JOURNAL OF DEVELOPMENT EFFECTIVENESS, 2018
https://doi.org/10.1080/19439342.2018.1452778

Routledge
Taylor & Francis Group

ARTICLE

Check for updates

Teacher pay and student performance: evidence from the Gambian hardship allowance

Todd Pugatch[a,b] and Elizabeth Schroeder[a]

[a]Department of Economics, Oregon State University, Corvallis, OR, USA; [b]IZA, Bonn, Germany

ABSTRACT
More than two dozen developing countries have implemented policies to increase teacher compensation in rural schools. 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 test scores. However, we find suggestive evidence that student performance improved at the top of the test score distribution and fell at the bottom. 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.

ARTICLE HISTORY
Received 25 April 2017
Accepted 10 March 2018

KEYWORDS
Teacher compensation; rural schools; Gambia; program evaluation; regression discontinuity

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, that is, regardless of qualifications or performance. The policy affects more than 26,000 students and 1,100 teachers. This paper evaluates the effect of the programme on student achievement.

An arbitrarily assigned 3-km cut-off determining hardship status provides exogenous variation to identify the programme's impact. We evaluate it using the universe of administrative records from an achievement test administered to all Gambian students in grades 3 and 5. Despite the sizeable salary premium and resulting increases in teacher qualifications, we find no effects of the hardship allowance on average student performance. These null results persist when pooling the sample across genders, grades, or time periods. 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 enrolment, effort, student composition across schools, 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 socio-economically advantaged students

CONTACT Elizabeth Schroeder liz.schroeder@oregonstate.edu
Supplemental data for this article can be accessed here.

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.

Efforts to upgrade teacher quality such as the Gambian hardship allowance are gaining popularity in developing countries.[1] Among the Sustainable Development Goals (SDGs) adopted by the United Nations (UN) in 2016 is to 'substantially increase the supply of qualified [certified] teachers' by 2030, with an accompanying target of raising the percentage of trained teachers in primary and secondary schools (United Nations 2016, 20). Because rural schools are more likely than urban schools to have a low percentage of qualified teachers (Mulkeen and Chen 2008), policymakers frequently focus their efforts on these hard-to-staff areas. We found documentation of 40 policies to recruit teachers to rural areas in 29 developing countries, with increased salary the modal incentive; a full list appears in supplemental appendix Table SA1. Such efforts are likely to intensify in the coming years under the adoption of the UN SDGs.

In addition to recruiting new teachers, salary increases aim to improve the effectiveness of teachers already present in hardship schools. Even when pay and performance are not linked formally in the employment contract, as is the case in The Gambia, increased pay could raise productivity through an efficiency wage effect. Teachers might respond to social norms of gift exchange by reciprocating the salary premium with increased effort (Akerlof 1982), or they might shirk less because of an increased opportunity cost of unemployment or transfer to a non-hardship school (Shapiro and Stiglitz 1984). The latter channel is likely to be particularly important for civil servants such as teachers, who tend to be unionised and enjoy substantial labour protection, and for whom salary increases may substitute for lack of credible threats for dismissal or effective monitoring of effort. Policy organisations implicitly endorse the efficiency wage view in calling for unconditional pay increases for teachers. A recent UNICEF report stated, 'Pay levels of frontline school teachers and health workers are crucial to outcomes for children' (UNICEF 2010, 1).

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. By contrast, we isolate exogenous and large differences in teacher salary across the 3-km threshold determining the hardship allowance.

This paper contributes to the literature in three ways. First, among the many policies we found that use salary increases to recruit teachers to remote schools, none (to our knowledge) has been evaluated using a credible identification strategy. We show, both here and in a companion study (Pugatch and Schroeder 2014), that the hardship allowance increased the flow of qualified teachers to remote schools and reduced qualified teacher-pupil ratios. In this paper, we explore whether these changes led to gains for students, an essential question for schools that are hard to staff (McEwan 1999; Kremer et al. 2005; Glewwe and Kremer 2006; Urquiola 2006).

Second, we contribute to the growing literature on the link between teacher salary and student performance. Although much recent attention has focused on the effectiveness of pay-for-performance teacher contracts (Lavy 2002; Lavy 2009; Glewwe, Ilias, and Kremer 2010; Muralidharan and Sundararaman 2011; Woessmann 2011; Fryer 2013), such programmes are rare. Unconditional salary increases like the one we analyse are easier to implement politically and therefore arguably more reflective of the policy environment in most countries.

The closest counterparts to our study are Urquiola and Vegas (2005), Greaves and Sibieta (2014), and De Ree et al. (2017), which examine arbitrary and unconditional differences in teacher salary in Bolivia, England, and Indonesia, respectively. Like our work, none of these studies find evidence of increased student performance in response to higher salaries. In addition to contributing an evaluation in an African country, our study complements theirs by focusing on long-run effects, 8 years after introduction of the policy.[2] Although the efficiency wage effect on initial recipients is more likely to have faded out in our case, the longer time horizon allows for new teachers to enter the profession, opening another potential channel for improvement if these new entrants are more capable or motivated. Together, the three studies help to build an externally valid case for the minimal benefits that result from a costly and popular policy.

Third, because the Gambian hardship allowance succeeded in upgrading 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. 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 2014b). 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; Kirabo, 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 literature review of the effect of school inputs on education outcomes in the developing world, Glewwe et al. (2013) 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 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,' that is, 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 (2013), Glewwe et al. (2013), McEwan (2015), and Ganimian and Murnane (2016) survey the literature, and find that the highest quality studies, which rely on experimental variation in the presence of contract teachers from randomised control trials, show positive effects of contract teachers on student outcomes. An exemplar of this 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 (embodied in the Sustainable Development Goal mentioned previously) that increases in qualified teachers should improve school outcomes. Because of their ties to the local community or perceived need to distinguish themselves, unqualified teachers may be better positioned to help students learn.[4]

What explains our null results for average student performance? They are particularly surprising given that the policy increased teaching inputs as intended. The weak link between formal qualifications and effective teaching has already been mentioned. Another possibility is that salary increases fail to motivate teachers in the absence of credible sanctions for poor performance from parents or administrators, effective competition among schools, or contractual links with student performance, all of which would be consistent with both our results and those previously cited by

Urquiola and Vegas (2005) and De Ree et al. (2017). Finally, it is possible that the policy worked poorly for some students but well for others, such that the average null effect hides important heterogeneity. Indeed, the only persistent patterns we find in the data are gains for students at the top of the test score and socio-economic status distributions.

In the next section, we describe the hardship allowance and the Gambian education system. Section III presents a theoretical framework, a corresponding empirical specification, and a discussion of parameter interpretation. Section IV describes the data, Section V presents results, and Section VI concludes the study.

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 six numbered administrative regions. Regions 1 and 2 surround the capital city, Banjul, and encompass the urban areas containing the most economic opportunities as well as better-performing schools. The remaining regions are more economically disadvantaged, and are numbered in increasing distance from the capital, with Region 6 being the most remote from the urban areas. The country's main highways run east-west on either bank of the Gambia River dividing the country, and extend through Region 6. Thus, within each region, there are schools clustered along the main road, as well as schools that are more than 35 or 40 km away from the main road (except for Region 3, where no school is more than 23 km from the main road).

The hardship allowance policy, in place since 2005, was designed to compensate teachers in schools that are far from a main road. The policy 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 km from the main highways. Figure 1 depicts the policy. The hardship allowance is 30% of salary in Regions 3 and 4, 35% in Regions 5, and 40% in Region 6, to account for the added burdens of working in schools that are more removed from the capital. Teachers receive the allowance regardless of whether they have completed the 3-year certification programme offered by Gambia College, though qualified (certified) teachers earn a base salary 2.5 times greater than unqualified teachers. The programme is financed by the World Bank and costs an average of US$23 per teacher per month, or US$350,000 per year. The allowance is large in proportion to teacher salaries, which average US$67 per month without the allowance, or to Gambian per capita GDP of US$43 per month.

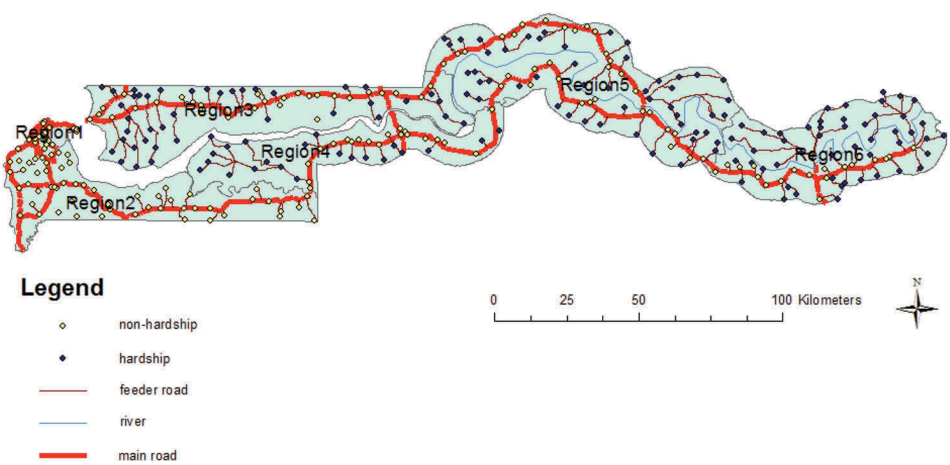

Figure 1. Gambia Hardship Allowance, (2012).
Source: Pugatch and Schroeder (2014), Figure 1

The most common means for students to travel to school are walking or bicycling; in a survey of fifth graders (discussed in detail below), 68% of students in non-hardship areas and 74% of students in hardship areas reported walking or bicycling to school. The most common alternatives are getting a ride from friends or family or waiting by the road for a ride, with a very small number of students riding a bus or other paid vehicle. Travel times are correspondingly long. In non-hardship areas, 50% of students take more than half an hour to get to school, and 28% take an hour or more. In hardship areas, the respective percentages are 53 and 33. These long travel times likely prevent most students from attending schools other than their closest options; since we are unaware of any enforced regulations requiring students to attend their assigned schools, however, we explore potential sorting as a result of the policy below. Travel options are similar for teachers, although some of the most remote rural schools provide teachers with living quarters near the schools.

Teachers may request assignment to particular regions or schools. Postings are not entirely voluntary, however, as the Ministry's regional offices review teacher allocations to schools each year and make the final assignments. The hardship allowance could be expected to both attract new teachers to the profession, and to increase the willingness of existing teachers to work in hardship schools. Evidence suggests that, due to the pre-existing disparity between the numbers of qualified teachers in hardship and non-hardship areas, the Ministry granted requests of teachers who were willing to locate in hardship areas. 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. Although we lack data on teacher tenure and movement between schools, a back-of-the-envelope calculation finds that the increase in qualified teacher labour supply due to the hardship allowance was sufficient to account for the estimated increase in qualified teachers in hardship areas.

This paper addresses whether these increases in observable teacher 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, Science, and Integrated Studies (local culture), 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 standardised 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

A. Theoretical framework

The goal of the hardship allowance was to improve the quality of remotely located schools. Ideally, the policy would eliminate the gap in school quality between hardship and non-hardship schools. Figure 2 provides a stylised depiction of the relationship between school quality Q and distance from the main road d.[5] The negative slope of the solid line indicates that average school quality declines smoothly as schools become more remote.

Let T be an indicator for whether a school lies above the 3-km threshold distance from the main road. Adopting potential outcomes notation, let Q_1 be school quality with the hardship allowance and Q_0 be school quality without the allowance. In principle, each school has values of Q_1 and Q_0, regardless of actual treatment status. In practice, only one of these values is observed because a

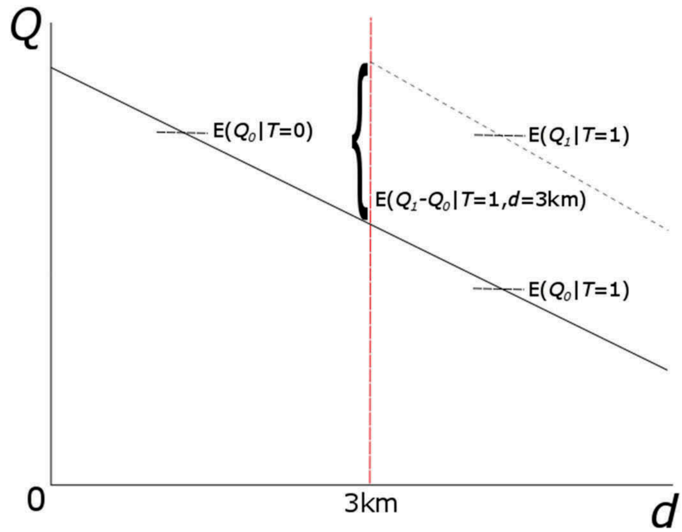

Figure 2. Theoretical effect of hardship allowance.

school can be in only one treatment state. In the absence of the policy, there is a gap in average school quality between schools on either side of the 3-km cut-off, given by the differences between $E(Q_0|T = 0)$ and $E(Q_0|T = 1)$. The goal of the hardship allowance is to remove this gap by bringing average quality in remote schools up to $E(Q_1|T = 1)$.

As researchers, we would like to estimate the average treatment effect of the hardship allowance on hardship schools (the ATT), given by the difference between $E(Q_1|T = 1)$ and $E(Q_0|T = 1)$. We are unable to estimate this parameter because schools are observed in only one treatment state. However, if the distribution of school characteristics other than treatment is continuous across the 3km threshold, we can approximate $E(Q_0|T = 1, d = 3\text{km})$ using the non-hardship schools below the cut-off. The resulting parameter, $E(Q_1\text{-}Q_0|T = 1, d = 3\text{km})$, is the causal effect of the hardship allowance for schools at the cut-off. As shown in the figure, we expect this parameter to be positive, because schools close to the cut-off should have no systematic differences other than hardship status. Note that this positive treatment effect at the cut-off is expected even if a comparison of all hardship and non-hardship schools reveals no average quality difference, so that $E(Q_0|T = 0)$- $E(Q_1|T = 1) = 0$, as would be the case if the policy succeeds in eliminating the quality gap.

B. Empirical specification

We estimate the effect of the hardship allowance with a regression discontinuity design, using as identifying variation the policy granting hardship status to schools that were at least 3 km from a main road. 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.

Because hardship status is not perfectly predicted by the 3-km cut-off, we use a fuzzy regression discontinuity design. 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*:

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

where *D* is an indicator for hardship status and *T* is an indicator for the school being located at least 3 km from the main road. We condition on a flexible function of the road distance, *f(d)*, which we model as a polynomial, as well as a vector of student characteristics X, including age, age squared, and gender. 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 standardised test score in English or math, denoted y. The reduced-form specification measures the effect of crossing the distance threshold on the outcome:

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

where all explanatory variables are as in (1). Here, ρ measures the intent to treat (ITT).

We also estimate the local average treatment effect (LATE) using instrumental variables (IV) estimation:

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

In Equation (3), we instrument for treatment status D using the threshold indicator T. The IV estimate of θ has the interpretation of the average treatment effect for schools induced into treatment by being just across the threshold. 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.

Consistent estimation of both the ITT and the LATE requires that conditional on $f(d)$, the threshold indicator T is uncorrelated with the regression error terms. In extensive discussions with the Ministry of Basic and Secondary Education, we verified that no other policies are determined by the 3-km threshold (or any other distance threshold, for that matter). Even in the absence of confounding policies, however, the use of a running variable based on spatial location poses additional challenges. Although the policy rule allows for a comparison between schools of different treatment status close to the cut-off, their locations in relation to one another, or to geographic features not captured by distance from the main road, could bias our estimates of the treatment effect.

For instance, if hardship and non-hardship schools close to the cut-off are also close to each other, teachers and students might sort non-randomly into each type of school. Competition between schools could also affect outcomes directly even in the absence of deliberate sorting. Similarly, the proximity of other primary schools, regardless of their treatment status, to schools near the cut-off could influence outcomes. Although these competitive forces might properly be considered part of the treatment effect in general equilibrium, they nonetheless would violate the stable unit treatment value assumption (SUTVA) required for causal inference. In other words, the characteristics of the local education market matter.

Another set of concerns relates to geographic endowments. Suppose, for instance, that most non-hardship schools close to the cut-off are in the western half of Region 3, near the capital, while most hardship schools close to the cut-off are in the eastern half of Region 6, the most remote region of the country. Or suppose that non-hardship schools close to the cut-off tend to be near the Gambia River, whereas hardship schools close to the cut-off tend to be far from the river but near the Senegalese border. Then comparing schools across the cut-off would confound the treatment effect with these geographic (dis)advantages. Although visual inspection of the map in Figure 1 does not suggest that these situations exist in the data, they are nonetheless appropriate to consider.

To address the possibility of spatial confounders, we decompose the error term in Equations (1)–(3) into three components:

$$\varepsilon_{is} = g(M_s) + h(E_s) + \upsilon_{is} \tag{4}$$

where M is a vector of education market characteristics, E is a vector of geographic endowments other than distance from the main road, and υ is an idiosyncratic error. The functions g and h allow for flexible functional forms. To capture market characteristics M, we include distance to the nearest school of opposite treatment status, the number of schools in the cluster, and a dummy for Basic Cycle School (grades 1–9). To capture geographic endowments E, we include distance to the capital Banjul, distance to the nearest major body of water (river or coast), and region dummies. Each variable in M and E enters as a quadratic (except the dummies, which enter separately). Our modified identifying assumption is that $f(d)$, $g(M)$, and $h(E)$ capture spatial correlation in the error term, thereby isolating exogenous variation in the distance cut-off T.[6] By including smooth functions of several geographic features, our approach resembles the multidimensional spatial discontinuity method used by Dell (2010).

C. *Parameter interpretation*

Before presenting results, it is worth reflecting further on how to interpret the parameter estimates we obtain. Consider the following education production function:

$$y = Q(T, X, C) \tag{5}$$

where the school quality function Q maps teaching inputs T, non-teaching school inputs X, and child inputs C into a student's test score y. Because the hardship allowance alters teaching inputs, our interest is in how changes in T alter y.[7] Following Glewwe and Muralidharan (2016); hereafter GM), consider the following parameters:

$$\frac{\partial y}{\partial T} = \frac{\partial Q}{\partial T} \tag{6}$$

$$\frac{dy}{dT} = \frac{\partial Q}{\partial T} + \frac{\partial Q}{\partial X}\frac{dX}{dT} + \frac{\partial Q}{\partial C}\frac{dC}{dT} \tag{7}$$

The parameter in (6), a partial derivative, captures how test scores change due to a change in teaching inputs, holding all other school and student inputs constant. GM call this the production function parameter, because it captures changes in output when nothing else about the education production process changes. The parameter in (7), a total derivative, captures how test scores change due to a change in teaching inputs, allowing school and student inputs to adjust endogenously. GM call this the policy parameter, because it captures the effect of a change in inputs inclusive of all behavioural changes induced by the policy. Note that these parameters are equal when school and student inputs do not respond to the change in teaching inputs (so that $\frac{dX}{dT} = \frac{dC}{dT} = 0$).

As GM argue, both parameters are relevant. The production function parameter helps to understand the nature of education production, while the policy parameter helps to understand the effects of policy changes. It is therefore important to consider which parameter one estimates in an impact evaluation. In our context, we have already demonstrated that teaching inputs changed in response to the policy (Pugatch and Schroeder 2014); it is reasonable to think that non-teaching inputs, student effort, and student composition changed as well.[8] Additionally, the long time horizon since the policy was enacted (most of our analysis uses data from 2012, while the policy began in 2005) suggests adjustments along several margins have already occurred in response to the policy.

Accordingly, we interpret all estimates as the policy parameter. As the name implies, this parameter is the most relevant for policy design, as it captures the many adjustments likely to occur when enacting a policy. The relevant counterfactual to the policy is therefore an *ex ante* equivalent school in which the policy was not enacted. Our empirical analysis thus begins by estimating the overall effect of the policy on student outcomes, that is, the left-hand side of (7). In addition to average student outcomes, we estimate how the policy affected outcomes throughout the test score distribution. Later, we will attempt to decompose this overall effect into mechanisms that correspond to the terms in the right-hand side of (7).

The effect of teaching inputs, $\frac{\partial Q}{\partial T}$, can operate through two channels[9]:

(1) *Improvement in teacher quality*. An influx of qualified teachers could raise the average skills of teachers in hardship schools.
(2) *Enhanced teacher motivation*, that is, 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.[10]

The remaining mechanisms, corresponding to $\frac{\partial Q}{\partial X}\frac{dX}{dT}$ and $\frac{\partial Q}{\partial C}\frac{dC}{dT}$, are policy-induced changes in:

(3) *Student composition*
(4) *School quality*

We will test for the presence of these mechanisms by interacting with the distance threshold indicator with proxies for each. We also test for changes in school inputs and student composition, to explore how closely our estimates of the policy parameter might align with the production function parameter.

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-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 therefore should reflect any lasting effects of the policy. We also prefer 2012 data because all fifth graders in that year completed a demographic questionnaire, allowing us to test for student sorting in response to the policy and for heterogeneous treatment effects. We later check if results change when using data for earlier years.

In addition to standardised 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 data on hardship status and enrolment. 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. 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, leaving a final sample of 244 schools. Figure 1 shows a map of this dataset (schools in Regions 1–2 are included on the map for illustration only).

Table 1. School & student characteristics, by hardship status.

	Non-hardship (1)	Hardship (2)	Difference (2) − (1)
Number of schools	96	148	52
Number of students	29,723	26,682	−3,041
Enrolment			
Total	309.6	180.3	−129.3***
Male	148.0	86.7	−61.3***
Female	161.6	93.6	−68***
% Female	0.52	0.52	0.00
Teachers			
Total	12.6	7.8	−4.8***
Qualified	7.6	5.1	−2.5***
% Qualified	0.61	0.65	0.05**
Pupil-teacher ratio	29.5	26.9	−2.6
Location			
Distance from main road (km)	0.9	11.8	10.9***
Distance to Banjul	143.3	163.1	19.8*
Distance to major body of water	3.9	5.0	1.1***
Distance to nearest school of Opposite treatment status	5.6	8.6	3.1***
Student performance			
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***
Student characteristics (grade 5)			
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 multigrade classroom	0.44	0.52	0.08***
Absent 6 or more times last month	0.10	0.10	0.01
Travels at least 1 h to school	0.25	0.28	0.04***

Table shows school-level means for enrolment 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 standardised distribution of national scores. Sample limited to government-run Lower Basic and Basic Cycle schools in Regions 3–6 only. Enrolment from Basic Cycle schools counts only students in grades 1–6. Per cent qualified teachers weighted by number of teachers at school. Female enrolment percentage and pupil-teacher ratios weight by student enrolment. Distance from main road is travel distance; all other distances are Euclidean. All distances in kilometres Source: EMIS and NAT results, 2012.

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. There are several notable differences between the two groups. Hardship schools have nearly 130 fewer students and 5 fewer teachers on average. Nonetheless, hardship schools have a higher proportion of qualified teachers and a lower pupil-teacher ratio, consistent with a reallocation of teaching inputs in response to the policy. Hardship schools are significantly more remote, located on average 11.8 km from a main road, compared to 0.9 km for non-hardship schools. Large average distances between schools of opposite treatment status – 5.6 km for non-hardship schools and 8.6 km for hardship schools – likely make it difficult for students to switch without a costly residential move. As our sample excludes the more urban Regions 1 and 2, students in both hardship and non-hardship schools have negative average test 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.

Further analysis of the relationship between teacher qualifications and student performance within our sample reveals no statistically significant association between the proportion of qualified teachers at a school and test scores. The finding persists regardless of whether we examine the unconditional correlation or when conditioning on the same controls as our main specification in Equation (2) (results in Table SA2 of the supplemental appendix). These findings align with the widespread conclusion in the literature that teacher qualifications fail to predict student performance (Darling-Hammond, Berry, and Thoreson 2001; Hanushek 2003; Rivkin, Hanushek, and Kain 2005; Kirabo, Rockoff, and Staiger 2014; Pugatch 2017). Taken at face value, they suggest that policies to increase the present of qualified teachers will fail to improve student performance. However, these findings are correlations, not causal impact estimates, whereas the causal effect is the relevant policy question. Moreover, the hardship allowance could increase student performance through an efficiency wage effect.

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 socio-economic 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 programme's impact. Moreover, the likelihood that students in hardship and non-hardship schools also differ in unobservable characteristics makes it essential to use an identification strategy, such as regression discontinuity, that accounts for both observable and unobservable differences.

V. Results

A. *First stage*

As discussed in Section III, we use a fuzzy regression discontinuity design because treatment status, given by government assignment of a hardship designation, does not align perfectly with the 3-km distance threshold.[11] 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 cut-off.[12] While measurement error in the running variable would bias our results towards a weaker first stage,[13] 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 (1), where we have collapsed the data by school because this is the relevant level of variation in hardship status.[14] In column (1), we include a quartic in distance from the main road and other spatial controls as in Equation (4). The distance threshold coefficient of 0.43 indicates a 43-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.[15] 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

Table 2. Regression discontinuity, stage 1.

	Full sample							Discontinuity samples		Excludes schools
								1–5km	2–4km	Opened since 2005
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Distance ≥3 km	0.43	0.42	0.43	0.44	0.44	0.43	0.61	0.57	0.70	0.43
	(0.15)***	(0.14)***	(0.14)***	(0.14)***	(0.14)***	(0.14)***	(0.31)*	(0.28)**	(0.60)	(0.15)***
Observations	244	244	244	244	244	244	244	69	33	228
R-squared	0.88	0.88	0.88	0.88	0.89	0.89	0.91	0.69	0.66	0.89
F-stat on distance cut-off	8.7	8.4	9.2	10.4	10.3	9.3	3.9	4.2	1.4	8.6
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
Spatial controls	x	x	x	x	x	x	x	x	x	x
Region and school type fixed effects				x	x	x	x	x	x	x
2003 Census controls					x	x	x			
Polynomial varies on either side of cut-off							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). 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. Spatial controls are quadratics in number of schools in cluster, distance to nearest school of opposite treatment status, distance to Banjul, and distance to nearest major body of water (linear controls only for discontinuity samples). Controls from 2003 Census from nearest settlement to school: log population, employment/population ages 18+, per cent with access to electricity, per cent illiterate, per cent Muslim, per cent of Mandinka, Fula, and Wollof ethnicities. *significant at 10%; **significant at 5%; ***significant at 1%.

somewhat, increasing to 0.61 and remaining significant at 10%. When limiting the sample to schools closest to the threshold, as in columns (8)-(9), the distance threshold remains of similar magnitude.

Column (4), which has the strongest instrument (F = 10.4) among the specifications with a relevant first stage, is our preferred specification.[16] A graph of the first stage appears in Figure 3, showing the probability of government-assigned hardship status within bins of the running variable and predicted hardship status from this preferred specification.

Valid causal inference in a regression discontinuity design also requires lack of manipulation of the running variable in order to secure favourable 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 cut-off. Given the external financing of the programme, a disproportionate number of new schools just beyond the cut-off 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 4 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 km (p-value = 0.24), using the suggested bin width and bandwidth.[17] 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 continuity across the threshold distance in the distributions of observable characteristics prior to treatment. In supplemental appendix Table SA3, 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], to prevent pre-test bias). We find no significant coefficients on the distance threshold, indicating balance in pre-treatment characteristics.[18]

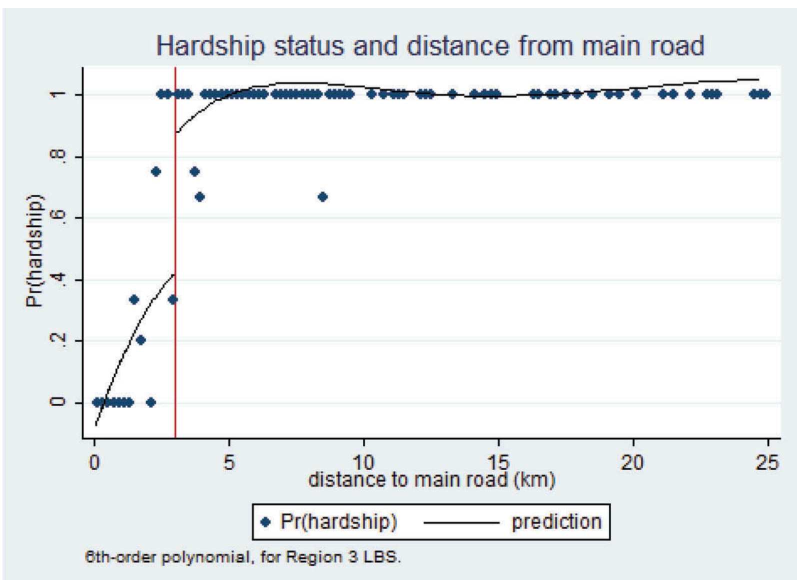

Figure 3. Regression discontinuity, first stage.

Figure shows mean of government-assigned hardship status within bins defined by distance from main road (bandwidth = 0.2 km). Line is predicted hardship status from first stage regression, as in column (4) of Table 2. Source: Pugatch and Schroeder (2014), Figure 2.

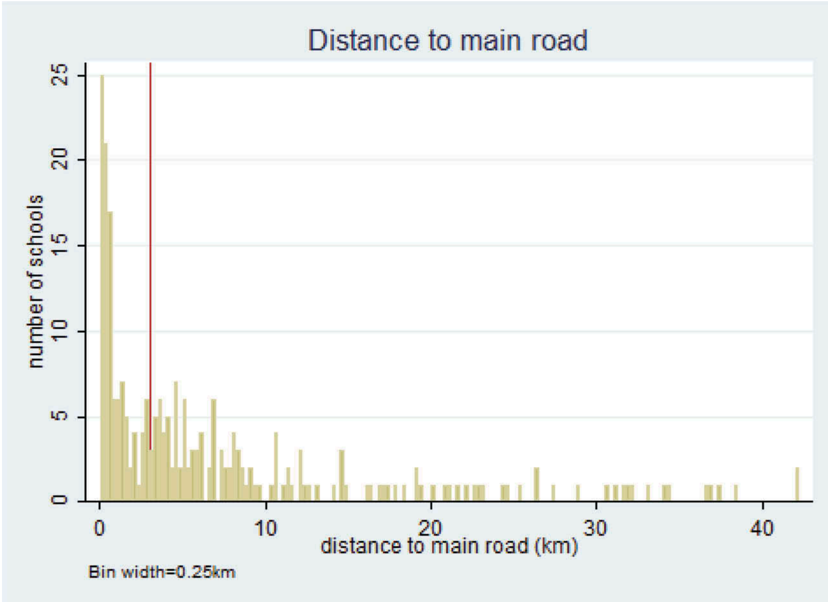

Figure 4. Density of distance from schools to main road.
Pugatch and Schroeder (2014), Figure 3

B. Teacher and student response

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 analyses whether the programme improved student performance as a result. We first document that the hardship allowance improved teaching inputs as intended. In Table A1, we find that the policy increased the quantity of teachers and qualified teachers. It also increased the quantity of female teachers and qualified teachers, an area of particular concern given the low proportion of female teachers in hardship schools (16%) and the government's focus on improving education for girls (Gajigo 2016; Blimpo, Gajigo, and Pugatch 2016; Giordono and Pugatch 2017) . The overall effect was to lower the pupil-qualified teacher ratio as intended. In Pugatch and Schroeder (2014), we show that these gains need not have come at the expense of non-hardship schools. By 2012, increased inflows of qualified teachers induced by the programme were sufficient to account for all the gains experienced by hardship schools, so that substitution of teachers between hardship and non-hardship schools would not have been necessary.[19]

In addition to altering teaching inputs, 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-enrolment, 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 earlier. 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.

To check whether enrolment changed in response to the hardship allowance, we regressed school enrolment on the distance threshold or (instrumented) hardship status, as in Equations (2) and (3). The point estimates, reported in Table A2, are positive for all students and for boys and girls separately (columns 1–3), but too imprecise to conclude that the hardship allowance increased enrolment. 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 effect of the hardship allowance on test scores are not skewed by any confounding effect 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 socio-economic 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 summarise 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 socio-economic 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).

In columns (4)–(6) of Table A3, we replace the socio-economic 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 socio-economic status index, and therefore proxy for the student's academic effort (Jackson 2012; Gershenson 2016). 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 estimate measures whether the policy changed student effort. We find no evidence that the policy had this effect.

In columns (7)–(9) of Table A3, we look for evidence of changes in school quality other than the teacher characteristics considered previously. 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, we find no evidence that the hardship allowance altered enrolment, the composition of students, or measures of school quality besides teaching inputs.[20] This means that the policy parameter we estimate in this paper (Equation [7]) is likely of similar magnitude as the production function parameter (Equation [6]).

C. Student performance

Table 3 presents reduced-form and instrumental variables results (corresponding to estimation of Equations (2) and (3), respectively) for student z-scores from the National Assessment Test.[21] In Panel A, column (1), the coefficient of 0.01 means that Grade 3 students from a school located just beyond the distance threshold scored 0.01 standard deviations better in English than those from a school located just inside the threshold. The estimate is precise enough to rule out intent-to-treat effects larger than 0.2 standard deviations, a common threshold for large effects of an education intervention. The corresponding instrumental variables estimate in Panel B is 0.02. 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. 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.[22] As the reduced-form

Table 3. Student performance.

	English						Math					
	Grade 3			Grade 5			Grade 3			Grade 5		
	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: reduced form												
Distance ≥3 km	0.01	0.01	0.00	0.21	0.21	0.19	−0.04	−0.05	−0.05	0.20	0.19	0.20
	(0.10)	(0.13)	(0.10)	(0.18)	(0.22)	(0.15)	(0.12)	(0.14)	(0.12)	(0.16)	(0.21)	(0.15)
R-squared	0.06	0.06	0.05	0.06	0.03	0.08	0.07	0.07	0.06	0.06	0.05	0.07
Panel B: Instrumental variables												
Hardship allowance	0.02	0.02	0.01	0.59	0.61	0.52	−0.09	−0.11	−0.11	0.56	0.57	0.54
	(0.22)	(0.29)	(0.21)	(0.59)	(0.76)	(0.48)	(0.24)	(0.30)	(0.24)	(0.54)	(0.69)	(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	12.6	9.8	14.2	6.7	5.6	7.3	12.6	9.7	14.2	6.7	5.6	7.3
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 1 January 2012 based on date of birth, and spatial controls as in first stage. Robust standard errors in parentheses, clustered by school. *significant at 10%; **significant at 5%; ***significant at 1%.

estimates are more precise, we can rule out effects of 0.2 standard deviations from crossing the threshold in additional subgroups: Grade 3 girls' English, Grade 3 math, and Grade 3 girls' math. The instrumental variables estimates are also most precise for those subgroups, but have larger confidence intervals that include more substantial effects.

A 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. Again we fail to find discernible effects of the hardship allowance on student learning, and can rule out effects larger than 0.2 standard deviations for several outcomes. The lack of significant results in this augmented dataset is striking, because sample sizes have increased nearly fivefold. Additional attempts to increase statistical power included pooling the 2012 results for grades 3 and 5, pooling math and English scores, and pooling all subject scores, but none produced statistically significant results. We omit these results for brevity, but present them in Tables SA6-SA8 of the supplemental appendix.

Another possible explanation for our null results is that the policy effect changed over time. Because the policy was enacted in 2005, any initial boost to productivity from the hardship allowance could have run its course by 2012, the year used in most of our analysis. This explanation would be consistent with teachers becoming accustomed to the hardship allowance and setting their effort in reference to this new 'fair pay' norm (Akerlof 1982; Akerlof and Yellen 1990). To check if the effect changed over time, in Table SA8, Panel B of the supplemental appendix we use data for 2008–2012 and allow the treatment effect to vary by year (we omit the main effect of the distance threshold, so that the interaction terms reported can be interpreted as the ITT for each year). Although a few coefficients are statistically significant, notably for Grade 5 English, no consistent pattern emerges. The results appear too tenuous to conclude that the policy improved average outcomes.[23]

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, time periods, and attempts to increase statistical power is notable. Moreover, broadening our inquiry from cognitive skills to measures of student effort (Table A3; see previous subsection) also failed to find any evidence of programme impact.

D. Changes in test score distributions

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 of the test-score distribution. Recent work on both U.S. social programmes and interventions in developing countries suggests that mean effects often represent the average of very heterogeneous treatment effects (e.g. Bitler, Gelbach, and Hoynes 2006; Glewwe, Kremer, and Moulin 2009). Thus, we address the possibility that the hardship allowance changed the distribution of test scores by improving scores at some points in the distribution, while lowering them at other points.

As a first step in this analysis, we limit the sample to schools between 2 and 4 km from the main road and compare the full distribution of test scores at schools above and below the 3-km threshold. Figure 5 presents Q-Q plots of these distributions. Each point represents a centile (for example, the median), with the x-coordinate giving that centile for students below the threshold, and the y-coordinate giving that centile for students above the threshold. If there were no difference in the distributions, the centiles would be identical and all points would lie on the 45-degree line. Instead, we see that the lower centiles are quite similar, although in several cases the centiles of students in hardship areas are in fact lower than those in non-hardship areas. At higher ends of the distribution, however, the centiles of test scores beyond the 3-km threshold are higher than those of students below the threshold. We obtain similar results when plotting the smoothed cumulative distribution functions of test scores for schools between 2 and 4 km from the main road (shown in Figure SA2 of the supplemental appendix). These findings are consistent with the hardship allowance improving test scores at the top of the distribution. For grade 3, the results are also consistent with the policy reducing scores at the bottom of the distribution.

However, a comparison of unconditional distributions as in Figure 5 may give an inaccurate impression of the effect of the hardship allowance on the test score distribution if there are systematic differences between schools on either side of the cut-off other than the policy. To address this issue, we estimate reduced-form quantile treatment effects throughout the test score distribution using the nonparametric estimator of Frandsen, Frölich, and Melly (2012), which relies on the discontinuity in programme assignment for identification. We provide details on the estimator and present results in supplemental appendix Figure SA3. While the estimates are quite noisy, they reinforce the patterns found in the Q-Q plots and are consistent with the hardship allowance benefitting students at the top of the distribution but harming those at the bottom.

The pattern in the test-score quantiles is consistent with other evidence from developing countries that interventions intended to improve student performance have benefited only the students who were already performing the best. Glewwe, Kremer, and Moulin (2009) find that textbooks provided to rural Kenyan schools did not improve average test scores, but did have a positive impact for students who had high scores on earlier tests; they interpret this as the result of curricula oriented towards academically strong students, specifically students who are proficient in English. Pritchett (2013) argues more broadly that curricula in many developing countries are suited only for the top students, leaving the rest to fall progressively behind. Banerjee et al. (2007) and Duflo, Dupas, and Kremer (2011) find evidence that supports this pattern in India and Kenya, respectively, where they argue that pedagogy and curricula cater to the top students, rather than the new population of students brought in by increases in enrolment. The effects we find at the top and bottom of the distribution of test scores could potentially be the result of the additional qualified teachers focusing their lessons on stronger students.

E. Mechanisms

In sum, we have found no evidence that the hardship allowance changed average student performance, some evidence of improved test scores at the top of the distribution, but (in grade

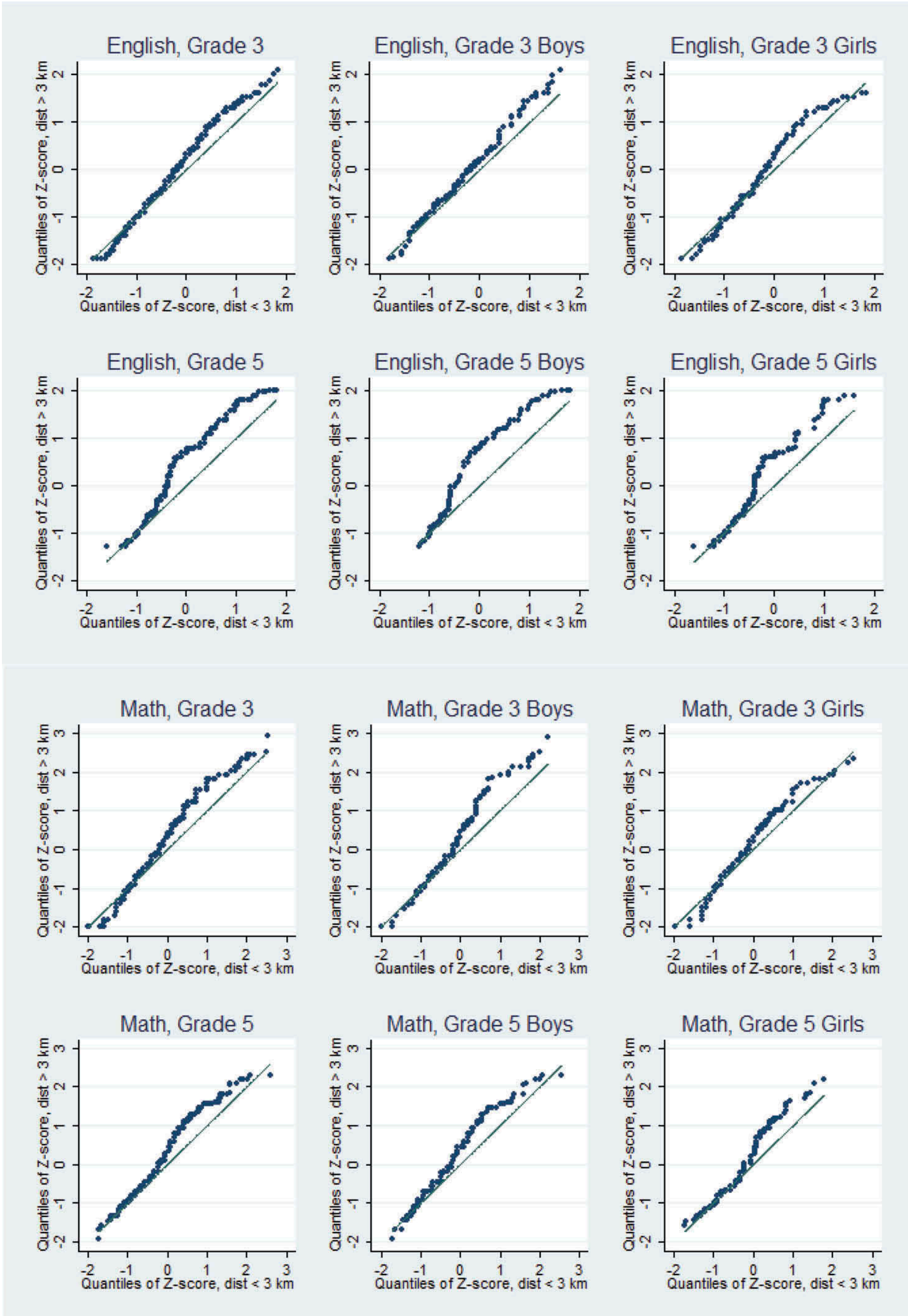

Figure 5. Quantile-quantile plots, schools 2–4 km from main road.

3) lower scores for those at the bottom of the distribution. We now return to the mechanisms behind these results, as discussed in Section III.C. Given the improvements in teaching inputs found in Table A1 and the apparent heterogeneity in effects on students across different schools, each of the hypothesised mechanisms emerge as potential explanations for our results:

- *Improvement in teacher quality*. An influx of qualified teachers could reorient classrooms to focus on the best students at the expense of the weakest.
- *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.[24]
- *Enhanced teacher motivation*. To demonstrate higher productivity in response to the hardship allowance, teachers might focus on the best-performing students, for whom learning gains could come at lowest marginal cost of teacher effort.
- *School quality*. If teaching and non-teaching inputs are complements in the education production function (5), then only higher quality schools might gain from increases in teaching inputs.

We look for evidence of each of these mechanisms by augmenting the reduced form Equation (2) with an interaction term between the indicator for the 3-km distance threshold and a proxy for each mechanism, as well as the main effect of the proxy and interactions between it and the polynomial terms of the running variable. The coefficient on the interaction term between the proxy and the distance threshold 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, that is, the coefficient on the interaction term will be zero. Non-zero coefficients on interaction terms would suggest the presence of mechanisms through which the programme exerted an effect.

The proxies we use are the pupil-qualified teacher ratio at a school, which measures teaching inputs; 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 captures aspects of school quality not included elsewhere; and the socio-economic status index described in Section V.B, which summarises the student's relative advantages.

Before discussing the results of this exercise, we note its limitations. The coefficient on the interaction term between a proxy variable and the distance threshold captures heterogeneity in student performance between schools just beyond the threshold with different values of this characteristic. 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 proxy 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, that is, using 'bad control' (Angrist and Pischke 2008). We therefore consider the next set of results as providing descriptive evidence on the mechanisms underlying the policy parameter we estimate, rather than being causal.

Table 4 presents results of the exercise. We limit the sample to Grade 5 because the student survey used to create the socio-economic status index exists only for that grade. For Grade 5 English, male students in schools located just beyond the threshold distance perform better the higher their socio-economic status (statistically significant at 10%). The coefficient magnitude of 1.02 means that in a school just across the 3-km threshold, a student with the highest socio-economic status according to our index (an index value of 1) is predicted to score 1.02 standard deviations higher in English than a student with the lowest socio-economic status (an index value of 0). This is a striking discrepancy, and it is important to emphasise that it does not merely reflect average differences between students of high and low socio-economic 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-socio-economic 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 0.72 in math for all students. This result is

Table 4. Student performance and heterogeneity.

	English			Math		
	All	Boys	Girls	All	Boys	Girls
	(1)	(2)	(3)	(4)	(5)	(6)
Distance ≥3 km	0.16	0.23	0.08	−0.44	−0.48	−0.46
	(1.50)	(1.88)	(1.27)	(1.30)	(1.48)	(1.37)
Distance ≥3 km interacted with:						
Pupil-qualified teacher ratio	0.01	0.01	0.01	0.004	0.01	0.001
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Hardship %	−0.62	−1.92	0.26	1.36	0.39	2.05
	(4.41)	(5.54)	(3.74)	(4.01)	(4.55)	(4.18)
z(2011)	0.95	1.07	0.88	0.86	1.02	0.80
	(0.64)	(0.78)	(0.55)	(0.61)	(0.70)	(0.58)
Socio-economic status index	0.42	1.02	−0.02	0.72	1.06	0.51
	(0.40)	(0.55)*	(0.45)	(0.41)*	(0.58)*	(0.46)
Observations	6,379	2,986	3,393	6,367	2,979	3,388
R-squared	0.13	0.12	0.16	0.12	0.12	0.13
Mean of dependent variable	−0.32	−0.23	−0.41	−0.25	−0.16	−0.33
p-value on interaction terms	0.15	0.04	0.32	0.16	0.10	0.48

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. Hardship % 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 socio-economic status index, see notes to Table A3. 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 1 January 2012 based on date of birth, and spatial controls as in first stage. Robust standard errors in parentheses, clustered by school. *significant at 10%; **significant at 5%; ***significant at 1%.

also consistent with the literature that finds stronger effects of education interventions on students who started out most advantaged. The corresponding coefficients for the subsample of girls, however, are smaller and not statistically significant.

Coefficients on other interaction terms – pupil-qualified teacher ratio, hardship salary premium, and a school's lagged z-score – are generally of the expected sign, 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 potentially heterogeneous treatment effects of the policy.[25]

We close this discussion of mechanisms by considering whether the gains at the top of the distribution found in Figure 5 reflects differences among students or schools. We re-estimate the main reduced-form specification in (2), splitting the sample by median socio-economic status index and by 2011 average English and math z-score. Because the socio-economic 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 socio-economic status in Panel A and school quality in Panel B.[26] Given the power concerns in the main results, we might not expect much precision in subsamples, but in fact several coefficients are statistically significant in a manner consistent with the distributional results. In Panel A, students in schools beyond the distance threshold from above-median socio-economic status households score significantly better in math, on the order of 0.32–0.42 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 effects for above-median socio-economic status students are significant only at the 10% level, an examination of the test-score distributions supports this

Table 5. Grade 5 results, by student socio-economic status and school quality.

	English						Math					
	Below median			Above median			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: socio-economic status												
Distance ≥3 km	0.11	−0.01	0.18	0.31	0.41	0.23	0.06	−0.04	0.13	0.36	0.42	0.32
	(0.17)	(0.24)	(0.13)	(0.21)	(0.25)	(0.20)	(0.16)	(0.21)	(0.16)	(0.19)*	(0.23)*	(0.19)*
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	0.06	0.04	0.07	0.08	0.06	0.12	0.07	0.06	0.07	0.07	0.07	0.08
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 ≥3 km	0.01	0.03	−0.02	0.27	0.22	0.28	0.05	0.05	0.05	0.18	0.16	0.18
	(0.13)	(0.15)	(0.12)	(0.29)	(0.37)	(0.24)	(0.14)	(0.19)	(0.17)	(0.25)	(0.29)	(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	0.07	0.06	0.06	0.07	0.05	0.10	0.06	0.07	0.05	0.07	0.05	0.09
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 socio-economic status index (Panel A) or school's position in quality distribution. Socio-economic status index calculated as 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 1 January 2012 based on date of birth, and spatial controls as in first stage. Robust standard errors in parentheses, clustered by school. *significant at 10%; **significant at 5%; ***significant at 1%.

interpretation. Supplemental appendix Figure SA4 presents the Q-Q plots for each group. The figure shows that the differences in the test-score distributions are driven by students with higher socio-economic status. Although the socio-economic 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. Our identification strategy includes a rich set of controls for potential confounding factors associated with a school's spatial location and passes a battery of specification checks. Although the policy improved teaching inputs, we failed to find any effects on average student performance. These null results persist across genders, grades, and time periods. While low power is a concern, in many specifications we can rule out test score effects of 0.2 standard deviations or larger. We also fail to find effects of the policy on enrolment, per cent of female students enrolled, student socio-economic characteristics, student effort, or school quality beyond teaching resources.

Despite the lack of average impact, we find suggestive evidence of heterogeneity in the effect of the hardship allowance. Our results point to increased learning at the top of the test score distribution and decreased achievement at the bottom. Test score gains appear to be driven by high socio-economic status students in schools just beyond the distance threshold, rather than by quality differences across schools. Our findings suggest that only socio-economically advantaged students benefitted from the hardship allowance, perhaps because the teachers recruited via the programme are better able to connect 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 could help make the programme both more equitable and beneficial overall.

Of course, these rather discouraging results come with the usual caveats inherent to regression discontinuity designs. The treatment effects we estimate are necessarily local to schools near the 3-km cut-off, whereas the policy aimed to improve conditions in all hardship schools. We cannot say whether our results reflect the impact of the policy for the most remote schools.

Nonetheless, if teaching and non-teaching inputs are complements rather than substitutes, and if hardship schools near the cut-off are of higher average quality than those located farther away, then we should expect to see the largest gains precisely at the cut-off. That we fail to find positive results at the cut-off across a range of subjects, grades, and time periods make it unlikely that the programme is raising average student performance overall, though this remains an open question.

If anything, gains from the policy appear to be concentrated among the most capable students. Our results therefore echo Ganimian and Murnane (2016), whose survey of the most rigorously identified evaluations of schooling interventions in developing countries noted how difficult it is to 'induce teachers to maximise their efforts to teach all students well' (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 programme 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 most Gambian students. Tying salary increases more directly to student performance could help to establish this link.

Notes

1. Although Gambian teachers in hardship schools receive the salary premium regardless of qualifications, the larger base salary of qualified teachers provides a stronger incentive to serve in hardship areas compared to unqualified teachers.
2. De Ree et al. (2017)use data from 2–3 years after programme introduction in Indonesia, while Urquiola and Vegas (2005) use recent reclassifications of Bolivian schools from rural to urban status without a corresponding reduction in the rural salary premium. Greaves and Sibieta (2014) do not discuss the duration of the policy they examine.
3. McEwan (2015) also reviews the literature, focusing only on evidence from randomised 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 specialised in-service training programmes, 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 programme 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.
5. The figure is inspired by DiNardo and Lee (2004).
6. We find no evidence of discontinuities in education market characteristics or geographic endowments at the 3-km cut-off (results available upon request), but nonetheless include them in the analysis to take a conservative approach to identification.
7. Although teaching inputs T differ from the threshold indicator T of the preceding sections, the abuse of notation is deliberate because the regression discontinuity design is based on the marginal effect of crossing the threshold.
8. The direction and magnitude of these changes are not obvious. Although we would expect better teachers and students to sort into hardship schools near the cut-off, their ability to do so is constrained by personnel policies (for teachers) and the availability of substitute schools nearby (for students). For non-teaching inputs,

it is not clear whether administrators would increase resources to hardship schools (to take advantage of complementarities in teaching and non-teaching inputs, for instance) or reduce resources (to equalise differences between hardship and non-hardship schools).

9. Formally, if teaching inputs $T = T(q,m)$, where q is teacher quality and m is teacher motivation, then the total change in teaching inputs is the sum of changes in these channels: $dT = \frac{\partial T}{\partial q} dq + \frac{\partial T}{\partial m} dm$.
10. One channel through which increased teacher effort could affect student outcomes is time spent teaching, if the policy reduced teacher absenteeism. While our data do not allow us to address this issue, anecdotal evidence suggests that teacher absenteeism is not a significant problem in the Gambia. Each cluster of schools has a monitor who regularly checks on and records teacher attendance. Teachers with unexcused absences receive warning letters and their pay is withheld if absences continue. Officials told us that this policy is very effective. The increased pay would, however, increase the cost of missing a day of work.
11. 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 enrolment, not number of teachers as in the earlier study. We use different weights 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.
12. We use GIS software to measure distance between schools and the main road as the minimum travel distance along all feeder roads originating at the school. MoBSE calculates distance only for schools whose hardship status is in doubt, using vehicle odometer readings which are not centrally recorded, precluding us from making a direct comparison.
13. Although the classical errors-in-variables formula for attenuation bias does not apply to binary regressors such as the 3-km threshold, dummy variables measured with error also lead to attenuation bias (Aigner 1973).
14. 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.
15. We include controls for log population, employment/population ratio for ages 18+, per cent with access to electricity, per cent illiterate, per cent Muslim, and per cent of Mandinka, Fula, and Wollof ethnicities (the three largest ethnic groups).
16. 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 choosing polynomial order.
17. 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 favour of an increase in hardship schools means that the setting satisfies this condition.
18. Figure SA1 of the supplemental appendix graphs the results of Table SA3. Each panel of the figure shows the mean of a characteristic from the 2003 Census within bins of 0.2km distance from the main road. No discontinuities across the 3km threshold are apparent in the figure.
19. Table A1, column (5) further shows that these gains were not achieved by sending additional teacher trainees to non-hardship schools. We cannot rule out substitution of individual teachers, however, as we do not observe employment histories.
20. We also checked for changes in travel time to school, which would be expected to increase if students assigned to non-hardship schools were traveling extra distances to be able to attend schools that had received the allowance. We find no effects of crossing the threshold on measures of travel time or means of transportation to school. See supplemental appendix Table SA4.
21. 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 analyses the full sample at the school level.
22. Nonparametric estimation of the outcomes in Table 3 using local linear regression and the optimal bandwidth Imbens and Kalyanaraman (2012) produce qualitatively similar results as those reported here, with no statistically significant estimates of the coefficient on the distance threshold. See Table SA5 of the supplemental appendix.
23. In an additional robustness check, we include Region 2 in the sample. Region 2 was originally considered for a 20% hardship allowance but ultimately excluded from the programme for lack of funds, making it particularly apt as a falsification exercise. We find no significant differences in school performance across the 3-km cut-off when limiting the sample to Region 2, with standard errors very large. Nor do we find any differential effects between the programme regions (3–6) with Region 2 in a combination RD/difference-in-differences analysis. These results increase our confidence that the policy did not produce average test score gains in the programme regions. See Table SA9 of the supplemental appendix.
24. These explanations are also broadly consistent with Pritchett's (2013) argument that curricula in many developing countries are suited only for the top students. Another possibility related to each of the first two explanations is that teachers and students match to each other based on quality. We think this is unlikely, because the small size of hardship schools makes it difficult for such matches to occur. In fact, it is more likely

that students learn in a multigrade classroom (52% of students in hardship schools, as reported in Table 1) than there would be multiple classrooms in the same grade for a teacher to switch. Nonetheless, we cannot completely rule out student-teacher matching on unobservable characteristics; even studies from the U.S. using massive datasets with detailed information on students and teachers struggle with this issue (Chetty, Friedman, and Rockoff 2014a; Rothstein 2014).

25. As an alternate test of the efficiency wage hypothesis, we interact the distance threshold with a full set of region indicators. We fail to find any differential effects of the policy across regions, consistent with the hardship premium interaction in Table 4. Results appear in supplemental appendix Table SA10.
26. We determine median socio-economic 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.

Acknowledgements

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 Hydara, Nathalie Lahire, Palamin Mbowe, Aidan Mulkeen, Momodou Sanneh, Ryoko Tomita, and Bassirou Touré. We thank Paul Ferraro, Laura Kawano, Owen Ozier 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.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Todd Pugatch is an Associate Professor of Economics at Oregon State University and a Research Fellow at IZA. He completed a Ph.D. in Economics at the University of Michigan in 2011. His research areas are development economics, labor economics, and the economics of education. His recent work focuses on education policy in the Gambia and Rwanda, youth unemployment in South Africa, the effect of Mexican immigration on labor markets in the U.S., and promoting student success in higher education in the U.S.

Elizabeth Schroeder is an Associate Professor of Economics at Oregon State University. She conducts research in development and applied micro-econometrics, including nonparametric and semiparametric estimation. Research topics include evaluating the impact of education and credit policies, as well as other microeconomic applications.

References

Aigner, D. J. 1973. "Regression with a Binary Independent Variable Subject to Errors of Observation." *Journal of Econometrics* 1 (1): 49–59. doi:10.1016/0304-4076(73)90005-5.

Akerlof, G. A. 1982. "Labor Contracts as Partial Gift Exchange." *The Quarterly Journal of Economics* 97 (4): 543–569. doi:10.2307/1885099.

Akerlof, G. A., and J. L. Yellen. 1990. "The Fair Wage-Effort Hypothesis and Unemployment." *The Quarterly Journal of Economics* 105 (2): 255–283. doi:10.2307/2937787.

Angrist, J. D., and J.-S. Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. 1st ed. Princeton, NJ: Princeton University Press.

Banerjee, A. V., S. Cole, E. Duflo, and L. Linden. 2007a. "Remedying Education: Evidence from Two Randomized Experiments in India." *The Quarterly Journal of Economics* 122 (3): 1235–1264. doi:10.1162/qjec.122.3.1235.

Bitler, M. P., J. B. Gelbach, and H. W. Hoynes. 2006. "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments." *The American Economic Review* 96 (4): 988–1012. doi:10.1257/aer.96.4.988.

Blimpo, M. P., O. Gajigo, and T. Pugatch. 2016. "Financial Constraints and Girls' Secondary Education: Evidence from School Fee Elimination in the Gambia." *The World Bank Economic Review* (October): lhw051. doi:10.1093/wber/lhw051.

Boone, P., I. Fazzio, K. Jandhyala, C. Jayanty, G. Jayanty, S. Johnson, V. Ramachandrin, F. Silva, and Z. 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. http://www.nber.org/papers/w18971.

Bourdon, J., M. Frolich, and K. 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. doi:10.1111/%28ISSN%291467-985X/issues.

Chetty, R., J. N. Friedman, and J. E. Rockoff. 2014a. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104 (9): 2593–2632. doi:10.1257/aer.104.9.2593.

Chetty, R., J. N. Friedman, and J. E. Rockoff. 2014b. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* 104 (9): 2633–2679. doi:10.1257/aer.104.9.2633.

Darling-Hammond, L., B. Berry, and A. Thoreson. 2001. "Does Teacher Certification Matter? Evaluating the Evidence." *Educational Evaluation and Policy Analysis* 23 (1): 57–77. doi:10.3102/01623737023001057.

De Ree, J., K. Muralidharan, M. Pradhan, and H. Rogers. 2017."Double for Nothing? Experimental Evidence on an Unconditional Teacher Salary Increase in Indonesia†." *The Quarterly Journal of Economics*, November. doi:10.1093/qje/qjx040.

Dell, M. 2010. "The Persistent Effects of Peru's Mining 'Mita.'." *Econometrica* 78 (6): 1863–1903. doi:10.3982/ECTA8121.

DiNardo, J., and D. S. Lee. 2004. "Economic Impacts of New Unionization on Private Sector Employers: 1984-2001." *The Quarterly Journal of Economics* 119 (4): 1383–1441. doi:10.1162/0033553042476189.

Duflo, E., P. Dupas, and M. Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review* 101 (5): 1739–1774. doi:10.1257/aer.101.5.1739.

Frandsen, B. R., M. Frölich, and B. Melly. 2012. "Quantile Treatment Effects in the Regression Discontinuity Design." *Journal of Econometrics* 168 (2): 382–395. doi:10.1016/j.jeconom.2012.02.004.

Fryer, R. G. 2013. "Teacher Incentives and Student Achievement: Evidence from New York City Public Schools." *Journal of Labor Economics* 31 (2): 373–407. doi:10.1086/667757.

Gajigo, O. 2016. "Closing the Education Gender Gap: Estimating the Impact of Girls' Scholarship Program in the Gambia." *Education Economics* 24 (2): 167–188. doi:10.1080/09645292.2014.985287

Galiani, S., and R. Perez-Truglia. 2013. "School Management in Developing Countries." In *Education Policy in Developing Countries*, edited by P. Glewwe. University of Chicago Press.

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*. Banjul, Gambia: Ministry of Basic and Secondary Education.

Ganimian, A. J., and R. J. Murnane. 2016. "Improving Education in Developing Countries Lessons from Rigorous Impact Evaluations." *Review of Educational Research* 86 (3): 719–755. doi:10.3102/0034654315627499.

Gershenson, S. 2016. "Linking Teacher Quality, Student Attendance, and Student Achievement." *Education Finance and Policy* 11: 125–149. doi:10.1162/EDFP_a_00180.

Giordono, L., and T. Pugatch. 2017. "Non-Tuition Costs, School Access and Student Performance: Evidence from the Gambia." *Journal of African Economies*. doi:10.1093/jae/ejw033.

Glewwe, P., E. A. Hanushek, S. Humpage, and R. Ravina. 2013. "School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010." *Education Policy in Developing Countries* 4 (14,972): 13.

Glewwe, P., and K. Muralidharan. 2016. "Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications." *Handbook of the Economics of Education* 5: 653-743.

Glewwe, P., and M. Kremer. 2006. "Schools, Teachers, and Education Outcomes in Developing Countries." *Handbook of the Economics of Education* 2: 945–1017.

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

Glewwe, P., N. Ilias, and M. Kremer. 2010. "Teacher Incentives." *American Economic Journal: Applied Economics* 2 (3): 205–227.

Goyal, S., and P. Pandey. 2013. "Contract Teachers in India." *Education Economics* 21 (5): 464–484. doi:10.1080/09645292.2010.511854.

Greaves, E., and L. Sibieta. 2014. "Estimating the Effect of Teacher Pay on Pupil Attainment Using Boundary Discontinuities." IFS Working Papers. https://www.econstor.eu/handle/10419/101331.

Hanushek, E. A. 2003. "The Failure of Input-Based Schooling Policies." *Economic Journal* 113 (485): F64–98. doi:10.1111/1468-0297.00099.

Hanushek, E. A. 2011. "The Economic Value of Higher Teacher Quality." *Economics of Education Review* 30 (3): 466–479. doi:10.1016/j.econedurev.2010.12.006.

Imbens, G., and K. Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *The Review of Economic Studies* 79 (3): 933–959. doi:10.1093/restud/rdr043.

Jackson, C. K. 2012. "Non-Cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9th Grade Teachers in North Carolina." National Bureau of Economic Research. http://www.nber.org/papers/w18624.

Kirabo, J. C., J. E. Rockoff, and D. O. Staiger. 2014. "Teacher Effects and Teacher-Related Policies." *Annual Review of Economics* 6 (1): 801–825. doi:10.1146/annurev-economics-080213-040845.

Kling, J. R., J. B. Liebman, and L. F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75 (1): 83–119. doi:10.1111/ecta.2007.75.issue-1.

Kremer, M., N. Chaudhury, N. F. H. Rogers, K. Muralidharan, and J. Hammer. 2005. "Teacher Absence in India: A Snapshot." *Journal of the European Economic Association* 3 (2–3): 658–667. doi:10.1162/jeea.2005.3.2-3.658.

Lavy, V. 2002. "Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement." *Journal of Political Economy* 110 (6): 1286–1317. doi:10.1086/342810.

Lavy, V. 2009. "Performance Pay and Teachers' Effort, Productivity, and Grading Ethics." *The American Economic Review* 99 (5): 1979–2011. doi:10.1257/aer.99.5.1979.

Lee, D. S., and T. Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (2): 281–355. doi:10.1257/jel.48.2.281.

McCrary, J. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142 (2): 698–714. doi:10.1016/j.jeconom.2007.05.005.

McEwan, P. J. 1999. "Recruitment of Rural Teachers in Developing Countries: An Economic Analysis." *Teaching and Teacher Education* 15 (8): 849–859. doi:10.1016/S0742-051X(99)00025-6.

McEwan, P. J. 2015. "Improving Learning in Primary Schools of Developing Countries A Meta-Analysis of Randomized Experiments." *Review of Educational Research* 5 (3): 353–394. doi:10.3102/0034654314553127.

Mulkeen, A., and D. Chen. 2008. *Teachers for Rural Schools: Experiences in Lesotho, Malawi, Mozambique, Tanzania, and Uganda*. Africa Human Development Series. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/6423

Muralidharan, K., and V. Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India." *Journal of Political Economy* 119 (1): 39–77. doi:10.1086/659655.

Muralidharan, K., and V. Sundararaman. 2013. "Contract Teachers: Experimental Evidence from India." Working Paper 19440. National Bureau of Economic Research. http://www.nber.org/papers/w19440.

Pritchett, L. 2013. *The Rebirth of Education: Schooling Ain't Learning*. Washington, DC: Center for Global Development.

Pugatch, T. 2017."Is Teacher Certification an Effective Tool for Developing Countries?" *IZA World of Labor* April. doi:10.15185/izawol.349.

Pugatch, T., and E. Schroeder. 2014a. "Incentives for Teacher Relocation: Evidence from the Gambian Hardship Allowance." *Economics of Education Review* 41 (August): 120–136. doi:10.1016/j.econedurev.2014.04.003.

Rivkin, S. G., E. A. Hanushek, and J. F. Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica* 73 (2): 417–458. doi:10.1111/ecta.2005.73.issue-2.

Rothstein, J. 2014. "Revisiting the Impacts of Teachers." UC-Berkeley Working Paper. http://eml.berkeley.edu/~jrothst/workingpapers/rothstein_cfr.pdf.

Shapiro, C., and J. E. Stiglitz. 1984. "Equilibrium Unemployment as a Worker Discipline Device." *The American Economic Review* 74 (3): 433–444.

UNICEF. 2010. "Protecting Salaries of Frontline Teachers and Health Workers." http://www.unicef.org/socialpolicy/files/Protecting_Salaries_of_Frontline_Teachers_and_Healthcare_Workers_final%281%29.pdf.

United Nations. 2016. "Report of the Inter-Agency and Expert Group on Sustainable Development Goal Indicators." http://unstats.un.org/unsd/statcom/47th-session/documents/2016-2-IAEG-SDGs-Rev1-E.pdf.

Urquiola, M. 2006. "Identifying Class Size Effects in Developing Countries: Evidence from Rural Bolivia." *Review of Economics and Statistics* 88 (1): 171–177. doi:10.1162/rest.2006.88.1.171.

Urquiola, M., and E. Vegas. 2005. "Arbitrary Variation in Teacher Salaries: An Analysis of Teacher Pay in Bolivia." In *Incentives to Improve Teaching: Lessons from Latin America*, edited by E. Vegas, 187–211. Washington, DC: The World Bank.

Vegas, E., and D. L. Joost. 2003. "Do Differences in Teacher Contracts Affect Student Performance? Evidence from Togo." World Bank. http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2003/10/24/000160016_20031024103517/additional/310436360_20050276095003.pdf.

Woessmann, L. 2011. "Cross-Country Evidence on Teacher Performance Pay." *Economics of Education Review* 30 (3): 404–418. doi:10.1016/j.econedurev.2010.12.008.

Appendices

Table A1. Teacher response to hardship allowance.

	Number of teachers					Qualified	Pupil-	Pupil-
	Total	Female	Qualified	Qualified Female	Trainee	Proportion	Teacher Ratio	Qualified Ratio
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance ≥3 km	5.0	1.9	3.9	1.3	1.2	0.05	−10.3	−20.2
	(2.1)**	(0.7)**	(1.5)**	(0.6)**	(0.8)	(0.08)	(10.5)	(8.3)**
Observations	244	244	244	244	244	244	244	244
R-squared	0.45	0.48	0.40	0.49	0.33	0.11	0.26	0.19
Mean of dependent variable	14.1	3.8	9.5	2.5	2.7	0.64	30.5	50.6

Table shows results of regressions of school outcomes on distance threshold. Data from 2012, using same sample as main analysis. All regressions include a 6th-order polynomial in distance from school to main road, region and school type fixed effects, and quadratic spatial controls as in first stage. Regressions weighted by school enrolment. 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 enrolment.

	Enrolment				Test-takers			
	All	Boys	Girls	% female	All	Boys	Girls	% enrolled
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: reduced form								
Distance ≥ 3km	18.2	5.2	13.0	0.02	9.9	4.0	5.8	0.01
	(31.3)	(15.9)	(16.6)	(0.02)	(8.3)	(4.6)	(4.3)	(0.03)
R-squared	0.22	0.25	0.2	0.31	0.3	0.3	0.2	0.14
Panel B: Instrumental variables								
Hardship allowance	47.2	13.5	33.7	0.05	25.6	10.5	15.1	0.03
	(80.3)	(39.8)	(43.5)	(0.04)	(22.3)	(11.6)	(12.3)	(0.06)
Observations	244	244	244	244	244	244	244	243
1st stage F-stat	7.7	7.7	7.7	10.4	7.7	7.7	7.7	10.4
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. Enrolment includes only grades 1–6 for Basic Cycle Schools. Test-takers refer to National Assessment Test (NAT), grades 3 and 5. All regressions include a 6th-order polynomial in distance from school to main road, region and school type fixed effects, and spatial controls as in first stage. Regressions for female enrolment percentage and per cent enrolled who take NAT are weighted by school enrolment; 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.

	Socio-economic status			Student effort			School quality		
	All	Boys	Girls	All	Boys	Girls	All	Boys	Girls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance ≥ 3km	−0.01	0.001	−0.01	−0.02	0.01	−0.04	−0.03	−0.07	−0.003
	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)
Observations	6,379	2,986	3,393	6,371	2,984	3,387	7,001	3,271	3,730
R-squared	0.03	0.04	0.03	0.02	0.03	0.03	0.02	0.01	0.02
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. Socio-economic 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, and spatial controls as in Stage 1. Robust standard errors in parentheses, clustered by school. *significant at 10%; **significant at 5%; ***significant at 1%.