

IZA DP No. 9458

Carrot *and* Stick?
Impact of a Low-Stakes School Accountability
Program on Student Achievement

Seokjin Woo
Soohyung Lee
Kyunghee Kim

October 2015

Carrot *and* Stick? Impact of a Low-Stakes School Accountability Program on Student Achievement

Seokjin Woo

Myongji University

Soohyung Lee

University of Maryland and IZA

Kyunghee Kim

Korea Institute for Curriculum and Evaluation

Discussion Paper No. 9458

October 2015

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: iza@iza.org

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

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

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

ABSTRACT

Carrot *and* Stick? Impact of a Low-Stakes School Accountability Program on Student Achievement¹

A key concern in the design of education policies relates to the structure of incentives in accountability systems. This paper examines a school accountability program that provides financial support to low-performing schools but has no direct punishment scheme for recipients who do not exhibit improvement. Although the program does not include high-stakes consequences, our estimates indicate that the program reduced the share of underperforming students by 18 percent. This paper's results suggest that to improve student achievement, a school accountability program does not need to set high-stakes consequences that potentially induce unwanted strategic behaviors on the part of school workers.

JEL Classification: C5, I2

Keywords: school accountability, student achievement, school performance, fuzzy regression discontinuity design

Corresponding author:

Soohyung Lee
Department of Economics
University of Maryland
3114 Tydings Hall
College Park, MD 20742
USA
E-mail: LeeS@econ.umd.edu

¹ We thank Lesley Turner and Ben Malin for detailed suggestions. All errors are on our own. Woo acknowledges the financial support by the 2013 Research Fund of Myongji University. This paper is forthcoming in *Economics Letters*. Online Appendix available at: www.soohyunglee.com/research

I. Introduction

One of the main concerns in education policy-making is how to properly structure incentives in school accountability programs (e.g., Figlio and Loeb, 2011, Neal, 2011). In the U.S., such accountability programs are implemented by the federal government (e.g., the No Child Left Behind Act of 2001) as well as by states and localities (e.g., the 2007 New York City accountability system that established school progress reports).² These accountability programs set performance goals that schools are required to meet and impose consequences on schools deemed to be low performing. Rarely do such programs provide low-performing schools with additional resources to meet these goals. Furthermore, whether or not performance goals are met can significantly affect school employees in terms of their monetary compensation, employment, and promotions.

While this high-stakes “carrot and stick” approach provides a significant incentive for school workers to meet performance goals, it may also lead to counterproductive actions that could jeopardize the educational system (e.g., manipulation of test scores and shifting of resources to test subjects from non-test subjects).³ Therefore, it is important to examine whether an accountability program that imposes weaker financial incentives - and thus imposes a lower risk of strategic behaviors - can improve school performance as much as the current high-stakes programs. The results of this paper suggest that it may be possible to improve student achievement without relying on heavy financial incentives.

We study a national accountability program called the “School For Improvement” (SFI herein) program that was introduced in South Korea in 2010. Similar to accountability programs in the U.S., the SFI program classifies schools according to

² For example, Hanushek and Raymond (2005) and Rockoff and Turner (2010) evaluate the impact of each program on student achievement and other outcomes.

³ Rouse et al. (2013) provide a recent survey of this literature.

student achievement and aims to help failing schools. Unlike the U.S. programs, the SFI program provides all failing schools with substantial financial resources that they can use to increase human resources by hiring additional teachers' aides, acquiring consulting services for teaching skills, and offering additional classes after school. The SFI program does not provide financial incentives to school workers based on post-period student performance; rather, it uses indirect pressure from peers and parents by publicly announcing schools' performance.

The SFI program identifies a school as failing when at least 20 percent of its students are classified as "underperforming" based on the baseline assessment in 2009. Due to this selection rule, we use a regression discontinuity (RD) design to estimate the impact of the SFI program on student achievement. We use a nationwide administrative dataset containing reliable measures of student performance in 2009 and 2010. We find that the SFI program leads to a significant 3.7 percentage point (18 percent) reduction in the percentage of students classified as underperforming and significantly improves student achievement in all test subjects (Reading, Math, English, Science, and Social Studies). These results are robust to various specifications and alternative empirical models.

Our results suggests that a well-defined low-stakes accountability program that provides failing schools with additional resources may be able to improve student achievement, just like a high-stakes accountability program, while imposing a lower risk of inducing unintended behavior on the part of school workers. Therefore, this paper calls for further studies on designing the proper level of incentives for school improvement.

II. Institutional Background and Data

Since 2009, the South Korean government has conducted a national test every year to evaluate all students enrolling in grade 6 (the final grade of primary school), grade 9 (the final grade of middle school) or grade 11 (the second year of high school).⁴ In March 2010, the government launched an accountability program called School For Improvement (SFI) to reduce the share of “underperforming” students – those considered to lack a basic understanding of the curriculum. We examine the impact of the SFI in the middle school setting because, unlike the high school setting, students cannot choose the middle school they attend: they are randomly assigned to one within their school districts, which provides a clean identification setting (see details in Lee et al. 2014). Compared to the primary school setting, the middle school curriculum is well aligned with the contents covered by the national college entrance exams, which is important to predict a student’s success in college and other labor market outcomes.

The SFI provides each participating school approximately 70,000 US dollars per year, and the school can use it to hire additional teachers’ aides, offer additional after-school classes, or secure consulting services.⁵ Regarding the consulting services, the Korean Ministry of Education dispatches to a school a team of professionals who identify students at risk, evaluate lectures, observe the dynamics among school workers, and then provide solutions to address the issues that may account for the school’s low performance (see details in Kim et al, 2012). The SFI program does not impose explicit consequences when a school does not show any improvement. However, the program uses rather gentle

⁴ The test, called the National Assessment of Educational Achievement (NAEA), is administrated by the Korean Ministry of Education. For middle school students, the NAEA consists of five test subjects – Reading, Math, English, Science, and Social Studies. See details in Lee et al (2014).

⁵ If this financial subsidy were entirely used to hire teachers’ aides, the average school could have hired 53 workers (on a basis of 20 hours per week), which is approximately the same as the average number of full-time teachers in a given school.

incentives by making the information on each school's performance publicly available on the government's website. See details in Online Appendix Section B.

Our data includes school-by-year observations in 2009 and 2010 of all middle schools in South Korea. We observe the share of 9th grade students classified as underperforming, which is measured by the average of the share of students who are classified as lacking a basic understanding of a given test subject (Reading, Math, English, Science, and Social Studies). We supplement the data with a separate school-by-year dataset called the Korea Education & Research Information Service (KERIS) from the Korean Ministry of Education. The KERIS includes school-level characteristics, such as student-teacher ratio, whether a school was established by a private entity, and teacher characteristics. See the online appendix for a detailed description of the data.

All middle schools that had at least 20 percent underperforming students on the 2009 national test were required to participate in the SFI program, while the rest could join the program voluntarily. Consistent with this selection rule, all of the schools whose initial share of underperforming students was greater than the threshold participated in the SFI, while 6.7 percent of the non-targeted schools participated in the program. In Figure 1 (Panel A), we plot the share of middle schools that participated in the SFI depending on their share of underperforming students on the 2009 national test; bins are 1 percentage point wide. This figure shows a distinct jump in the probability of SFI participation at the threshold value of 20 percent (significant at the 1 percent level), which allows us to employ the fuzzy regression discontinuity approach.

III. Econometric Framework and Identification Strategy

We employ the fuzzy regression discontinuity design in the spirit of previous work on the causal impact of school accountability policies on student achievement (e.g., Chiang,

2009, Rochoff and Turner, 2010, Dee, 2012, Rouse et al., 2013). We use the discontinuous relationship between a school's treatment status and its average share of underperforming students in the baseline year to identify the causal impact of SFI eligibility on student achievement. Specifically, we estimate local OLS regression models by examining schools whose baseline performance in 2009 belongs to local bandwidths around the threshold, where we set the optimal bandwidth based on the method suggested by Imbens and Kalyanaraman (2012).

Our identifying assumption is that the distribution of unobservables with respect to the running variable is continuous through the threshold, which implies local random assignment. Following the standard practice of examining the plausibility of the identification strategy (e.g., Lee and Lemieux, 2010), we find that the identification assumption is likely to hold in our setting. First, schools are unlikely to manipulate their performance because the national tests were created and are supervised by the Korean Ministry of Education, and over 98 percent of students enrolled in middle schools took the national tests. Second, we find a discontinuity at the threshold in terms of the outcome variable in 2010 but not in terms of predetermined school characteristics or of the density of observations. Furthermore, our results are robust to the inclusion or exclusion of covariates. Finally, alternative choices of bandwidth generate results similar to our baseline estimates. See the Online Appendix Section D for additional information and formal test results of our identifying assumption.

V. Results

We first present graphical evidence of the positive effect of the SFI on school performance. In Figure 1 (Panel B), we plot the average share of underperforming students in 2010 – the year after the program was implemented – for each bin of baseline

performance. For easy comparison, we include two solid lines in the graph that are the estimated quadratic polynomial for both sides of the threshold.⁶ The figure shows a discontinuous decrease at the threshold, suggesting that schools required to participate in the SFI program (located right side of the threshold) have smaller shares of underperforming students in 2010, relative to schools that are less likely to receive the SFI (located on the left side of the threshold). These findings are robust in terms of selection of bin sizes (see the Online Appendix Section E).

Next, we estimate local OLS regression models. In Table 1, we present the impact of SFI eligibility on the average share of underperforming students in a school (Panel A). Our baseline results, reported in column 1, show that the SFI reduces the share of underperforming students by a statistically significant 3.7 percentage points (an 18.5 percent reduction for the schools at the cutoff). This result remains fairly stable when we include school characteristics and school district fixed effects (column 2) or use alternative bandwidths (columns 3 and 4).⁷ Furthermore, we find positive effects of the SFI on all test subjects reported in Panels B to F. The SFI program reduces the share of students who lack a basic understanding of each test subject by 3 to 6 percentage points. See the Online Appendix Section E. for additional robustness checks, including 2SLS estimates.

Our results show that the SFI program delivers outcomes in line with the policy goal – reducing the share of underperforming students in a school. We additionally examine the average test scores and find that the SFI program has a positive impact on

⁶ The results are comparable when we use a local linear model. Following Gelman and Imbens (2014), we avoid high-order polynomials.

⁷ School characteristics include share of female students, share of students in welfare programs, single-sex school or not, established by a private entity or not, school size, pupils per teacher, share of teachers with master's degree or PhD, and school district fixed effects.

average test scores in all subjects although the magnitude varies by subject. This result suggests that the SFI program may not have much adverse impact on non-targeted students, different from the US setting that Neal and Schanzenbach (2010) examine. Finally, we find that the SFI program is more effective in schools that received a higher financial subsidy per student, suggesting that increased financial subsidy, rather than public pressure, may account for the improvement in school performance. See details in Online Appendix Sections E.2 and E.3.

IV. Conclusion

This paper examines the impact of a school accountability program that provides financial support to failing schools but has no direct punishment scheme for schools that do not exhibit improvement. Although the program does not provide explicit economic incentives for school personnel, we find that it had sizable positive impacts on students' achievement the year after it was launched. Our results imply that it may be possible to “nudge” schools, without employing any aggressive incentive scheme to improve school accountability. Thus, policymakers may want to experiment with such a program, especially if there is severe opposition, from school workers and other relevant parties, to high-stakes accountability schemes.

This paper also contributes to the large literature examining the impact of school resources on student outcomes. While the results on the estimated impacts are mixed,⁸ recent studies point out that school resources may increase student performance when they are devoted to human resource improvement instead of to school materials (see Caneiro et al., 2015). This finding suggests that the significant impact of Korea's SFI

⁸ See the literature review in De Witte et al. (2014) for developed countries, and Caneiro et al. (2015) for developing countries.

program may be accounted for by the fact that the Korean schools are allowed to use the financial support only for human resources improvement.

REFERENCES

- Carneiro, Pedro, Oswald Koussihouede, Nathalie Lahire, Costas Meghir, Corina Mommarts. 2015. "Decentralizing Education Resources: School Grants in Senegal," IZA Working Paper, No. 8960.
- Chiang, Hanley. 2009. "How Accountability Pressure on Failing Schools Affects Student Achievement." *Journal of Public Economics*, 93(9-10),1045-1057.
- Dee, Thomas. 2012. "School Turnarounds: Evidence From the 2009 Stimulus," NBER working paper No. 17990.
- De Witte, Kristof, Benny Geys, and Catharina Solondz. 2014. "Public Expenditures, Educational Outcomes and Grade Inflation: Theory and Evidence from a Policy Intervention in the Netherlands," *Economics of Education Review*, 40, 152-166.
- Figlio, David, and Susanna Loeb. 2011. "School Accountability," *Handbook of the Economics of Education*, 3, 384-421.
- Gelman, Andrew, and Guido Imbens. 2014. "Why High-Order Polynomials Should Not Be Used in Regression Continuity Designs," NBER Working Paper. 20405
- Hanushek, Eric A. and Margaret E. Raymond. 2005. "Does School Accountability Lead to Improved Student Performance?," *Journal of Policy Analysis and Management*, 24(2), 297-327.
- Imbens, Guido, and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79(3), 933-959.
- Kim, Kyunghye, Heekyung Kim, Wonkyun Noh, Kyungah Sang, Jinah Shin, Hyekyung Chung, Seokjin Woo, Sungchang Ryu, Jungah Hahn, Sherri Lauver, Carla McClure, Miranda Cairns, and Brian Fu. 2012. "School Accountability Programs: South Korea and United States," Korea Institute for Curriculum and Evaluation, Policy Brief, in Korean.
- Lee, David S., and Thomas Lemiue. 2010. "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48, 281-355.

- Lee, Soohyung, Lesley J. Turner, Seokjin Woo, and Kyunghee Kim. 2014. "All or Nothing? The Impact of School and Classroom Gender Composition on Effort and Academic Achievement," NBER Working Paper No.
- Neal, Derek. 2011. "The Design of Performance Pay in Education," *Handbook of the Economics of Education*, 4, 495-550.
- Neal, Derek, and Diane Whitmore Schanzenbach. 2010. "Left Behind by Design: Proficiency Counts and Test-Based Accountability," *Review of Economics and Statistics*, 92(2), 263-283.
- Rockoff, Jonah, and Lesley J. Turner. 2010. "Short-Run Impacts of Accountability on School Quality," *American Economic Journal: Economic Policy*, 2, 119-147.
- Rouse, Cecilia E., Jane Hannaway, Dan Goldhaber, and David Figlio. 2013. "Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure," *American Economic Journal: Economic Policy*, 5(2): 251-81.

Table 1: Summary Statistics

	<u>Initial Share <20%</u>		<u>Initial Share ≥20%</u>	
	2,937 schools		104 schools	
	2009	2010	2009	2010
	(1)	(2)	(3)	(4)
Panel A. Share of under-performing students	0.069	0.054	0.247	0.091
Panel B. Share of students lacking a basic understanding of				
- Reading	0.044	0.031	0.182	0.058
- Math	0.112	0.061	0.337	0.102
- English	0.052	0.042	0.197	0.069
- Science	0.067	0.075	0.254	0.127
- Social Studies	0.068	0.061	0.267	0.100

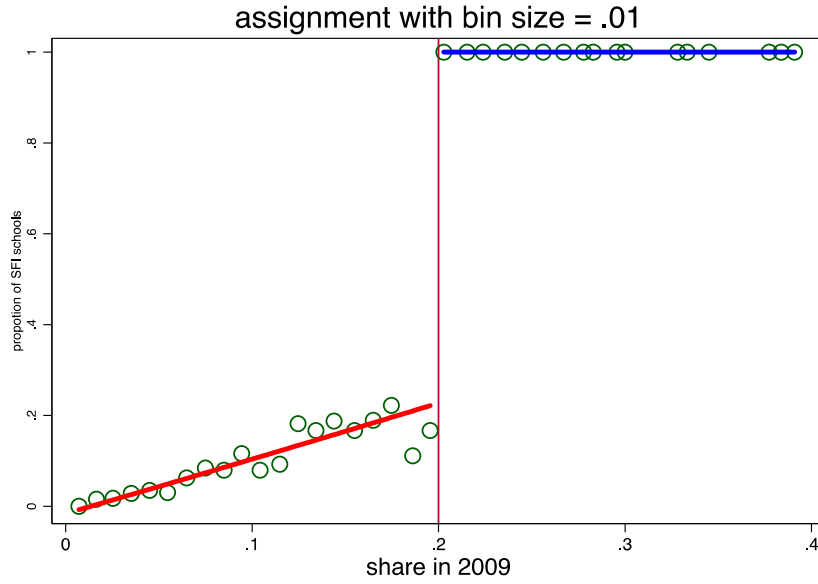
Table 2: Impact of SFI Eligibility on the 2010 School Performance

	(1)	(2)	(3)	(4)
Panel A. Share of underperforming students				
SFI program	-0.037*** (0.009)	-0.025*** (0.010)	-0.037*** (0.013)	-0.046*** (0.012)
Obs. In bandwidth	2,483	1,804	1,042	3,039
Panel B. Share of students lacking a basic understanding of				
<i>- Reading</i>				
SFI program	-0.029*** (0.011)	-0.018** (0.009)	-0.028** (0.012)	-0.031*** (0.010)
Obs. In bandwidth	1,494	1,204	642	2,731
<i>- Math</i>				
SFI program	-0.043*** (0.013)	-0.030** (0.013)	-0.040*** (0.015)	-0.044*** (0.012)
Obs. In bandwidth	2,946	2,704	1,509	3,040
<i>- English</i>				
SFI program	-0.030** (0.013)	-0.020 (0.013)	-0.030** (0.015)	-0.028*** (0.010)
Obs. In bandwidth	959	965	444	1,830
<i>- Science</i>				
SFI program	-0.057** (0.024)	-0.032** (0.015)	-0.056* (0.029)	-0.059** (0.023)
Obs. In bandwidth	653	1,535	331	1,234
<i>- Social Studies</i>				
SFI program	-0.043** (0.019)	-0.025* (0.015)	-0.047** (0.021)	-0.042** (0.018)
Obs. In bandwidth	532	615	287	936
Other controls	No	Yes	No	No
Bandwidth	Optimal	Optimal	70% of optimal	130% of optimal

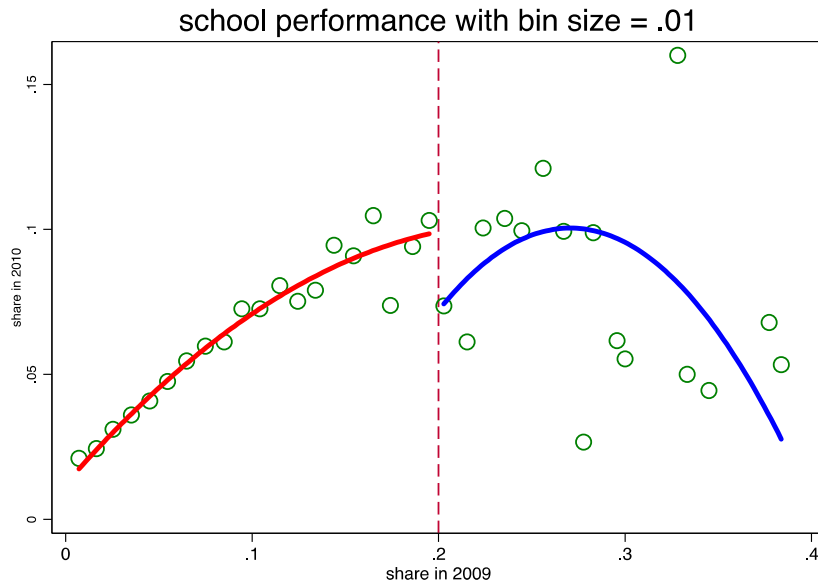
Notes: The unit of observations is school. SFI program is 1 if a school had at least 20 percent of underperforming students on the 2009 national test. Optimal bandwidth is chosen based on the procedure developed by Imbens and Kalyanaraman (2012). “Other controls” include share of female students, share of students in welfare programs, whether the school was established by a private entity, whether the school is single-sex, school size, pupils per teacher, teacher’s educational attainment, and school district fixed effects. * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors are reported in parentheses.

Figure 1 SFI Program and Effects

Panel A. Initial Share of Underperforming Students and Likelihood of Being Treated



Panel B. Shares of Underperforming Students Pre- and Post-Program



Note: The x-axis is the share of underperforming students in 2009. The y-axis in Panel A is the fraction of schools in the SFI program, and that in Panel B is the share of underperforming students in 2010. Quadratic polynomials are used to fit the data.