This draft is preliminary and incomplete.
Please do not cite or quote without permission.

Evaluating a test-based public subsidy program for low-cost private schools: Regression-discontinuity evidence from Pakistan

Felipe Barrera-Osorio
World Bank
fbarrera@worldbank.org

Dhushyanth Raju
World Bank
draju2@worldbank.org

Date last revised: October 15, 2008

Abstract

In this paper, we estimate the causal effects of a public subsidy program targeted at low-cost private schools in Pakistan on school size, student learning, and schooling inputs. Program benefits comprise of a monthly per-student subsidy as well as annual monetary bonuses to teachers and schools based on school performance in a specially-designed independent academic test. Both initial (program entry) and continued benefit eligibility are tied to, among other things, achieving stipulated minimum student pass rates in the test, thus allowing the application of regression-discontinuity (RD) methods to identify and estimate program impacts. Data on the latest two rounds (phase-3 and phase-4) of program entrants are used in the study. Modeling the entry process of phase-4 program applicants as a sharp-RD design, we find evidence of large positive impacts on the number of students, teachers, classrooms, and blackboards emerging within a short treatment period. In contrast, given the presence of crossovers arising from test retaking, modeling the entry process of phase-3 program applicants as a partially fuzzy-RD design, we find no evidence of significant program impacts. We posit that the latter finding is likely due to weak identification arising from a small jump in the probability of treatment assignment at the minimum pass rate.

JEL classification codes: I21; I28

Keywords: education; Pakistan; regression-discontinuity design; subsidies; private schools

We thank the staff of the Punjab Education Foundation, in particular, Usman Tariq Bhatty, Mian Kashif Ijaz, Shahid Hafiz Kardar, Allah Bakhsh Malik, Haroon Raheem, and Huma Rizvi, for extensive discussions on program design and implementation and assistance with the program administrative data; and Nabiha Asghar, Humayun Javed, Kashif Mahmood, and Sehar Sarah Sikander Shah for assistance with the follow-up data collection. Financial support for the follow-up data collection from the World Bank's Education Program Development Fund (EPDF) and Bank-Netherlands Partnership Program (BNPP) Trust Funds is gratefully acknowledged. The findings, interpretations, and conclusions expressed herein are our own and do not necessarily represent the views of the World Bank, its Board of Directors, or any of its member countries. All remaining errors are our own.

Introduction

In this paper, we estimate the causal effects of a recently-instituted public subsidy program to low-cost private schools in the province of Punjab, Pakistan. Coined the Foundation-Assisted Schools (FAS) program and administered by the Punjab Education Foundation (PEF), a semi-autonomous organization, the program's main stated objectives are to increase school participation among children from socioeconomically-disadvantaged backgrounds *as well as* their achievement levels. As described later, these aims are particularly salient given the serious shortfalls in these outcomes in the country.

The underlying rationale behind this program is to leverage—essentially via public financing—the growing low-cost private educational sector in Pakistan to increase equiTable access to schooling and improve student achievement more efficiently than can be achieved through the country's public school system. As such, the program falls under the rubric of public-private partnerships (PPPs) in education, a set of interventions which is increasingly perceived in international policy circles as a promising mechanism for attaining key educational goals (World Bank 2008). In addition, opportunities for introducing PPP programs of medium to large scale are arising in various developing countries (e.g., India, Kenya, Nigeria) as the private educational sector matures and becomes an important player in service delivery.

The FAS program was initiated on a pilot basis by PEF in November 2005; since then it has been rapidly expanded in phases to cover more schools. As of June 2008, the program covers 1,082 low-cost private schools (primary, middle, and secondary levels[1]) with roughly 474,000 students across 18 out of 35 districts in Punjab. The original benefit was a monthly per-student subsidy of 300 rupees (US$4.3[2]); the receipt of this benefit was strictly tied to, among other things, tuition-free schooling for all enrolled children and a minimum student pass rate in a specially-designed written academic test offered periodically by PEF. Subsequently, annual monetary bonuses to teachers in program schools that meet a stipulated higher student pass rate and to program schools that achieve the highest pass rate in the test in each district were also introduced.

These main program design features represent key innovations, some more cutting edge than others. To provide a reference point, the standard subsidy program is designed to directly

[1] Primary schools are composed of grades 1-5. Middle schools are composed of grades 1-8 or 6-8. Secondary schools are composed of grades 1-10, 6-10, or 9-10.
[2] Exchange rate used: 71 Pakistani rupees per US dollar (June 29, 2008).

finance educational inputs (Gauri and Vawda 2003). In view of this, a first innovation is the enrollment-related subsidy with largely free rein on how the school uses the subsidy amount.[3] A second innovation is tying benefit receipt to a minimum level of test performance. This feature categorizes the FAS program among test-based school accountability programs characterized by high-stakes testing. Such programs constitute a rare band in the current spectrum of education programs, particularly in developing countries.[4] A third innovation is the use of test performance rewards in the form of group bonuses for teachers and competitive bonuses for schools.[5]

To the best of our knowledge, credible evidence on the impacts of public subsidies to support the establishment or operation of private schools is limited. Among existing studies, one in particular is pertinent as it examines a subsidy program basically of similar design in Pakistan. Kim et al. (1999) study the impact of a program that offered a low temporary per-girl student subsidy conditional on free girls' schooling to establish and operate private primary schools; this program was randomly offered to a subset of poor neighborhoods lacking public girls' primary schools in an urban area in the province of Balochistan. Using a difference-in-differences approach, they find that the program substantially increased girls' as well as boys' school participation in treatment neighborhoods, and that these increases were obtained at lower costs than would have been possible through the public school system.[6] For other well-identified studies of subsidy programs (as well as other types of PPP interventions), see the recent review by World Bank (2008).

Further, to the best of our knowledge, credible evidence on the effects of high-stakes testing and test-based school accountability programs is also limited, and largely come from the United States. In general, the existing body of evidence on the effects of high-stakes testing on student achievement is mixed. The evidence on strategic behavior in response to high-stakes testing is however more clear-cut. As summarized by Jacob (2002, 2007), the literature shows

---

[3] The FAS program is not the first of kind in this respect in Pakistan—the country has experimented with enrollment-related subsidies in the past, though not at this scale (see, e.g., Kim et al. 2003).

[4] One country where test-based accountability programs have come to heavily populate the educational landscape is the United States; the most prominent example is the country's No Child Left Behind (NCLB) program (Jacob 2007).

[5] From an evaluation standpoint, while it would have been useful to examine the individual effects of these innovations, it is not possible in this case as the FAS program represents a composite treatment.

[6] Kim et al. (1999b) examine a variant of the urban subsidy program which offered a per-girl student subsidy to set up and run community schools in rural communities. Although they find that participation increased for girls as well as boys (though less so), the evidence is less credible as the program was not randomly assigned to eligible communities, and the results are based on a reflexive comparison.

that schools respond to high-stakes testing by, for example, excluding poorer students from testing, putting in more effort and time into test preparation, cheating on tests, and altering testing conditions. There is also emerging evidence that the achievement gains measured in high-stakes tests are not generalizable to other assessments, suggesting that "test score inflation" may be at play; some of the examples of gaming listed above have also been shown to be behind cases of test score inflation.

Consistent with the objectives of the program, the primary question we ask in this study is: What is the causal effect of the FAS program on the number of students (overall and separately by gender) and average student learning in program schools? A secondary question we ask is: What is the causal effect of the FAS program on inputs in program schools such as the number of teachers, classrooms, blackboards, and toilets and on student-teacher and student-classroom ratios? To answer the above questions, we use school-level baseline (pre-treatment) and follow-up (post-treatment) data on school characteristics and student test performance obtained from program administrative records, phone interviews of school administrators, school surveys, and independent student learning assessments.

Although FAS program was not randomly assigned to eligible-schools, thus precluding experimental estimates of program impacts, we can arrive at nonexperimental estimates of the causal effects of the program as the treatment assignment mechanism fits a regression-discontinuity (RD) framework well. The FAS program has enrolled schools over time in rounds or phases. In the last two entry phases, namely phases 3 and 4, an independent written academic test constructed by PEF, called the Short-Listing Quality Assurance Test (SLQAT), was conducted as the final step in the program entry screening process. In order for schools to enter the program in these two phases, they had to apply to the program, pass a qualitative assessment conducted by PEF inspection teams, and then pass the SLQAT. If the school achieved the required minimum student pass rate (the cutoff) in the SLQAT, then the school becomes eligible for the program, and not otherwise. Further, in practice, all schools which became eligible, also elected to participate in the program. At that time of follow-up data collection, as phase-4 was the last entry phase, schools that took the phase-4 SLQAT were either untreated or treated based on their SLQAT pass rate relative to the cutoff—that is, the probability of treatment jumps from zero to one at the cutoff, thus yielding a sharp regression-discontinuity design. On the other hand, schools that took the phase-3 SLQAT and failed had another opportunity to take the test in phase 4—some phase-3 SLQAT "failers" took the phase-4 SLQAT and passed. Thus, the probability of treatment for phase-3 SLQAT takers also jumps at the cutoff but by less than one, as there is a positive probability of receiving the treatment for phase-3 SLQAT failers. This yields a special case of a fuzzy regression-discontinuity design.

Given these designs, the treatment effects of the program are identifiable and estimable. In the case of the sharp design, under some mild regularity conditions, the average causal effect of the program at the cutoff point is identified. In the case of our fuzzy RD design, it turns out that under the same mild regularity conditions, the average causal effect of treatment on the *untreated* at the cutoff is identified. Under both designs, treatment effects at the cutoff are estimated nonparametrically using local linear regressions. While, to its benefit, when the method is applied to the right settings, regression-discontinuity framework yields internal valid estimates, the generalizability of these estimates are likely to be limited as they are in principle only valid for narrow subpopulation: in our case for schools with pass rates near the cutoff.

Our findings on the causal impacts of the FAS program are mixed. For phase-3 schools, applying a partially-fuzzy design, we find no evidence of program impacts at the cutoff on our outcomes of interest. We posit that this is due to weak identification, a clear symptom of which was the inordinately large estimated standard errors associated with our treatment parameter estimates. In contrast, for phase-4 schools, applying a sharp design, we find robust evidence that the FAS program significantly increased the number of students, teachers, classrooms, and blackboards. The mean treatment effects at the cutoff were sizeable: our conservative estimates indicate that program schools expanded by roughly 85 students, and 3-4 teachers, classrooms, and blackboards. These treatment estimates are also large as a percentage of mean baseline values for these outcomes for phase-4 schools near the cutoff, and are particularly noteworthy given that phase-4 program schools were exposed to only 10 months of treatment before the endline phone interview data were gathered. Finally, our cost-effectiveness calculations show that the FAS program should be counted among the cheapest interventions in developing countries for raising enrollment levels.

The remainder of the paper is organized as follows. Section 2 describes the educational context in Pakistan. Section 3 describes the FAS program in detail. Section 4 lays out our identification and estimation strategies. Section 5 describes the data and provides some summary statistics for SLQAT-taking schools. Section 6 presents our impact and cost-benefit findings.

Section 7 summarizes and interprets our main findings, as well as provides some concluding comments. These comments comprise of a discussion of external validity and one specific potential threat to internal validity: treatment spillovers to non-program low-cost private schools in the vicinity of program schools.

2. The educational context

The FAS program was conceived and introduced in an educational landscape defined by three features. One: equitable access to schooling and student attainment and achievement were and continue to be acute, persistent challenges. Two: the public sector, which remains the dominant provider of education, suffers from chronic weaknesses which impair its ability to effectively address these challenges. And three: in the wake of this public sector failure, the private sector has strongly stepped in, growing dramatically in size and reach, and now represents a major policy opportunity for addressing these challenges. These three features are dealt with in turn.

The educational situation of Pakistan is poor in absolute terms, relative to other countries in its region, and relative to developing countries at its level of per capita income (see, e.g., x for a recent comparative picture). Given present trends, Pakistan is unlikely to meet the United Nation's Millennium Development Goal (MDG 2) of universal primary education by 2015. As Table 1 shows, estimates based on household survey data gathered in 2004/05, just before the FAS program was initiated, indicates large deficits in key measures. In Pakistan as a whole, only 58% of children ages 6-17 years were enrolled in a formal school (grade 1+)—this share drops to 50%, 51%, and 42% when we restrict the sample to girls, children from rural households, and children from the poorest (fifth quintile) households, respectively. In terms of attainment rates, while the vast majority of individuals ages 20-24 who entered formal school (grade 1+) completed primary school (grade 5), only 51% completed secondary school (grade 10). In terms of achievement, student assessments of students in grades 4 and 8 in a nationally-representative sample of public schools indicate that mean scores were less than 50%; relative to this benchmark, the shortfalls were particularly large in language and mathematics (Government of Pakistan 200x; Government of Pakistan 200x). On the positive side, some of these indicators lie on an upward trajectory—as Table 1 shows, between 1998/99-2004/05, school participation rates have increased for all samples, and particularly for the poor. Completion rates have also increased over this period albeit marginally.

The educational patterns and trends in Punjab largely mirror those for all-Pakistan; the province's performance relative to all-Pakistan is however mixed. As Table 1 shows, Punjab had slightly higher levels of participation in 2004/05 and better progress over the period 1998/99-2004/05 than all-Pakistan. On the other hand, completion rates for Punjab trail all-Pakistan, particularly secondary school completion: in 2004/05, secondary school participation was 45%, six percentage points lower than for all-Pakistan. In terms of student achievement, the national public school student assessments show that Punjab also had mean scores less than 50% in all tested subjects; compared to the rest of the country, Punjab preformed similarly in mathematics and marginally better in language (Government of Pakistan 200x). The absolute low learning performance in Punjab is corroborated by Das et al. (2006) and Andrabi et al. (2007) who show that learning levels of third graders in private and public schools in a selected sample of villages are far below curriculum standards across all tested subjects.

The public school system has however been unable to mount an effective campaign to improve educational outcomes due to several interrelated problems. Leaving aside the issue of the amount of available resources,[7] a first problem is how resources are allocated: in all provinces, over 90% of the education budgets are directed towards recurrent expenditures, the overwhelming share of which is teacher salaries (Husain et al. 2003). Consequently, resources for physical infrastructure upkeep and development, the provision of basic amenities, furniture, and basic teaching and learning technologies, and teacher training and professional development are scarce. In addition, resources are not efficiently and equitably allocated across educational inputs and levels of education. A second issue is resource absorptive capacity: even the limited resources allocated towards development expenditures are only partly utilized. A third issue is resource wastage: substantial resources are used up without with little positive return (*need refs for all*).

Underlying all of these problems is the issue is weak governance: the public school system lacks accountability and monitoring systems. In particular, it lacks an incentive system that promotes the legitimate and efficient use of resources for achieving key educational goals articulated in terms of clearly-defined standards and targets. The combination of these factors creates for a debilitating mix, with little sustained political will and administrative capacity and

[7] Public expenditure on education as a share of national income is less than 2%, which is significantly lower than the average shares in the South Asia region and in the developing world as a whole, and half of the target set by UNESCO for countries at its income level (UNESCO 200x).

knowhow to remedy the situation. The net result is a highly cost-ineffective public school system, with a poor record of progress in terms in enrolling children, retaining them in school, and educating them.

In contrast, the private school sector has grown dramatically in terms of the number of institutions and share of enrollment. Further, responding to the broad demand for greater access and better quality, this sector has evolved in character, becoming less elite and more of a mass system, increasingly reaching low-income and rural households.[8] As Andrabi et al. (2006) show using private school census data from 2000, there was an exponential increase in the number of private schools over the 1990s, with over 50% of existing private schools established on or after 1996. Further, they find that while existing schools set up before 1990 were predominantly in urban areas, the distribution since then has become increasingly rural. In line with this increase in institutions, the share of enrollment in private school has also increased. As Table 1 shows, in 2004/05, 25% and 27% of children ages 6-17 were enrolled in private schools in all-Pakistan and Punjab, respectively; these shares represent significant increases from 1998/1999. The increases were particularly dramatic in rural areas and for households in the poorest quintile. Similar evidence is provided by Andrabi et al. (2006) who show that the growth rate over the 1990s in private school enrollment was highest among low-income households nationally, and among middle-income households in rural areas, whereas the growth rate in public school enrollment was negative in both urban and rural areas and across the household income distribution. Finally, Andrabi et al. (2006) also show that fees in private schools are generally low: median annual fees per student in 2000 were 960 and 751 rupees in rural and urban areas, respectively, and account for a small percentage of mean annual household expenditure.

Available evidence also suggests better learning outcomes in private schools relative to pubic schools. Using data from a sample of villages with at least one private school from three districts in Punjab, Das et al. (2006) and Andrabi et al. (2007) show large and significant gaps in average learning levels of third-graders between private schools and public schools across all tested subjects; they also find that these gaps remain after controlling for relevant child, family, and village characteristics.

[8] It is conceivable that much of the growth and metamorphosis in the private sector was in direct response to the rigidities and shortcomings in the public school system.

3. The Foundation-Assisted Schools program

This section describes the main design and implementation features of the Foundation-Assisted Schools (FAS) program. We begin by providing a brief overview of the agency that administers the program, the Punjab Education Foundation (PEF). We then describe, in turn, the FAS program's coverage, benefits, and benefit maintenance rules. A key benefit maintenance rule is a minimum level of school performance in a specially-constructed written academic test offered semi-annually by PEF called the Quality Assurance Test (QAT). A pared-down version called the Short-Listing Quality Assurance Test (SLQAT) is also administered by PEF to screen in applicant schools into the program. Given the primary importance of school performance on these tests for initial and continued program qualification, we describe the main design and implementation features of these tests. We conclude the section by describing the program entry rules. Our identification strategy is based on parts of this information.

Institutional background: The FAS program is administered by PEF, a publicly-funded semi-autonomous statutory organization established in 1991 which serves as the main institutional mechanism for PPPs in education in Punjab. The organization's primary aims are to enable socioeconomically-disadvantaged households to access private education and to raise the quality of education in low-cost private schools (the FAS program has identical objectives). To this end, it employs an array of instruments such as start-up and operational subsidies, vouchers to households in poor localities, school management and teacher training, and specialist teacher services to promote private education. The FAS program is PEF's largest program. In the 2007-08 fiscal year, PEF spent 1.1 billon rupees (US$15.8 million) on FAS benefits.[9] This amount accounted for 61% of total expenditures in that year.

Program coverage and timeline: The FAS program was initiated in November 2005 on a pilot basis in 54 schools in five districts in Punjab. Since then, the program has been expanded in phases to cover additional districts as well as more schools within program districts. (See Table 2 for a timeline of the FAS program highlighting the start-dates of each of the phases as well as the dates of other key events.) At present, the program has completed four entry phases (the pilot phase represents phase 1), and, as of June 2008, covers 1,082 private primary, middle, and

[9] FAS program administrative costs are unavailable as PEF does not disaggregate administrative costs by program. We however know that total administrative costs were less than 1% of total expenditures in fiscal year 2007-08, suggesting that FAS program administrative costs are negligible relative to current FAS program benefit costs.

secondary schools in 18 out of the 35 districts in Punjab.[10] Of these, 945 schools (87%) are located in just seven districts (see Table 3 for a district- and phase-wise disaggregation of the current number of program schools as well as Figure 1 for the location of program districts in Punjab). This number represents a significant share of all private schools in these seven districts: using the 2005 National Education Census (NEC), a survey-based census of schools in Pakistan, we estimate that the program presently covers 21% of private schools in these districts (see Table 4).[11]

Program location: The program was initially designed to be targeted at districts ranked lowest in terms of adult literacy rates, based on data obtained from the 2003-04 Multiple Indicators Cluster Survey. However, in phases 1 and 2, this targeting decision was not applied as PEF was limited in its institutional and logistical capacity at that time to effectively administer the program in poorly-ranked districts that had physical environmental challenges and limited transportation and accommodation options (Malik 2007). In contrast, in phases 3 and 4, the targeting decision was effectively applied. Consequently, as shown in Table 3, 51% and 89% of current program schools are located in districts ranked among the bottom-quarter and bottom-half, respectively.

Program school characteristics: Table 5 presents the distribution of current program schools by selected characteristics. Combining the four phases, the mean school size is 351 students. The majority of schools are middle level (59%), coeducational (83%), registered with the local government authorities (87%), and rural (55%). Disaggregating program schools by phase, the distribution of schools by these characteristics are roughly comparable except in the case of school level: in contrast to the pattern in the aggregated sample, current program schools that entered in phases 1 and 2 were mainly secondary schools (65-73%). School size also appears to be monotonically increasing by phase; the mean size of current program schools that entered in phase 1 is 561 students, whereas that of current program schools that entered in phase 4 is 351 students. While there may be multiple competing explanations for this pattern, length of exposure to the FAS program is clearly one of them.

[10] The FAS program also currently covers three higher secondary schools (schools that contain classes 11-12).

[11] NEC's coverage of private schools might be incomplete as, according to the survey documentation, private schools were identified in the field by interviewers with the assistance of local officials. It is conceivable that private schools that are, for example, unregistered or very small, or operate in obscure locations are more likely to be missed by interviewers. Thus, the estimate of 21% can be viewed as an upper-end estimate of FAS program coverage of private schools.

Program benefits: The FAS program offers three types of monetary benefits. These benefits were introduced at different times over the course of the program's life. Initiated in November 2005 as the program's original benefit, the first type of benefit is an enrollment-related subsidy: the school receives a monthly per-student subsidy amount of 300 rupees (US$4.3) up to a maximum of 750 students (i.e., the total amount is capped at 225,000 rupees or US$3,214).[12] Given a mean school size of 215 students at the time of application in phase-3 and phase-4 schools, the mean monthly subsidy payment is roughly 64,500 rupees (US$921). The determination of the subsidy amount was guided by a survey conducted by PEF in 2005 in selected districts which showed that the vast majority of private schools that operate in rural areas and disadvantaged urban neighborhoods charge between 50-400 rupees per month (US$0.7-5.7) in tuition and fees—based on this information, the subsidy amount was set at the upper-segment of this price range (Malik 2007). The subsidy benefit is paid on a monthly basis for all twelve months of the year. To facilitate timely and regular payments, starting in August 2007, the benefit amounts have been transferred to schools' bank accounts electronically.

The second type of benefit is a teacher bonus for a high level of school test performance: once every academic year, a maximum of five teachers in each program school where at least 90% of students in tested classes obtain a score of 40% or higher in the QAT receive an award of 10,000 rupees (US$143) each. This is a substantial bonus for the teaching workforce in program schools: using the available data on maximum monthly teacher salaries at the time of application to the program for phase-3 and phase-4 program schools, we estimate that the bonus amount translates to roughly 370% of mean maximum monthly teacher salaries. This bonus size is approximately an order of magnitude larger than teacher bonus programs in general (see Glewwe et al. 2003). In practice, the bonuses are offered to teachers who taught the classes that were tested in the QATs. As with the subsidy payments, the bonuses are transferred electronically to the personal bank accounts of the award recipients. PEF announced the teacher bonus benefit to program schools in December 2006. To date, two rounds of teacher bonuses have been awarded: in January 2007 based on the QAT 2 results and in February 2008 based on the QAT 4 results.

The third and final benefit is a competitive school bonus for top school test performance: once every academic year, the program school in each district which has the highest share of

[12] In September 2008, the monthly per-student subsidy was raised to 350 rupees (US$5) for primary and middle schools, and to 400 rupees (US$5.7) for secondary schools. The enrollment cap for the subsidy amount was originally set at 500 students and has been raised incrementally over time.

students with a score of 40% or higher[13] in the QAT is awarded 50,000 rupees (US$714). This bonus is sizeable as well: given a mean school size of 215 students at the time of application in phase-3 and phase-4 program schools, the bonus amount translates to roughly 76% of the mean monthly subsidy payment to schools. Again, the bonuses are transferred electronically to the bank accounts of the award recipients. PEF announced the school bonus benefit to program schools in February 2007. To date, one round of the school bonuses has been awarded: in February 2008 based on the QAT 4 results.

Benefit maintenance rules: Once schools qualify for and join the FAS program, benefit maintenance requires (1) a minimum level of total enrollment of 100 students; (2) eliminating tuition and fees[14] to all students (and, importantly, publicly announcing this by placing a PEF-issued signboard outside the school gate); and (3) achieving the minimum student pass rate of 67% in the QAT. These conditions are applied stringently. If the school fails to comply with the first two conditions at any time, it is disqualified with immediate effect. If the school fails to achieve the QAT-related condition over three (consecutive??) attempts, the school is also disqualified.

There are also other conditions for benefit maintenance. These include (1) maximum student-teacher and student-classroom ratios of 30:1; (2) conducting only one class in a classroom in any period; (3) registering with school with the District Registration Authority within a year of program entry; (4) maintaining or upgrading the quality of the school's physical infrastructure (e.g., adequate classroom space; properly-constructed walls, floors, and roofing; sufficient ventilation; sufficient artificial or natural light); and (5) adequate furniture and teaching tools (e.g., benches, desks, and blackboards).[15] These additional conditions are applied more leniently—typically, when PEF detects a violation among this subset of conditions, schools are provided with a grace period within which to comply. However, once a program school is disqualified, irrespective of the reason, the school is permanently disqualified.

Table 6 presents the number of program schools that were disqualified, disaggregated by phase. The reason for disqualification is predominantly serial failure to satisfy the minimum performance level on the QAT. Further, all schools disqualified for this reason entered the

[13] Ties are broken by looking at mean test scores among the relevant schools.
[14] Fees for board examinations are exempt.
[15] Benefit maintenance is also tied to limiting school administration costs to 9% of total operational costs, and limiting the use of the subsidy to specified areas such as paying for building and furnishing classrooms, teacher salaries and training, and learning technologies. This condition has not been applied to date.

program in phases 1 and 2. This reason for this inter-phase pattern is straightforward: only phase-1 and phase-2 schools have been tested three times using the QAT and hence have the necessary observations for this condition for continued benefit eligibility to be applied, whereas, to date, phase-3 and phase-4 schools have been tested using the QAT twice and once, respectively. Notwithstanding, the number of schools that have been disqualified for any reason is negligible: only 28 out of the 1,111 program entrants (2.5%) have exited the program over time. More importantly for our analysis, only three schools that entered the program in phases 3 and 4 have exited the program for any reason. Thus, program dropout is (so far) not an issue with these school samples.

Direct impact channels: The structure of the program (both the benefits and benefit maintenance conditions) delineated above can be expected to have a positive effect on participation, equity, and student achievement via several direct channels. First, setting the monthly subsidy as a linear function of the number of children enrolled incentivizes program schools to draw in additional students. Second, setting the per-student subsidy payment low makes it more likely that only private schools which already cater to low-income households and/or operate in rural areas and disadvantaged urban neighborhoods choose to participate in the program. Third, tying the receipt of program benefits to the elimination of tuition and fees (which puts the program school's price at an advantage over other local non-program competitor private schools and at par with public schools) is likely to raise the demand for schooling, particularly among households for whom the direct cost of schooling serves as a major impediment to school participation.[16] Further, given the presence of strong gender-related biases in Pakistan, to the extent that the direct cost of schooling is a greater impediment for girls than boys, eliminating this cost is likely to have a larger effect on girls' school participation relative to boys'. Fourth, typing the continued receipt of program benefits to a minimum level of test performance introduces a stick incentive which potentially raises student achievement. Fifth and final, offering test performance-based bonuses to teachers and schools introduces a carrot incentive which also potentially raises student achievement.

Some of these same design features can, however, also introduce countervailing incentives, which may arrest gains in participation, equity, and achievement. For example, while

[16] An increase in enrollment in program schools may not translate into a commensurate increase in participation among children in areas where program schools operate. The lower price of schooling in program schools is likely to induce both displacement and diversion effects: some share of the new demand for enrollment in the program school will likely come from students already enrolled in other schools or children that where initially *considering* enrolling elsewhere.

the per-student subsidy may encourage schools to expand enrollment, if the entrants are drawn from the out-of-school child population, they are likely to possess lower learning abilities. Consequently, enrolling such children may diminish the school's ability to meet the minimum pass rate requirement for continued benefit receipt. Further, to the extent that lower learning ability is associated with individual and family characteristics such as poverty status, gender, disability status, and ethnicity, if program schools systematically screen in the more able from the pool of potential entrants to protect their ability to meet the minimum pass rate requirement, this would impair equity gains. What's more, this tension between more students and meeting the minimum pass rate requirement is likely to grow stronger over time for schools as the pool of entrants is tapped from the top-down in terms of learning ability.

Tying continued program eligibility to a minimum level of test performance can also operate independent of the interplay with the per-student subsidy to introduce undesirable incentives. The minimum pass rate requirement introduces a high-stakes test situation wherein program schools face high-powered incentives to screen in (enroll) strong learners and screen out (release) poor learners as well as other forms of strategic behavior such as spending more teaching and preparation time on subjects and learning skills tested on the test and increased student effort on the test day which may not necessarily translate into general and long-lasting improvements in learning. In fact, at the extreme, it is plausible that strategic inclusionary and exclusionary practices can result in the minimum pass rate requirement being satisfied simply by changing the composition of students in terms of preexisting learning levels (pure selection) than by improving the learning levels for a given student composition (pure influence). Moreover, using a minimum threshold rather than tying test performance to, for example, (improvements in) the average performance of students creates incentives for schools to concentrate teaching efforts on children whose marginal performance matters for meeting the minimum threshold; as a result, high-performing and extremely low-performing students are likely to be relatively neglected. The extent to which the minimum threshold produces such behavior is however potentially attenuated by the performance-based bonuses for teachers and schools which introduce incentives for schools to extend their teaching efforts to students lower down in the learning (ability) distribution. These incentives however can work to reinforce high-stakes testing in promoting test score inflation.

The structure of the program can also be expected to directly affect investments in the quantity and quality of school inputs and resources as well. For example, increases in enrollment induced by the per-student subsidy have to be met by increases in the number of classrooms and teachers if the stipulated maximum student-teacher and student-classroom ratio conditions are to be complied with. The maximum ratios also induce program entrants with preexisting ratios in excess of the maximums to invest in additional classrooms and teachers. In addition, the physical infrastructure and learning environment quality conditions encourage program schools to ensure the proper design and construction of infrastructural expansions, and to invest in teaching tools (e.g., blackboards) and amenities (e.g., toilets), in step with enrollment growth. These input-related conditions also encourage schools to schedule investments in school resources and amenities to either lead or accompany enrollment increases. However, given that these conditions are not stringently applied, it suggests that PEF might tolerate investments in school inputs that lag enrollment increases, although by not too long.

The Quality Assurance Tests: The Short-Listing Quality Assurance Test (SLQAT) and the Quality Assurance Test (QAT) are used to assess the initial and continued eligibility of schools respectively, based on the share of students who satisfy largely arbitrarily-set minimum levels of learning. Specifically, a school is accepted for the program if at least 67% of its tested students receive 33% or higher on the SLQAT. A school remains in the program if at least 67% students in the tested grades receive 40% or higher on the QAT.

The tests are designed as follows.[17] First, a team of PEF subject specialists develop subject- and grade-specific content lists that a school would be expected to cover in an academic year. This exercise is undertaken by studying a sample of common syllabi and textbooks used by low-cost private schools. Second, based on these content lists, the subject specialists develop a multiple-choice question bank that spans almost the full range of Bloom's (1956) taxonomy of learning levels. The subject areas covered in the tests are English, Urdu, mathematics, and science (general science in grades 1-8 and biology, chemistry, and physics separately in grades 9-10). The tests are produced for grades 3, 4, 6, 7, and 9.

The SLQAT constitutes a pared-down version of the QAT. It is intentionally designed to be simpler than the QAT. While the QAT tests five levels of cognitive learning: knowledge, comprehension, application, analysis, and synthesis, the SLQAT tests only the two lowest levels:

[17] Sample test papers are available at http://www.pef.edu.pk/ModelPapers.htm (Last accessed: September 30, 2008) .

knowledge and comprehension. Whereas the QAT tests learning in all covered subject areas, the SLQAT tests learning in English, Urdu, and mathematics only. The QAT and SLQAT have time limits of 65 and 55 minutes, respectively.

The SLQAT is administered as follows. PEF offers the SLQAT once the applicant school has cleared the physical inspection screening (this screening stage is described later). Each school is tested separately, and, depending on the school's level and the number of grades it contains, two to three grades are tested. Which grades are tested in a given school are not disclosed in advance and are randomly selected by the subject specialists. Further, multiple test papers are prepared for each grade, and which specific test papers are offered in a given school are also randomly determined. The SLQAT is transported in sealed envelops and opened only in the presence of the school administrator and teachers. At that time, both the PEF inspection team and the school learn which grades are selected for the SLQAT. All students present in school in the grades selected for the SLQAT are expected to take the test. Test invigilation is coordinated by a subject specialist assigned to the inspection team. The completed tests are transported back by the inspection team to PEF headquarters and graded by the subject specialists. The SLQAT scores for all tested students are pooled together across grades before the school's SLQAT pass rate is calculated. The school passes the SLQAT if at least 67% of tested students score 33% or higher on the test. To date, two rounds of the SLQAT have been offered: as part of the entry process in phases 3 and 4 (see Table 2 for dates).

To date, five QATs have been conducted (see Table 2 for the dates). The first and second QATs were administered by PEF. Starting from the third QAT however, for logistical and credibility reasons, the administration of the QAT was competitively outsourced to Punjab University and the Bahawalpur Board of Intermediate and Secondary Education. The QAT is offered twice every academic year in September/October and March/April. Given these timings, one-third of the syllabus is tested in the fall round and the full syllabus is tested in the spring round. The test papers are developed by PEF and provided to the contracted organizations to administer.

The general test administration procedures outlined above with respect to the SLQAT are followed with the QAT as well. There are, however, some important differences. Unlike with the SLQAT, the school receives formal advance notices of the date of the QAT. In addition, all students are expected to be in school on the day of the test—this is verified by looking at the previous month's enrollment statement sent by the school to PEF for the payment of the subsidy and the school's enrollment records on the day of test. If a student in a grade selected for the QAT is not present on the day of the test, that student receives a zero on the test (this decision was made to discourage schools from selecting the test-taking pool). The tests are graded by the Punjab University Education Testing Service. As with the SLQAT, the individual student scores are pooled together across grades to calculate the school score. Only the school scores are communicated to PEF. The school passes the QAT if at least 67% of students in the tested grades score 40% or higher.

Program entry rules: Program entry follows a stepwise process. First, schools apply to the program when a call for applications is issued in newspapers by PEF. Application eligibility is restricted to existing private primary, middle, and secondary schools with a minimum enrollment of 100 students from the districts listed in the call.[18] Except in a few cases, schools that submit completed applications by the announced deadline are considered. Second, PEF inspection teams visit the schools to verify the data provided in the applications and assess the local reputation of the school and the quality of the physical infrastructure and schooling environment. A points and weighting scheme is provided by PEF to the inspection teams—this scheme primarily helps ensure that the same school attributes are considered across schools and by different inspection teams. However, the screening exercise is largely subjective: whether a school qualifies depends principally on qualitative impressions gathered by the inspection team on a range of variables. Schools that applied to the program in phases 1 and 2 were subject to this inspection-based screening only.

Schools that applied to the program in phases 3 and 4 were subject to a subsequent round of screening. Conditional on passing the inspection screening, all students present on the day of the inspection screening in selected grades in the school were offered the SLQAT. The school is considered eligible for the FAS program if at least 67% of its tested students receive 33% or

[18] PEF originally expected that, given that the subsidy replaces student tuition and fees, only schools which charge students a monthly tuition of 300 rupees or less would be interested in the program, given that they stood to gain additional revenue from the program. Hence, they also stipulated that application eligibility was limited to schools that charged 300 rupees or less in tuition. However, interestingly, they found that schools that charge (significantly) more than that amount were interested in the program as well. A potentially important reason behind this behavior—corroborated by anecdotal evidence gathered by us during field visits—is that the regular (risk-free, at least in the short-term) monthly subsidies from PEF resolved the uncertainty associated with collecting tuition and fees from students. That is, the school's revenue generated by the subsidy was at least as high (exceeded the certainty equivalent) as the *expected* revenue from collecting tuition and fees from students.

higher on the SLQAT. The percentage on the test itself (33%) is not explicitly related to some notion of a minimum level of academic learning and the SLQAT was not intentionally introduced to target the program at schools with relatively higher initial levels of student learning. Rather, in light of the problems that program schools who entered in phases 1 and 2 faced in complying with the QAT condition for benefit maintenance, PEF introduced the SLQAT at program entry to reduce the size of the learning "step" that screened-in schools have to climb in order to comply with the QAT condition.

Table 7 presents the numbers of schools in each stage of the FAS program entry process as reported by PEF. Of the 1,070 and 1,430 schools that were inspected in phases 3 and 4, respectively, 799 (75%) and 872 (61%) schools were offered the SLQAT, suggesting that inspection screening did effectively exclude schools from further consideration. Of the schools that were offered the SLQAT, 514 (64%) and 431 (49%) schools achieved the minimum pass rate and became eligible for the FAS program.[19] In terms of treatment uptake, PEF reports that 482 and 425 program-eligible schools in phases 3 and 4 respectively signed the formal program participation agreement; these numbers translate into treatment uptake rates of roughly 94% and 98% in the two phases. Given this, when we develop our empirical strategy, treatment uptake is stylized as perfect.

The school pass rates on the SLQAT and the application of a minimum pass rate of 67% provide an observable, quantitative criterion on which the program eligibility status of schools that applied to the program and cleared the inspection screening in phases 3 and 4 is determined. This information is crucial for developing our identification strategy. Separate strategies will however need to be designed for the phase-3 and phase-4 school samples. This is due to the fact that the call for applications has been repeated over time, and hence, ineligibility for the program at entry in a given phase is not necessarily a permanently fixed state. In the case of schools that sought program entry in phase 4, ineligibility is a fixed state only because phase 5 has not been initiated by PEF yet.[20] On the other hand, schools that sought program entry in phase 3 and failed to achieve the minimum pass rate in the phase-3 SLQAT had the opportunity to re-seek entry into the program in phase 4—some of the phase-3 SLQAT failers achieved the minimum pass

[19] Our own tallies based on the SLQAT data slightly deviate from the above numbers: we find that 796 and 856 schools took the SLQAT and, out of these, 511 and 432 schools achieved the minimum pass rate in phases 3 and 4, respectively.

[20] To be precise, applications for phase 5 have been accepted by PEF but the screening of schools have been put on hold indefinitely.

rate in the phase-4 SLQAT and became eligible for the program then. The next section formally lays out our identification and estimation strategies.

4. Empirical strategy

Identification

Following the exposition in van der Klaauw (2008) and Todd (2006), let y_i denote the outcome of interest (e.g., enrollment size, average student test performance) in school *i*, and let the indicator variable $d_i \in \{0,1\}$ denote treatment assignment, where one denotes that the school receives the FAS program, and zero otherwise. Further, let y_{0i} and y_{1i} denote the potential outcomes of school i in the untreated and treated states, respectively. The actual outcome observed for school i is given by

$$y_i = d_i y_{1i} + \left(1-d_i\right) y_{0i} = y_{0i} + \left[y_{1i} - y_{0i}\right] d_i = y_{0i} + \alpha_i d_i, \tag{1}$$

where α_i denotes the treatment effect for school *i*.

Unless treatment is randomly assigned, simply comparing average outcomes of treated schools to untreated schools, $E\left[y_{i1} \mid d_i = 1\right] - E\left[y_{i0} \mid d_i = 0\right]$, would not yield an unbiased estimate of the average *causal* effect of the treatment: $E\left[\alpha_i\right]$. In general,

$$E\left[y_{1i} \mid d_i = 1\right] - E\left[y_{0i} \mid d_i = 0\right] = E\left[\alpha_i \mid d_i = 1\right] - \left\{E\left[y_{0i} \mid d_i = 1\right] - E\left[y_{0i} \mid d_i = 0\right]\right\}. \tag{2}$$

The first term on the right-hand side of (2) is the average treatment effect on the treated and the second term is the selection bias arising from differences in potential average untreated outcomes between treated and untreated schools. Selection is plausible in our case. For example, if, ex ante, FAS program schools have higher levels and better quality inputs and resources or have students with higher learning abilities relative to non-program schools, then the selection bias term will be positive and lead to an upward bias in our estimate of the average treatment effect.

This selection bias problem can however be overcome by using the institutional feature that FAS program eligibility is ultimately determined by the student pass rate obtained by the school in the SLQAT relative to the known pass rate cutoff of 67%. That is, schools that obtain pass rates equal to or greater than the cutoff are eligible for program participation, and not otherwise. Given that nearly all schools that become eligible to participate in the FAS program also choose to participate in the program, in practice, the cutoff also determines actual program participation. Thus, program participation status is assigned based on the decision rule

$$d_i(z_i) = 1\{z_i \geq c\}, \tag{3}$$

where z_i denotes school i's pass rate which is perfectly observed (z will be more generally referred to as the assignment variable), c the known distinct cutoff pass rate, and $1\{\cdot\}$ an indicator function. This design for treatment assignment based on a completely deterministic function is referred to as a sharp regression-discontinuity design (Trochim 1984).

Now the SLQAT pass rate z might be correlated with our outcomes of interest. For example, SLQAT pass rates are expected to be positively correlated with mean student test outcomes. Thus, the treatment assignment mechanism is clearly not random and comparing schools that receive the treatment ($d_i(z_i) = 1$) to schools that do not ($d_i(z_i) = 0$) will yield a biased estimate of the average treatment effect. If however we consider schools with pass rates near the cutoff to be comparable or exchangeable, then the treatment assignment mechanism in the neighborhood of the cutoff can be viewed as if it was almost random. More formally, let $e > 0$ denote an arbitrarily small number. Comparing the outcomes of schools with pass rates just below the cutoff (hereafter referred to as marginal failers) with the outcomes of schools with pass rates at or just above the cutoff (hereafter referred to as marginal passers), gives us

$$E(y_{1i} \mid z_i = c+e) - E(y_{0i} \mid z_i = c-e) = E[\alpha_i \mid z_i = c+e] - \{E[y_{0i} \mid z_i = c+e] - E[y_{0i} \mid z_i = c-e]\}. \tag{4}$$

Under the assumptions that (1) the limit $\lim_{e \downarrow 0} E[y_{0i} \mid z_i = c+e]$ is well defined, (2) $E[y_{0i} \mid z_i = c]$ is continuous in the assignment variable z at the cutoff, that is, the conditional expectations of the outcome variable exhibits local smoothness in the absence of the treatment, and (3) the density of the assignment variable z is positive in the neighborhood of the cutoff, the difference in the mean outcome between marginal passers and marginal failers identifies

$$E[\alpha_i \mid z_i = c] = \lim_{e \downarrow 0} E(y_{1i} \mid z_i = c+e) - \lim_{e \downarrow 0} E(y_{0i} \mid z_i = c-e), \tag{5}$$

which represents the average treatment effect on the treated (ATT) at the cutoff (Hahn et al. 2001; Todd 2006).[21]

[21] The treatment parameter could also be interpreted as a marginal average treatment effect (MATE) (Heckman and Vytlacil 2005). This interpretation is useful when considering the treatment impact implications of marginally lowering the cutoff.

The phase-4 schools data fit a sharp design well, as their pass (eligibility) versus fail (ineligibility) status remains fixed at the present time. A variant of the above framework will however be required for phase-3 schools given that some phase-3 failers reapplied to the program in phase 4 and passed the SLQAT at that time. As a result, our set of phase-3 SLQAT takers can be divided into three distinct subsets: (1) phase-3 failers who did not enter the program anytime later, (2) phase-3 failers who entered the program later (crossover schools), and (3) phase-3 passers who entered the program at that time (as noted before, the treatment uptake rate by passers was near perfect, and there have been no dropouts to date). While the pass rate cutoff rule strictly determines program eligibility status for phase-3 SLQAT takers, the processes and factors behind the phase-4 application decision of phase-3 failers are likely to be nonrandom and unobservable to us. As a result, the eventual participation status of this group of schools is also considered to be driven partly by selection on unobservables.

Given the structure of the problem for phase-3 schools, the probability of FAS program participation as a function of the pass rate in the phase-3 SLQAT, $E(d_i \mid z_i) = \Pr[d_i = 1 \mid z_i]$, consists of a smaller jump than one at the cutoff. Despite this, the probability of program participation will still be discontinuous in the assignment variable z at the cutoff. This feature implies that the phase-3 schools data more appropriately fit a fuzzy regression-discontinuity design (Trochim 1984). Under this design, apart from the identifying assumptions for the sharp design, if we assume that treatment d_i is nondecreasing in the assignment variable z at the cutoff (local monotonicity) and that potential outcomes are independent of treatment conditional on the assignment variable z at the cutoff (local conditional independence), then the average causal effect of the treatment is identified by

$$E[\alpha_i \mid z_i = c] = \frac{\lim_{e \downarrow 0}[y_{1i} \mid z_i = c+e] - \lim_{e \downarrow 0}[y_{0i} \mid z_i = c-e]}{\lim_{e \downarrow 0}[d_{1i} \mid z_i = c+e] - \lim_{e \downarrow 0}[d_{0i} \mid z_i = c-e]}, \tag{6}$$

where the denominator of (6) is nonzero given the discontinuity in the probability of treatment at the cutoff (Hahn et al. 2001; Todd 2006).

This treatment parameter represents the regression-discontinuity analog to the local average treatment effect (LATE) (Imbens and Angrist 1994; Angrist et al. 1996). More explicitly, the treatment parameter can be interpreted as the average treatment effect for schools in the neighborhood of the cutoff who are induced to take up treatment when their pass rates are at or above the cutoff $(d_i(z_i)=1)$ but not so when they are below $(d_i(z_i)=0)$.[22]

The general fuzzy design setup above allows for partial (or imperfect) compliance for both treatment-eligible and treatment-ineligible groups, that is, it allows for both crossovers (treatment-ineligibles taking up treatment) and no-shows (treatment-eligibles not taking up treatment) (Bloom 1984). In our particular case, the problem is essentially one of one-way partial compliance: specifically, we have crossovers over time from treatment-ineligible to treatment-eligible status (with all treatment-eligible schools choosing to take up treatment). Hence, the probability of program participation for phase-3 SLQAT passers is one, while the probability for failers lies between zero and one. Battistin and Rettore (2008) describe the other one-way partial compliance case, where treatment-ineligible agents are not exposed to treatment and treatment-eligible agents self-select into treatment. They formally show that the LATE at the cutoff can be identified simply under the conditions required for identifying the ATT at the cutoff under the sharp design, thus labeling their framework as a "partially" fuzzy design. Their framework straightforwardly applies to our case of one-way partial compliance as well. However, whereas in their specific setup, they identify the ATT at the cutoff ($\lim_{e \downarrow 0} E[\alpha_i \mid z_i = c+e]$), by applying the smoothness condition to obtain the counterfactual average outcome for the untreated: $\lim_{e \downarrow 0} E[y_{i1} \mid z_i = c-e]$ —we instead identify the average treatment effect on the untreated ($\lim_{e \downarrow 0} E[\alpha_i \mid z_i = c-e]$) (Duflo et al. 2007).

Estimation

Choice of estimator: As Nichols (2007b) notes, the regression-discontinuity treatment parameters can be estimated in several ways. For example, in the case of the sharp design, a first way is to simply compare average outcomes between marginal treated and marginal untreated schools. A second way is to estimate the average treatment effect via parametric regression of the outcome on various powers of the assignment variable *z*, the program participation indicator variable *d*, and interactions of all *z* terms with *d* for the full sample of schools—the estimated coefficient on

[22] In the terminology used by Angrist et al. (1996), the fuzzy regression-discontinuity design treatment effect represents the LATE at the cutoff for compliers. The local monotonicity assumption required for identification rules out defiers.

the treatment indicator d will then yield the average treatment effect at the cutoff. This is the most common strategy adopted in applied studies using regression-discontinuity designs (McCrary and Royer 2006). An important concern with the global smoothing approach however is that misspecification of the polynomial function of the assignment variable z in the regression can yield inconsistent estimates of the treatment effect; further, the estimates may be sensitive to the behavior of the regression away from the cutoff (Imbens and Wooldridge 2008). Given that our interest lies in estimating the effect at a single point using observations in its neighborhood, a local smoothing approach using nonparametric regression is relatively better suited to the problem at hand. In addition, potential misspecification bias is generally less of a problem with the latter approach (van der Klaauw 2007).[23]

However, while nonparametric methods generally perform well in terms of the asymptotic bias of the estimator when the point of interest is an interior point on the support of z, many perform poorly when the point is a boundary point, which is always the case with regression-discontinuity problems. As Hahn et al. (2001) note, the inferior performance of nonparametric estimators such as the kernel regression arises because the bias converges to zero at a slower rate at boundary points than at interior points. Consequently, the resulting bias can be serious in finite samples.

Local polynomial regression has been shown to have better boundary properties than standard kernel regression (Fan 1992; Fan and Gijbels 1996), and consequently is attractive for use in regression-discontinuity estimation. Specifically, it has better convergence rates than and significant bias-reduction relative to standard kernel regression. The estimator achieves this superior performance by directly accounting for the shape of the conditional expectation near the boundary (Porter 2002). In addition, as noted by Galdo et al. (2007) the local polynomial estimator is found to be robust to different data design densities.

Following Hahn et al. (2001), we opt for local linear regression (i.e., a local polynomial of order one). In the case of the sharp design which applies to our phase-4 SLQAT takers, the application of local linear estimation would simply entail individually estimating the conditional expectation of the outcome y at the cutoff from below (denoted by $-$) and above the cutoff

[23] Notwithstanding, as McCrary and Royer (2006) delineate, the parametric approach offers some advantages over the nonparametric approach such as ease of implementation of the method as well as accompanying diagnostics. The parametric approach also allows the inclusion of control variables which potentially improves efficiency.

(denoted by +) and then subtracting the two estimates. More formally, let $\hat{\alpha}_y^+$, $\hat{\alpha}_y^-$, $\hat{\beta}_y^+$, and $\hat{\beta}_y^-$ denote the solutions to the following minimization problems:

$$\left(\hat{\alpha}_y^+, \hat{\beta}_y^+\right) = \arg\min \frac{1}{n} \sum_{i=1}^{n} k\left(\tfrac{z_i - c}{h}\right) d_i \left[y_i - \alpha - \beta\left(z_i - c\right)\right]^2 \tag{7}$$

$$\left(\hat{\alpha}_y^-, \hat{\beta}_y^-\right) = \arg\min \frac{1}{n} \sum_{i=1}^{n} k\left(\tfrac{z_i - c}{h}\right)\left(1 - d_i\right)\left[y_i - \alpha - \beta\left(z_i - c\right)\right]^2, \tag{8}$$

where $k(\cdot)$ is the kernel, the weighting function, and $h > 0$ the bandwidth or the window width in which the kernel function is applied. Then, under the sharp design, a consistent estimate of the ATT at the cutoff is given by

$$\hat{\alpha}_y = \hat{\alpha}_y^+ - \hat{\alpha}_y^-. \tag{9}$$

In the case of the fuzzy design, the application of local linear regression would entail estimating the ratio of two differences. Given our particular problem related to phase-3 SLQAT takers, here, we would individually estimate the conditional expectations of the outcome y at the cutoff c from above and below (the estimations of $\hat{\alpha}_y^+$ and $\hat{\alpha}_y^-$ are shown above), and the conditional expectation of treatment d from below only (as the conditional expectation of treatment d from above is defined as one). Given this, let $\hat{\alpha}_d^-$ and $\hat{\beta}_d^-$ denote the solutions to the following minimization problem:

$$\left(\hat{\alpha}_d^-, \hat{\beta}_d^-\right) = \arg\min \frac{1}{n} \sum_{i=1}^{n} k\left(\tfrac{z_i - c}{h}\right)\left(1 - d_i\right)\left[d_i - \alpha - \beta\left(z_i - c\right)\right]^2 \tag{10}$$

Then, under our particular partially fuzzy RD design, a consistent estimate of the LATE at the cutoff is given by

$$\frac{\hat{\alpha}_y}{\hat{\alpha}_d} = \frac{\hat{\alpha}_y^+ - \hat{\alpha}_y^-}{1 - \hat{\alpha}_d^-}. \tag{11}$$

To perform statistical inference, empirical standard errors for (9) and (11) are obtained using the standard i.i.d. nonparametric bootstrap with paired sampling $\left(y_i, z_i\right)$, resampling from the original data 500 times.[24]

Choice of kernel and bandwidth: Implementation of local linear estimation requires the specification of the kernel k and bandwidth h. We opt for the triangular kernel

[24] Standard errors for the estimators can also be obtained analytically (see, for example, Porter 2003 and Imbens and Lemieux 2008 for options).

($k = \max\left\{0, 1 - \left|\frac{z_i - c}{h}\right|\right\}$) given that it is boundary optimal and thus well suited to regression-discontinuity problems (Cheng et al. 1997). As Imbens and Lemieux (2008) note, while more sophisticated kernels are available, they do not provide any significant gain in asymptotic bias reduction. Further, they note that, in general, parameter estimates appear to be robust to the choice of kernel; however, if the parameter estimates are found to be sensitive to the choice of kernel type, they claim that it likely reflects sensitivity to the choice of bandwidth. Consequently, following their guidance, we fix our selected kernel type and investigate the robustness of our findings to alterative bandwidth choices.

The choice of bandwidth size is a relatively more important decision given the well-known trade-off between estimation bias and variance. Consensus in the literature on valid and robust methods for bandwidth selection for regression-discontinuity problems is presently lacking. This is, in part, due to two reasons. First, standard automated procedures for bandwidth selection for nonparametric regression such as plug-in and cross-validation methods have been developed with the estimation of functions at interior points in mind; the nature of the trade-off between bias and variance at interior points may differ from the nature of the trade-off at boundary points (Ludwig and Miller 2005). Second, the literature on the development of bandwidth selectors customized to the regression-discontinuity context is an emergent one (Imbens and Lemieux 2007; McCrary and Royer 2006). In addition, the relative finite sample performances of the proposed bandwidth selection methods for regression-discontinuity designs have yet to be assessed.

Given this, we discount the specific method used to select the optimal bandwidth size and instead emphasize checking the robustness of our findings to reasonable deviations in bandwidth size from the optimal bandwidth choice. To select the optimal bandwidth size, we simply rely on the default bandwidth choice in Nichols' (2007a) `rd` program for Stata which assigns positive weight to at least 30 observations on each side of the cutoff, and applies the same bandwidth size for the local linear estimators of the conditional mean outcomes (and probabilities of treatment) above and below the cutoff (i.e., $h^{+} = h^{-} = h$). In our case, depending on the round of data examined, the default choice translates to a bandwidth size of three or four percentage points. We believe that this choice is on the conservative side, and satisfactorily balances restricting estimation to a local neighborhood around the cutoff against having sufficient statistical power to yield informative estimates of treatment effects. Notwithstanding, we examine whether our findings (both estimates of the treatment effects and their associated standard errors) are sensitive to selected increases in bandwidth size. We however do not check for robustness using bandwidth sizes smaller than the `rd` program's default size as, given our data, the resulting estimates are likely to be noisy.

5. Data and sample

Baseline data: The baseline data come from the application records and SLQAT test records collected and maintained by PEF. As mentioned in Section 3, the first step in seeking entry into the FAS program is to apply to the program when a call for applications is issued in newspapers. Interested schools obtained the application form from PEF's website[25] or by contacting PEF headquarters directly. In phase 3, fully-completed forms received by PEF before the announced deadline, as well as application forms filled out by schools solicited by PEF inspection teams on their way to inspect applicant schools were considered. In phase 4, only fully-completed forms received by PEF before the announced deadline were considered; all other applications were rejected. According to PEF, the number of unique applications received in phases 3 and 4 were ???? and ????, respectively.

The application form gathers data on school characteristics (e.g., location, gender type, level, physical infrastructure characteristics), enrollment numbers by class and gender, and individual-level teacher and administrative staff characteristics (e.g., monthly salary, qualifications, tenure). However, not all of the information is entered into a computer database—the practice to date has been to construct a school-level application database which contains all school characteristics, total school enrollment separately by gender, total number of teachers and administrative staff separately by gender, and the minimum and maximum monthly teacher salaries in the school. Separate databases are maintained for phase-3 and phase-4 applicants.

The school application computer database serves as the source of baseline data on the following outcomes measured at the school-level: number of students, teachers, classrooms, blackboards, and toilets. The school-level outcomes of student-teacher and student-classroom ratios are constructed by us using the data on total number of students, teachers, and classrooms.

[25] See: http://www.pef.edu.pk/PDF%20Material/FASAppFormEnglish.pdf (Last accessed: July 30, 2008).

This database also serves as the source of baseline data on key school characteristics such as location, level, gender type, and registration status.

As mentioned in Section 3, all schools that meet the criteria for a qualifying application are subject to a qualitative assessment by PEF inspection teams. As part of this assessment, the data provided by the school in the application form are verified. These inspection data would have been useful for checking the accuracy of the application data; they were however collected in paper form and not entered into a computer database. Consequently, these data are unavailable for the purposes of this study.[26]

Schools that clear the qualitative assessment take the SLQAT, which serves as the source of data on the school's student pass rate (our treatment assignment variable *z*) and our baseline school-level outcome measure of learning: average student performance on the SLQAT. Separate computer databases have been constructed by PEF for phase-3 and phase-4 SLQAT-takers. These databases contain the total test score for each student that took the test.[27] In these databases, the student-level test score data are organized by school and, within school, by grade. The school identification information provided in the databases comprise of the school's name and location (tehsil and district). Although PEF constructed the student pass rate for each school, we used the student-level data in order to construct our own measure of school SLQAT pass rates as well as average SLQAT scores. The match rate between their calculation of school SLQAT pass rates and ours' in both phases is nearly perfect at 99.5%. While it makes little difference, we use our measure of school SLQAT pass rates as the treatment assignment variable

Constructing a single computer database for our analysis required that the application data are linked with the SLQAT data at the school level. Unfortunately, the same unique school identifier variable was not used across databases. Consequently, the application databases were linked to the SLQAT databases using an iterative visual-matching process. First, schools were

[26] Conversations with PEF staff indicate that inspection teams found a high incidence of inaccuracies in the application data, particularly with respect to student enrollment numbers. We have no reason to expect however that extent and degree of such inaccuracies systematically differ between schools with pass rates below and above the cutoff, or, more importantly, between schools with pass rates marginally below and above the cutoff. That is, we expect that the extent and degree of inaccuracies to be continuous in the assignment variable at the cutoff.

[27] While it would be interesting to examine whether baseline mean test scores differed between marginal failers and passers in each of the tested subjects, given that we only have the total test score received by each student, this is not possible.

matched across databases using the district name and school name variables.[28] Exact matching failed in a number of cases as PEF did not maintain consistency in the spelling, word ordering, and completeness of the school's name across the two databases; hence, matching on school name frequently required matching on key words and word patterns. In cases where we suspected that the combination of district name and school name (even with keyword and word pattern matching) did not yield a unique school record in a database, the set of matching variables was extended to include school address. This extension helped resolved most uncertainties. On the basis of this exercise, 94% and 97% of school records in the SLQAT databases were linked with school records in the application databases for phase-3 and phase-4, respectively. Only the linked school records are used in our analysis, which yield sample sizes of 747 and 830 schools in phase-3 and phase-4, respectively. Though not entirely accurate, hereafter, we refer to these samples as the full SLQAT-taking samples.

Endline phone interview data: As a stopgap to the endline school survey and student assessment data expected to be conducted before the end of the calendar year, school-level data on the contemporaneous number of students, teachers, classrooms, blackboards, and toilets were collected via phone interviews by a team of independent interviewers from schools with SLQAT pass rates between $+/-$ 15 percentage points of the cutoff in both phases. This was made possible as all schools had to provide telephone numbers when they applied to the FAS program, and these data were entered into the school application computer databases for both phases. The school sample sizes within this pass rate range from the cutoff are 268 and 319 schools in phases 3 and 4, respectively. These sizes represent 36% and 38% of the respective full SLQAT-taking samples. Hereafter, we refer to these samples as the neighborhood SLQAT-taking samples.

Endline outcome data for both phase-3 and phase-4 neighborhood samples were collected in October 2008, roughly 14 and 10 months respectively after schools received their first subsidy payment. The treatment exposure period covers at least half of the 2007-08 academic year and a couple of months into the 2008-08 academic year (the academic year of private schools typically begins in August). The data were collected from either the school owner or the school headmaster (principal). The contemporaneous program participation status of schools in the

[28] Although information on the school's tehsil was also available in both the application and test databases, we found that this information was error-ridden. Consequently, this information was not used in the cross-database matching exercise.

neighborhood samples were determined by us using PEF's records of FAS schools as of September 2008.

Despite repeated attempts by the interviewers, not all schools in the neighborhood samples were contactable over the phone. Unit nonresponses were largely due to wrong numbers or unpicked-up calls. Once contacted, however, there were no cases of refusals; further, all contacted schools provided the full set of data requested (i.e., there were no cases of item nonresponse). Roughly 28% and 26% of schools have missing data due to unit nonresponse in the phase-3 and phase-4 neighborhood samples, respectively. These large percentages clearly reduce the study's statistical power but do not necessarily introduce bias.

To investigate the presence of systematic bias, we examine the simple correlation between unit nonresponse and selected covariates, separately by phase (see Table 8). The covariates considered comprise of indicator variables for whether the school's SLQAT pass rate is at or above the cutoff, which interviewer collected the phone data, as well as the school's registration status, level, gender type, and location all measured using baseline data. It appears that the probability of nonresponse is not significantly associated with an above-cutoff SLQAT pass rate.[29] In addition, with a few exceptions—which could arise simply due to random chance given that we test multiple individual hypotheses—the probability of nonresponse does not appear to be significantly associated with the selected covariates. These findings hold for both neighborhood samples. Thus, the evidence suggests we cannot reject the null hypothesis that the pattern of nonresponse is random. In light of this evidence, we assume that the nonresponse rate is continuous in the assignment variable at the cutoff.

Sample

Table 9 presents statistics on the distribution of schools by salient characteristics measured at baseline such as level, gender type, registration status, and location. These statistics are provided disaggregated by phase and sample (full vs. neighborhood). First and foremost, across both phases and samples, the distributional pattern across characteristics is similar; for example, the majority of schools are (1) middle schools (63-72%, depending on the phase and sample), (2) coeducational (82-87%), (3) officially-registered (81-88%) and (4) rural (53-59%). Second, the distributional patterns across characteristics appear to more similar across samples for a given

[29] Even if nonresponse rates are similar for schools below and above the cutoff, it is still possible that the selection process for nonresponding schools below the cutoff differs from the selection process for nonresponding schools above the cutoff (Duflo et al. 2007).

phase than across phases for a given sample. These inter-phase differences are more pronounced when we look at the distribution of schools by location (urban vs. rural and across districts) and registration status.

Table 10 presents means and standard deviations for selected outcomes measured at baseline, again separately by phase and for the full and neighborhood SLQAT-taking samples. In the phase-3 full SLQAT-taking sample (column 1), we find that schools on average have 253 students; 10 teachers, classrooms, and blackboards; and 3 toilets. Student-teacher and student-classrooms ratios are 25:1 and 28:1, respectively; ex ante, these levels are below the stipulated maximums for FAS program benefit maintenance. Finally, the mean SLQAT pass rate and score are 73% and 45%, respectively. Comparing the means between phases in the full sample, we find that a number of outcomes have statistically-significantly different values (see column 3). Specifically, the mean values for phase-4 schools appear to be lower in magnitude, suggesting smaller schools and weaker schools in terms of learning relative to phase-3 schools. A plausible explanation for the learning differences between phases is positive selection in entry timing: better schools are likely to enter the program before poorer schools; thus, the quality of the applicant pool is likely to deteriorate over time. Relatedly, a plausible explanation for the size differences is queuing-for-quality: to the extent that schools with higher levels of achievement are more attractive providers, they are likely to be larger. However, with the sole exception of the mean number of female students, these differences disappear when we restrict our attention to the neighborhood samples—this suggests that these specific samples are similarly composed.

6. Results

Before presenting our estimates of the mean impacts of the FAS program, we discuss results from a couple of key model specification tests. First, we examine whether there is a discontinuous change in the conditional probability of treatment at the cutoff, as the applicability of the regression-discontinuity design hinges on this feature of the data. Second, we examine whether the density function of the SLQAT pass rate exhibits local smoothness at the cutoff. This test is particularly useful if data are unavailable to directly test whether the conditional mean untreated outcomes at the cutoff exhibit local smoothness. In such cases, if the density test rejects local smoothness, it can cast some doubt on whether the regression-discontinuity estimates are internally valid. However, in our case, we can directly test the underlying identifying assumption by using our baseline data on outcomes, and in so doing, assess the reliability of the density test as well.

Discontinuity in the probability of FAS treatment at the cutoff: Figure 5 plots the local linear regression functions estimated separately above and below the cutoff for the phase-3 and phase-4 neighborhood samples. The jump in the probability of treatment is clearly visible for both samples. As expected, for phase-3 SLQAT takers, the probability of treatment is one above the cutoff and positive but less than one below. Further, we observe that the probability of treatment is increasing as we approach the cutoff from below. A plausible and likely explanation for this pattern is that phase-3 marginal failers closer to the cutoff have a greater chance of success in the phase-4 SLQAT relative to marginal failers farther away. For phase-4 SQLAT takers, as expected, the probability of treatment jumps discontinuously from zero to one at the cutoff. These patterns in the conditional probabilities of treatment support our selections of fuzzy and sharp designs for the phase-3 and phase-4 data, respectively.

Local smoothness in the design function of SLQAT pass rates: It is important to point out at the outset that local smoothness in the density function of the assignment variable at the cutoff is neither a necessary nor sufficient condition for identification under the regression-discontinuity design. However, if local smoothness in the density is rejected, it may suggest manipulation of the assignment variable z, which potentially invalidates the identifying condition of local smoothness in conditional mean untreated outcomes at the cutoff.

Manipulation is plausible in our case as the pass rate cutoff for program participation was advance public knowledge and schools and perhaps even program managers may have an interest in a particular treatment outcome. As McCrary (2008) notes, not all forms of manipulation can produce identification problems, drawing a distinction between partial and complete manipulation. Partial manipulation occurs when the value of z is partly determined by factors under the control of the agent but also partly by a random component—generally, this form of manipulation is not a potential threat to validity. On the other, complete manipulation occurs when the value of z is fully under the control of the agent—generally, this form represents a potential threat to validity. We argue however that the pass rate on the SLQAT is determined by both a deterministic component by the school as well as a stochastic component (e.g., the day the SLQAT is offered at the school, the classes which are tested, and the specific SLQAT papers offered at the school)—thus the pass rate that the school *actually* received on the

SLQAT may be subject to partial manipulation, which is typically benign. However, PEF staff exercise complete control over the calculation of SLQAT pass rates, as well as their reporting; this, thus, introduces a risk of complete manipulation. It is plausible that PEF staff might face incentives to manipulate the pass rates of marginal passers downwards (to just below the cutoff) if, for example, the program budget limits the number of schools that the FAS program can be offered to. Alternatively, sympathetic PEF staff might manipulate the pass rate of marginal failers upwards (to just above the cutoff), so that these schools become eligible to participate in the program. So long as the manipulation is monotonic, that is, failers are made into passers, or vice versa, but not both, a test of the discontinuity in the density of z at the cutoff will be informative. Given that we use SLQAT pass rates that we calculate based on the individual student test score data as our measure of the assignment variable, manipulation would have to occur at the individual student level.

To assess whether the density of pass rates exhibits local smoothness at the cutoff, as a preliminary investigation, we simply visually inspect a frequency histogram of schools at each SLQAT pass rate value in the neighborhood sample for both phases (see Figure 3). It appears from the histograms that the number of schools at the cutoff value of 67% is much larger than the number of schools just below at 66% in both phases. For example, in the case of the phase-3 sample, there is just one school with a pass rate of 66%, whereas there are 15 schools with a pass rate of 67%. However, if we look at the number of schools with a pass rate of 65%, it is roughly equivalent to the number of schools at the cutoff. This frequency pattern is not unique to the immediate neighborhood around the cutoff—in general, the histograms show several sharp frequency peaks and troughs, at times, at adjacent pass rates.

As a confirmatory investigation, we implement a test proposed by McCrary (2008) by separately estimating kernel density regressions below and above the cutoff. Figure 4 plots the kernel density functions and Table 10 presents the density discontinuity parameter and bootstrapped standard error estimates at the cutoff of 67% for alternative bandwidth choices. For both phases, we find a statistically-significant discontinuity in the pass-rate density at the cutoff. Further, the significance of the finding is robust to increases in the estimator's bandwidth size from the optimal choice (depending on the sample, the optimal size is either three or four percentage points).

Local smoothness in baseline mean outcomes: We examine whether baseline mean outcomes at satisfy local smoothness at the cutoff in two ways: by (1) examining differences in simple means between marginal failers and passers for alternative neighborhood sizes and (2) estimating the discontinuity at the cutoff via local linear regressions. Table 11 presents the simple baseline mean outcomes for marginal failers and differences from these means for marginal passers for two neighborhood sizes: $+/-$ 15 percentage points and $+/-$ 5 percentage points. Across both phases, by and large, there appears to be no statistically-significant differences in baseline mean outcomes for marginal failers and passers. The one exception is the mean SLQAT score; we find consistent evidence that the mean score for marginal passers is significantly higher than for marginal failers. A likely explanation for this finding is that the neighborhood sizes are not small enough to counteract the influence of the strong positive correlation between the mean SLQAT scores and the SLQAT pass rate. Our findings based on local linear regressions should be more immune to this problem.

Table 12 presents discontinuity estimates of baseline mean outcomes at the cutoff using local linear regressions, separately by phase and for alternate bandwidth sizes (depending on the sample, the optimal bandwidth size is three to four percentage points). In addition, Figures 6a-6c depict the estimated local linear regression functions using the optimal bandwidth size. In general, the evidence suggests that we can reject local smoothness more frequently than what our simple comparisons of mean outcomes between marginal failers and passers would have led us to assert, as well as more frequently than would be expected by random chance given standard significance levels. Notwithstanding, none of our findings of local discontinuities are robust to the alternate bandwidth sizes. Given their sensitivity, we discount these findings.

Impact estimates: We now turn to our estimates of the mean effects of the FAS program. Table 12 presents our impact estimates based on local linear regressions, separately by phase. In addition, Figure 6 depicts histograms of mean values of outcomes by SLQAT pass rates for the phase-4 neighborhood sample which is subject to a sharp design. It is accompanied by Figure 7 which depicts the estimated local linear regression functions using the optimal bandwidth choice for the same sample.

Turning first to the phase-3 school sample (columns 1-3 in Table 12) which is subject to a partially-fuzzy design, we find no evidence of significant program impacts across the outcomes of interest we examine. The magnitude of the treatment parameter estimates also appear to be quite sensitive to bandwidth size. In addition, the empirical standard errors associated with the point estimates are very large, indicating weak identification.

Turning next to the phase-4 school sample (columns 4-6), we find evidence of significant positive effects on the number of students, teachers, classrooms, and blackboards; furthermore, their significance is robust to the choice of bandwidth size. The figures show a discernible structural change in the mean levels for these outcomes marginally above and below the cutoff. However, as with the fuzzy estimates for the phase-3 sample, the magnitude of these effects also appears to be somewhat sensitive to bandwidth choice. In particular, the estimates drop fairly sharply when we increase the bandwidth size from the optimal choice of 3 percentage points to 4.5 percentage points. Taking the most conservative estimates, the mean impacts at the cutoff on the number of students, teachers, classrooms, and blackboards are 85, 3.4, 4, and 2.8, respectively. In relative terms, given the baseline means for these outcomes in the phase-4 neighborhood sample (see Table 10), these effects translate into impacts of roughly 37%, 37%, 47% and 27%, respectively. These effects are substantial, and particularly so given the short treatment exposure we investigate: the endline phone interview data were collected only some 10 months after phase-4 schools received their first subsidy payment.

Cost-effectiveness: Using the conservative estimates of mean program effects for the phase-4 sample, a first way that we estimate the program's cost-effectiveness is by deriving the annual rupee cost of a program gain of one additional student. Given a baseline mean school size of 232 students for schools in the phase-4 neighborhood sample (which we treat as the number of children that would have attended the FAS program school even in the absence of the treatment) and an annual subsidy amount of 3,600 rupees per student, the figure is 13,426 rupees (US$189) per additional student per year. A second way we calculate the program's cost-effectiveness is by deriving the annual per-student cost of increasing enrollment by 1 percent. Using an annual subsidy amount of 3,600 per student and an enrollment impact estimate of 37%, our estimated ratio is 97 rupees (US1.4). This figure compares extremely favorably with the cost-effective ratios of other educational interventions across the developing world that produced enrollment gains; in fact, the FAS program's ratio is among the lowest (Evans and Ghosh 2008).

7. Conclusion

To summarize, our findings on the causal impacts of the FAS program are mixed. For phase-3 schools, applying a partially-fuzzy design, we find no evidence of program impacts at the cutoff on our outcomes of interest. We posit that this is due to weak identification, a clear symptom of which was the inordinately large estimated standard errors associated with our treatment parameter estimates. In contrast, for phase-4 schools, applying a sharp design, we find robust evidence that the FAS program significantly increased the number of students, teachers, classrooms, and blackboards. The mean treatment effects at the cutoff were sizeable: our conservative estimates indicate that program schools expanded by roughly 85 students, and 3-4 teachers, classrooms, and blackboards. These treatment estimates are also large as a percentage of mean baseline values for these outcomes for phase-4 schools near the cutoff, and are particularly noteworthy given that phase-4 program schools were exposed to only 10 months of treatment before the endline phone interview data were gathered. Finally, our cost-effectiveness calculations show that the FAS program should be counted among the cheapest interventions in developing countries for raising enrollment levels.

References

Andrabi, Tahir, Jishnu Das, Asim Ijaz Khwaja, Tara Vishwanath, Tristan Zajonc and LEAPS team. 2007. Pakistan Learning and Educational Achievements in Punjab Schools (LEAPS): Insights to inform the education policy debate.

Andrabi, Tahir, Jishnu Das, Asim Ijaz Khwaja. 2006. A dime a day: The possibilities and limits of private schooling in Pakistan. World Bank Policy Research Working Paper 4066.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91(434):444-455.

Battistin, Erich, and Enrico Rettore. 2008. Ineligible and eligible non-participants as a double comparison group in regression-discontinuity designs. *Journal of Econometrics* 142:715-730.

Bloom, Benjamin S. 1956. *Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain.* New York: Longman.

Bloom, Howard S. 1984. Accounting for no-shows in experimental evaluation designs. *Evaluation Review* 8(2):225-246.

Cheng, Ming-Yen, Jianqing Fan, and James S. Marron. 1997. On automatic boundary corrections. *Annals of Statistics* 25(4):1127-1160.

Das, Jishnu, Priyanka Pandey, and Tristan Zajonc. 2006. Learning levels and gaps in Pakistan. World Bank Policy Research Working Paper 4067.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2006. Using randomization in development economics research: A toolkit. Harvard University Center for International Development Working Paper No. 138.

Evans, David, and Arkadipta Ghosh. 2008. Prioritizing educational investments in children in the developing world. RAND Working Paper 587.

Fan, Jianqing. 1992. Design adaptive nonparametric regression. *Journal of the American Statistical Association* 87:998-1004.

Fan, Jianqing, and Irene Gijbels. 1996. *Local polynomial modeling and its applications*. New York: Chapman & Hall.

Galdo, Jose, Jeffrey Smith, and Dan Black. 2007. Bandwidth selection and the estimation of treatment effects with unbalanced data. IZA Discussion Paper No. 3095.

Glewwe, Paul, Nauman Illias, and Michael Kremer. 2003. Teacher incentives. National Bureau of Economic Research Working Paper No. 9671.

Government of Punjab, Pakistan (Planning and Development Department). 2004. *Punjab District-based Multiple Indicators Cluster Survey 2003-04*. Lahore: Government of Punjab.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 69(1):201-209.

Heckman, James J., and Edward Vytlacil. 2005. Structural equations, treatment effects, and econometric policy evaluation. *Econometrica* 73(3):669-738.

Husain, Fazal, Muhammad Ali Qasim, and Khalid Hameed Sheikh. 2003. Analysis of public expenditure on education in Pakistan. Munich Personal RePEc Archive Paper No. 2722.

Imbens, Guido W., and Jeffrey M. Wooldridge. 2008. Recent developments in the econometrics of program evaluation. IZA Discussion Paper No. 3640.

Imbens, Guido W., and Thomas Lemieux. 2008. Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142(2):615-635.

Imbens, Guido W., and Joshua D. Angrist. 1994. Identification and estimation of local average treatment effects. *Econometrica* 62:467-476.

Jacob, Brian A. 2007. Test-based accountability and student achievement: An investigation of differential performance on NAEP and state assessments. National Bureau of Economic Research Working Paper No. 12817.

Jacob, Brian A. 2002. Accountability, incentives, and behavior: The impact of high-stakes testing in the Chicago Public Schools. National Bureau of Economic Research Working Paper No. 8968.

Kim, Jooseop, Harold Alderman, and Peter F. Orazem. 1999a. Can private school subsidies increase schooling for the poor? The Quetta urban fellowship program. *World Bank Economic Review* 13(3):443-466.

Kim, Jooseop, Harold Alderman, and Peter F. Orazem. 1999b. Evaluation of the Balochistan Rural Girls' Fellowship Program: Will rural families pay to send girls to school? World Bank Education Department Working Paper 22983.

Ludwig, Jean, and Douglas L. Miller. 2005. Does Head Start improve children's life chances? Evidence from a regression discontinuity design. National Bureau of Economic Research Working Paper 11702.

Malik, Allah Bakhsh. 2007. *Freedom of choice: Affordable quality education in public private partnership*. Lahore, Pakistan: Maqbool Academy Press.

McCrary, Justin. 2008. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142(2):698-714.

McCrary, Justin, and Heather Royer. 2006. The effect of female education on fertility and infant health: Evidence from school entry policies using exact date and birth. National Bureau of Economics Research Working Paper 12329.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2006. Teacher incentives in developing countries: Experimental evidence from India. Mimeo. Harvard University.

Nichols, Austin. 2007a. `rd`: Stata module for regression discontinuity estimation. http://ideas.repec.org/c/boc/bocode/s456888.html. Last accessed:

Nichols, Austin. 2007b. Causal inference with observational data. Mimeo. http://pped.org/stata/ciwod.pdf. Last accessed:

Porter, Jack. 2002. Asymptotic bias and optimal convergence rates for semiparametric kernel estimators in the regression discontinuity model. Harvard Institute of Research Working Paper No. 1989.

Thistlethwaite, D., and D. Campbell. 1960. Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology* 51:309-317.

Trochim, William. 1984. *Research design for program evaluation: The regression-discontinuity approach*. Beverly Hills CA: Sage Publications.

van der Klaauw, Wilbert. 2007. Regression-discontinuity analysis. Forthcoming in *The New Palgrave Dictionary of Economics*.

van der Klaauw, Wilbert. 2008. Breaking the link between poverty and low student achievement: An evaluation of Title I. *Journal of Econometrics* 142(2):731-756.

Table 1. Selected education statistics for Pakistan and Punjab province, 1998-99 and 2004-05

Sample	1998/99		2004/05		% change, 1998/99-2004/05	
	(1) Pakistan	(2) Punjab	(3) Pakistan	(4) Punjab	(5) Pakistan	(6) Punjab
Current school participation rate, 6-17 year olds						
All	49.0	50.9	57.8	61.9	17.9	21.5
Rural	42.7	45.4	51.4	57.0	20.4	25.6
Female	40.4	44.8	50.3	57.1	24.5	27.3
Poorest quintile	28.4	30.2	42.1	45.5	48.3	50.6
Completion rate, 20-24 year olds who ever attended school						
Primary	89.1	87.8	93.1	92.6	4.6	5.4
Secondary	49.5	42.5	51.2	45.3	3.6	6.6
Participation share in formal private schools, 6-17 year olds						
All	21.3	24.5	25.2	27.4	18.3	12.1
Rural	11.4	15.0	15.7	19.6	37.7	30.4
Female	23.7	27.0	27.1	28.4	14.2	5.4
Poorest quintile	6.7	9.0	10.0	13.4	49.0	48.2

Notes: Statistics for Pakistan actually represent statistics for the four provinces of Pakistan (unit data for the territories were not available). All statistics are corrected for sampling weights. The age group 6-17 is used instead of the official ages for primary and secondary schooling (5-15) as this range is more consistent with the data. Delayed entry and completion characterize Pakistani education—consequently, completion rates are estimated for the age group 20-24, the upper age segment in the international classification for youth.

Data sources: 1998-99 Pakistan Integrated Household Survey (PIHS); 2004-05 Pakistan Social and Living Standards Measurement Survey (PSLM). Field work for the 2004-05 PLSM was conducted between September 2004-March 2005, preceding the FAS program start date.

Table 2. FAS program timeline

Phase 1 application deadline	October 2005
Phase 1 school agreements signed	November 2005-January 2006 (52/54)
Phase 1 first monthly subsidy payment	January 2006
QAT 1	March/April 2006
Phase 2 application deadline	June 2006
Phase 2 school agreements signed	September-December 2006 (143/150)
Phase 2 first monthly subsidy payment	October 2006
QAT 2	October/November 2006
First payment of annual teacher bonus	January 2007
QAT 3	March/April 2007
Phase 3 application deadline	April 2007
Phase 3 school agreements signed	July-August 2007 (473/482)
Phase 3 first monthly subsidy payment	August 2007
Phase 4 application deadline	July 2007
QAT 4	October/November 2007
Phase 4 school agreements signed	November 2007 (424/424)
Phase 4 first monthly subsidy payment	December 2007
Second payment of annual teacher bonus	February 2008
First payment of annual school bonus	February 2008
QAT 5	March/April 2008
Phase 5 application deadline	April 2008

Notes: The statistics in parentheses represent the number of schools that signed agreements in those months relative to the number of schools that entered the program in that phase.

Table 3. District- and phase-wise frequency distribution of program schools

District	Adult literacy rate (10+ years)	Adult literacy rate rank (Low number = high rank)	Phase 1	Phase 2	Phase 3	Phase 4	Total
Rawalpindi	78	1					
Lahore	*70* [a]	2	7	11			18
Sialkot	70	3	7	9			16
Chakwal	69	4	9	17			26
Jhelum	68	5					
Gujranwala	67	6					
Gujrat	65	7		15			15
Faisalabad	60	8					
Narowal	60	9		6			6
Sargodha	58	10					
Toba Tek Singh	58	11					
Attock	57	12					
Mandi Bahauddin	57	13					
Mianwali	56	14		12			12
Hafizabad	55	15					
Sahiwal	54	16					
Khushab	52	17	10	19			29
Sheikhupura	50	18	2				2
Khanewal	49	19					
Jhang	47	20			44	48	92
Layyah	46	21					
Multan	46	22			58	80	138
Bahawalnagar	44	23		14	115	43	172
Okara	43	24					
Vehari	43	25		1			1
Kasur	42	26					
Pakpattan	42	27					
Rahim Yar Khan	42	28					
D. G. Khan	40	29					
Bahawalpur	37	30	10	16	114	88	228
Bhakkar	37	31		11			11
Lodhran	37	32			42	49	91
Muzaffargarh	36	33		1	72	81	154
Rajanpur	34	34			35	35	70
Nankana Sahib [b]	--	--		1			1
		Number of districts	6	13	7	7	18
		Number of program schools (as of June 2008)	45	133	480	424	1082

(dark shading)	High presence
(light shading)	Low presence
Bold	Program district

Notes: District-level adult literacy rates obtained from Government of Punjab (2004) based on data from the 2003-04 Multiple Indicators Cluster Survey (MICS). [b]The literacy rate for Nankana Sahib was unavailable. [a]The statistic for Lahore district is a simple mean of the literacy rates for towns and cantonments in Lahore.

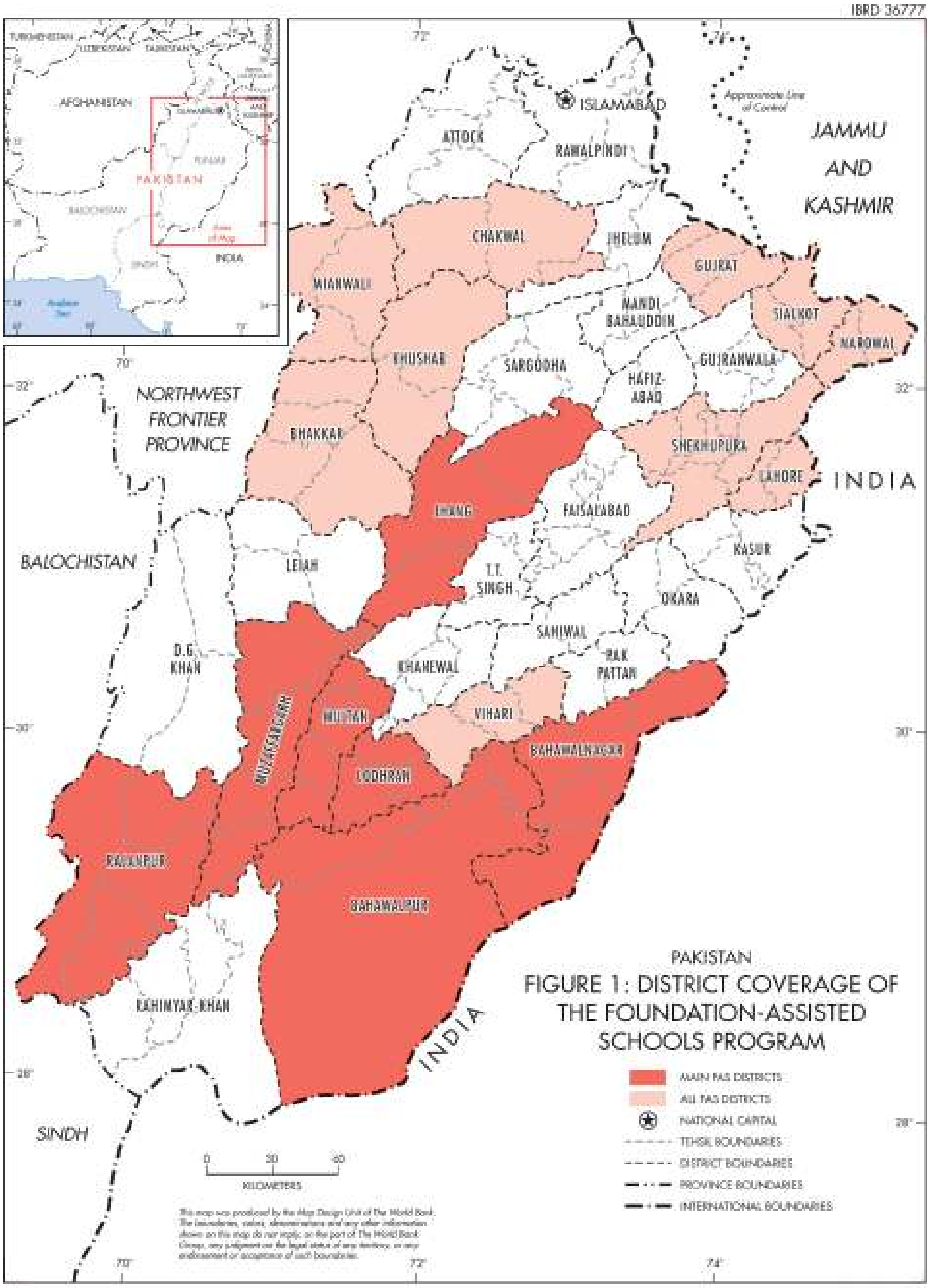
IBRD 36777
ISLAMABAD
Approximate Line of Control
JAMMU AND KASHMIR
ATTOCK
RAWALPINDI
CHAKWAL
JHELUM
GUJRAT
MIANWALI
MANDI BAHAUDDIN
SIALKOT
NAROWAL
KHUSHAB
SARGODHA
GUJRANWALA
HAFIZ-ABAD
NORTHWEST FRONTIER PROVINCE
BHAKKAR
SHEKHUPURA
LAHORE
INDIA
JHANG
FAISALABAD
BALOCHISTAN
LEIAH
KASUR
T.T. SINGH
OKARA
SAHIWAL
D.G. KHAN
KHANEWAL
PAK PATTAN
MUZAFFARGARH
MULTAN
VIHARI
BAHAWALNAGAR
LODHRAN
RAJANPUR
BAHAWALPUR
RAHIMYAR KHAN
SINDH
INDIA
PAKISTAN
FIGURE 1: DISTRICT COVERAGE OF THE FOUNDATION-ASSISTED SCHOOLS PROGRAM
MAIN FAS DISTRICTS
ALL FAS DISTRICTS
NATIONAL CAPITAL
TEHSIL BOUNDARIES
DISTRICT BOUNDARIES
PROVINCE BOUNDARIES
INTERNATIONAL BOUNDARIES
KILOMETERS
AFGHANISTAN
PAKISTAN
INDIA
SEPTEMBER 2008

Table 4. Program coverage of private schools in selected FAS districts

Program district	(1) Number of program schools	(2) Number of primary, middle and secondary private schools	(3) Program share
Bahawalnagar	172	447	0.38
Bahawalpur	230	871	0.26
Jhang	93	686	0.14
Lodhran	93	284	0.33
Multan	139	1,411	0.10
Muzaffargarh	157	690	0.23
Rajanpur	71	226	0.31
Total	955	4,615	0.21

Notes: Numbers of private schools provided by Corinne Siaens using the 2005 National Education Census (NEC) data. Column (2) reports numbers of relevant schools strictly classified as private in the NEC.

Table 5. Mean characteristics of current FAS program schools

Characteristic	(1) Phase 1	(2) Phase 2	(3) Phase 3	(4) Phase 4	(5) All phases
Total number of students	561.40	547.42	373.83	241.66	351.18
Level					
Primary	0.02	0.05	0.05	0.11	0.07
Middle	0.24	0.31	0.60	0.69	0.59
Secondary	0.73	0.65	0.35	0.20	0.34
Gender type					
Coeducational	0.69	0.86	0.83	0.82	0.83
Female	0.20	0.11	0.09	0.11	0.11
Male	0.11	0.03	0.07	0.07	0.07
Registration status					
Registered	0.91	0.97	0.89	0.81	0.87
Unregistered	0.09	0.03	0.11	0.19	0.13
Location					
Urban	0.27	0.38	0.42	0.33	0.38
Slum (*katchi abadi*)	0.09	0.08	0.06	0.09	0.07
Rural	0.64	0.55	0.52	0.58	0.55
N	45	133	480	424	1082

Notes: The statistics exclude the three higher secondary schools that are program schools.

Table 6. FAS program school participation status by phase

Phase	(1) Entrants	(2) Disqualified, all reasons	(3) Disqualified, serial QAT failure	(4) Current participation
1	54	9	7	45
2	150	16	13	133
3	482	2	0	480
4	425	1	0	424
Total	1,111	28	20	1,082

Notes: Disqualification also includes voluntary exits.

Table 7. PEF-reported numbers of schools at key program entry stages

Application stage	Phase 3	Phase 4
Applications received by PEF	--	--
Applications considered by PEF	1,069	1,428
Schools inspected by PEF	1,069	1,428
Schools tested using the SLQAT	799	872
Schools selected into the FAS program	514	431

Table 8. Correlates of nonresponse in post-treatment phone interview data

	Phase-3 neighborhood sample			Phase-4 neighborhood sample		
	(1)	(2)	(3)	(4)	(5)	(6)
	Below *c*	Above *c*	Full	Below *c*	Above *c*	Full
Number of nonresponses			76			83
Nonresponse rate			0.28			0.26
Above cutoff			0.03			0.07
			(0.06)			(0.05)
Interviewer ID 1						
Interviewer ID 2						
Interviewer ID 3						
Baseline characteristics						
Registered			-0.04			0.00
			(0.09)			(0.07)
Coeducational			0.01			0.06)
			(0.08)			(0.07)
Female			0.07			-0.04
			(0.10)			(0.09)
Male			-0.19			-0.08
			(0.10)			(0.11)
Primary			0.15			-0.18**
			(0.12)			(0.08)
Middle			-0.05			0.16***
			(0.06)			(0.06)
Secondary			0.00			-0.10
			(0.07)			(0.07)
Urban			-0.01			-0.04
			(0.06)			(0.05)
Slum			-0.18**			0.00
			(0.07)			(0.08)
Rural			0.07			0.03
			(0.05)			(0.05)
Bahawalnagar			-0.05			-0.04
			(0.06)			(0.09)
Bahawalpur			-0.05			-0.07
			(0.06)			(0.06)
Jhang			0.13			-0.01
			(0.09)			(0.07)
Lodhran			-0.09			0.10
			(0.09)			(0.08)
Multan			0.00			0.07
			(0.08)			(0.06)
Muzaffargarh			0.08			0.05
			(0.08)			(0.06)
Rajanpur			-0.05			-0.12
			(0.11)			(0.09)

Notes: * denotes statistical significance at the 10% level; ** at the 5% level; and *** at the 1% level. Heteroskedasticity-robust standard errors are reported in parentheses.

Table 9. Distribution of schools by selected baseline characteristics, full and neighborhood samples

Characteristic	Full SLQAT-taking sample $z \in [0,100]$		Neighborhood SLQAT-taking sample, $z \in [52,82]$	
	(1) Phase 3	(2) Phase 4	(3) Phase 3	(4) Phase 4
Level				
Primary	0.09	0.12	0.07	0.12
Middle	0.63	0.72	0.71	0.70
Secondary	0.28	0.16	0.22	0.18
Gender type				
Coeducational	0.82	0.87	0.87	0.86
Female	0.12	0.08	0.10	0.09
Male	0.07	0.05	0.04	0.06
Registration status				
Registered	0.88	0.81	0.88	0.83
Unregistered	0.12	0.19	0.12	0.17
Location type				
Urban	0.35	0.30	0.38	0.30
Slum	0.10	0.11	0.09	0.11
Rural	0.54	0.59	0.53	0.59
District				
Bahawalnagar	0.21	0.11	0.20	0.09
Bahawalpur	0.24	0.20	0.25	0.21
Jhang	0.13	0.11	0.14	0.14
Lodhran	0.08	0.12	0.07	0.09
Multan	0.13	0.16	0.13	0.20
Muzaffargarh	0.14	0.20	0.14	0.17
Rajanpur	0.07	0.10	0.06	0.09
N	747	830	268	319

Table 10. Summary statistics of baseline outcome measures, full and neighborhood samples

Outcome measure	Full SLQAT-taking sample, $z \in [0,100]$			Neighborhood SLQAT-taking sample, $z \in [52,82]$		
	(1) Phase 3 Mean (SD)	(2) Phase 4 Mean (SD)	(3) (2)-(1) Difference (SE)	(4) Phase 3 Mean (SD)	(5) Phase 4 Mean (SD)	(6) (5)-(4) Difference (SE)
Number of students	252.85	222.63	-30.22***	240.82	232.27	-8.56
	(155.70)	(106.68)	(6.79)	(134.67)	(108.05)	(10.21)
Number of male students	147.5	135.23	-12.28**	138.04	141.61	3.57
	(106.35)	(80.96)	(4.84)	(88.25)	(83.55)	(7.18)
Number of female students	110.88	91.8	-19.08***	105.12	96.05	-9.07*
	(78.62)	(55.00)	(3.49)	(69.89)	(58.91)	(5.45)
Mean SLQAT score	45.43	40.17	-5.26***	40.82	40.47	-0.34
	(14.42)	(13.11)	(0.70)	(6.14)	(5.48)	(0.48)
SLQAT pass rate	73.06	64.43	-8.62***	68.42	67.86	-0.56
	(22.97)	(24.16)	(1.19)	(9.13)	(8.66)	(0.74)
Number of teachers	10.22	9	-1.22***	9.82	9.32	-0.5
	(5.03)	(3.69)	(0.22)	(4.30)	(3.77)	(0.34)
Number of classrooms	9.56	8.46	-1.10***	8.9	8.64	-0.26
	(4.71)	(3.79)	(0.22)	(3.78)	(3.71)	(0.31)
Number of blackboards	9.93	9.07	-0.86***	9.39	9.3	-0.09
	(5.03)	(3.97)	(0.23)	(4.11)	(3.90)	(0.33)
Number of toilets	3.24	2.98	-0.27***	3.09	2.99	-0.09
	(1.79)	(1.93)	(0.09)	(1.59)	(1.73)	(0.14)
Student-teacher ratio	25.06	25.67	0.6	24.98	25.59	0.6
	(9.27)	(9.51)	(0.47)	(9.59)	(8.26)	(0.75)
Student-classroom ratio	27.51	28.13	0.62	28.3	28.35	0.05
	(12.63)	(12.13)	(0.63)	(13.99)	(11.47)	(1.07)

Notes: * denotes statistical significance at the 10% level; ** at the 5% level; and *** at the 1% level. SD denotes standard deviation; SE standard error; and z the treatment assignment variable.

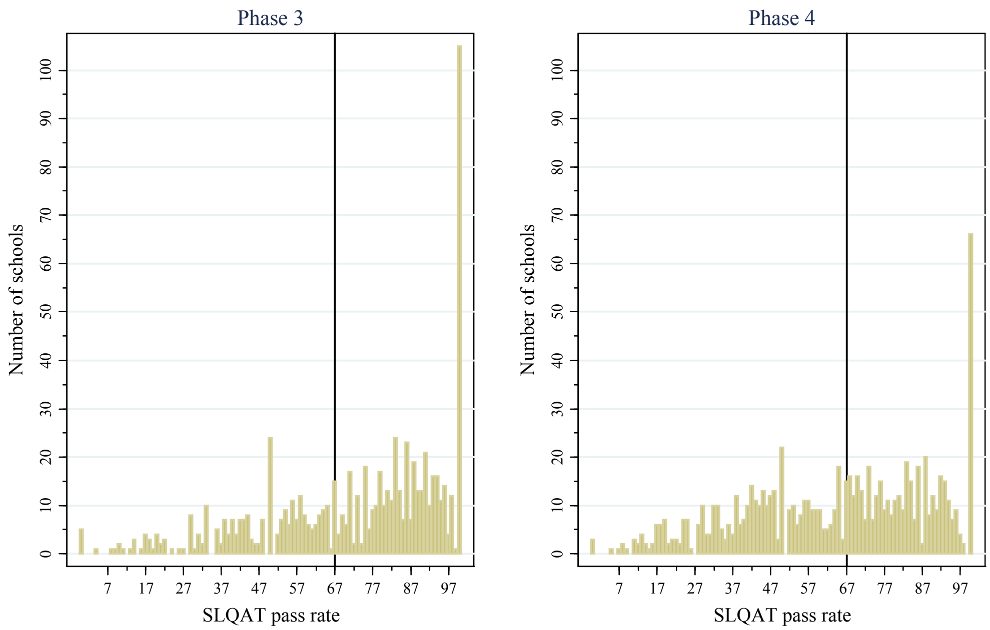

Fig. 2.
Frequency histogram of SLQAT pass rates, full sample

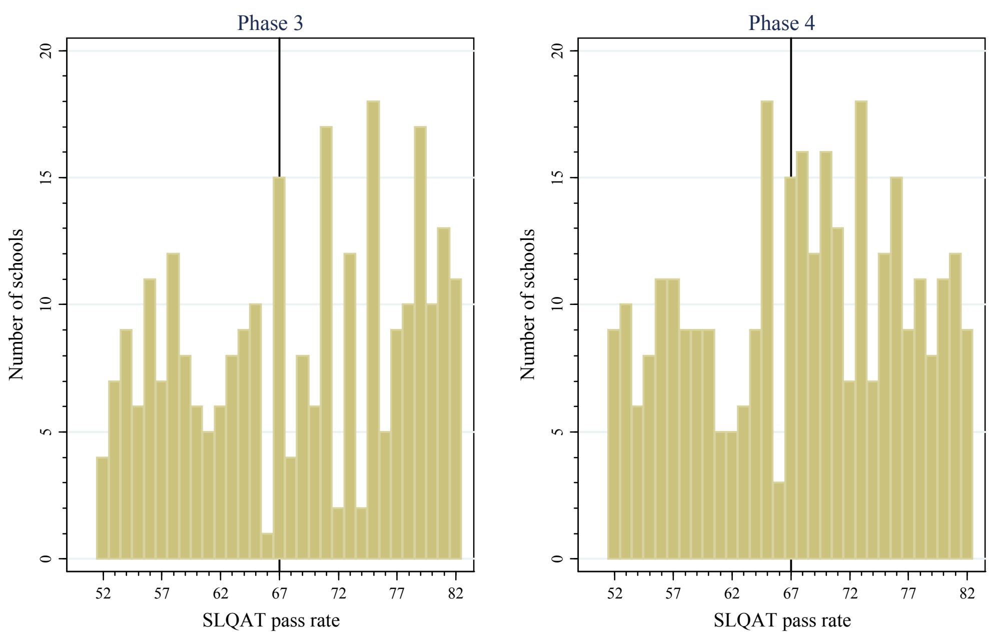

Fig. 3.
Frequency histogram of SLQAT pass rates, neighborhood sample

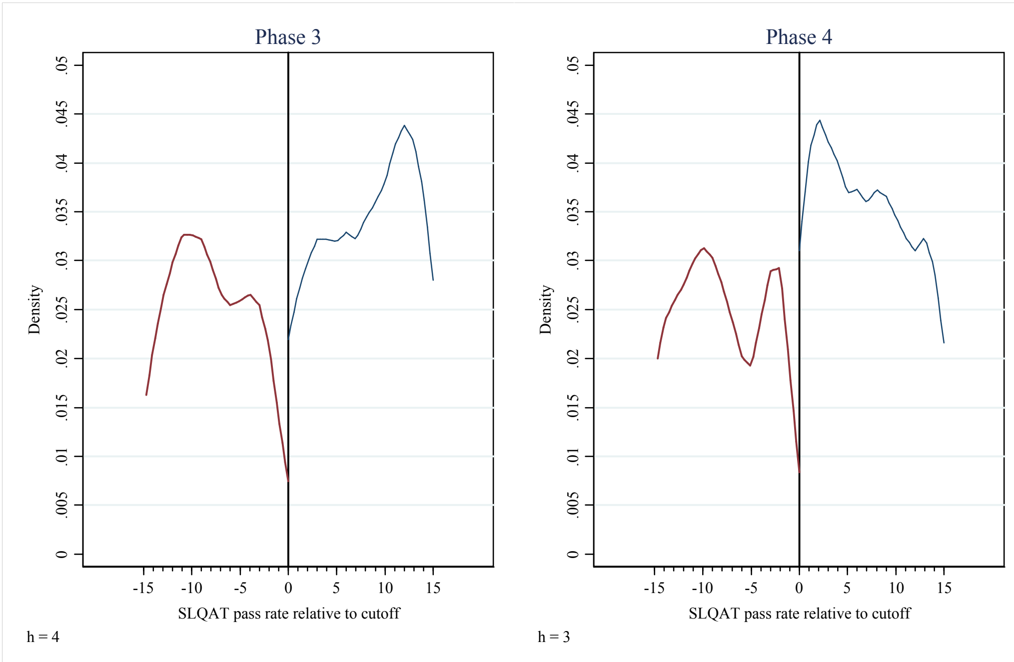

Fig. 4.
Density of SLQAT pass rates, neighborhood sample

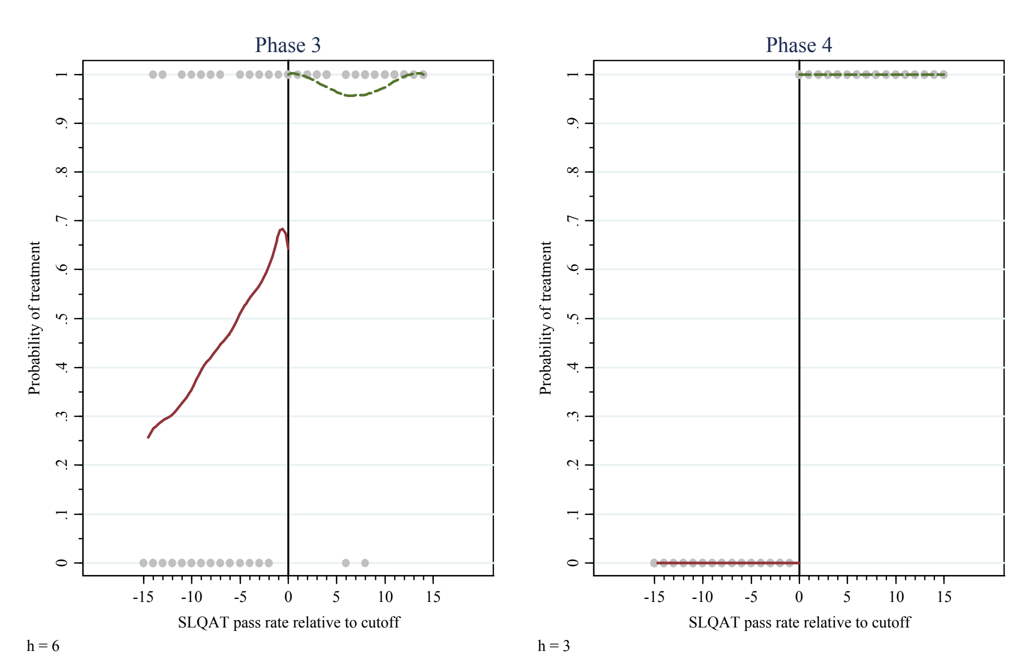

Fig. 5.
Probability of treatment, neighborhood sample

Table 10. Local discontinuity estimates in the density of SLQAT pass rates
Kernel regression with triangular kernel

Cutoffs	Phase 3			Phase 4		
	dh	$1.5\times dh$	$2\times dh$	*dh*	$1.5\times dh$	$2\times dh$
$c = 57\%$	0.0027	0.0016	0.0012	0.0032	0.0018	0.0012
	(0.0020)	(0.0015)	(0.0012)	(0.0020)	(0.0016)	(0.0012)
$c = 67\%$	0.0052***	0.0038***	0.0031***	0.0087**	0.0066**	0.0056***
	(0.0017)	(0.0013)	(0.0011)	(0.0041)	(0.0031)	(0.0024)
$c = 77\%$	0.0054**	0.0045***	0.0041***	0.0015	0.0007	0.0006
	(0.0023)	(0.0017)	(0.0014)	(0.0019)	(0.0015)	(0.0012)

Notes: * denotes statistical significance at the 10% level; ** at the 5% level; and *** at the 1% level. Bootstrapped standard errors based on 200 replications are presented in parentheses. The default bandwidth (*dh*) varies between 3-4 percentage points.

Table 11. Differences in pre-treatment mean outcomes between SLQAT marginal passers and marginal failers

Outcome measure	Phase 3				Phase 4			
	$z \in [52,82]$		$z \in [62,72]$		$z \in [52,82]$		$z \in [62,72]$	
	(1) Marginal failers Mean (SD)	(2) Marginal passers Diff (SE)	(3) Marginal failers Mean (SD)	(4) Marginal passers Diff (SE)	(5) Marginal failers Mean (SD)	(6) Marginal passers Diff (SE)	(7) Marginal failers Mean (SD)	(8) Marginal passers Diff (SE)
Number of students	242.32	-2.52	268.5	-48.17	229.68	4.33	254.83	-28.01
	(138.80)	(16.93)	(188.03)	(37.21)	(99.26)	(12.03)	(114.57)	(21.27)
Number of male students	137.23	-0.78	162.06	-35.21	141.75	-0.24	161.98	-19.04
	(99.54)	(11.48)	(147.88)	(27.27)	(82.45)	(9.55)	(109.97)	(19.80)
Number of female students	105.09	0.04	106.44	-11.13	87.93	7.58	92.85	-7.44
	(68.52)	(8.67)	(65.47)	(15.60)	(44.37)	(6.26)	(48.03)	(9.53)
Mean SLQAT score	36.26	7.67***	38.55	2.42***	36.63	6.41***	39.05	1.50**
	(3.76)	(0.56)	(3.68)	(0.87)	(3.87)	(0.49)	(2.66)	(0.61)
Number of teachers	9.91	-0.15	9.82	-0.79	9.23	0.14	9.80	-0.63
	(4.22)	(0.53)	(5.41)	(1.12)	(3.36)	(0.42)	(4.03)	(0.74)
Number of classrooms	8.88	0.04	8.41	0.01	8.53	-0.04	9.61	-1.22*
	(3.83)	(0.47)	(4.10)	(0.84)	(3.83)	(0.43)	(4.12)	(0.74)
Number of blackboards	9.29	0.16	8.74	-0.02	9.17	0.11	10.22	-1.16
	(4.10)	(0.51)	(4.59)	(0.96)	(3.72)	(0.44)	(4.34)	(0.78)
Number of toilets	3.08	0.01	2.74	0.23	2.98	-0.06	3.37	-0.32
	(1.66)	(0.20)	(1.83)	(0.37)	(1.74)	(0.20)	(2.44)	(0.45)
Student-teacher ratio	24.67	0.53	25.95	-0.53	25.48	0.18	26.57	-1.10
	(9.28)	(1.18)	(8.10)	(2.16)	(7.90)	(0.93)	(7.32)	(1.47)
Student-classroom ratio	28.69	-0.66	31.97	-4.87	28.18	0.28	27.44	1.55
	(13.58)	(1.72)	(15.61)	(3.32)	(10.69)	(1.29)	(8.77)	(2.04)

Notes: * denotes statistical significance at the 10% level; ** at the 5% level; and *** at the 1% level. z denotes the treatment assignment variable; SD standard deviation; and SE standard error.

Table 12. Local smoothness in pre-treatment conditional mean outcomes
Local linear regression with triangular kernel and bandwidth h

Outcome measure	Phase 3			Phase 4		
	(1)	(2)	(3)	(4)	(5)	(6)
	$h = 4\%$ pts	$h = 6\%$ pts	$h = 8\%$ pts	$h = 3\%$ pts	$h = 4.5\%$ pts	$h = 6\%$ pts
Number of students	-97.05	14.32	-39.12	-78.43	-132.88*	-107.76**
	(142.69)	(91.80)	(62.11)	(85.17)	(68.85)	(54.54)
Number of male students	25.27	54.81	0.83	-50.27	-102.17*	-66.13
	(87.42)	(56.82)	(38.33)	(67.53)	(52.90)	(46.52)
Number of female students	-122.23*	-39.22	-38.85	-11.28	-23.37	-27.15
	(66.47)	(39.55)	(28.61)	(50.05)	(29.77)	(23.58)
Mean SLQAT score	6.45**	4.04**	2.47	0.42	-1.14	-1.40
	(3.15)	(2.03)	(1.72)	(2.78)	(2.06)	(1.54)
Number of teachers	-0.72	1.35	-0.59	4.37*	-1.41	-1.78
	(3.66)	(2.52)	(1.75)	(2.55)	(3.25)	(2.02)
Number of classrooms	0.31	1.78	0.61	2.43	-2.17	-2.44
	(2.90)	(2.13)	(1.51)	(2.17)	(2.67)	(1.78)
Number of blackboards	-0.74	1.71	0.69	0.67	-2.98	-2.95*
	(3.50)	(2.19)	(1.62)	(1.69)	(2.39)	(1.77)
Number of toilets	1.36	1.45**	0.93	-1.32	-2.13	-2.19*
	(1.23)	(0.73)	(0.58)	(1.91)	(1.54)	(1.23)
Student-teacher ratio	-8.53	-1.37	-0.15	-22.37*	-11.91	-8.70
	(8.62)	(4.09)	(3.37)	(11.77)	(8.70)	(5.99)
Student-classroom ratio	-14.37	-3.8	-4.19	-12.26	-5.73	-3.12
	(12.16)	(7.34)	(5.83)	(9.35)	(7.05)	(5.61)

Notes: * denotes statistical significance at the 10% level; ** at the 5% level; and *** at the 1% level. Bootstrapped standard errors based on 500 replications are reported in parentheses.

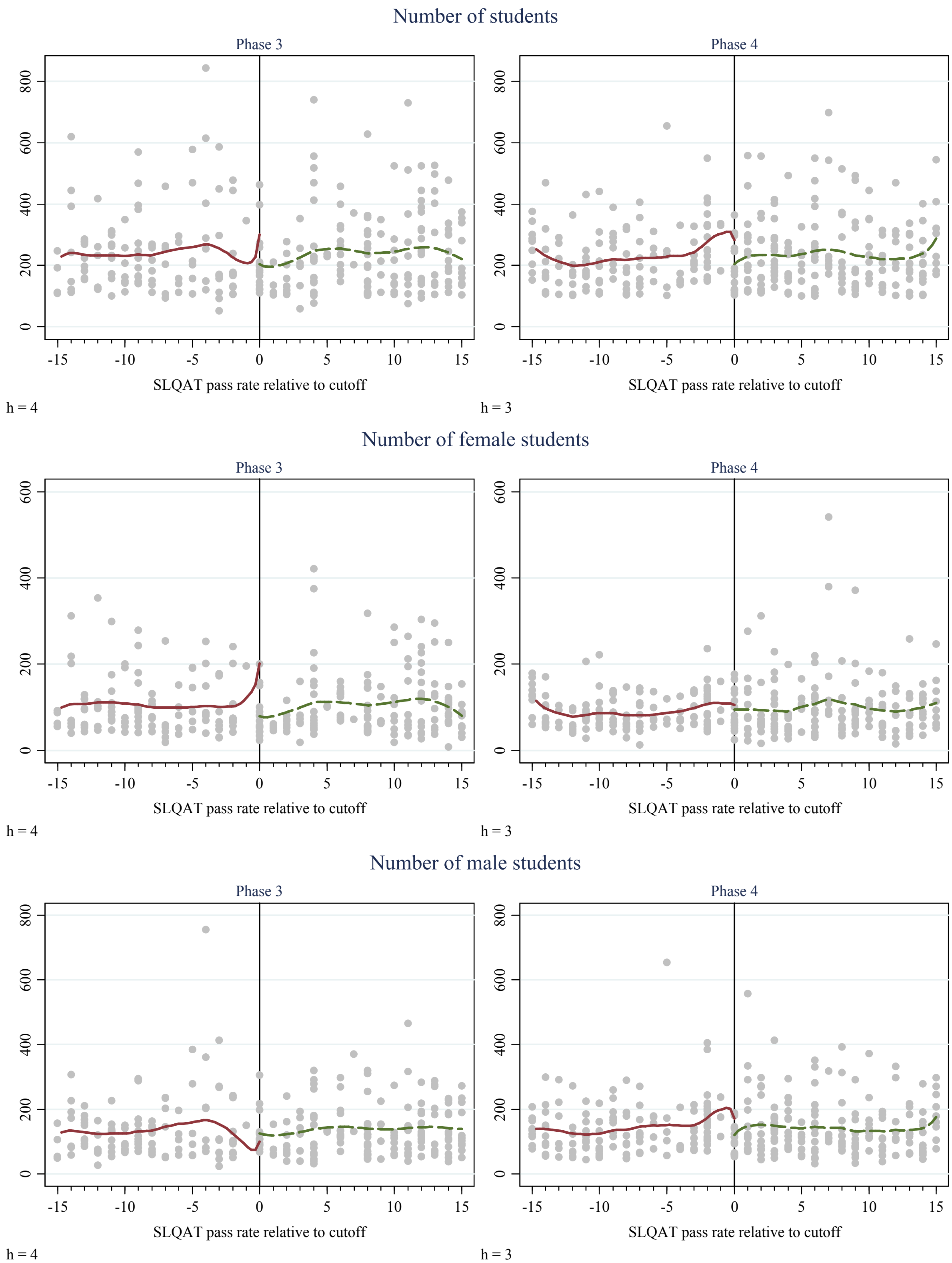

Fig. 6a
Local smoothness in pre-treatment conditional mean outcomes

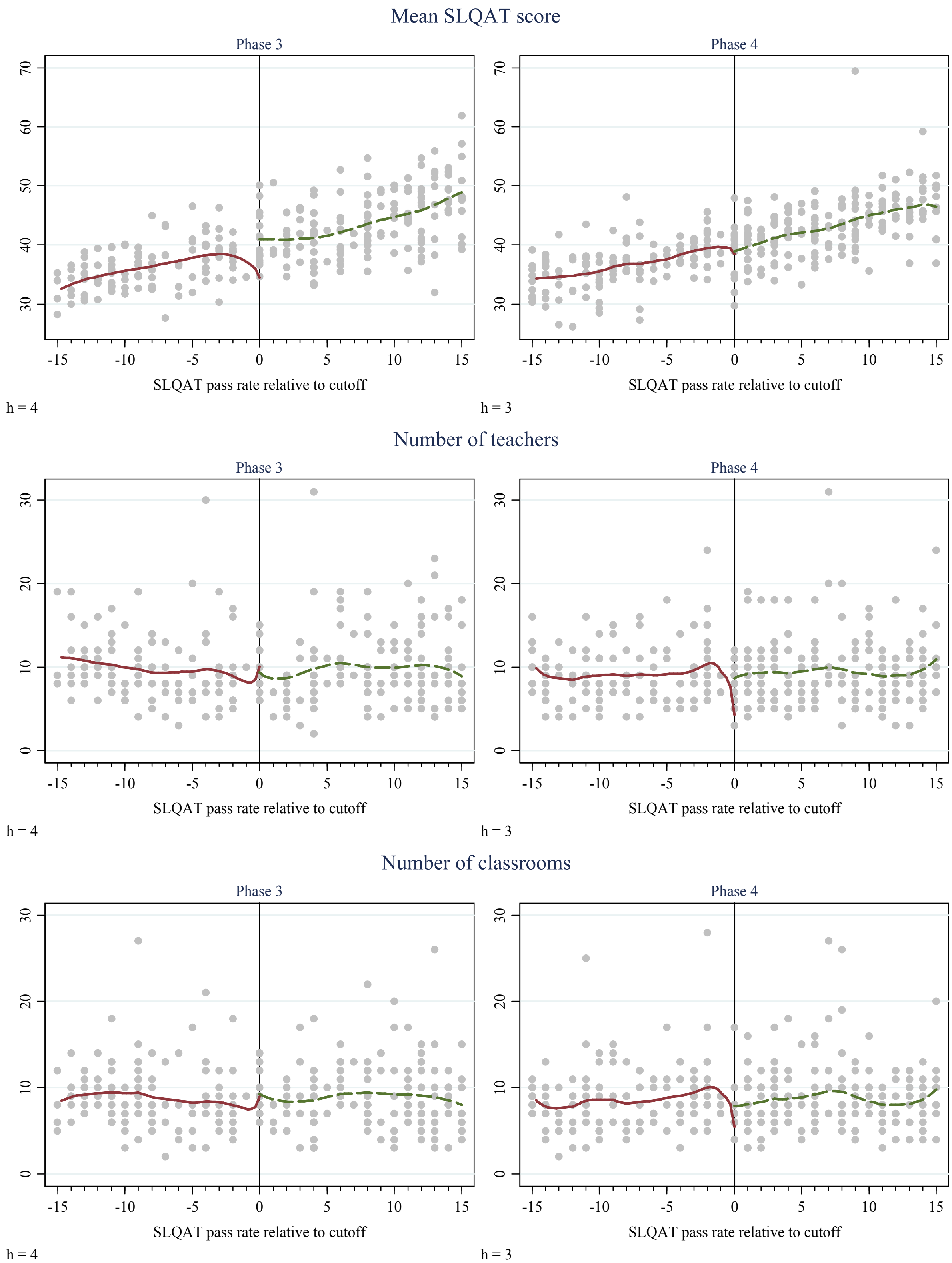

Fig. 6b.
Local smoothness in pre-treatment conditional mean outcomes (cont.)

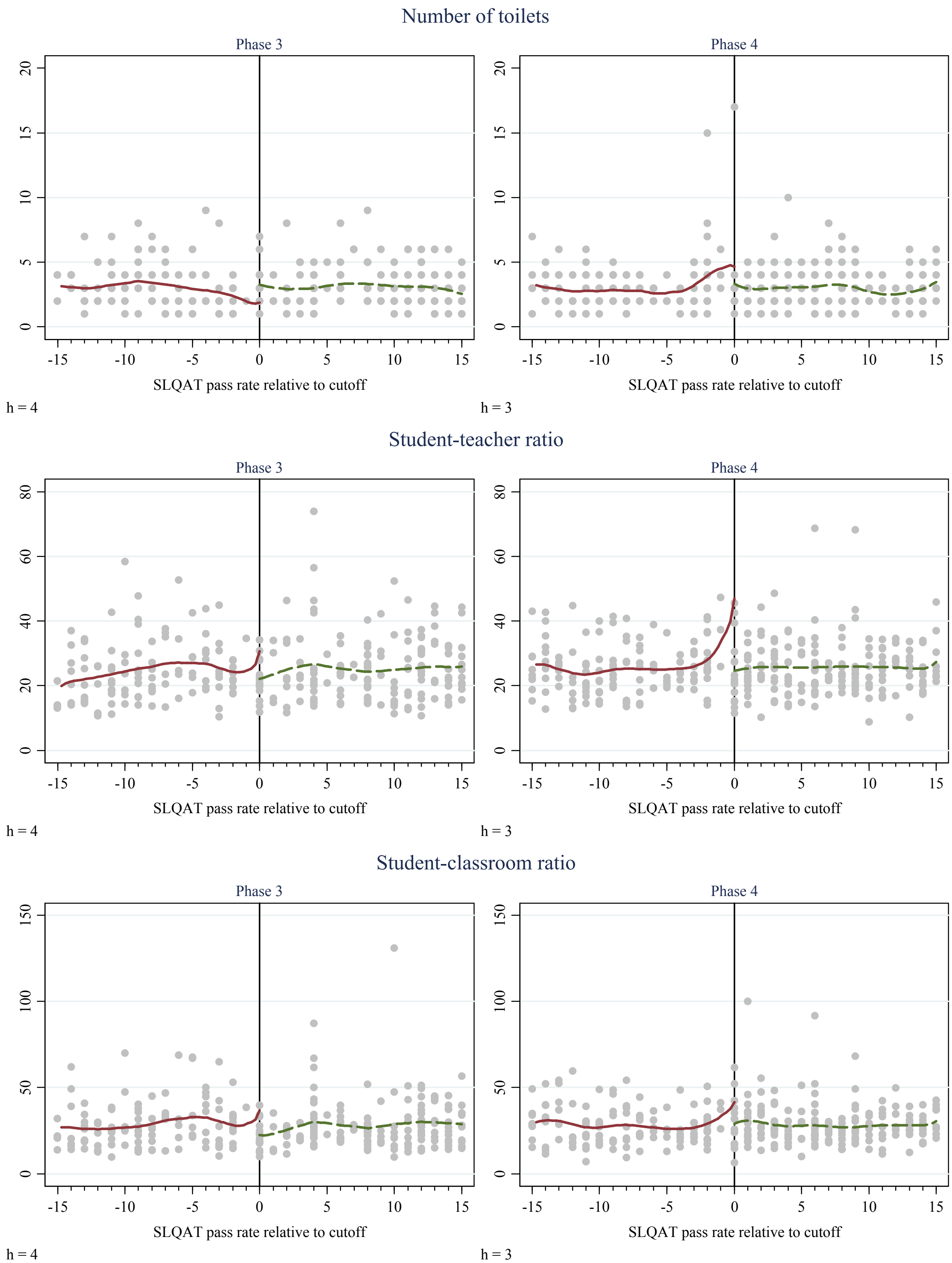

Fig. 6c
Local smoothness in pre-treatment conditional mean outcomes (cont.)

Table 13. Discontinuity estimates of post-treatment conditional mean outcomes
Local linear regression with triangular kernel and bandwidth h

Outcome measure	Phase 3 *Partially fuzzy RD estimates*			Phase 4 *Sharp RD estimates*		
	(1) $h = 6\%$ pts	(2) $h = 9\%$ pts	(3) $h = 12\%$ pts	(4) $h = 3\%$ pts	(5) $h = 4.5\%$ pts	(6) $h = 6\%$ pts
Number of students	180.3	-29.03	-28.03	121.50***	85.20**	88.84**
	(8.09e+14)	(18002.95)	(6230.31)	(46.20)	(42.77)	(41.32)
Number of teachers	6.46	3.34	3.09	5.06**	3.39*	3.46*
	(3.96e+13)	(331.89)	(42.19)	(2.32)	(1.92)	(1.82)
Number of classrooms	2.63	-0.75	1.43	9.82**	4.55*	4.00**
	(1.77e+13)	(121.08)	(95.78)	(3.97)	(2.38)	(2.02)
Number of blackboards	--	--	--	6.51**	3.14*	2.83*
				(2.59)	(1.85)	(1.72)
Number of toilets	--	--	--	0.08	-0.37	-0.14
				(1.26)	(0.95)	(0.85)
Student-teacher ratio	0.39	-6.17	-5.07	1.15	-0.85	-1.04
	(97.98)	(86.51)	(114.35)	(3.77)	(2.77)	(2.54)
Student-classroom ratio	11.48	1.45	-4.77	-29.2	-11.39	-8.12
	(88.45)	(131.10)	(59.86)	(22.17)	(9.89)	(7.24)

Notes: * denotes statistical significance at the 10% level; ** at the 5% level; and *** at the 1% level. Bootstrapped standard errors based on 500 replications are reported in parentheses.

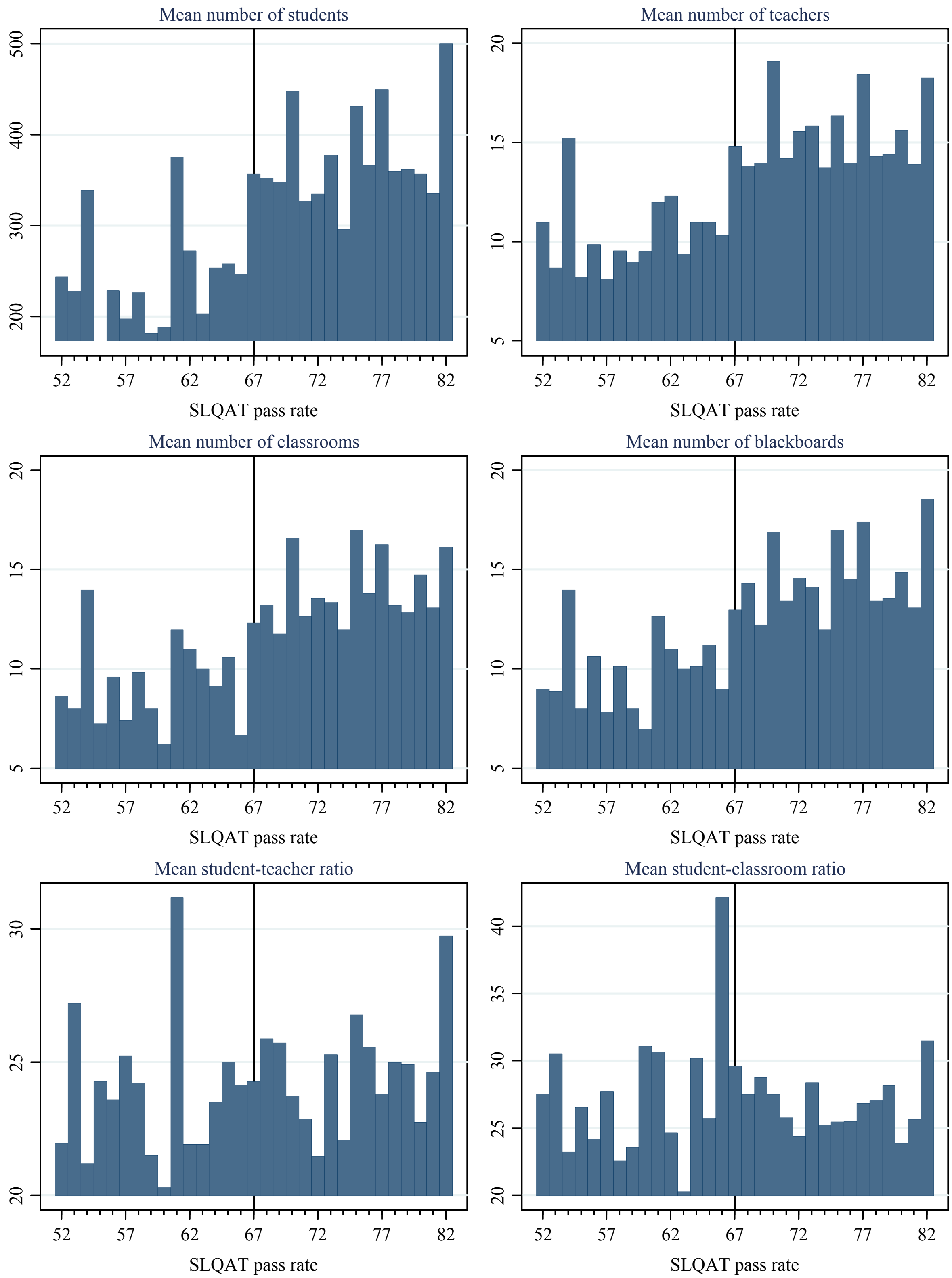

Fig. 7.
Histogram of post-treatment mean outcomes

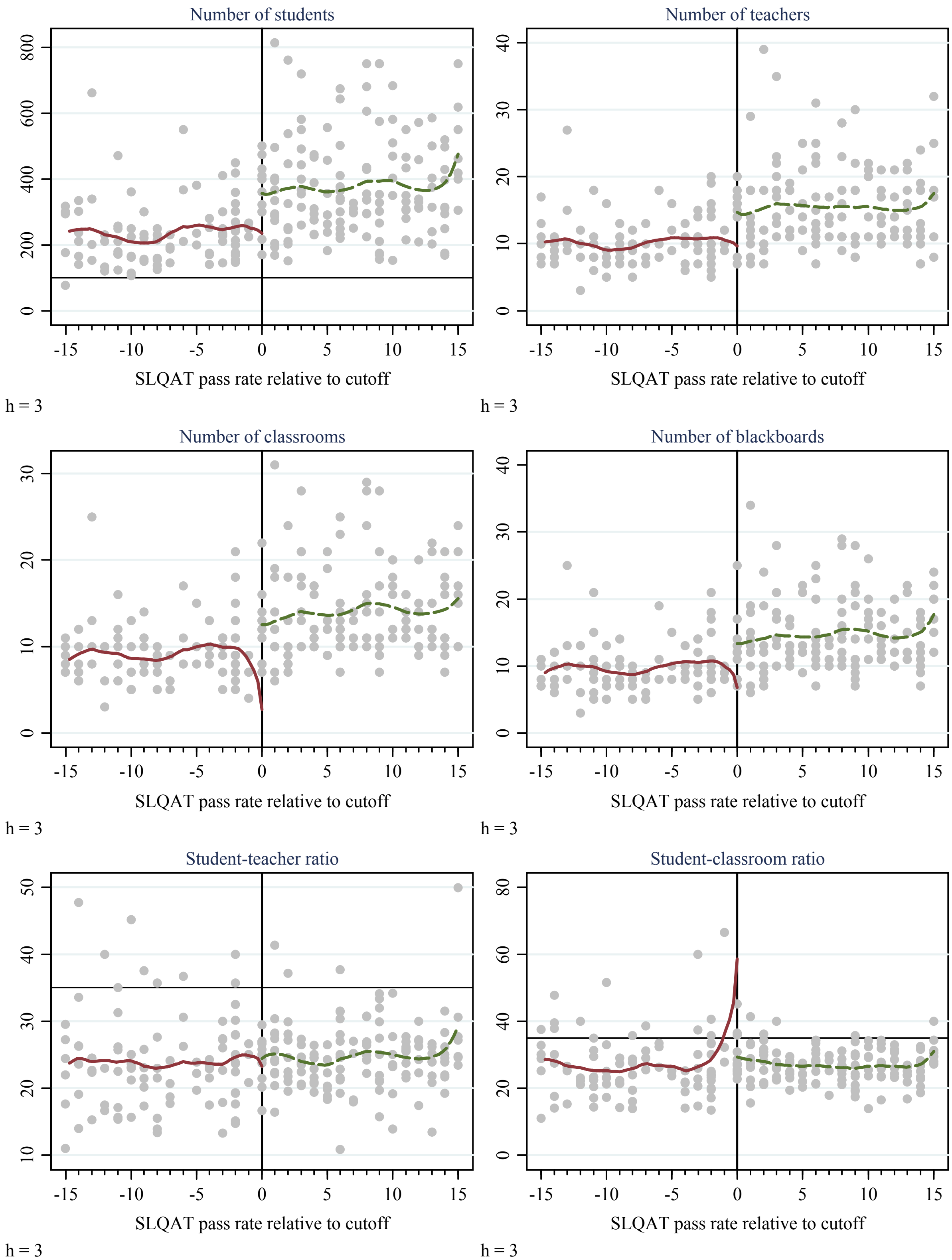

Fig. 8.
Local discontinuities in post-treatment conditional mean outcomes