Asia Pacific Education Review
https://doi.org/10.1007/s12564-020-09667-5

Promoting student achievement in high school using school funding: evidence from quantile regression discontinuity design

Dongsook Han[1] · Sangheon Kim[2] · Heeran Park[3] · Hosung Sohn[4]

Received: 17 August 2020 / Revised: 26 December 2020 / Accepted: 28 December 2020

Abstract
Even though there are many quasi-experimental research in recent literature, there is still no consensus on whether an increase in school funding improves student achievement. Leveraging a natural experiment in South Korea, this study exploits the discontinuity in school funding rules to identify the impact of increased funding on the test scores of high-school students in a national assessment exam. The setting provides a useful context to study the effect of school funding because students typically attend largely similar schools that follow a standardized curriculum, thus eliminating the possibility of the results being contaminated by idiosyncratic variation in school-level characteristics. This study reports mean regression discontinuity estimates as well as quantile regression discontinuity estimates using a procedure suggested by Frandsen et al. (J Econom 168:382–395, 2012). The findings reveal that an increase in school funding, which is equal to approximately 300,000 won per student, results in improved exam performance, particularly in mathematics. Contrary to the stated purpose of the program, however, the evidence suggests that students in the middle and top of the ability distribution gained the most from the intervention, rather than students who are at the highest risk of failing.

Keywords School funding · Regression discontinuity design · Quantile regression discontinuity · Student achievement

Introduction

The effect of school resources on student achievement has been one of the most comprehensively examined topics in education finance literature. A review paper by Häkkinen et al. (2003) identifies many studies that have explored this topic, and many other academic works have also been published since the date of their study. Despite the size and scope of the existing literature on this topic, however, the degree to which increase in school funding contributes to improvements in student achievement remains unresolved. This is especially true for high-school students because the vast majority of the literature examines the effect of funding on academic outcomes for elementary school students.

On the empirical side, the main threat to identifying a causal effect of school funding on student achievement is the concern that school resources may be correlated with unobserved student characteristics that are directly related to student achievement. For instance, this might be the case if high achieving students are able to sort themselves into well-funded or otherwise high performing schools. This is an important concern in a country where school funding is primarily determined at the local level and is a locational amenity that is reflected, to a large extent, in the price of real estate (Bogart and Cromwell 1997; Goodman and Thibodeau 1998; Downes and Zabel 2002; Clapp et al. 2008). While previous research attempts to solve this confounding problem using cross-sectional data and control function methods, recent approaches in empirical analysis have sought to achieve causal identification using natural experiments leveraging either difference-in-differences around a particular intervention or administrative discontinuities that induce random variation in school funding.

While the older, predominantly cross-sectional literature, which is summarized in a series of reviews by Hanushek

✉ Hosung Sohn
sohnhosung@gmail.com

1 Research Center for State-Owned Entities, Korea Institute of Public Finance, Seoul, South Korea

2 Graduate School of Public Administration, Seoul National University, Seoul, South Korea

3 Institute of Public Policy and Administration, Chung-Ang University, Seoul, South Korea

4 School of Public Service, Chung-Ang University, Seoul, South Korea

Published online: 09 February 2021

(1986, 1994, 1997), tends to find little evidence of there being a relationship between school funding and student achievement, more recent quasi-experimental literature have produced more mixed results.[1] These literature can be divided into three primary strands: (i) natural experiments that leverage panel data around the timing of school finance reforms, (ii) natural experiments that utilize an administrative discontinuity that produces a nonlinearity in school funding rules, and (iii) randomized experiments. With regard to the first type of natural experiments—i.e., natural experiments that involve analyzing the effect of changes in school financing—three recent studies are worth noting. The first is Guryan (2001), which exploits variation induced by an education finance "equalization" plan in Massachusetts to estimate the effect of increased spending on schools in traditionally underfunded districts. The findings of this study reveal that there were improvements in test scores for fourth graders but not for eighth graders. The second is Papke's (2005) study, which examined a similar equalization plan implemented in Michigan, and the findings show that the plan resulted in improvements in math test scores for fourth and seventh graders, but there was no significant effect on verbal scores. The last one is Card and Payne's (2002) study, which used national-level data to analyze the effect of education finance reforms more generally. Their findings reveal that when spending is equalized across school districts, the distribution of test score outcomes is also equalized, suggesting that funding is an important factor in enhancing student achievement.

The second type of natural experiment exploits an administrative rule that is used to distribute funding to individual schools. These studies typically utilize a regression discontinuity design to compare schools that, on the basis of prior test scores, are eligible to receive enhanced funding with schools that are not eligible to receive the funding. The intuition of the design is based on the discontinuity in the policy variable, eligible and ineligible schools should be relatively similar, and therefore, selection bias is negligible. One such example of a regression discontinuity study is that of Jacob and Lefgren (2004), who examined the effect of an accountability policy in Chicago that tied the provision of remedial education and summer school programs to student achievement using a cutoff; they found evidence of positive effects for third graders but not for sixth graders. Similarly, Chay et al. (2005) sought to examine the effect of an administrative cutoff in the assignment of extra resources for poorly performing schools in Chile. Their findings revealed that the increased funding had significant effects on fourth graders. More recently, van der Klaauw (2008) exploited a reform in New York City that provides additional funding to low performing schools, and they found no evidence of improved test scores.

Despite the emergence of a rich quasi-experimental literature, several gaps still exist. First, while a number of studies address the effect of funding on academic outcomes for younger students, few recent studies examine the effect of funding on high-school student's test scores. This is primarily attributed to the fact that older students take exams selectively, which leads to a sample selection issue. Given the persistence of high dropout rates, however, the effect of increased funding on the performance of older students remains a first order question.[2] Parallel literature has considered the effect of school quality more broadly, taking into account the effect of high versus low-quality education regimes. The most notable study in this literature is that of Card and Krueger (1992), who used earnings data from the U.S. Census and the state in which an individual received the majority of education to examine the effect of educational quality on subsequent earnings.

Second, the internal validity of prior research depends critically on identifying assumptions that are difficult to verify. With respect to natural experiments regarding school finance reform, identification relies on the assumption that the timing of reforms is random and that no other policy interventions coincide with the reforms. Likewise, with respect to regression discontinuity designs, estimates represent the causal effects if students and schools that are above and below the threshold are identical in all respects. While there are strong reasons to believe that students on either side of the discontinuity are identical, it is less plausible that this is also the case for the schools that the students attend. This is because in most countries, many educational policies are set locally (at the school board or even the school-level), and as a result, there is a great deal of idiosyncratic variation among schools. In theory, one could test for the smoothness of school-level variables as a function of the running variable. In practice, however, there are often too few schools within a reasonable bandwidth of the discontinuity to generate sufficiently powerful tests. The second problem is that students may have the ability to sort endogenously into schools based on their test scores.

The educational setting in South Korea offers a key advantage to studies on the effect of school funding on

[1] Verstegen and King (1998), citing among other research, and a study by Hedges et al. (1994), note that the best evidence from among the older studies indicates an association between school funding and student achievement.

[2] A parallel literature has considered the effect of school quality more broadly, considering the effect of high versus low quality education regimes. The most notable study in this literature is that of Card and Krueger (1992) who use earnings data from the U.S. Census and the state in which an individual received the majority of his education to assess the effect of educational quality on subsequent earnings.

student outcomes. This advantage lies in the fact that, in South Korea, teacher salaries, curricula, and a variety of other school-level policies are heavily standardized and are, as this study will show, similar across schools. Moreover, in many cases, students are almost randomly assigned to schools within neighborhoods. As a result, self-selection of students and idiosyncratic variation in school characteristics are, to a greater degree, negligible in the context of South Korea. As this study will show, variations in school-level characteristics are extremely smooth as a function of the running variable that is used in South Korea to determine whether a school is eligible for increased funding.

This study examines the effect of a 2010 initiative implemented throughout South Korea to provide a onetime lump sum payment to schools with a large number of students who failed one or more national assessment examinations administered in the 10th grade. Schools for which more than 20% of students failed key examinations in 2009 received funding in the order of 300,000 won per student, roughly a 3% increase in per pupil funding for 2010. Schools below the 20% threshold were not eligible to receive the funding. All schools above the threshold received the funding while no school below the threshold received it. This study reports regression discontinuity estimates of the effect of the increased funding focusing on students who attend schools with exam failure rates in the neighborhood of the 20% threshold. The findings reveal that exam failure rates in schools just above the threshold were approximately 6% points lower in mathematics and 3 to 5% points lower in English. Leveraging a new procedure suggested by Frandsen et al. (2012), this study reports the quantile regression discontinuity estimates of the effect of the treatment on students at various points of the ability distribution. While the intended goal of the program is to reduce failure rates, the findings show that students in the middle and upper end of the ability distribution benefited the most from the program.

Institutional background

In 2008, the Ministry of Education, Science, and Technology of South Korea conducted its first nationwide assessment of educational achievement for students of all educational levels. The test was named the National Assessment of Educational Achievement (NAEA) and is comparable to the United States' National Assessment of Educational Progress. Under this program, every student in elementary, middle, and high school is tested, at the same day, on five subjects: verbal, mathematics, English, social studies, and science studies.[3]

[3] South Korea's education system is as follows: elementary school (6 years), middle school (3 years), and high school (3 years).

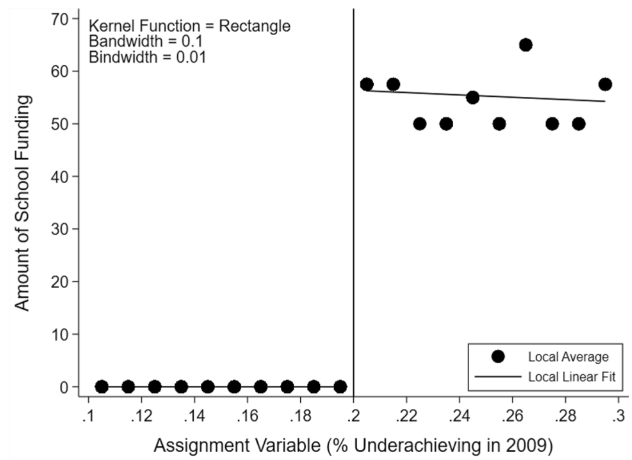

Fig. 1 Average amount of school funding by the share of underachieving students

The purpose of the NAEA is to ascertain the number of students who do not meet the basic academic standards set by the Korean government and cushion schools with a high share of underachieving students with the necessary funding to boost their performance.

To identify underachieving students, the Ministry of Education, Science, and Technology first assigns each student to one of the following four categories; (i) high achieving, (ii) normal achieving, (iii) elementary achieving, and (iv) underachieving. Next, the Ministry calculates the share of underachieving students for each school in each subject to identify poor-performing schools and provide funding for them. If the average percentage of failing students across the five subjects exceeds a specific threshold (20% for high schools), the Ministry provides funding for that school. The amount of funding varies by the number of students in each school. Specifically, the amount that each poor-performing school receives is 50,000,000 won if the number of students in a school is below 300 and 80,000,000 won if the number of students in a school is equal to 300 or above. Figure 1 plots the average amount of 2010 funding as a function of the share of underachieving students. As can be seen from the figure, a large discontinuity in the amount of school funding is observed at the 20% threshold that determines the eligibility for funding. On average, schools received approximately 300,000 won per student.

The share of underachieving students that is used to determine poor-performing schools varies depending on the school-level. For elementary schools, the cutoff is 5%; that is, when the share of underachieving students is equal to or greater than 5%, an elementary school receives funding. For middle and high schools, the minimum share required for a school to receive funding is 20%. Therefore, every school in which the share of underachieving students exceeds these cutoffs receives funding. In addition, schools

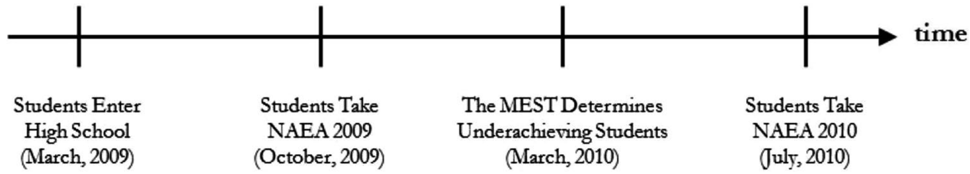

Fig. 2 Timeline

that receive funding are required to hire additional teachers (e.g., interns and retired teachers) and run after-school teaching and teacher training programs to promote students' academic achievement.

Figure 2 presents a specific timeline for NAEA 2009 and 2010, along with the timing of the funding provision. In high school, the school year begun in March 2009. Students who joined high school in 2009 took the NAEA in October 2009. Based on this NAEA, the Ministry of Education, Science, and Technology was able to identify low performing schools. The students again took the NAEA in July 2010. Consequently, this research examines whether, within a reasonable neighborhood of the 20% threshold, schools that received funding in March 2010 performed better than schools that did not receive the funding.

Data

To analyze the effect of increased school funding, the researchers used administrative records of students' test scores as well as answers on survey questionnaires retrieved from every principal and student. The survey questionnaires include a rich set of school-level characteristics such as class size and pupil-to-teacher ratio. These questionnaires also include information on teacher characteristics (e.g., the share of teachers with a master's degree) as well as student characteristics (e.g., percentage of students living in poverty). Using this information, one can test for the validity of a regression discontinuity design by examining whether there are any discontinuities in baseline covariates as a function of the running variable.

The sample used for the analysis is students in high school. Students in elementary and middle school were not used because students who took the NAEA between 2009 and 2010 are different in these two school levels. To be more specific, elementary school students in sixth grade and middle school students in third grade took the NAEA in 2009, and in 2010, students in the same grades were tested. Therefore, for these two periods, different students took the NAEA. On the other hand, high-school students in the first grade took the NAEA in 2009, and in 2010, these same students are again took the NAEA as second grade students.[4] As a result, high-school students are more suitable for analyzing the effect of school funding on students' academic achievement. The sample was also restricted to students attending high schools not located in Seoul because there are almost no schools located in Seoul that received funding.

Table 1 presents descriptive statistics for the sample used in this paper. Panel A presents the share of students who were classified as underachieving students in NAEA 2010 based individual subjects. On average, approximately 4 to 5% of students failed in verbal, math, and English tests. The standard deviations were 5 to 6% points. Note, however, that the average share of students who failed in all subjects is 0.5%.

In Table 1, Panel B presents the descriptive statistics for the baseline covariates used in the analysis. For student-level characteristics, the baseline average scores (average scores in NAEA 2009) are provided first. For the entire sample, the average share of female students is 0.47, with a standard deviation of 0.11. The baseline failure rates in social studies, verbal, math, science, and English are also presented. The highest failure rate was recorded in social studies and the lowest failure rate in verbal. On average, the failure rate is 2.5% for verbal, 9.4% for social studies, 7.3% for math, 8.7% for science, and 4.5% for English. As a proxy for students' family background, information on the share of students living in poverty as well as students receiving free lunch, an indicator that the student's family is living at least close to the poverty level, is presented. On average, 4.9% of the students are living in poverty and 12.3% of the students are receiving free lunch. Finally, as a proxy for a student's family situation, the share of students living with both parents was calculated. The share is 0.868, which signifies that approximately 13% of the students are not living with both parents.

Next, information on several key variables, each of which is measured at the school-level, is presented. First, data on the average class size as well as the pupil-to-teacher ratio is presented.[5] On average, there are approximately 35 students

[4] Note that the grade numbers start over again at each school level in South Korea, which is different from the US.

[5] The class size variable that we used in our analysis indicates the number of students in a homeroom class. In South Korea, every student is assigned a homeroom class within a school, and the class size variable denotes the average number of students in a homeroom class. The pupil-to-teacher ratio variable is obtained by dividing the total

Table 1 Descriptive statistics

Variables	Mean	Standard Deviation	Min	Max
Panel A: Outcome variables				
Share of students failed in verbal	0.045	0.054	0.000	0.440
Share of students failed in math	0.056	0.060	0.000	0.429
Share of students failed in English	0.045	0.058	0.000	0.632
Share of students failed in all subjects	0.005	0.010	0.000	0.105
Panel B: Covariates				
Baseline average verbal score	366.517	5.446	346.351	380.664
Baseline average social studies score	358.989	4.445	346.086	372.603
Baseline average math score	361.095	5.956	349.456	379.783
Baseline average science score	360.666	4.786	346.621	379.462
Baseline average English score	362.942	7.254	346.737	386.014
Baseline average total score	361.931	5.293	347.050	377.363
Baseline failure rates in verbal	0.025	0.473	0.000	0.579
Baseline failure rates in social studies	0.094	0.123	0.000	0.948
Baseline failure rates in math	0.073	0.089	0.000	0.600
Baseline failure rates in science	0.087	0.107	0.000	0.845
Baseline failure rates in English	0.045	0.076	0.000	0.722
Share of female students	0.471	0.109	0.078	0.957
Share of students in poor families	0.049	0.042	0.000	0.322
Share of free lunch students	0.123	0.120	0.000	1.000
Share of students living with parents	0.868	0.074	0.000	0.984
Average class size	34.961	5.765	9.875	49.167
Pupil-to-teacher ratio	6.565	3.295	1.681	60.600
Share of teachers with master's degree	0.409	0.161	0.032	1.000
Share of new teachers	0.050	0.077	0.000	0.500
Amount of funding (in million won)	55.844	11.959	50.000	80.000

The total number of students used for analysis is 380,649 and total number of schools used for analysis is 1234. The shares of failed students presented in Panel A are derived from the 2010 NAEA exams. All the means are estimated using school-level data. For the "Amount of funding" variable, the estimation is based on schools that received school funding. The class size variable that we used in our analysis indicates the number of students in a homeroom class. The pupil-to-teacher ratio variable is obtained by dividing the total number of students (in grade 10 to 12) in a school by the total number of teachers in a school

per class, with an overall pupil-to-teacher ratio of 6.6. Overall, the percentage of newly-hired teachers (those with fewer than 2 years of experience) is 5%. On the other hand, approximately 41% of the teachers possess at least a master's degree. Finally, the average amount of school funding that a selected school received is about 55,844,000 won, which is about $55,000 (based on an exchange rate $1 = 1000 won).

Footnote 5 (continued)

number of students (in grade 10 to 12) in a school by the total number of teachers in a school.

Empirical methods

This study employed a regression discontinuity design to estimate the effect of school funding on student achievement because the provision of school funding is "discontinuously" determined by a simple rule (i.e., the percentage of underachieving students in each school). Depending on the probability of receiving a treatment, regression discontinuity designs are classified into two types: the sharp regression discontinuity and the fuzzy regression discontinuity design. If the probability of receiving a treatment jumps from 0 to 1 as one passes through the assignment rule, one needs to use the sharp regression discontinuity design. In this study, provision of school funding (T_s) is a deterministic function of the percentage of underachieving students:

$T_s = 1\left(X_{is} \geq 0.2\right),$

where X_{is} denotes the percentage of underachieving students in school s that student i attends. Therefore, this study employed a sharp regression discontinuity design. Under the sharp regression discontinuity setting, the average treatment effect (τ_i) is

$$\tau_i = \lim_{x \to 0.2^+} E\left[Y_i | X_{is} = x\right] - \lim_{x \to 0.2^-} E\left[Y_i | X_{is} = x\right], \tag{1}$$

where Y_i is an outcome variable (e.g., students' percentile ranks). To estimate the conditional expectation function in Eq. (1), this study used a local linear regression estimator that minimizes the following:

$$\min_{\beta_l, \gamma_l} \sum_{0.2-h \leq X_{is} < 0.2} \left[Y_i - \beta_l - \gamma_l (X_{is} - 0.2)^2\right] K\left(\frac{X_{is} - x}{h}\right) \tag{2}$$

and

$$\min_{\beta_r, \gamma_r} \sum_{0.2 \leq X_{is} < 0.2+h} \left[Y_i - \beta_r - \gamma_r (X_{is} - 0.2)^2\right] K\left(\frac{X_{is} - x}{h}\right) \tag{3}$$

In Eqs. (2) and (3), β_l and β_r indicate intercepts at the left and right of the 0.2 cutoff, and γ_l and γ_r are the corresponding slope coefficients. To estimate Eqs. (2) and (3), researchers have to make choices on two key parameters: the bandwidth, h, and the kernel function, $K(\cdot)$. For the kernel function, the following triangle kernel was used:

$K(u) = (1 - |u|)1(|u| \leq 1),$

where u, in this study, is

$\frac{X_{is} - x}{h}.$

Fan and Gijbels (1996) show that a triangle kernel is optimal for a local linear regression estimator at the boundary.

Likewise, the choice of a bandwidth is important in a regression discontinuity design because it determines the sample that will be used for the local linear regression on either side of the 0.2 cutoff, and the regression discontinuity estimator can be sensitive to the choice of a bandwidth. If the regression discontinuity estimator is highly sensitive to the choice of a bandwidth, one cannot reliably conclude from the estimated treatment effect that there is, indeed, a consistent treatment effect. In the analysis to follow, therefore, we provide regression discontinuity estimates estimated from the two bandwidth choices and show the sensitivity of the regression discontinuity estimates to the choice of a bandwidth.

Note that the estimated average treatment effect for Eq. (1) does not provide information on the distributional impact of the treatment. In the current setting, the effect of school funding might be different depending on the achievement of students. That is, school funding may have more desirable effects for students in the top quantiles than those in the bottom quantiles of the distribution of students' test scores (or vice versa). Furthermore, researchers might not find any average effect even though there are apparent impacts at various points in the distribution of an outcome. This implies that when the differential effects are averaged, the resulting effect may be zero. Hence, to retrieve the distributional effects of school funding, this study estimated the quantile regression discontinuity estimator developed by Frandsen et al. (2012). The estimator uses local distribution regression to estimate the following local quantile treatment effects:

$$\tau_{QTE} = Q^Y_{X_{is} \geq 0.2}(q) - Q^Y_{X_{is} < 0.2}(q), \tag{4}$$

where $Q^Y_{X_{is} \geq 0.2}(q)$ and $Q^Y_{X_{is} < 0.2}(q)$ denote the qth quantile of an outcome variable Y for the treated and the untreated group, respectively. It is worth noting that Eq. (4) is the effect of a treatment on the distribution, not the effect of a treatment on an individual. To consistently estimate the quantile treatment effect in Eq. (4), Frandsen et al. (2012) propose using local linear regression to estimate the distribution of outcomes. Using their method, this study provides quantile treatment effects for deciles ranging from 0.1 to 0.9 in 0.1 increments. As with the regression discontinuity estimator, which focuses on estimating the mean treatment effect, a researcher needs to choose a bandwidth for estimating the quantile treatment effect. Frandsen et al. (2012) propose a data-driven choice of a bandwidth. This study follows their suggestion in choosing the bandwidth and provide quantile regression discontinuity estimates based on their suggested choice of bandwidth.[6]

Results

Main results

A regression discontinuity design returns a plausible estimate of the causal effect of a treatment under certain conditions. In particular, the validity of the design hinges on the assumption that the only thing that changes discontinuously as a function of the running variable is whether an observational unit is treated. Depending on the extent to which the covariates that are theoretically related to student achievement are not smooth across the 20% threshold of the running variable, regression discontinuity estimates of the effect of

[6] For the quantile regression discontinuity estimator, we benefited from the *STATA* code provided by the authors.

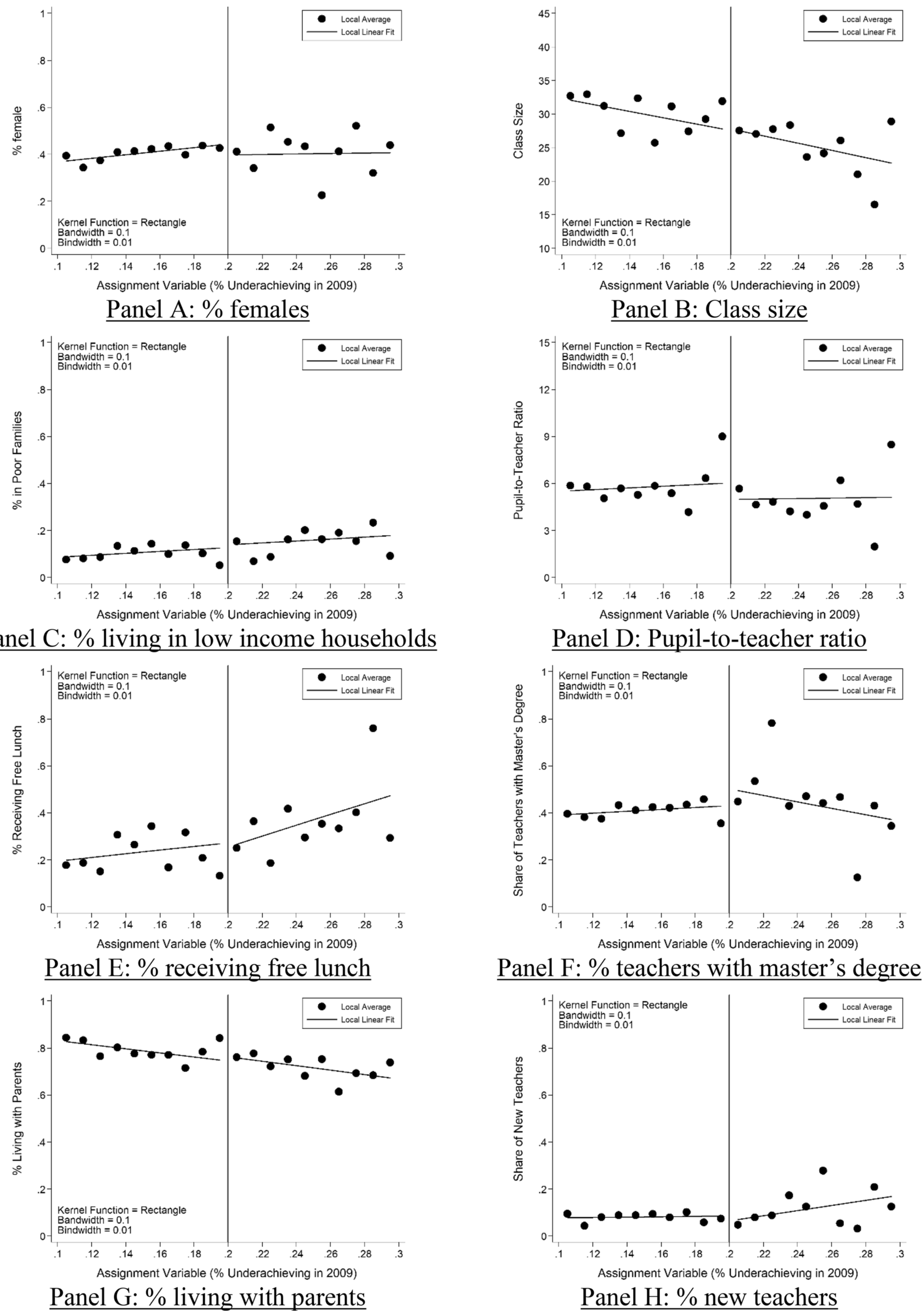

Fig. 3 Baseline covariates by assignment variable. **a** % females. **b** Class size. **c** % living in low income households. **d** Pupil-to-teacher ratio. **e** % receiving free lunch. **f** % teachers with master's degree. **g** % living with parents. **h** % new teachers

funding no longer plausibly approximate an experiment. As such, the discussion of this study's results begins with Fig. 3, which considers the smoothness of key covariates across the 20% policy threshold.

This study examined the following student characteristics, each of which is measured as the average among students at the school-level: (i) the share of female students, (ii) the share of students living in poverty, (iii) the share of students receiving free lunch, and (iv) the share of students living with both biological parents. This study also examined the following school-level variables: (i) class size, (ii) student-to-teacher ratio, (iii) the share of teachers with a master's degree, and (iv) the share of new teachers—those with less than 2 years of experience. Figures 2 to 8 plot student- and school-level data as a function of the running variable using a local polynomial regression of degree 1 (i.e., local linear regression), a rectangle kernel function, and a bandwidth and binwidth of 0.1 and 0.01, respectively. For the sake of consistency, this study follows these specifications in creating regression discontinuity design-type plots. Note that the figures are qualitatively similar even if other specifications such as a triangle kernel function and other bandwidth choices were used. In an effort to ascertain the transparency of the estimated effects, this study used a triangle kernel function with varying bandwidth choices to test for the discontinuity at the cutoff point. Regression lines are fitted separately for students and schools below and above the 20% policy threshold. An examination of all the panels in Fig. 3 reveals very little evidence as to whether any of these covariates are nonsmooth across the 20% threshold. The largest discontinuity estimate is for students living with a single biological parent, which is approximately a 2% point difference local to the threshold. As it will be demonstrated later, neither this discontinuity estimate nor any other rises to the level of statistical significance. Because the figures clearly show smoothness across the 20% cutoff, this study does not engage in interpreting all the figures for the sake of simplicity.

Figure 4 presents the same plots using data on student-level baseline exam scores in five subjects: verbal (Korean), social studies, mathematics, science, and English. Combining all these five subjects, panel f in Fig. 4 presents the baseline total test scores. As can be seen from all the figures, there is no visually clear break at the 20% cutoff point in these baseline student exam scores. If there was discontinuity in these baseline exam scores, the claim that the school funding contributed to promoting student achievement (i.e., the discontinuity in the outcome variables) could be questionable. Therefore, the results shown in Fig. 4 are favorable for the identifying assumptions of the regression discontinuity design adopted in this study. Figure 5 presents the baseline failure rates of the five subjects based on the assignment variable. Similar to the baseline exam scores, all the failure rates are extremely smooth across the assignment variable. Furthermore, there are no discernible discontinuities at the 20% cutoff that determines the eligibility for school funding. As a result, Fig. 5 lends support to the identifying assumptions of the regression discontinuity design.

Note that students took five subjects in NAEA 2009, and the share of underachieving students in each school is calculated based on the share of underachieving students in each of the five subjects. In NAEA 2010, however, the Korean government decided to administer only three subjects out of the five subjects; i.e., verbal, mathematics, and English. Therefore, because there are no scores on social studies and science, only these three subjects were analyzed to examine the effect of school funding on outcome variables.

Given that many of the baseline school- and student-level characteristics are balanced across the 20% cutoff point, identifying assumptions of the regression discontinuity design adopted in this study is met (i.e., no discontinuity in baseline characteristics at the cutoff point). Note, however, that graphical analyses presented in Figs. 3, 4, and 5 do not allow the statistical significance of the discontinuity at the cutoff point to be formally tested, so the formal results are presented much later.

Before formal statistical analyses are conducted, the graphical analyses of outcome variables are presented in Fig. 6. In Fig. 6, the same plots using data on student-level exam scores in three subjects—verbal in panel a, mathematics in panel b, English in panel c, and all subjects in panel d—are presented. As with all the other figures, regression lines are drawn using local linear regression, a rectangle kernel function, and bandwidth and binwidth of 0.1 and 0.01, respectively. In Fig. 6, the percentage of students in each bin who failed a given exam is plotted as a function of the running variable. Panel d in Fig. 4 plots the percentage of students who failed all the three exams as a function of the running variable. Referring to Fig. 6, there appears to be an approximate decline in the proportion of students who failed the verbal examination of 3% local to the 20% funding threshold. Panel b documents an approximate decline of 4% points (from 15 to 11%) in the number of students who failed the mathematics exam. Panel c provides evidence of a 3% points decline for English examination. Overall, the percentage of students who failed the exams dropped from 2% for students in schools that were just below the threshold to less than 1% for students in schools that were just above the threshold (see Panel d). Figure 7 presents identical plots using school-level data as opposed to student-level data. The results are remarkably similar. Regardless of whether student- or school-level data are analyzed, the effects of the funding are qualitatively large, representing a 25 to 50% reduction in the proportion of failing students for a given examination subject.

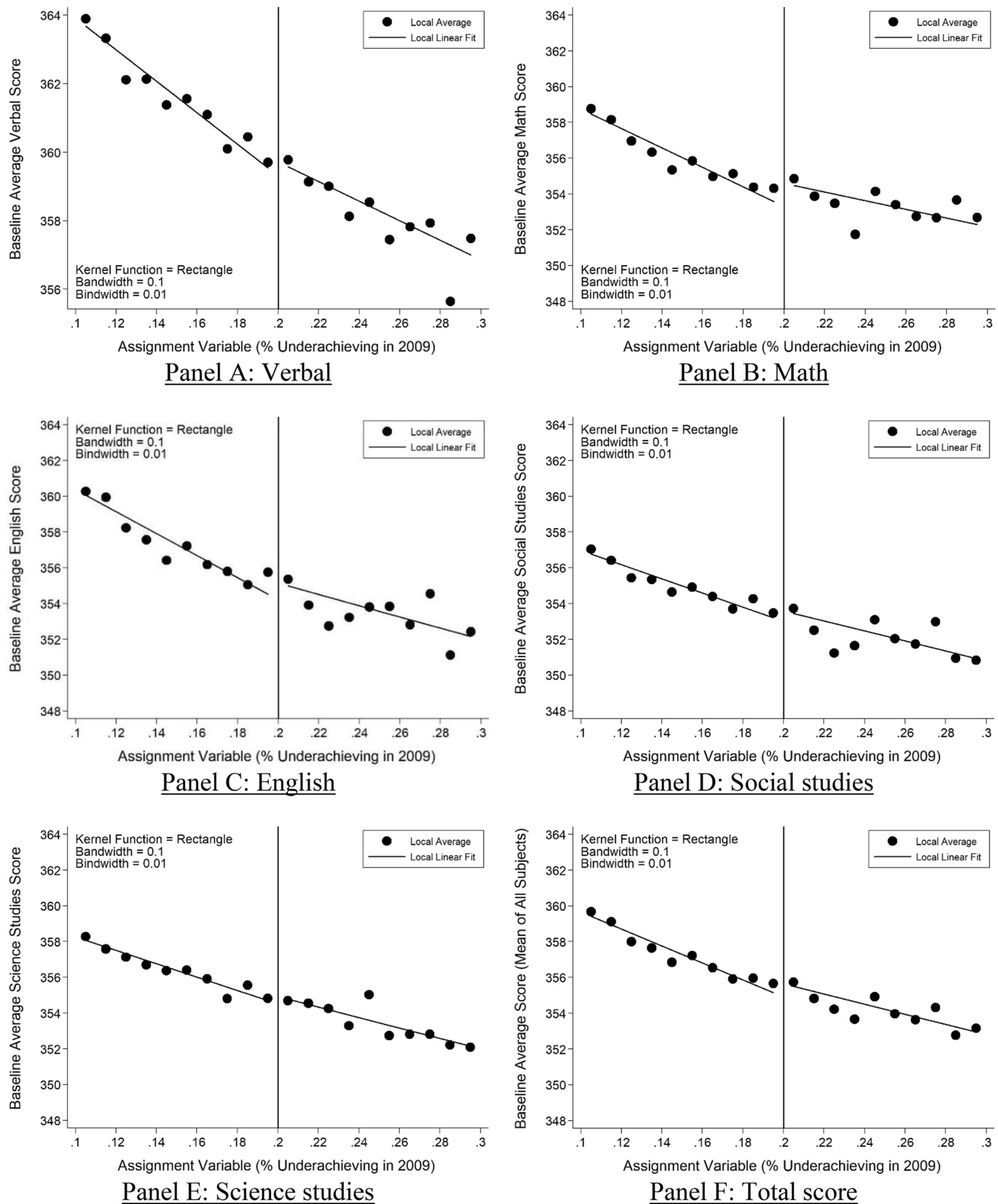

Fig. 4 Baseline test scores by assignment variable. **a** Verbal. **b** Math. **c** English. **d** Social studies. **e** Science studies. **f** Total score

Table 2 presents the same data in tabular form using local linear regression to compute regression discontinuity estimates along with standard errors. The table presents regression discontinuity estimates for outcomes using both school- and student-level data as well as regression discontinuity estimates for covariates measured at the school-level. The regression discontinuity estimate is the gap between the regression predictions for the bins local and the 20% threshold where "local" is defined using a particular bandwidth. To assess the extent to which estimates of the treatment effect are sensitive to the choice of a bandwidth, the regression discontinuity estimates and standard errors for two different bandwidths around the 20% threshold (0.05 and 0.10, respectively) are computed. In Table 2, the smoothness of the covariates as a function of the discontinuity in the running variable for each of the eight covariates and the six

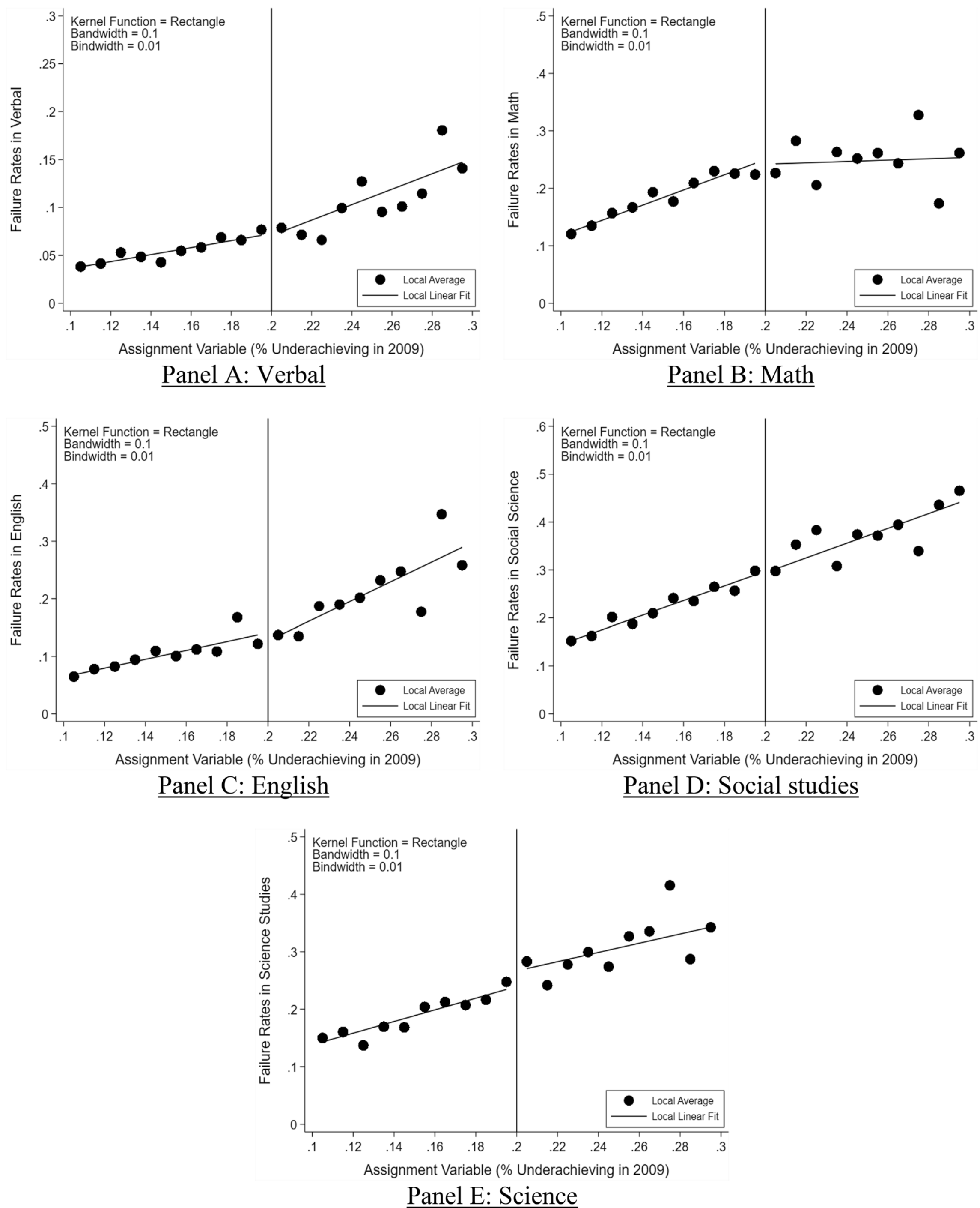

Fig. 5 Failure rates by assignment variable. **a** Verbal. **b** Math. **c** English. **d** Social studies. **e** Science

baseline test scores is assessed. For each of the three bandwidths, there is absolutely no evidence of any discontinuity in any of the six baseline test scores or in any of the eight student or school characteristics. Although the presence of discontinuities along unobserved covariates cannot be ruled out, the remarkable smoothness of the observed covariates lends considerable credence to the research design.

Table 3 computes the regression discontinuity estimates of the outcomes using student- and school-level data. Referring to Panel A of Table 3, which uses student-level data, there is evidence of a 6% point reduction in the proportion of students failing the mathematics exams, depending on whether a 0.05 or a 0.10 bandwidth is employed. Using a bandwidth of 0.10 shows that there is a 2.1 and 2.9% point

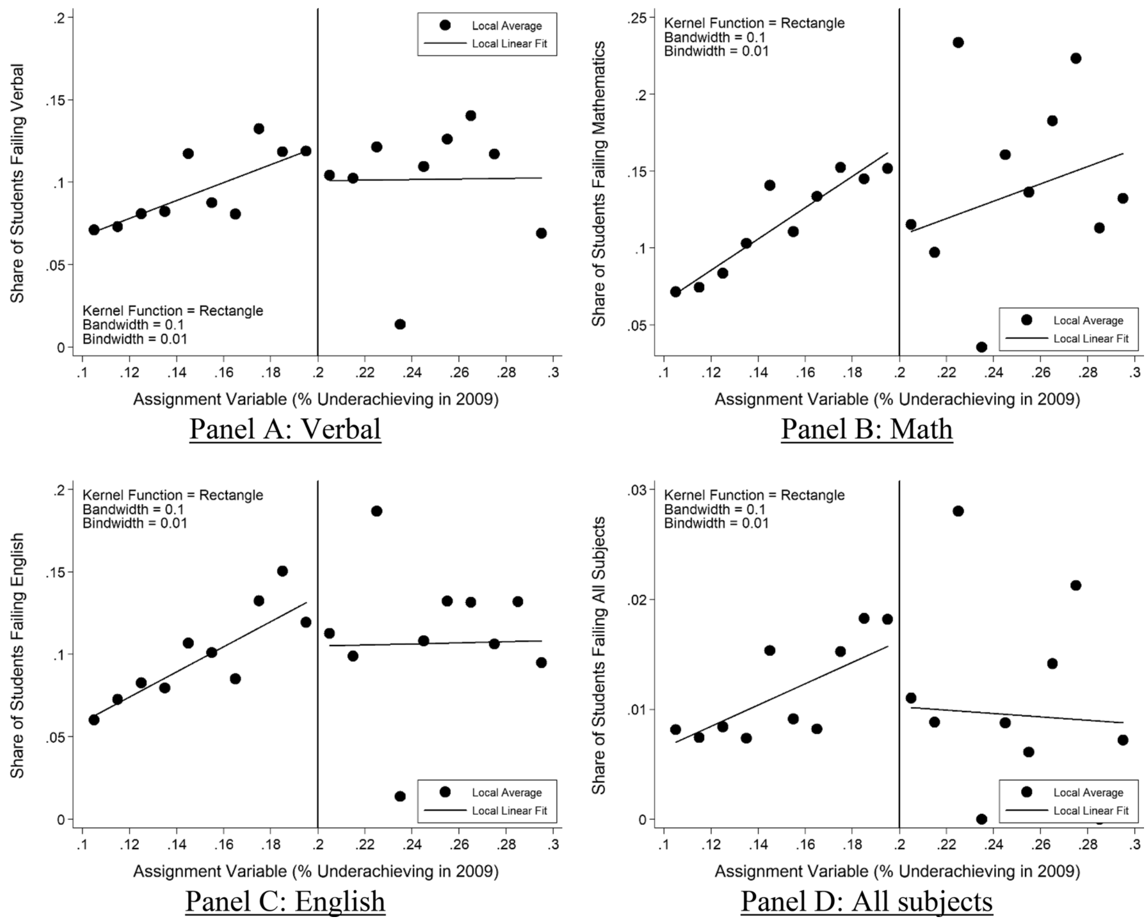

Fig. 6 Share of students failing by assignment variable (student-level data). **a** Verbal. **b** Math. **c** English. **d** All subjects

decline in the proportion of students failing verbal and English examinations, respectively, although these effects are estimated imprecisely. The effects on students failing all the exams are small (less than 1% point). Panel B of Table 3 presents the same estimates using school-level data. On the whole, the results are quite similar to those presented in Panel A of Table 3, although the standard errors are slightly smaller in particular for the analysis based on the bandwidth choice of 0.10. On average, the effect estimates for verbal, math, and English under the bandwidth choice of 0.10 turn out to be slightly larger. For example, the effect estimate for math is about 8.4% points, which is 2.5% points larger than those obtained for student-level data. Regardless, because the estimated effects on mathematics are all statistically and practically significant and stable across the specifications, it is clear that school funding was helpful in promoting the math achievement of low-performing students.

Quantile treatment effects

According to the Ministry of Education, Science, and Technology, the purpose of the funding program is to provide additional resources to schools with a high number of underachieving students to help them make adequate educational progress. Despite this program's mandate, schools are free to use the funding they are allocated in various ways; they don't receive any formal instructions as to how the funds should be used to address the needs of underachieving students. Considering the purpose for which the funding is allocated, it is important to ascertain whether students whose test scores are on the lower end of the distribution actually benefit from this program. To explore the distributional implications of the program, this study employs a novel procedure proposed by Frandsen et al. (2012), which is designed to generate regression discontinuity estimates for different quantiles of an outcome variable. The method compares test scores within a given quantile (or, in our case, decile) of the outcome variable for students in treated versus untreated schools. This exercise allows for determining which students benefited most from the funding enhancement.

Quantile regression discontinuity estimates are presented in Table 4, which reports treatment effects and standard errors (in parentheses) for each decile of the distribution for a given examination subject. For example,

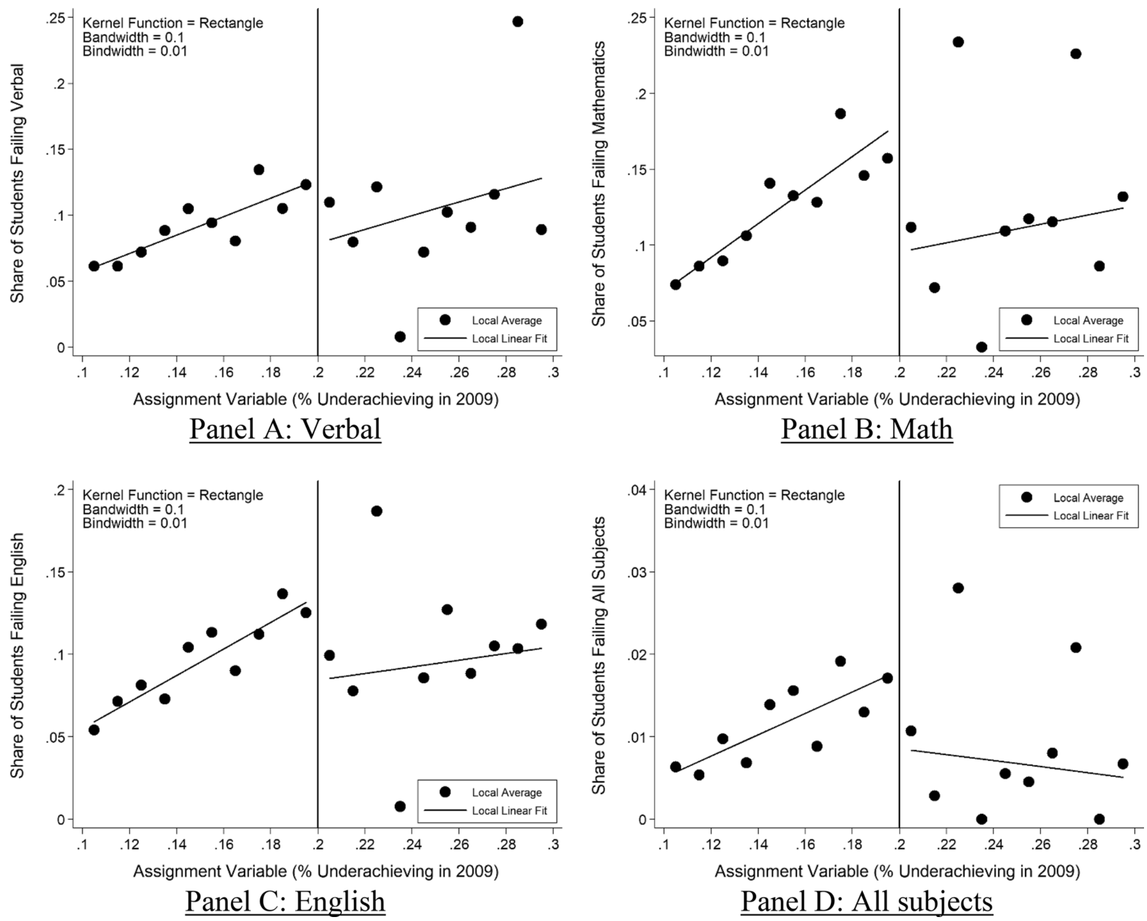

Fig. 7 Share of students failing by assignment variable (school-level data). **a** Verbal. **b** Math. **c** English. **d** All subjects

Table 2 Statistical tests of balance in baseline covariates

Variables	Bandwidth (0.05)		Bandwidth (0.10)	
	RD Estimate	Standard Errors	RD Estimate	Standard Errors
Share of female students	− 0.005	0.043	− 0.004	0.035
Share of students in poor families	0.051	0.034	0.030	0.025
Share of students receiving free lunch	0.036	0.046	0.008	0.037
Share of students living with parents	− 0.035	0.050	− 0.022	0.036
Average class size	− 2.886	2.224	− 0.828	1.905
Pupil-to-teacher ratio	− 1.341	2.129	− 1.006	1.567
Share of teachers with master's degree	− 0.039	0.129	− 0.001	0.107
Share of new teachers	0.035	0.092	0.017	0.065
Baseline average verbal score	0.390	0.661	0.258	0.489
Baseline average social studies score	0.258	0.768	0.035	0.576
Baseline average math score	0.608	0.977	0.732	0.731
Baseline average science score	0.004	0.577	0.137	0.445
Baseline average English score	0.100	1.010	0.307	0.736
Baseline average total score	0.272	0.732	0.294	0.538

"RD" denotes regression discontinuity. Bandwidth of 0.05 indicates that the regression discontinuity estimates are based on observations within the range of 0.15 and 0.25 of the running variable value (i.e., the share of underachieving students). Bandwidth of 0.10 indicates that the regression discontinuity estimates are based on observations within the range of 0.15 and 0.30 of the running variable value

Table 3 Regression discontinuity estimates on outcome variables

Variables	Bandwidth (0.05)		Bandwidth (0.10)	
	RD estimate	Standard errors	RD estimate	Standard errors
Panel A: student-level data				
% failed in verbal	− 0.025	0.034	− 0.021	0.026
% failed in math	− 0.060**	0.026	− 0.059***	0.021
% failed in English	− 0.024	0.032	− 0.029	0.024
% failed in all subjects	− 0.009	0.007	− 0.006	0.005
Panel B: school-level data				
% failed in verbal	− 0.020	0.029	− 0.040*	0.021
% failed in math	− 0.070***	0.027	− 0.084***	0.021
% failed in English	− 0.021	0.035	− 0.048**	0.022
% failed in all subjects	− 0.007	0.007	− 0.009*	0.005

"RD" denotes regression discontinuity estimates. "S.E." indicates standard errors clustered at the school level. Bandwidth of 0.05 indicates that the regression discontinuity estimates are based on observations within the range of 0.15 and 0.25 of the running variable value (i.e., the share of underachieving students). Bandwidth of 0.10 indicates that the regression discontinuity estimates are based on observations within the range of 0.10 and 0.30 of the running variable value

***, **, and *Indicate statistical significance at the 1%, 5%, and 10% level

Table 4 Quantile regression discontinuity estimates

Quantile	Regression discontinuity estimates by subject			
	Verbal	Math	English	Total
0.1	0.014***	0.000	0.009**	0.013***
	(0.004)	(0.005)	(0.004)	(0.004)
0.2	0.019***	0.044***	0.009*	0.021***
	(0.005)	(0.006)	(0.005)	(0.004)
0.3	0.031***	0.052***	0.022***	0.029***
	(0.007)	(0.008)	(0.006)	(0.006)
0.4	0.049***	0.055***	0.036***	0.049***
	(0.009)	(0.009)	(0.007)	(0.008)
0.5	0.060***	0.055***	0.024***	0.061***
	(0.010)	(0.011)	(0.009)	(0.009)
0.6	0.056***	0.082***	0.050***	0.069***
	(0.012)	(0.012)	(0.012)	(0.011)
0.7	0.048***	0.078***	0.051***	0.069***
	(0.014)	(0.013)	(0.015)	(0.013)
0.8	0.050***	0.098***	0.043***	0.069***
	(0.015)	(0.015)	(0.015)	(0.016)
0.9	0.049***	0.081***	0.050***	0.041***
	(0.018)	(0.019)	(0.017)	(0.018)
Number of observations	201,885	201,885	201,885	201,885

Robust standard errors are in parentheses

***, **, and *Indicate statistical significance at the 1%, 5%, and 10% level, respectively

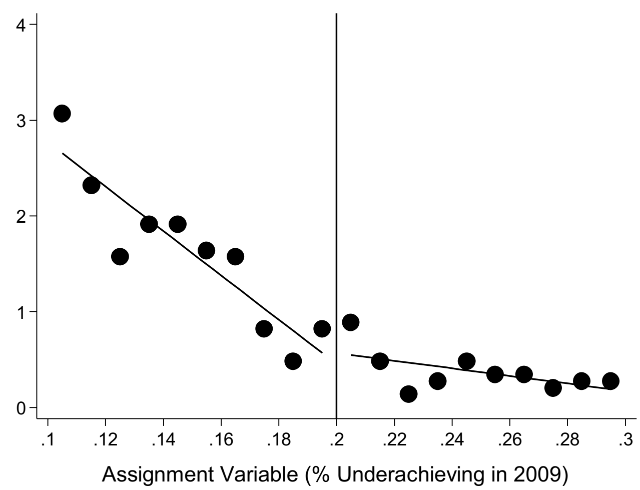

Fig. 8 Density of the assignment variable

referring to the first row of Table 4, students in the bottom decile of the verbal examination in treated schools scored 1.4% points higher than students in the bottom decile of the verbal examination in untreated schools. Likewise, students in the bottom decile of the English examination in treated schools scored 0.9% points higher than students in untreated schools. Overall, the evidence in Table 4 suggests that the largest treatment effects were witnessed among middle to high ability students. This is especially true for the mathematics exam, where students in the top two deciles improved by approximately 8 to 10% points. This is also true for verbal and English examinations, in

Table 5 Results of falsification tests

	Bandwidth = 0.05				Bandwidth = 0.10			
	Verbal	Math	English	Total	Verbal	Math	English	Total
Effect	− 0.006	− 0.032	− 0.003	− 0.008	− 0.025	− 0.055	− 0.028	− 0.009
Standard error	0.012	0.013	0.012	0.004	0.009	0.010	0.009	0.003
Rank	5/20	2/20	6/20	2/20	1/20	1/20	1/20	1/20

"1/20" indicates that the true treatment effect is larger than all of the placebo treatment effects

which students in the middle of the distribution accrued the largest gains from the funding.

Robustness

This study conducted several robustness checks designed to test the sensitivity of the findings to choices made during the research process as well as verify the soundness of the underlying research design. First, in Fig. 8, the visual results from the "McCrary test" for heaping on either side of the threshold of the running variable are presented (McCrary 2008). This test checks for a particular violation of the assumptions of a regression discontinuity design in which agents have knowledge of an administrative discontinuity and can change their behavior to end up on a given side of the discontinuity. In this context, such a violation might refer to a scenario in which students (or their parents) have advance knowledge of which schools will receive additional funding and accordingly can sort their children into the schools that were under-performing in the prior year. The discussions with Ministry of Education, Science, and Technology officials indicate that this would be an impossibility as the list of underachieving schools is not made public until long after parents would have to make the choice to send their child to a different school. Moreover, several facts imply that such manipulation is highly unlikely. First, answers of the NAEA exams were not disclosed until students received their final score. Second, the 20% eligibility cutoff was also not announced until students received final scores. As such, it was impossible for a school to game the share of underachieving students because the school did know the value of a threshold. Third, the range of scores for each grade category was not also determined a priori, so a school cannot manipulate students' grade. All in all, we argue that manipulation is impossible given the reasons mentioned above.

Because this assumption is central to achieving identification, however, it is nevertheless worth testing empirically. Had parents been able to manipulate the running variable, then an unusually large mass of students would "heap" just above the 20% threshold to take advantage of the fact that schools with slightly more than 20% of failing students would receive excess funding. The McCrary test checks for such a violation by testing whether the frequency of students is smooth across the policy threshold. Figure 8 provides exceptionally strong visual evidence that the frequency of students is not sensitive to the 20% mandate.[7]

Second, a falsification test was conducted to assess the degree to which estimated regression discontinuity effects around the 20% funding discontinuity are large relative to a series of placebo regression discontinuity estimates. In particular, for each subject, this study estimated the regression discontinuity treatment effect for twenty different administrative cutoffs, ranging from 1 to 20%. Because the only value of the running variable that resulted in an actual change in funding was the 20% threshold, regression discontinuity estimates using the 20% cutoff should be among the largest. The remaining regression discontinuity estimates, each of which uses a threshold of between 1 and 19% (in 0.01 increments), should produce no evidence of a nonzero treatment effect. This falsification test, which is merely the formalization of the "rule of thumb" that researchers should check for important discontinuities throughout the distribution of the running variable, is important for several reasons. First, the existence of meaningful treatment effects at different points along the distribution of the running variable undermines the validity of the regression discontinuity design insofar as the administrative discontinuity is assumed to provide a unique shock to the outcome variable. When it comes to test scores that rise at other values of the running variable at which no policy discontinuity is evident, it would be reasonable to conclude that the increase in test scores at the 20% mark may be spurious.

Third, a standard issue in interpreting the results of educational interventions designed to raise performance among the lowest scorers is the possibility of mean reversion. A particular concern is that students who scored poorly in a given year received a poor draw from their underlying distribution of ability and, as such, were likely to improve their scores in the absence of an intervention. The regression discontinuity design has been proven to be effective in obviating the mean reversion issue (Chay et al. 2005), and this

[7] We also implement the formal test suggested by McCrary (2008) which compares the height of the bars of a histogram local to the discontinuity. We fail to reject that the frequency is smooth across the threshold.

falsification test is also helpful in addressing this concern indirectly by checking to see whether the regression discontinuity estimates using similar administrative thresholds (e.g., 18% and 19%) also generate negative treatment effects that are as large as those generated using the 20% threshold. Because mean reversion should be as large an issue at 18 or 19% as at 20%, failure to detect negative treatment effects using these thresholds suggests a lack of important mean reversion effects.

Table 5 presents the results of this exercise. It begins by replicating the regression discontinuity coefficients and standard errors for two different choices of bandwidths—0.05 and 0.10—and all integer choices of thresholds of the running variable from 1 to 20%. Next, the rank of the size of the true treatment effect relative to the size of the nineteen placebo treatment effects is presented. A rank of "1/20" indicates that the true treatment effect is larger than all the placebo treatment effects. Table 5 presents remarkably strong evidence that the regression discontinuity estimates reported in Table 3 are not spurious. Although the estimates for verbal and English using a bandwidth of 0.05 are only the fifth and sixth largest in the distribution of placebo treatment effects, these estimates were very small and not significant to begin with. On the other hand, using the 0.05 bandwidth, the true treatment effects for mathematics as well as for all the three exams are the second largest among the distribution of counterfactual estimates. Moreover, using a bandwidth of 0.10, which produces consistent evidence of significant treatment effects in each case, the true treatment effect is found to be larger than all of the placebo treatment effects. In a number of cases, the true effect is considerably larger. These results can be interpreted as evidence in favor of a causal effect of funding at the 20% threshold.

Finally, if students who did poorly at a school in 2009 move to another school that is not underachieving, then this would bias the estimated result (i.e., noncompliance bias). In addition, if students who failed in the 2009 exams dropped out of treated schools relatively more compared with those in untreated schools, then this would also result in biased effect estimates (attrition bias). To examine whether such biases are pervasive in the current research setting, the share of students who moved out from a school in 2010 and that of students who dropped out from a school in 2010 were examined. On average, the share of students who moved out from a school in 2010 was found to be very small—less than 3%. In addition, the share of students who dropped out of school was found to be very small—less than 4%. Moreover, the shares of these two variables are very smooth across the running variable, and no statistically significant discontinuity at the 20% cutoff (− 0.008 for the first share and − 0.010 for the second share) was observed. Therefore, the bias that may arise from noncompliance and attrition is less likely in the current research setting.

Conclusion

This study leverages an administrative discontinuity in the disbursement of school funding in South Korea to study the effect of funding on student performance, particularly the tendency of students to fail key national assessment examinations. With respect to existing literature, this study is most similar to the studies of Jacob and Lefgren (2004), which studied the effect of increases in school funding in Chicago; Chay et al. (2005), which studied the effect of increased school funding in Chile; and van der Klaauw (2008) which considers the effect of school funding in New York City. Each of these studies utilized a regression discontinuity design to exploit a discontinuity in school funding rules, with the first two studies finding evidence of positive effects. Each of these studies considers the effect of school funding on the academic achievements of primary school aged children. This paper takes a different route by choosing to use a regression discontinuity design to examine the effect of school funding rules on the educational achievement of older (high-school aged) students. Moreover, it extends the regression discontinuity framework, leveraging a recent methodological advance pioneered by Frandsen et al. (2012) to report quantile regression discontinuity estimates that identify the regression discontinuity effects for students throughout the distribution of student performance.

Having established that there is evidence in favor of a discontinuous decline in the rate of exam failure at the 20% threshold, the paper attempts to interpret these estimates. In particular, given that qualifying schools received approximately 300,000 won per student, it is natural to consider whether the program was cost effective. In the sample used in this study, the average school is comprised of 334 students and, on average, 7.3% of these students failed the math NAEA in 2009. The regression discontinuity estimates indicate potentially important declines in exam failure rates, particularly for math. For math, the exam failure rates declined by approximately 6% points, off a base of 7.3%. Thus, an average school that had approximately 26 failing students in 2009 would have had just five failing students in math in 2010. In addition, the point estimates suggest meaningful declines in the failure rates for the verbal and English exams as well, although the results are not significant at conventional levels.

With regard to cost-effectiveness analysis, an average school would have received approximately 56,000,000 won as a result of the intervention. The 56,000,000 won brought, on average, 21 students above the exam failure threshold in math, a cost of approximately 2,700,000 won per an additional passing student. For several reasons, this is likely an upper bound on the true number. In particular, the quantile regression discontinuity estimates presented in Table 4 indicate that the additional funding was not well targeted. Notably, in all the three subjects, but most notably in math, the largest gains accrued to students at the top of the distribution. Hence, the funding generated gains throughout the spectrum of students.

This study is not without limitations. First, because this study uses a regression discontinuity design, the generalizability of the estimated results may not be strong. Regression discontinuity estimates are derived from the sample of schools near the 20% cutoff, and as such, the findings of this study are local to the schools that are struggling. That is, the findings are internally valid; they may not be externally valid. Therefore, this study refrains from drawing a strong conclusion that school funding would be helpful for students in other achievement distributions such as high achieving students. Second, this study's effect estimates are based on a short time frame. Schools received funding in March 2010 and the test was conducted in July 2010, so the funding was spent over a period of about 4.5 months. Following discussion with officials at the Ministry of Education, this study concludes that although the time frame was short, the schools that received the funding were able to promote student achievements within a very short time frame because the schools focused their efforts on failed students. That is, the short term effects was possible due to the school's selective focus on failed students. There is a concern, however, that such selective focus may induce a so-called "teaching to the test" issue raised by Lazear (2006). While many argue that students would still benefit even if they are learning material that is specific to exam contents, it is highly desirable to examine the long term effects of school funding. Unfortunately, this study was not able to obtain the administrative data for 2011 and beyond and, therefore, it could not evaluate the effects over a long period of time, which presents another limitation.

References

Bogart, W. T., & Cromwell, B. A. (1997). How much more is a good school district worth. *National Tax Journal, 50*(2), 215–232.

Card, D., & Krueger, A. B. (1992). Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of Political Economy, 100*(1), 1–40.

Card, D., & Payne, A. A. (2002). School finance reform, the distribution of school spending, and the distribution of student test scores. *Journal of Public Economics, 83,* 49–82.

Chay, K. Y., McEwan, P. J., & Urquiola, M. (2005). The central role of noise in evaluating interventions that use test scores to rank schools. *American Economic Review, 95*(4), 1237–1258.

Clapp, J. M., Nanda, A., & Ross, S. L. (2008). Which School attributes matter? The influence of school district performance and demographic composition on property values. *Journal of Urban Economics, 63,* 451–466.

Downes, T. A., & Zabel, J. (2002). The impact of school characteristics on house prices: Chicago 1987–1991. *Journal of Urban Economics, 52,* 1–25.

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications*. New York: Chapman and Hall.

Frandsen, B. R., Frölich, M., & Melly, B. (2012). Quantile treatment effects in the regression discontinuity design. *Journal of Econometrics, 168,* 382–395.

Goodman, A. C., & Thibodeau, T. G. (1998). Dwelling age heteroskedasticity in repeat sales house price equations. *Real Estate Economics, 26*(1), 151–171.

Guryan, J. (2001). Does money matter? Regression-discontinuity estimates from education finance reform in Massachusetts. NBER working paper No. w8269, National Bureau of Economic Research.

Häkkinen, I., Kirjavainen, T., & Uusitalo, R. (2003). School resources and student achievement revisited: New evidence from panel data. *Economics of Education Review, 22,* 329–335.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature, 24*(3), 1141–1177.

Hanushek, E. A. (1994). A jaundiced view of adequacy in school finance reform. *Educational Policy, 8*(4), 460–469.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis, 19*(2), 141–164.

Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher, 23*(3), 5–14.

Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics, 86*(1), 226–244.

Lazear, E. P. (2006). Speeding, terrorism, and teaching to the test. *Quarterly Journal of Economics, 121*(3), 1029–1061.

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

Papke, L. E. (2005). The effects of spending on test pass rates: Evidence from Michigan. *Journal of Public Economics, 89,* 821–839.

van der Klaauw, W. (2008). Breaking the link between poverty and low student achievement: An evaluation of title 1. *Journal of Econometrics, 145,* 21–42.

Verstegen, D. A., & King, R. A. (1998). The relationship between school spending and student achievement: A review and analysis of 35 years of production function research. *Journal of Education Finance, 24*(2), 243–262.