (Chernozhukov, Fernández-Val, and Galichon 2010). Standard errors are clustered by school, computed by 200 bootstrap replications.

IV. Data

Data for the analysis are from 2012 and come from several sources. Student test scores come from the National Assessment Test (NAT), which has been administered to all students in Grades 3 and 5 since 2008. The Gambian office of the West African Examinations Council (WAEC) provided subject-

⁵ The method is similar to that of Canay (2011), who develops a two-step quantile regression estimator which partials out a unit-specific fixed effect in the first stage. The identifying assumption is that this fixed effect shifts the locations of all quantiles equally. We obtain similar results when using the Frandsen, Frolich and Melly estimator directly without first obtaining residuals, suggesting the Canay (2011) assumption is innocuous in our case.

specific scores for all students taking the NAT. Because the policy began in 2005, the absence of pre-treatment outcome data prevents us from using a longitudinal method such as difference-in-differences to identify the policy's effect. We limit our focus to 2012 because it is the most recent year available, and because all fifth graders in that year completed a demographic questionnaire that we use to supplement the test score data.

In addition to standardized test data, we also use information from the annual census of schools, the Education Management Information System (EMIS), conducted by The Gambian Ministry of Basic and Secondary Education (MoBSE). This census contains annual data on hardship status and enrollment. We use school locations provided by MoBSE and a map of the road network provided by the Gambia Bureau of Statistics (GBOS) to calculate travel distance from each school to the nearest main road. In addition, we have data on pre-treatment characteristics of the nearest village to each school from the 2003 Census, conducted by GBOS. All datasets other than the Census are administrative, and each contains the universe of its units.

The hardship policy applies to primary-school teachers in government-run schools, which include Lower Basic (primary) schools and Basic Cycle (combined primary and lower secondary) schools. Our sample includes both types of schools in Regions 3-6. The data on the Basic Cycle schools do not distinguish teachers of primary grades (1-6) from teachers of secondary grades (7-9), who do not receive the hardship allowance. We therefore include a dummy variable to control for Basic Cycle schools in all regressions to account for any systematic differences.

Travel distance from each school to the main road is the running variable in the regression discontinuity design. School locations are estimated by MoBSE officials, leading us to exclude schools with obvious measurement error in distance. We drop 42 schools whose map locations do not match their districts (a political boundary roughly equivalent to a U.S. county) as listed in the EMIS, as well as eight schools for which we could not find information on the nearest village in the 2003 Census, the final sample contains 244 schools. Remaining measurement error in distance should attenuate our estimates towards zero. Figure 1 shows a map of this dataset (schools in Regions 1-2 are included on the map for illustration only).

The dataset contains 148 hardship schools enrolling 29,723 students in grades 1-6, and 96 non-hardship schools with 26,682 students in grades 1-6. Table 1 shows sample means of a variety of school characteristics and tests for differences between hardship and non-hardship schools. Several notable

differences illustrate the differences between the two groups. Hardship schools are significantly more remote, located on average 11.8 kilometers from a main road, compared to 0.9 kilometers for non-hardship schools. Hardship schools also have nearly 130 fewer students on average. As our sample excludes the more urban Regions 1 and 2, students in both hardship and non-hardship schools have negative average z-scores, indicating that they perform worse than the national average. Hardship schools perform slightly worse than non-hardship schools within the sample, with differences of less than 0.1 standard deviations, but statistically significant in Grade 5.

National Assessment Test data record only a student's scores by subject, sex, and age. An important exception, however, is a questionnaire administered to all students in Grade 5 in 2012, which asked a battery of questions related to demographic and socioeconomic characteristics. Students in hardship schools are significantly less advantaged than their non-hardship peers, according to this survey. They are less likely to have a parent who completed primary school, more likely to report having no books at home or learning in a multi-grade classroom, and more likely to travel at least one hour to school. These differences highlight the inappropriateness of any empirical strategy that relies on simple comparisons between hardship and non-hardship schools in assessing the program's impact. Moreover, the likelihood that students in hardship and non-hardship also differ in unobservable characteristics makes it essential to use an identification strategy that accounts for both observable and unobservable differences.

V. Results

A. First Stage⁶

As discussed in Section 3, we use a fuzzy regression discontinuity design because treatment status does not align perfectly with the 3-kilometer distance threshold. Fourteen of the 244 schools in the estimation sample have hardship status that fails to correspond to their distance from the main road: 11 hardship schools fall within 3km, while 3 non-hardship schools are above this cutoff. MoBSE claims that our distance measures are incorrect and that they have faithfully applied the 3-kilometer threshold to all schools. They measure distance using vehicle odometer readings which are not centrally

⁶ Although this section closely follows the discussion of the first-stage regression discontinuity results in Pugatch and Schroeder (2014), numerical differences arise in the results because we weight by student enrollment, not number of teachers as in the earlier study. The different weights are chosen because our focus here is on student outcomes, so any school-level analysis should be representative of the student population, whereas Pugatch and Schroeder (2014) focus on teacher outcomes.

recorded, precluding us from making a direct comparison. If MoBSE is correct, then our running variable contains classical measurement error, biasing our results towards a weaker first stage. A more serious concern would be if any misclassification of hardship schools reflected manipulation by interested parties. We asked MoBSE officials, school administrators, and teachers throughout the country about this issue, but heard no reports of successful manipulation of hardship status. Nonetheless, we conduct statistical tests for evidence of manipulation and report results below.

Table 2 presents estimates of the first stage equation (3), where we have collapsed the data by school because this is the relevant level of variation in hardship status.⁷ In column (1), we include a quartic in distance from the main road but do not include any other additional controls. The distance threshold coefficient of 0.44 indicates a 44-percentage point greater likelihood that a school located just beyond this threshold will be a hardship school. This effect size is almost equivalent to the 47% of all schools that are hardship, and is statistically significant at the 1% level. In columns (2)-(3), we increase the polynomial order to 5 and 6, respectively, but the coefficient barely changes. Column (4) adds school type and region fixed effects, and column (5) adds several controls from the nearest village in the 2003 Census, prior to introduction of the hardship allowance in 2005.⁸ In column (6), we increase the polynomial order to 7. None of these modifications changes the coefficient on the distance threshold much. Column (7) allows the 7th-order polynomial to vary on both sides of the threshold. Now the coefficient changes somewhat, increasing to 0.56 and remaining significant at 10%. When limiting the sample to schools closest to the threshold, as in columns (8)-(9), the distance threshold continues to predict hardship status.

Column (4), which has the strongest instrument ($F=11.6$) among the specifications with a relevant first stage, is our preferred specification.⁹ A graph of the first stage appears in Figure 2, showing the probability of hardship status within bins of the running variable and predicted hardship status from this preferred specification.

⁷ We weight all first-stage regressions in Table 2 by the number of students enrolled in grades 1-6. We cluster standard errors by the cluster, the sub-regional administrative units for schools, of which there are 33 in the data.

⁸ We include controls for log population, employment/population ratio for ages 18+, percent with access to electricity, percent illiterate, percent Muslim, and percent of Mandinka, Fula, and Wollof ethnicities (the three largest ethnic groups).

⁹ Column (4) is also preferred among the specifications in which the F statistic exceeds 10 based on the Akaike and Bayesian Information Criteria (AIC and BIC). Lee and Lemieux (2010) suggest AIC as a guide to choice of polynomial order.

Valid causal inference in a regression discontinuity design also requires lack of manipulation of the running variable in order to secure favorable treatment status. If schools were strategically opened or closed on one side of the distance threshold, then the remaining schools may not be comparable across the cutoff. Given the external financing of the program, a disproportionate number of new schools just beyond the cutoff would be consistent with manipulation. Since adoption of the hardship allowance in 2005, 16 new schools in the dataset were constructed, split equally between hardship and non-hardship. (No schools were closed in that period.) Removing newly opened schools from the data does not change the first-stage results, as shown in column (10) of Table 2. A more formal test of manipulation of the running variable looks for “bunching” in the density around the threshold. Figure 3 presents a histogram of distances from schools to the main road, showing no clear evidence of bunching. Using the McCrary (2008) test to check formally, we fail to reject the null hypothesis of no manipulation at the threshold distance of 3 kilometers (p -value=0.24), using the suggested bin width and bandwidth.¹⁰ The results hold when halving the reference bandwidth (p -value=0.10), as also suggested in McCrary (2008).

As further checks on the validity of the first stage, we look for consistent application of the hardship criterion by substituting straight-line distance from schools to the main road as the running variable. Using this alternative measure, the coefficient on the distance threshold falls to 0.25, as shown in column (1) of Table A1. This magnitude is less than 60% of the first-stage estimates presented in Table 2, and significant only at the 10% level, helping to confirm that travel distance was the true criterion determining hardship status. We also look for continuity across the threshold distance in the distributions of observable characteristics prior to treatment. In the remaining columns of Table A1, we replace the first-stage dependent variable with a series of characteristics from the 2003 Census (matched to a school’s nearest village, and using different variables from those included in Table 2, column [5] in order to prevent pre-test bias). We find no significant coefficients on the distance threshold, indicating balance in pre-treatment characteristics.

B. Second Stage

The goal of the hardship allowance was to upgrade the quality of remotely located schools by providing incentives for teachers to locate and remain employed there. This paper analyzes whether the

¹⁰ The test requires monotonicity in the direction of manipulation of the running variable relative to its value in the absence of treatment. The likelihood that any manipulation of the running variable would occur only in favor of an increase in hardship schools means that the setting satisfies this condition.

program improved student performance as a result. Table 3 presents reduced-form and instrumental variables results in Panels A and B (corresponding to estimation of (2) and (3), respectively) for student z-scores from the National Assessment Test, separately for Math and English, Grades 3 and 5, and all students, males, and females.¹¹ In Panel A, column (1), the coefficient of 0.07 means that Grade 3 students from a school located just beyond the distance threshold scored 0.07 standard deviations better in English than those from a school located just inside the threshold. The corresponding instrumental variables estimate in Panel B is 0.15. Neither estimate is statistically distinguishable from zero.

Examining the remaining columns of Table 3, we find no effect of the hardship allowance on student performance at conventional significance levels, regardless of subject, grade, or sex of the student. The only exception is English for Grade 5 girls, who experience a 0.25 standard deviation intent to treat increase that is statistically significant at 10%. Across all specifications, point estimates are considerably larger in Grade 5 than Grade 3, but standard errors are too large to draw firm conclusions from this pattern.¹²

A potentially important confounding factor when interpreting the results on student learning presented in Table 3 is that the hardship allowance may have changed the number and composition of students attending each type of school. *A priori*, it is unclear in which direction any such effects would operate. Students could be induced into hardship schools from non-enrollment, in which case there would be negative selection, as the marginal non-enrolled student is likely to be of lower preparation and ability than previously enrolled students. Students could also be induced into hardship schools from non-hardship or private schools due to the upgrading of teacher quality documented in Pugatch and Schroeder (2014). This effect would reflect positive selection, as households who switch their children's school due to a perception of higher quality are likely more advantaged than average. If both selection effects operate simultaneously, then they may offset each other and lead to the null effects we find on student performance.

¹¹ First stage *F* statistics reported in Table 3 differ from those in Table 2 because Table 3 focuses on subsamples of students taking a particular test, while Table 2 analyzes the full sample at the school level.

¹² Gelman and Imbens (2014) criticize parametric estimation of regression discontinuity designs using polynomials in the running variable. Nonparametric estimation of the outcomes in Table 3 using local linear regression produce qualitatively similar results as those reported here, with no statistically significant estimates of the coefficient on the distance threshold (results not shown but available upon request).

To look for evidence of such selection, we first explore whether the number of students enrolled changed as a result of the hardship allowance. Table A2 presents results of school-level enrollment counts regressed on the distance threshold or (instrumented) hardship status, as in equations (4) and (5). The point estimates are positive for all students and for boys and girls separately (columns 1-3), but estimates are too imprecise to conclude that the hardship allowance increased enrollment. Nor is there evidence that the hardship allowance changed the proportion of female students enrolled (column 4). Columns (5)-(8) show similar results for the count of test-takers in Grades 3 and 5, suggesting that the test results of Table 3 are not skewed by any effect of the hardship allowance on the quantity of students in hardship schools.

We use data from the survey administered to all Grade 5 students taking the NAT to look for evidence of changes in student composition due to the hardship allowance. In Table A3, we present results from replacing the left-hand side of the reduced-form equation (2) with various student characteristics. We first construct an index representing a student's socioeconomic characteristics by aggregating several indicator variables for which a value of one represents a relatively advantaged background: whether the student has two siblings or fewer; speaks English at home; mother completed primary school; father completed primary school; has more than 10 books at home; travels less than an hour to school; has help on schoolwork available at home; and attended a nursery. The index is calculated as the proportion of responses equal to one from this list. Working with this index serves two purposes: to summarize in a single variable the rich information contained in the survey, and to guard against Type I errors due to multiple comparisons, in the spirit of Kling, Liebman, and Katz (2007). Using the socioeconomic status index as the outcome, the reduced-form coefficient on the distance threshold is not significantly different from zero (columns 1-3 of Table A3).¹³

In columns (4)-(6) of Table A3, we replace the socioeconomic status index with an index of student effort constructed in analogous fashion, and composed of indicators for whether the student never repeated a grade, was absent less than 6 days last month, attends extra class after school, or receives private tutoring. These characteristics are more likely to be under the control of the student than those entering the socioeconomic status index, and therefore measure (albeit imperfectly) the student's academic effort. Because the index should be distributed continuously across the distance threshold under the null hypothesis that the hardship allowance had no effect, the intent to treat

¹³ We have also run versions of these regressions with each of the component measures of the index as an outcome variable. Although the distance threshold coefficient is sometimes statistically significant at conventional levels in these specifications, the *p*-values fail to survive a Bonferroni correction for multiple hypothesis tests.

estimate measures whether the policy changed student effort.¹⁴ We find no evidence that the policy had this effect.

Finally, in columns (7)-(9) of Table A3, we look for evidence of changes in school quality other than the teacher characteristics considered in Pugatch and Schroeder (2014). The school quality index is the proportion equal to one of the following indicator variables: not in a multi-grade classroom, receive food at school, English class taught in English (rather than a local language), math class taught in English, use a textbook in English class, use a textbook in math class. We find no evidence of a change in school quality across the distance threshold using this index. In sum, in Table A3 we find no evidence that the hardship allowance altered the composition of students or measures of school quality not considered in our earlier work.

Although Tables A2-A3 help us rule out changes in student quantity, composition, and school quality as confounding factors influencing our main results, another potential explanation for the lack of significant effects found in Table 3 could be low statistical power. We attempt to address this by augmenting the dataset with all years since 2008, when the NAT was first administered, and present results in Table A4. Again we see no discernible effects of the hardship allowance on student learning. Grade 3 girls now experience an increase in English performance that is significant at 10%, but the effect for Grade 5 found in Table 3 is no longer significant (standard errors are now clustered by school-year). The lack of significant results in this augmented dataset is striking, because sample sizes have increased nearly fivefold, and increased first stage F-statistics should also make IV coefficients more precise.¹⁵

Findings in this section are admittedly noisy, and do not conclusively show that the average effect is indeed zero. However, their persistence across various subsamples, specifications, and attempts to increase statistical power is notable. Moreover, broadening our inquiry from cognitive skills to measures of student effort also failed to find any evidence of program impact.

C. Quantile treatment effects

A limitation of the analysis presented thus far is our exclusive focus on mean performance, whereas the hardship allowance could improve learning outcomes at other points in the distribution. To

¹⁴ These measures of effort may be viewed as proxies for non-cognitive skills in the spirit of Jackson (2012) and Gershenson (2014), who estimate the effect of teachers on student attendance.

¹⁵ Another attempt to increase statistical power by pooling the 2012 results for grades 3 and 5 also produced no statistically significant results (not shown but available upon request).

address this possibility, we re-analyze reduced-form outcomes by estimating quantile treatment effects, beginning with the (conditional) 5th percentile and continuing in increments of 5 until the 95th percentile. Figure 4 depicts results. Each graph shows the line of zero treatment effect (red); quantile regression estimates from 5th through the 95th percentiles in increments of 5 (solid black line); and 95% confidence interval for quantile regression coefficients (gray shaded area). In calculating confidence intervals, all standard errors are clustered by school.

The first notable characteristic of the quantile regression estimates is that they are noisy. With few exceptions, confidence intervals for quantile regression estimates of the intent to treat (ITT) contain zero. The most precise results are for Grade 3, which has a slightly higher number of students. For both Grade 3 math and Grade 3 English, the gradient of point estimates slopes upward, with negative and significant effects at the lowest quantiles, and positive and significant effects at the top quantiles.¹⁶ The point estimates are large, with the ITT exceeding one standard deviation in each direction when significantly different from zero. The results suggest that the hardship allowance benefited students at the top of the distribution but harmed those at the bottom. These offsetting effects help explain why the mean effects described in the preceding subsection are not significantly different from zero.

This upward-sloping pattern can also be observed in other groups, most notably in the math scores for the boys subsample and the subsample of Grade 3 girls. While the effects are not significant in any of the other panels, they are suggestive of a pattern. Figure 5 shows the smoothed densities of test scores for schools near the three-kilometer threshold, with the solid line denoting schools below the distance threshold and the dashed line showing schools above the threshold. While the distributions are generally centered around the same z-scores, the distributions of scores for more-remote schools tend to have fatter tails. The greater presence in remote schools of students with very high and low scores is consistent with the hypothesis that the hardship policy helped students at the top of the distribution and left behind those at the bottom.

A related question is whether this widening inequality of outcomes in response to the hardship allowance occurred across or within schools. An increase in within-school inequality would indicate an

¹⁶ Point estimates are statistically significantly different from zero only when pooling boys and girls in the same sample to increase power, as shown in Figure 4. For Grade 3 math, estimates are statistically significant for quantiles .15-.35 and .60-.95; for Grade 3 English, estimates are statistically significant for quantiles .05-.25 and .60-.95.

increased focus on the top students within the same classroom. To explore this possibility, we collapse test results by school and run the following reduced-form regression discontinuity specification:

$$y_{q,s} = \alpha + \rho T_s + f(d_s) + \gamma BCS_s + \delta_r + \varepsilon_{q,s} \quad (5)$$

where $y_{q,s}$ is the test score earned by the student scoring at quantile q within school s ; BCS is an indicator for being a Basic Cycle (combined primary and lower secondary) school; and all other notation is as in (4). We run a series of these regressions beginning with the 5th percentile of scores and continuing in increments of 5 until the 95th percentile.¹⁷ In the absence of the policy, schools just above and below the distance threshold should have the same within-school distribution of scores. Finding a pattern similar to that of Figure 4 would therefore reflect a widening of within-school inequality.

Results from estimation of (5) appear in Figure A1, using the same graphical elements as Figure 4, plus dashed black lines for the effect at the mean of y and corresponding 95% confidence interval. The results are very noisy, with nearly universal inclusion of a zero effect estimate in the confidence interval for each quantile. The gradient of results varies among downward-sloping, flat, and upward-sloping, with no clear pattern emerging. This stands in contrast to the results in Figure 4, for which nearly all gradients sloped upward, and which had more precise estimates. The results here suggest that the widening distribution of student test scores in response to the hardship allowance was not driven by increasing within-school inequality, but instead by differences across schools.¹⁸

D. Discussion

In sum, we have found no evidence that the hardship allowance changed average student performance; and some evidence of improved math and English scores for the best students in Grade 3, but lower scores for those at the bottom of the distribution. Given earlier evidence that the hardship allowance improved teacher quality (Pugatch and Schroeder 2014) and the apparent heterogeneity in effects on students across different schools, two potential channels emerge immediately to explain our results:

¹⁷ Because the unit of observation in these regressions is the school, we can no longer cluster standard errors at this level. Instead, we cluster standard errors by the cluster, a sub-regional education administrative unit of which there are 33 in the data.

¹⁸ Replacing the left-hand side of (5) with the difference between a school's 90th and 10th percentile test scores, an alternative summary measure of within-school inequality, also produced no statistically significant results (not shown but available upon request).

1. Improvement in teacher quality. An influx of qualified teachers could reorient classrooms to focus on the best students at the expense of the weakest.
2. Student composition: students from more advantaged backgrounds might be more likely to benefit from hardship schools, while less advantaged students might lag behind. This would be the case if, for instance, the curriculum accelerates under a new teacher recruited via the hardship allowance.¹⁹

Additional potential mechanisms include:

3. Enhanced teacher motivation, i.e., an efficiency wage effect. Because all teachers receive the hardship allowance regardless of qualifications, this effect would operate for all teachers, not only those who are certified or are new to a hardship school.
4. Peer effects: the presence of students from more advantaged backgrounds helps other students perform better.

We look for evidence of each of these mechanisms by augmenting the reduced form equation (4) with an interaction term between the indicator for the 3-kilometer distance threshold and an observable characteristic, as well as the main effect of the observable and interactions between it and the polynomial terms of the running variable. The coefficient on the interaction term between the observable characteristic and the distance threshold indicator describes heterogeneity in the intent to treat (ITT) according to that characteristic. If all schools respond to being just across the distance threshold equally, there will be no heterogeneity in the ITT, i.e., the coefficient on the interaction term will be zero. Non-zero coefficients on interaction terms would suggest the presence of particular mechanisms through which the program exerted an effect.

The observable characteristics we use are the percentage of qualified teachers at a school, which measures teacher quality; the hardship salary premium (30% for Regions 3-4, 35% for Region 5, and 40% for Region 6), which would suggest an efficiency wage effect; the school's mean z-score on English and math in the previous year, which proxies for the quality of a student's peers and other aspects of school quality not captured elsewhere; and the socioeconomic status index described in Section V.B, which summarizes the student's relative advantages.

¹⁹ These explanations are broadly consistent with Pritchett (2013), who argues that curricula in many developing countries are suited only for the top students, leaving the rest to fall progressively behind. We have addressed this question to some extent already by estimating changes in the within-school test score distribution in the previous subsection, but approach the question from a slightly different angle here.

Before discussing the results of this exercise, we note its limitations. The coefficient on the interaction term between an observable characteristic and the distance threshold indicator captures heterogeneity in student performance between schools just beyond the threshold with different values of this observable. If the variation in observables across the threshold is correlated with any unobserved attributes that influence student performance, then the coefficient on the interaction term will not consistently estimate the heterogeneity in the intent to treat parameter. Moreover, if the observable characteristic is itself influenced by being just beyond the threshold distance, as we have reason to believe in the case of hardship allowance receipt, then we are controlling for an intermediate outcome, i.e., using “bad control” (Angrist and Pischke 2008). We therefore consider the next set of results as providing descriptive evidence on the mechanisms we hypothesized, rather than being causal.

Table 4 presents results of the exercise. We limit the sample to Grade 5 because the socioeconomic status index cannot be constructed for Grade 3. For Grade 5 English, male students in schools located just beyond the threshold distance perform better the higher their socioeconomic status (statistically significant at 5%). The coefficient magnitude of 1.16 means that in a school just across the 3-kilometer threshold, a student with the highest socioeconomic status according to our index (an index value of 1) is predicted to score 1.16 standard deviations higher in English than a student with the lowest socioeconomic status (an index value of 0). This is a striking discrepancy, and it is important to emphasize that it does not merely reflect average differences between students of high and low socioeconomic status, which will be captured by the included main effect and interactions with the distance polynomial. Instead, this coefficient captures a discontinuous jump in performance across the hardship distance threshold for higher-socioeconomic status students, and therefore represents a heterogeneous intent to treat effect of the hardship allowance. A difference of similar magnitude and precision appears for male students in math, and a significant coefficient of 0.85 in math for all students.

Coefficients on other interaction terms—qualified teacher percentage, hardship salary premium, and a school’s lagged z-score—are generally positive as expected, but fail to produce statistically significant results. This helps rule out the competing explanations of the increase in teacher qualifications, teacher motivation through efficiency wages, and peer effects or other characteristics of school quality not otherwise captured, in explaining the treatment effect gradient found in Figure 4.

We close this discussion of mechanisms by returning to the question of whether the gains at the top of the distribution found in Figure 4 reflect differences among students or schools. To do so, we re-estimate the main reduced-form specification in (2), splitting the sample by median socioeconomic

status index and by 2011 average English and Math z-score. Because the socioeconomic status index reflects a student's household characteristics, regardless of school quality, while the lagged z-score reflects school quality using a cohort to which the student does not belong, splitting the sample in this way could help illuminate earlier findings.

Table 5 presents results, with the sample split by student socioeconomic status in Panel A and school quality in Panel B.²² In Panel A, students in schools beyond the distance threshold from above-median socioeconomic status households score significantly better in both English and Math, on the order of 0.33-0.44 standard deviations higher. In Panel B, we find no statistically significant differences according to school quality. This suggests that the heterogeneous treatment effects found earlier across test score quantiles are more reflective of differences among students than among schools. Although the socioeconomic status index will not map perfectly into a student's position in the test score distribution, the results nonetheless suggest that the best students benefitted from the hardship allowance, even if they did not attend the best schools.

VI. Conclusion

This paper has evaluated the effect of the Gambian hardship allowance on student performance, using a geographic discontinuity in the policy's implementation as a source of identifying variation. We failed to find any effects of the policy on average student performance. This lack of results persists across genders, grades, and when pooling samples across years to increase precision. We also fail to find effects of the policy on enrollment, percent of female students enrolled, student socioeconomic characteristics, student effort, or school quality (other than those on teacher characteristics found in earlier work).

Despite the lack of an average impact, we find evidence of heterogeneity in the effect of the hardship allowance. There is suggestive evidence of increased learning at the top of the test score distribution and decreased achievement at the bottom. Auxiliary analysis finds that these heterogeneous treatment effects are driven by differences among students in schools just beyond the distance threshold, rather than of a concentration of top- or bottom-performing students in particular schools. Our findings suggest that only socioeconomically advantaged students benefitted from the hardship allowance, perhaps because the teachers recruited via the program are better able to connect

²² We determine median socioeconomic status using student-level observations, while median 2011 scores are determined using school-level aggregates. The "below median" groups include those at the median. These definitions explain the differences in sample sizes reported in Table 5.

with these types of students. The flip-side of this result is the declines observed at the bottom of the test score distribution, suggesting that the changes induced by the hardship allowance left weaker students behind. An enhanced focus on these students would help make the program both more equitable and beneficial overall.

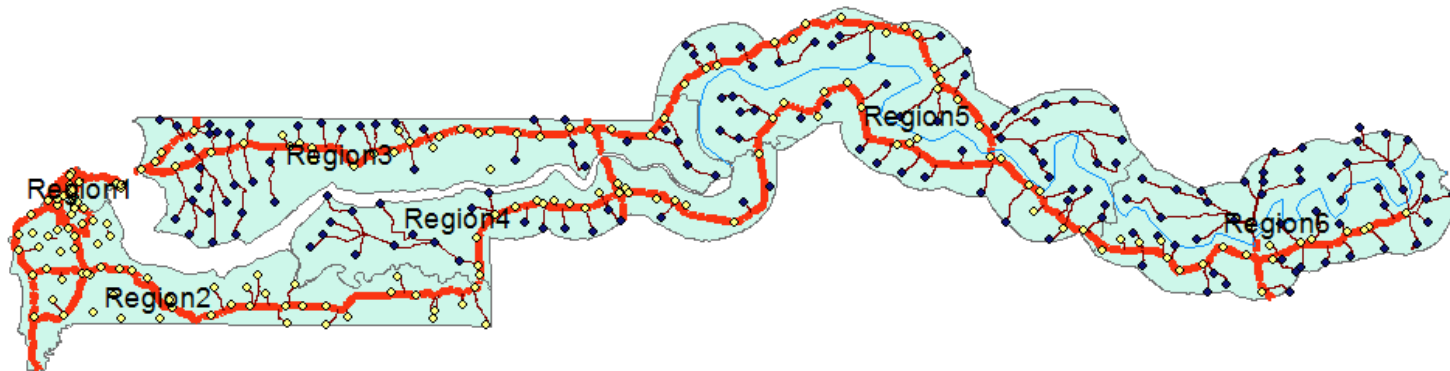
Although statistical noise hampers our main results (i.e., we do not have “precisely estimated zeroes”), they stem from a credible identification strategy that passes a battery of specification checks. Moreover, the lack of positive results across a range of subjects, grades, and subsamples make it unlikely that the program is raising average student performance, at least in schools near the 3km cutoff that are the focus of our regression discontinuity design. If anything, gains from the policy appear to be concentrated among the most capable students. Our results therefore echo Murnane and Ganimian (2014), whose survey of the most rigorously identified evaluations of schooling interventions in developing countries noted how difficult it is to “induce teachers to maximize their efforts to teach all students well” (p. 31). In our companion paper (Pugatch and Schroeder 2014), we estimated that the hardship allowance generated an overall increase of 140 qualified teachers in hardship schools, at a cost of US\$2,500 each. Although the program could benefit students through channels other than this increase in qualified teachers, its annual cost of US\$350,000 does not appear to translate into learning gains for the majority of Gambian students. Tying salary increases more directly to student performance could help to establish this link.

References

- Angrist, Joshua D., and Jorn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. 1st ed. Princeton University Press.
- Banerjee, Abhijit V., Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics* 122 (3): 1235–64.
- Boone, Peter, Ila Fazzio, Kameshwari Jandhyala, Chitra Jayanty, Gangadhar Jayanty, Simon Johnson, Vimala Ramachandrin, Filipa Silva, and Zhaoguo Zhan. 2013. *The Surprisingly Dire Situation of Children's Education in Rural West Africa: Results from the CREO Study in Guinea-Bissau (Comprehensive Review of Education Outcomes)*. National Bureau of Economic Research.
- Bourdon, Jean, Markus Frolich, and Katharina Michaelowa. 2010. "Teacher Shortages, Teacher Contracts and Their Effect on Education in Africa." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 173 (1): 93–116.
- Canay, Ivan A. 2011. "A Simple Approach to Quantile Regression for Panel Data." *The Econometrics Journal* 14 (3): 368–86.
- Chernozhukov, Victor, Iván Fernández-Val, and Alfred Galichon. 2010. "QUANTILE AND PROBABILITY CURVES WITHOUT CROSSING." *Econometrica* 78 (3): 1093–1125.
- Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood [†]." *American Economic Review* 104 (9): 2633–79.
- Darling-Hammond, Linda, Barnett Berry, and Amy Thoreson. 2001. "Does Teacher Certification Matter? Evaluating the Evidence." *Educational Evaluation and Policy Analysis* 23 (1): 57–77.
- De Ree, Joppe, Karthik Muralidharan, Menno Pradhan, and F. Halsey Rogers. 2014. "Double for Nothing? The Effects of Unconditional Teacher Salary Increases on Performance." <http://www.effective-states.org/wp-content/uploads/2014/06/Double-for-nothing.pdf>.
- Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review* 101 (5): 1739–74.
- Frandsen, Brigham R., Markus Frolich, and Blaise Melly. 2012. "Quantile Treatment Effects in the Regression Discontinuity Design." *Journal of Econometrics* 168 (2): 382–95.
- Fryer, Roland G. 2013. "Teacher Incentives and Student Achievement: Evidence from New York City Public Schools." *Journal of Labor Economics* 31 (2): 373–407.
- Galiani, Sebastian, and Ricardo Perez-Truglia. 2011. *School Management in Developing Countries*. SSRN Scholarly Paper ID 1972459. Rochester, NY: Social Science Research Network.
- Gambia Ministry of Basic and Secondary Education. 2013. *Results of the 2012 National Assessment of Pupil Achievement in English and Mathematics, Grades 3 and 5*.
- Gelman, Andrew, and Guido Imbens. 2014. *Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs*. Working Paper 20405. National Bureau of Economic Research.
- Gershenson, Seth. 2014. "Linking Teacher Quality, Student Attendance, and Student Achievement." Accessed October 20. http://www.admissions.american.edu/cas/economics/news/upload/Gershenson_Absences2.pdf.
- Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2010. "Teacher Incentives." *American Economic Journal: Applied Economics* 2 (3): 205–27.
- Glewwe, Paul W., Eric A. Hanushek, Sarah D. Humpage, and Renato Ravina. 2011. *School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010*. Working Paper 17554. National Bureau of Economic Research.

- Goyal, Sangeeta, and Priyanka Pandey. 2013. "Contract Teachers in India." *Education Economics* 21 (5): 464–84.
- Hanushek, Eric A. 2003. "The Failure of Input-Based Schooling Policies." *Economic Journal* 113 (485): F64–98.
- . 2011. "The Economic Value of Higher Teacher Quality." *Economics of Education Review* 30 (3): 466–79.
- Jackson, C. Kirabo. 2012. *Non-Cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9th Grade Teachers in North Carolina*. National Bureau of Economic Research.
- Jackson, C. Kirabo, Jonah E. Rockoff, and Douglas O. Staiger. 2014. "Teacher Effects and Teacher-Related Policies." *Annual Review of Economics* 6 (1): 801–25.
- Kingdon, Geeta, and Francis Teal. 2010. "Teacher Unions, Teacher Pay and Student Performance in India: A Pupil Fixed Effects Approach." *Journal of Development Economics* 91 (2): 278–88.
- Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75 (1): 83–119.
- Lavy, Victor. 2002. "Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement." *Journal of Political Economy* 110 (6): 1286–1317.
- . 2009. "Performance Pay and Teachers' Effort, Productivity, and Grading Ethics." *The American Economic Review* 99 (5): 1979–2011.
- Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (2): 281–355.
- McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142 (2): 698–714.
- McEwan, Patrick J. 2013. "Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments." *Unpublished Manuscript, Wellesley College*.
- Muralidharan, Karthik, and Venkatesh Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India." *Journal of Political Economy* 119 (1): 39–77.
- . 2013. *Contract Teachers: Experimental Evidence from India*. Working Paper 19440. National Bureau of Economic Research.
- Murnane, Richard J., and Alejandro J. Ganimian. 2014. *Improving Educational Outcomes in Developing Countries: Lessons from Rigorous Evaluations*. Working Paper 20284. National Bureau of Economic Research.
- Pritchett, Lant. 2013. *The Rebirth of Education: Schooling Ain't Learning*. Washington, D.C: Center for Global Development.
- Pugatch, Todd, and Elizabeth Schroeder. 2014a. "Incentives for Teacher Relocation: Evidence from the Gambian Hardship Allowance." *Economics of Education Review* 41 (August): 120–36.
- . 2014b. "Incentives for Teacher Relocation: Evidence from the Gambian Hardship Allowance." *Economics of Education Review* 41 (August): 120–36.
- Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica* 73 (2): 417–58.
- United Nations. 2014. "Introduction and Proposed Goals and Targets on Sustainable Development For the Post-2015 Development Agenda." United Nations.
- Vegas, Emiliana, and Joost De Laat. 2003. "Do Differences in Teacher Contracts Affect Student Performance? Evidence from Togo," World Bank.
- Woessmann, Ludger. 2011. "Cross-Country Evidence on Teacher Performance Pay." *Economics of Education Review* 30 (3): 404–18.

Figure 1: Gambia Hardship Allowance, 2012



Legend

- ◊ non-hardship
- ◆ hardship
- feeder road
- river
- main road

0 25 50 100 Kilometers



Figure 2: Regression Discontinuity, first stage

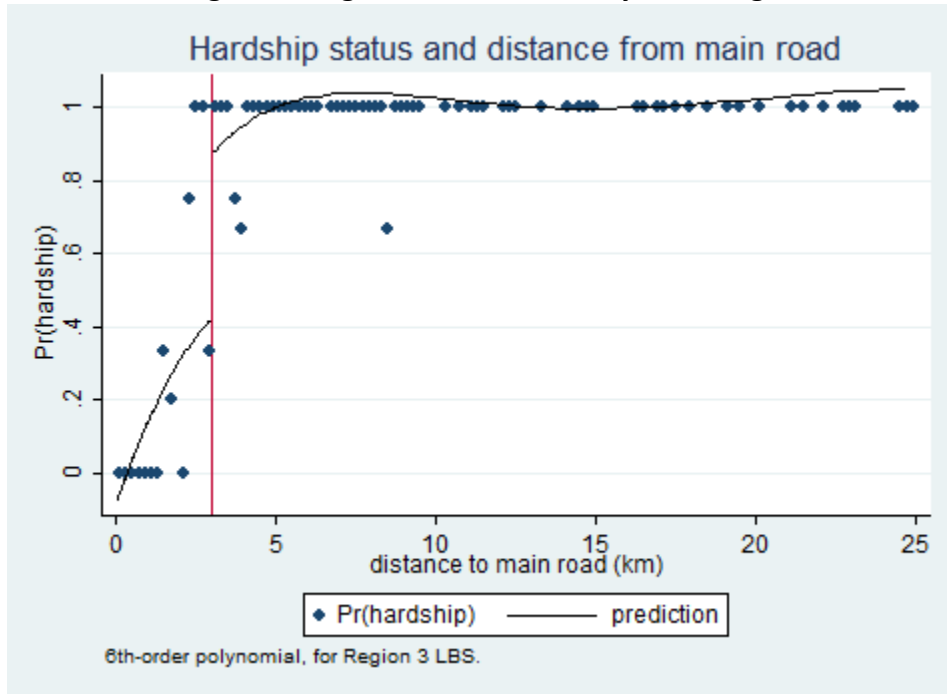


Figure shows mean of hardship status within bins defined by distance from main road (bandwidth=0.2km). Line is predicted hardship status from first stage regression, as in column (4) of Table 2.

Figure 3: Density of distance from schools to main road

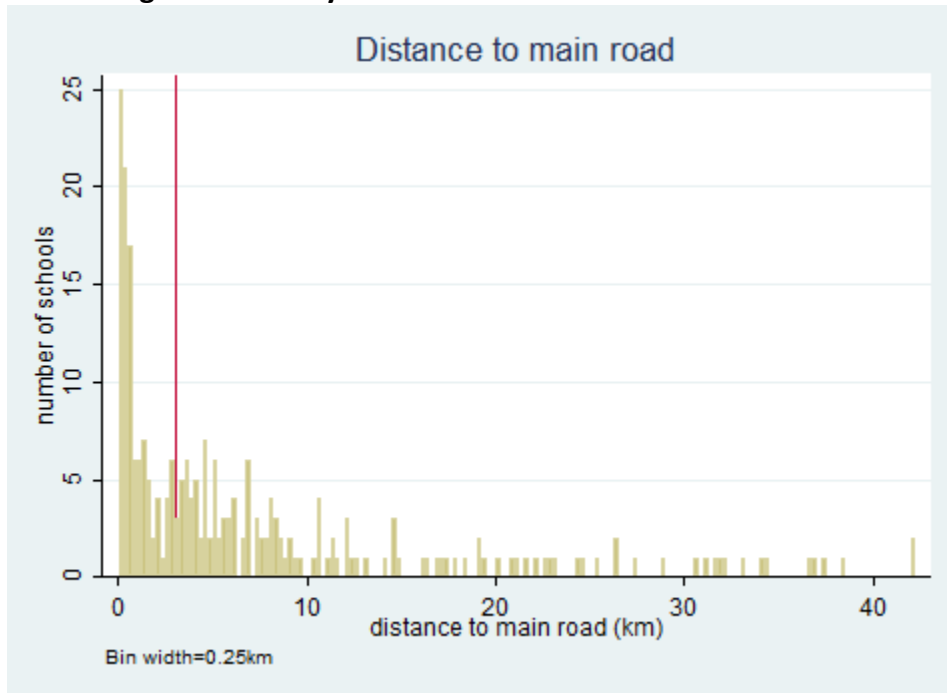


Figure 4: Quantile Treatment Effect Estimates

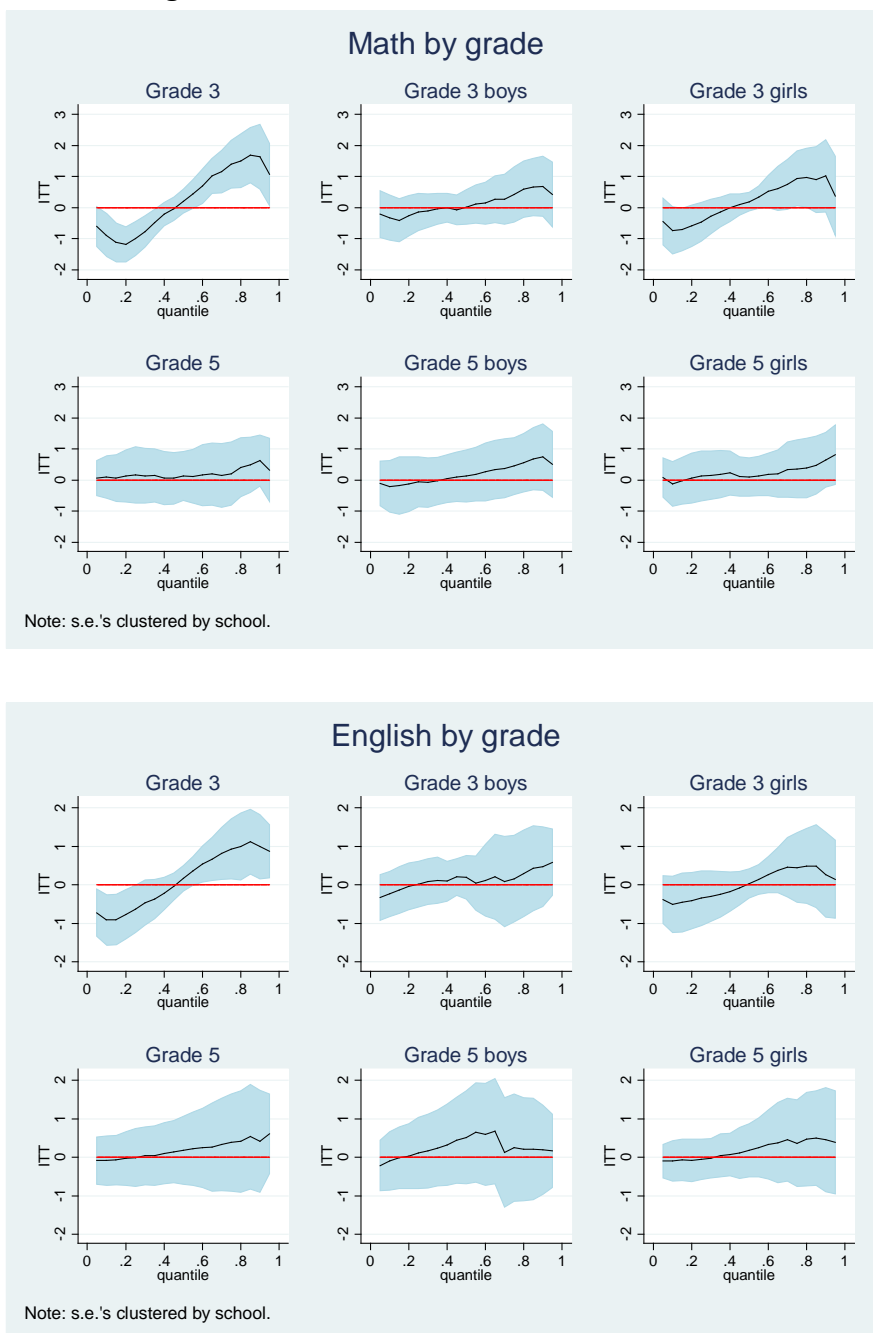


Figure shows results of reduced-form regression discontinuity design of test score on indicator for being in school beyond 3km distance threshold (intent to treat, or ITT). Each graph shows line of zero treatment effect (red); quantile regression estimates from 5th through 95th percentiles in increments of 5 (solid black line); and 95% confidence interval for quantile regression coefficients (gray shaded area). Standard errors clustered by school, computed by 200 bootstrap replications.

Figure 5: Test score densities

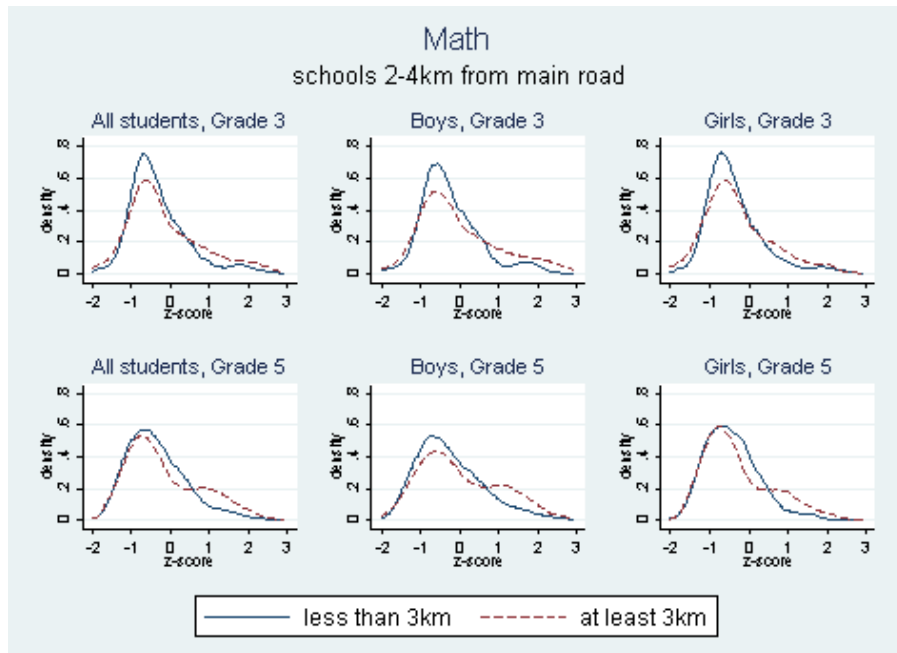
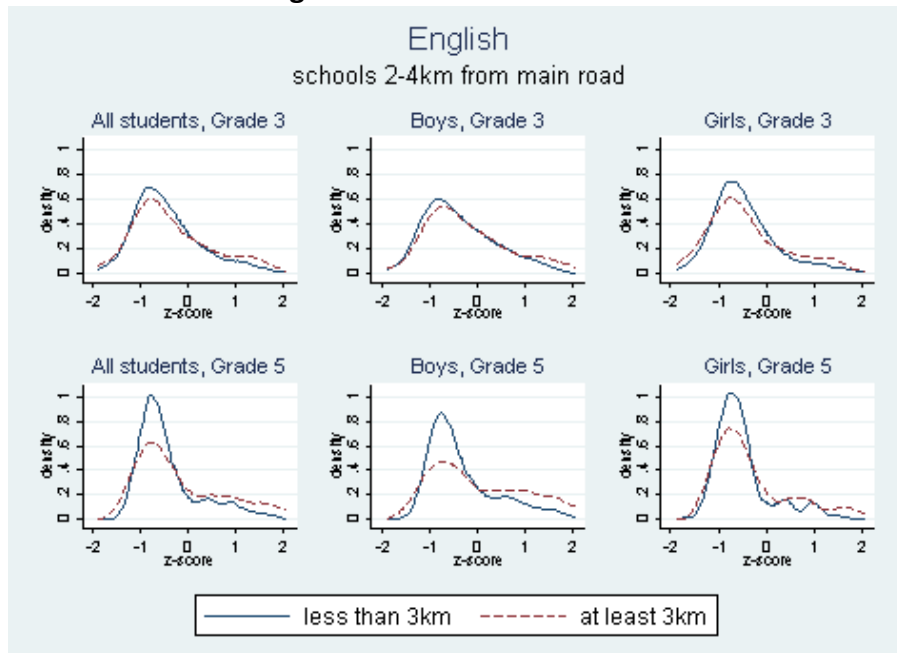
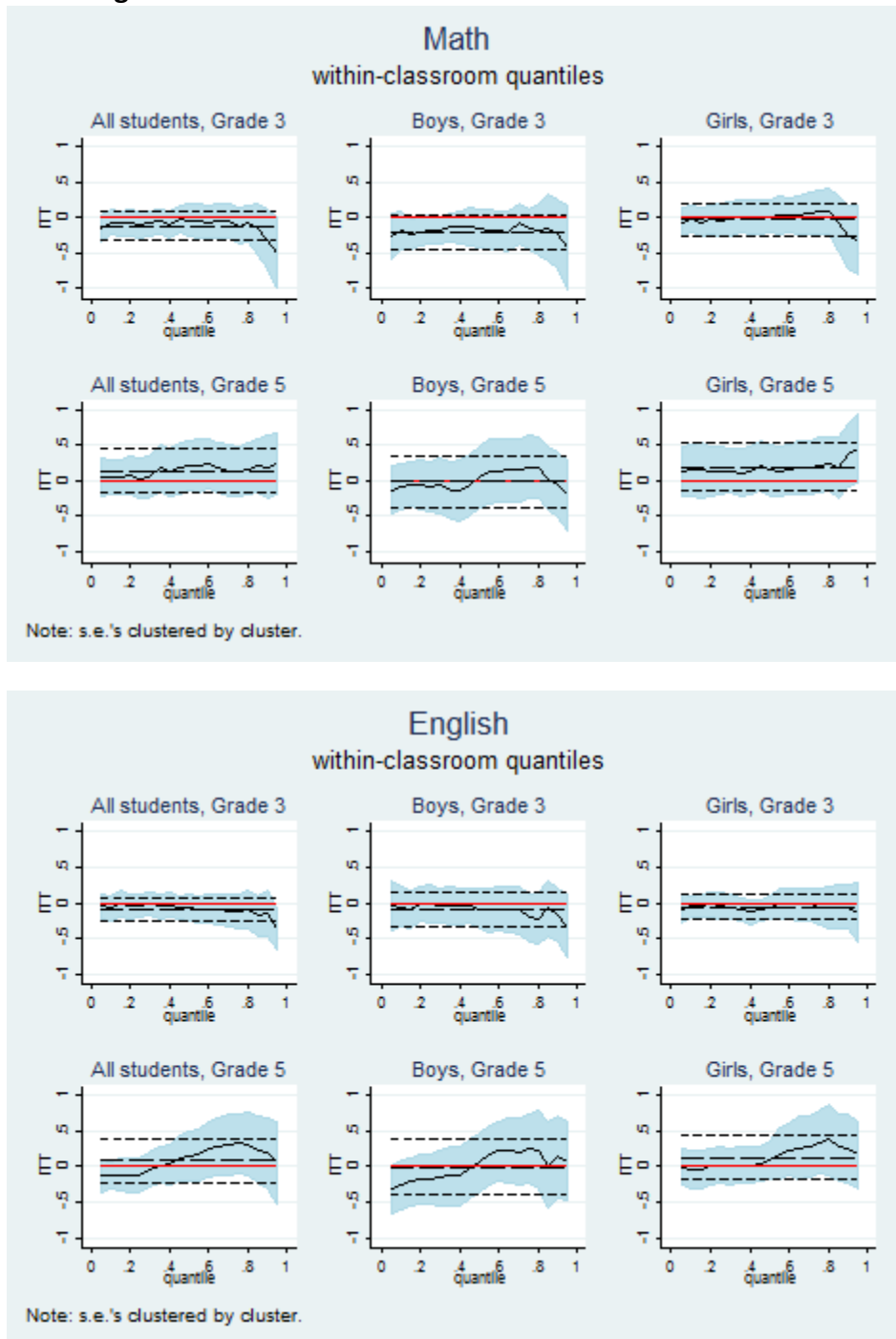


Figure A1: Within-school test score distribution estimates



Each graph shows line of zero treatment effect (red); within-school quantile regression estimates from 5th through 95th percentiles in increments of 5 (solid black line); 95th confidence interval for quantile regression coefficients (gray shaded area); within-school mean regression estimate (long dashed line) and its 95th confidence interval (short dashed lines). Standard errors clustered by cluster.

Table 1: School & student characteristics, by hardship status (2012)

Variable	Non-hardship	Hardship	Difference
Number of schools	96	148	52
Number of students	29,723	26,682	-3,041
Distance from main road (km)	0.9	11.8	10.9***
<u>enrollment</u>			
total	309.6	180.3	-129.3***
male	148.0	86.7	-61.3***
female	161.6	93.6	-68.0***
% female	0.52	0.52	0.00
pupil-teacher ratio	29.5	26.9	-2.6
<u>student performance</u>			
English, Grade 3	-0.34	-0.36	-0.02
English, Grade 5	-0.28	-0.34	-0.06***
Math, Grade 3	-0.29	-0.30	-0.01
Math, Grade 5	-0.21	-0.28	-0.07***
<u>student characteristics (Grade 5)</u>			
3 or more siblings	0.41	0.42	0.01
speaks English at home	0.09	0.09	0.00
mother completed primary	0.22	0.17	-0.05***
father completed primary	0.32	0.22	-0.09***
no books at home	0.28	0.33	0.04***
repeated at least one grade	0.42	0.45	0.03**
attends multi-grade classroom	0.44	0.52	0.08***
absent 6 or more times last month	0.10	0.10	0.01
travels at least 1 hour to school	0.25	0.28	0.04***

Table shows school-level means for enrollment or student-level means for performance and characteristics. Final column reports difference in means (hardship minus non-hardship), with significance levels 10%, 5% and 1% indicated by 1, 2, and 3 stars, respectively. Student performance is z-score, based on standardized distribution of national scores. Sample limited to government-run Lower Basic and Basic Cycle schools in Regions 3-6 only. Enrollment from Basic Cycle schools counts only students in grades 1-6. Female enrollment percentage and pupil-teacher ratios weight by student enrollment. Source: EMIS and NAT results, 2012.

Table 2: Regression Discontinuity, Stage 1

	full sample						discontinuity samples		excludes schools	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	1-5km	2-4km	opened since 2005
Distance ≥ 3km	0.44	0.43	0.44	0.45	0.44	0.43	0.56	0.69	1.04	0.44
	(0.15)***	(0.14)***	(0.14)***	(0.13)***	(0.13)***	(0.14)***	(0.32)*	(0.24)***	(0.52)*	(0.14)***
Observations	244	244	244	244	244	244	244	69	33	228
R-squared	0.87	0.87	0.88	0.88	0.89	0.88	0.9	0.66	0.53	0.89
F-stat on distance cutoff	8.9	8.8	9.8	11.6	11.4	9.8	3.0	8.1	3.9	9.7
Mean of dependent variable	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.56	0.68	0.47
Polynomial order	4	5	6	6	6	7	7	6	6	6
Region and school type fixed effects				x	x	x	x	x	x	x
Controls included					x					
Polynomial varies on either side of cutoff							x			

Regressions are linear probability models of school hardship status on travel distance from school to main road. I.e., regressions are Stage 1 of fuzzy RD design for treatment of hardship allowance. Robust standard errors in parentheses, clustered by cluster (sub-regional school administrative unit, of which there are 33 in sample). * significant at 10%; ** significant at 5%; *** significant at 1%. Sample is government-run Lower Basic and Basic Cycle Schools, Regions 3-6, in 2012. Sample excludes schools whose map location does not match district reported in EMIS. All regressions weighted by students enrolled in grades 1-6 at school. All regressions include polynomial in distance of indicated order. Regression controls included where indicated, from 2003 Census data on nearest settlement to school: log population, employment/population ages 18+, percent with access to electricity, percent illiterate, percent Muslim, percent of Mandinka, Fula, and Wolof ethnicities.

Table 3: Student results

	<u>English</u>						<u>Math</u>					
	<u>Grade 3</u>			<u>Grade 5</u>			<u>Grade 3</u>			<u>Grade 5</u>		
	all	boys	girls	all	boys	girls	all	boys	girls	all	boys	girls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
<u>Panel A: reduced form</u>												
Distance ≥ 3km	0.07	0.07	0.08	0.23	0.21	0.25	0.02	0.01	0.02	0.20	0.16	0.22
	(0.09)	(0.12)	(0.09)	(0.17)	(0.22)	(0.14)*	(0.12)	(0.13)	(0.12)	(0.16)	(0.20)	(0.15)
R-squared	0.04	0.03	0.04	0.04	0.02	0.05	0.04	0.04	0.04	0.05	0.03	0.05
<u>Panel B: Instrumental variables</u>												
hardship allowance	0.15	0.15	0.15	0.65	0.61	0.66	0.04	0.03	0.04	0.55	0.46	0.60
	(0.20)	(0.26)	(0.19)	(0.56)	(0.70)	(0.48)	(0.24)	(0.30)	(0.23)	(0.50)	(0.62)	(0.48)
Observations	7,587	3,513	4,074	6,558	3,066	3,492	7,591	3,515	4,076	6,550	3,059	3,491
1st stage F-stat	12.9	9.5	15.1	6.7	5.8	7.0	12.9	9.4	15.2	6.7	5.7	7.0
Mean of dependent variable	-0.35	-0.25	-0.44	-0.31	-0.21	-0.40	-0.30	-0.21	-0.37	-0.25	-0.15	-0.33

Table shows results of regressions of student outcomes on distance threshold or hardship allowance receipt, as indicated. Panel B uses distance threshold to instrument for hardship allowance. Dependent variables are student z-scores from National Assessment Test (NAT), 2012. z-score calculated relative to national average, including students in Regions 1-2. All regressions include a 6th-order polynomial in distance from school to main road, region and school type fixed effects, student's age and age squared, where age is exact age on Jan. 1, 2012 based on date of birth. Robust standard errors in parentheses, clustered by school. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4: Student performance and heterogeneity

	<u>English</u>			<u>Math</u>		
	<u>all</u>	<u>boys</u>	<u>girls</u>	<u>all</u>	<u>boys</u>	<u>girls</u>
	(1)	(2)	(3)	(4)	(5)	(6)
Distance \geq 3km	-0.07 (1.40)	-0.29 (1.69)	0.19 (1.22)	-0.89 (1.18)	-0.67 (1.33)	-1.11 (1.23)
<u>Distance \geq 3km interacted with:</u>						
QT%	0.44 (0.77)	0.17 (0.94)	0.67 (0.66)	0.35 (0.71)	-0.47 (0.81)	1.08 (0.71)
HS%	-0.03 (3.83)	0.36 (4.70)	-0.64 (3.43)	2.18 (3.61)	2.62 (4.00)	1.76 (3.91)
z(2011)	0.91 (0.68)	1.05 (0.83)	0.83 (0.60)	0.86 (0.65)	0.93 (0.74)	0.82 (0.64)
socioeconomic status index	0.55 (0.41)	1.16 (0.56)**	0.05 (0.45)	0.85 (0.41)**	1.19 (0.58)**	0.59 (0.46)
Observations	6,379	2,986	3,393	6,367	2,979	3,388
R-squared	0.13	0.11	0.15	0.11	0.11	0.12
Mean of dependent variable	-0.32	-0.23	-0.41	-0.25	-0.16	-0.33
p-value on interaction terms	0.14	0.08	0.21	0.08	0.11	0.15

Table shows results of regressions of student outcomes on distance threshold and interactions with observable characteristics. Dependent variables are student z-scores from National Assessment Test (NAT), Grade 5, 2012. z-score calculated relative to national average, including students in Regions 1-2. QT% is percent of school's teachers who are qualified. HS% is hardship salary premium (30% for Regions 3-4, 35% for Region 5, 40% for Region 6). QT% and HS% measured on (0,1) interval. z(2011) is school's average z-score on math and English in 2011. For definition of socioeconomic status index, see notes to Table 4. All regressions include a 6th-order polynomial in distance from school to main road; interactions between indicated observable characteristics and polynomial terms and main effect of observable characteristics; region and school type fixed effects; female dummy; student's age and age squared, where age is exact age on Jan. 1, 2012 based on date of birth. Robust standard errors in parentheses, clustered by school. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 5: Grade 5 results, by student socioeconomic status and school quality

	English						Math					
	at or below median			above median			at or below median			above median		
	all	boys	girls	all	boys	girls	all	boys	girls	all	boys	girls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: socioeconomic status												
Distance \geq 3km	0.09 (0.17)	-0.04 (0.23)	0.18 (0.13)	0.38 (0.20)*	0.45 (0.26)*	0.33 (0.18)*	0.01 (0.16)	-0.09 (0.20)	0.10 (0.16)	0.41 (0.19)**	0.44 (0.24)*	0.40 (0.17)**
Observations	3,548	1,630	1,918	2,831	1,356	1,475	3,543	1,626	1,917	2,824	1,353	1,471
R-squared	0.04	0.03	0.05	0.06	0.04	0.09	0.05	0.04	0.05	0.06	0.04	0.07
Mean of dependent variable	-0.34	-0.23	-0.43	-0.31	-0.22	-0.38	-0.25	-0.14	-0.33	-0.25	-0.17	-0.33
Panel B: school quality												
Distance \geq 3km	-0.05 (0.13)	-0.07 (0.18)	-0.03 (0.12)	0.24 (0.28)	0.15 (0.36)	0.29 (0.23)	0.01 (0.15)	-0.05 (0.20)	0.05 (0.16)	0.15 (0.24)	0.06 (0.29)	0.21 (0.23)
Observations	2,787	1,298	1,489	3,771	1,768	2,003	2,782	1,294	1,488	3,768	1,765	2,003
R-squared	0.05	0.04	0.05	0.03	0.01	0.06	0.06	0.06	0.04	0.04	0.01	0.05
Mean of dependent variable	-0.47	-0.37	-0.56	-0.19	-0.09	-0.28	-0.41	-0.32	-0.49	-0.12	-0.03	-0.21

Table shows results of regressions of student outcomes on distance threshold. Dependent variables are student z-scores from National Assessment Test (NAT), Grade 5, 2012. Sample split according to student's position in distribution of socioeconomic status index (Panel A) or school's position in quality distribution. Socioeconomic status index calculated as proportion proportion equal to 1 among the following indicators: two siblings or less, speak English at home, mother completed primary, father completed primary, more than 10 books at home, school travel time less than an hour, help on schoolwork available at home, attended nursery. School quality determined by average English and math z-score in Grades 3 and 5, 2011. All regressions include a 6th-order polynomial in distance from school to main road, region and school type fixed effects, student's age and age squared, where age is exact age on Jan. 1, 2012 based on date of birth. Robust standard errors in parentheses, clustered by school. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A1: Regression Discontinuity, Stage 1 robustness checks

	<u>straight-line distance</u>			<u>pre-treatment village characteristics</u>					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	hardship allowance	schooling above primary	access to piped water	access to flush toilet	marriage rate	polygamous marriage	born outside village	access to television	children born
Distance \geq 3km	0.25 (0.14)*	0.03 (0.05)	0.02 (0.02)	-0.01 (0.01)	0.00 (0.02)	0.04 (0.05)	0.03 (0.03)	-0.01 (0.09)	-0.03 (0.24)
Observations	244	244	244	244	244	244	244	244	244
R-squared	0.87	0.2	0.17	0.13	0.13	0.16	0.32	0.23	0.17
F-stat on distance cutoff	3.2	0.3	1.3	0.4	0.0	0.5	1.5	0.0	0.0
Mean of dependent variable	0.47	0.33	0.04	0.01	0.54	0.46	0.15	0.55	3.01

Table shows results of regressions of school hardship status (column 1) or village characteristics from 2003 Census on indicator for school distance more than 3km from main road. All regressions follow preferred Stage 1 specification and include 6th-order polynomial in distance, and region and school type fixed effects. Sample is government-run Lower Basic and Basic Cycle Schools, Regions 3-6, in 2012. Sample excludes schools whose map location does not match district reported in EMIS. Schools matched to nearest village to assign village characteristics. Column (1) replaces distance along road with straight-line distance. Column (9) dependent variable is average number of children born to women ages 12 and older. All regressions weighted by students enrolled in grades 1-6 at school. Robust standard errors in parentheses, clustered by cluster (sub-regional school administrative unit, of which there are 33 in sample). * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A2: Student enrollment

	<u>Enrollment</u>				<u>Test-takers</u>			
	all (1)	boys (2)	girls (3)	% female (4)	all (5)	boys (6)	girls (7)	% enrolled (8)
<u>Panel A: reduced form</u>								
Distance \geq 3km	27.1 (33.8)	11.0 (17.1)	16.0 (17.8)	0.01 (0.02)	14.3 (9.1)	6.7 (4.9)	7.6 (4.8)	0.02 (0.03)
R-squared	0.18	0.19	0.17	0.23	0.2	0.2	0.2	0.1
<u>Panel B: Instrumental variables</u>								
hardship allowance	72.7 (101.7)	29.6 (49.8)	43.1 (54.7)	0.02 (0.05)	38.3 (31.2)	18.0 (15.6)	20.3 (16.6)	0.05 (0.06)
Observations	244	244	244	244	244	244	244	243
1st stage F-stat	7.6	7.6	7.6	11.6	7.6	7.6	7.6	11.6
Mean of dependent variable	231.2	110.8	120.3	0.52	58.0	27.0	31.0	0.86

Table shows results of regressions of school outcomes on distance threshold or hardship allowance receipt, as indicated. Panel B uses distance threshold to instrument for hardship allowance. Enrollment includes only grades 1-6 for Basic Cycle Schools. Test-takers refers to National Assessment Test (NAT), grades 3 and 5. All regressions include a 6th-order polynomial in distance from school to main road, and region and school type fixed effects. Regressions for female enrollment percentage and percent enrolled who take NAT are weighted by school enrollment; all other regressions unweighted. Robust standard errors in parentheses, clustered by cluster (sub-regional school administrative unit, of which there are 33 in sample). * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A3: Student composition and school quality, Grade 5

	<u>Socioeconomic status</u>			<u>Student effort</u>			<u>School quality</u>		
	all (1)	boys (2)	girls (3)	all (4)	boys (5)	girls (6)	all (7)	boys (8)	girls (9)
Distance ≥ 3km	-0.002 (0.019)	0.007 (0.024)	-0.009 (0.020)	-0.023 (0.040)	0.007 (0.043)	-0.051 (0.044)	-0.029 (0.038)	-0.068 (0.045)	0.002 (0.042)
Observations	6,379	2,986	3,393	6,371	2,984	3,387	7,001	3,271	3,730
R-squared	0.02	0.03	0.03	0.02	0.02	0.02	0.01	0.01	0.01
Mean of dependent variable	0.41	0.41	0.41	0.66	0.67	0.66	0.62	0.62	0.61

Table shows results of regressions of outcome indices on distance threshold. Each index is measured as the proportion of affirmative responses to a collection of questions in National Assessment Test Grade 5 questionnaire, 2012. Socioeconomic status indicators: two siblings or less, speak English at home, mother completed primary, father completed primary, more than 10 books at home, school travel time less than an hour, help on schoolwork available at home, attended nursery. Student effort indicators: never repeated a grade, absent less than 6 days last month, attend extra class after school, receive private tutoring. School quality indicators: not in multi-grade classroom, receive food at school, English class taught in English, math class taught in English, English textbook, math textbook. All regressions include a 6th-order polynomial in distance from school to main road, region and school type fixed effects, dummy for female, and student's age and age squared. Robust standard errors in parentheses, clustered by school. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A4: Student results, 2008-2012

	<u>English</u>						<u>Math</u>					
	<u>Grade 3</u>			<u>Grade 5</u>			<u>Grade 3</u>			<u>Grade 5</u>		
	all	boys	girls	all	boys	girls	all	boys	girls	all	boys	girls
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	
<u>Panel A: reduced form</u>												
distance ≥ 3km	0.10	0.06	0.13	0.09	0.10	0.07	0.06	0.06	0.06	-0.01	0.01	-0.03
	(0.07)	(0.07)	(0.07)*	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)
R-squared	0.02	0.01	0.03	0.02	0.01	0.02	0.02	0.02	0.03	0.02	0.02	0.03
<u>Panel B: Instrumental variables</u>												
hardship allowance	0.26	0.15	0.34	0.25	0.28	0.21	0.16	0.14	0.16	-0.02	0.04	-0.07
	(0.19)	(0.19)	(0.20)*	(0.19)	(0.20)	(0.20)	(0.18)	(0.19)	(0.19)	(0.21)	(0.22)	(0.23)
Observations	35,650	16,779	18,871	31,330	14,738	16,592	35,553	16,748	18,805	31,381	14,764	16,617
1st stage F-stat	35.6	34.1	34.0	32.5	32.6	29.4	35.6	34.2	33.9	30.0	29.3	27.9
Mean of dependent variable	-0.27	-0.22	-0.31	-0.30	-0.24	-0.35	-0.26	-0.22	-0.30	-0.18	-0.14	-0.22

Table shows results of regressions of student outcomes on distance threshold or hardship allowance receipt, as indicated. Panel B uses distance threshold to instrument for hardship allowance. Dependent variables are student z-scores from National Assessment Test (NAT), 2008-2012. z-score calculated relative to national average, including students in Regions 1-2. All regressions include a 6th-order polynomial in distance from school to main road; year, region, and school type fixed effects; student's age and age squared, where age is exact age on Jan. 1 of test year based on date of birth. Robust standard errors in parentheses, clustered by school. * significant at 10%; ** significant at 5%; *** significant at 1%.