Is the Standardization of Compulsory Education School Helpful to Improve Students' Performance? An Empirical Analysis Based on Monitoring Data in Province A

Yanli Liu,[1] Jijun Yao,[1] Shike Zhou[2]

1. Nanjing Normal University, Nanjing 210024, China
2. Jiangsu Teaching and Research Office of Primary and Secondary Schools, Nanjing 210024, China

Abstract. *The standardization of compulsory education schools is an essential measure to promote education balance and improve education quality. However, when the implementation effect of this policy was evaluated, the improvement of input factors such as facilities and equipments, teaching faculty and funding were only focused on, but with less attention was paid to the impact of the standardization on students' academic performance. This kind of imbalance makes it difficult to fully grasp the real effects of standardization and its potential problems, which may lead to deviations in the direction of standardization construction. Therefore, this study used the monitoring data in province A that includes the standardized construction of compulsory education schools and academic quality of students to analyze the correlation between standardization construction and student performance through Ordinary Least Square (OLS). Further, a Regression Discontinuity Design (RDD) was used to analyze whether standardization construction is the cause of differences in student performance. The results showed that there was no significant correlation between standardization construction and elementary school students' performance, and negative correlation with middle school students' performance; no substential causal relationship exists between standardization construction and student performance. This is majorly due to the emphasis on the conditions of school-running was set too much, and then the student performance was neglected. Therefore, future evaluation should be gradually changed while deepening the standardization construction with more attention paid to the connotative development and efforts made to the improvement of the education quality.*

About the Author: *Yanli Liu, School of Education Science, Nanjing Normal University, Nanjing, China, E-mail: 1609604712@qq.com.*

Shike Zhou, School of Education Science, Nanjing Normal University, Nanjing, China, E-mail: jyszsk@163.com;

Correspondence to*: Jijun Yao, School of Education Science, Nanjing Normal University, Nanjing, China, E-mail: yaojijun_njnu@163.com.*

Introduction

I N 2010, China issued the "Outline of the National Middle- and Long-term Education Reform and Development Plan (2010-2020)", which proposed that "advancing the standardization of compulsory education schools and balancing the allocation of resources such as teachers, equipment, books, school buildings, etc." "Establish the national basic quality standards and monitoring system for compulsory education" (July 29, 2010) strive to continuously improve the standardization of compulsory education, promote education equity, improve education quality, and promote balanced development of compulsory education. In this context, local governments have successively issued relevant policy documents to vigorously promote the standardization of compulsory education schools. Take province A as an example, in 2015, the government of the province A released the "Standards for Provincial A Compulsory Education Schools (Trial)" (Provincial Government of A, 2015). Subsequently, the Education Department of Province A (2016) also developed the "Monitoring Indicators for the Standardization of Provincial A Schools for Compulsory Education (Trial)", through which the schools do self-assessment and self-test first, and then county-level and city-level review, and finally provincial review that formed the monitoring data of school standardization construction of province A.

The formulation and implementation of these measures have being greatly promoted the standardization construction of compulsory education schools ("standardization construction" for short). "The Compulsory Education School Standardization Construction Monitoring Report of Province A in 2017" showed that the standardization construction compliance rate has gotten significant improvment and the effect was remarkable. Up to 2017, "71.26% of schools in province A have basically reached the provincial standard", and the rate of compliance has increased by 15.3% compared with 2016; The average compliance rate of compulsory education schools in 13 districts and cities is 73.34% on average, 16% higher than that in 2016; in addition, the achievement of indicators had also been at a high level. In 2017, "the first-level indicators reached over 60%, three first-level indicators were close to 100%", "60% of the second-level indicators reached 80%", and "46 third-level monitoring points was above 80%" (Education Evaluation Institute of Province A, 2018). From the data above, we can see that the standardization construction of Province A has reached a high level.

Based on the experience of countries around the world, the construction of education standardization is an essential step to promote a balanced development of education and improve education quality. These achievements made by Province A in standardization construction were undoubtedly of great significance. However, in terms of specific indicators and construction content, most provinces, including A, currently focus on condition improvement in the process of standardization. Although this can improve school security, promote education equity, and provide a better learning environment, but as far as the ultimate goal of education policy implementation is concerned, it is obvious that it should be the development of students, rather than "input factors". When taking only the input such as facilities, equipment, teachers, and funds

as the first evaluation criteria, it will make us depend too much on the conditions for running schools, but not on students' development. However, it is hard to obtain the answer directly from the results of such evaluations.

Since 1980, the United States has promoted a gigantic "education standardization movement". During that process, the transformation of educational evaluation criteria had become the major reform point. Starting with Rose V. Council for A Better Education in Kentucky in 1989, a series of education lawsuits were passed and the standards for American education investment and accountability had gradually realized a change from "fair" to "adequate". The so-called "adequate education" was definitely different from the "fair" one, of which did not simply stop at judging whether the educational opportunities and resources were being equally allocated (This is exactly the starting and end points of most Chinese domestic education policy evaluations), but from the perspective of the extent to which the education provided by the education system for the educated has realized the development of the educated to assess the effectiveness of various educational measures. This is a typical result-oriented educational evaluation and resource allocation. However, the "adequate education" largely refers to students' academic performance that judges whether an education policy is effective, through looking at the improvement of students' performance (Yao, 2018; Li & Tao, 2015)

The reason for result-oriented education policy evaluation as an essential one is majorly due to "it is not enough to only focus on the fairness of resource allocation, it is necessary to combine the fair resource allocation to the quality of education, so as to achieve a balanced high-quality development" (Li & Zeng, 2002). Regarding China's education practice, studies on the result-oriented evaluation is yet adequate. Taking compulsory education school standardization construction as an example, the existing studies mainly focus on the discussion of standards, necessity and feasibility, construction principles, strategies, methodologies, and issues. But it remains unclear whether standardization promotes student performance. Today, China has basically realized "letting each child learn", but the urgent problem is how to realize "making each child learn better". So, the lack of evaluation research and practical exploration of output dimension is obviously not conducive to the completion of this core task.

Ideally, for output-oriented education policy evaluation, a comprehensive assessment of the impact of a policy on student performance, abilities, and attitudes should be made. However, most of these indicators are difficult to measure, so most of the output-oriented education policy evaluations use student performance as a proxy for output. In view of this, we, in the following content, will try to analyze the policy effects of standardization construction based on student performance. Through this type of analysis, we tried to answer whether standardization construction has improved the learning output situation represented by academic performance? Furthermore, we analyzed the differences in standardization construction and whether they met the standards, and explore whether it was the cause of academic differences between schools. Therefore, we first used the Ordinary Least Square (OLS) to study the correlation between the degrees of standardization construction, whether or not the standardization construction

meets the standards and the academic performance. Secondly, Regression Discontinuity Design (RDD) was used to infer causality between both these variables. Finally, the research conclusions were analyzed, discussed and related suggestions were put forward.

Research Design

Research Framework

Based on the available data, we took students in grades 3 and 8 in the compulsory education stage of province A as the research objects, used the academic performance of the students as the dependent variable, and standardize construction work as the independent variable. So we examined the impact of standardization construction on student development from the output side.

Based on this, we first studied the correlation between standardization construction and student performance. Specially, we used the standardized construction evaluation data of province A to describe the implementation effect of the policy by using the two variables: one is the standardized construction scores of each school (continuous variables) and whether the standardized construction had reached the standard (categorical variables). On this basis, OLS was used to analyze the relationship between the degree of standardization construction, the achievement of the standard, and student performance.

We need to point out that general regression analysis can help us figure out the relationship between independent and dependent variables, but it is difficult to do the causal inference. As mentioned earlier, we wanted to understand not only the relationship between standardization construction and student performance, but also whether this policy was responsible for differences in performance between schools. To this end, we used RDD to analyze whether there was a causal relationship between standardization construction and student performance.

RDD was first proposed by Thistlethwaite and Campbell (1960) in 1960 and it is a method for estimating processing effects in non-experimental environments. Using this method, they analyzed the impact of scholarships on students' future academic success. Students' scholarships depend on observable test scores. However, between the students who just met the award criteria and the students who almost reached the award criteria, in addition to whether they received scholarships (differentiated by the score standard), other characteristics are in line with the principle of randomness, so they are comparable. If the student's test score was greater than or equal to the cutoff value, then a scholarship was awarded to enter the experimental group; otherwise, the student was assigned into the control group. If there was a difference in the academic performance of the two groups, we could conclude that this difference is caused by the scholarship policy. Initially, this evaluation strategy did not receive much attention. However, since this method has advantages that other methods do not have in terms of causal inference, after the 1990s, more and more studies began to use this method to estimate policy effects in various backgrounds, making it gradually become an important method of causal inference in social science research (Lee & Lemieux, 2010). In the field of education,

some representative research results have also appeared. For example, Lee (2018) took Tennessee as an example and used Fuzzy Regression Discontinuity to study whether a grade-based bursary policy would affect students' bachelor's degree within six years. Li and Konstantopoulos (2016) used 2011 TIMSS data from 14 European countries, and used Instrumental Variable method and Regression Discontinuity Design to study the impact of class size on the mathematics performance of fourth grade students. Wei (2012) studied whether the "No One Child Left behind Act" improved the performance of students with disabilities based on RDD. But so far, the research on the use of RDD by Chinese scholars is relatively rare (Chen et al., 2014; Chu & Meng, 2017; Liu et al., 2016; Wang & Sun, 2015). In particular, within our scope, no researcher has used this method to carry out causal relationships between the standardization of compulsory education schools and student performance.

According to the "Monitoring Report on the Standardization Construction of Compulsory Education Schools in Province A in 2017", the schools with more than 80% of the monitoring points that have reached the standard are defined as "basically meeting the standardization construction standards" ("achieve the standard" for short) and the schools with less than 80% are defined as "not meeting the standardization construction standard"("below the standard" for short). This formed a quasi-experimental data structure at the policy level and provided the conditions for causal inference using RDD. We set the schools with "achieve the standard" as the treatment group, other schools as the control, and set the policy rate of reaching standard 80% as the threshold. In this way, we can observe whether there is a "cutoff" in the performance of schools near this critical value, so as to determine whether the standardized construction of the compulsory school as an exogenous shock variable has become the cause of the difference in student performance.

Model Set-up and Variable Description

Based on the above idea, we first used the OLS to explore the relationship between the degrees of standardization construction, whether or not the standardization construction meets the standards and student performance. The specific model is set as follows:

$$xxmean_i = \lambda_0 + \lambda_1 xzhachrate_i + \lambda_i W_i + \mu_i \qquad (1)$$
$$xxmean_i = \gamma_0 + \gamma_1 T_i + \gamma_i W_i + u_i \qquad (2)$$

Among them, *xxmean_i* represents the average score of the test results of students in i school; *xzhachrate_i* represents the degree of standardized construction of i school; T_i is a categorical variable, which indicates whether the standardized construction of i school has reached the standard, $T_i = 1$ represents reached the standard, $T_i = 0$ represents not reached; W_i is the control variable, it includes socio-economic variables such as the area to which the i school belongs, the per capita GDP of the city where the school located, the education demands of the city where the school located, and urban-rural categories; μ_i is the residual.

On this basis, we set the basic standards (that is, the rate exceeds or equal to 80%) defined in the standardization construction assessment of Province A as cutoff, and use Sharp Regression Discontinuity to infer causality. The specific study design is: according to whether each school's degree of standardization construction exceeds 80%, the following formula can be obtained:

$$T_i = \begin{cases} 1 \text{ if xzhachrate}_i \geq 80\% \\ 0 \text{ if xzhachrate}_i < 80\% \end{cases} \tag{3}$$

Where T_i is treatment variable, *xzhachrate$_i$* is assignment variable. If the degree of standardization construction xzhachrate$_i \geq 80\%$ of i school, then T_i is equal to 1; otherwise, T_i is 0. Whether the school enters the treatment group or the control group depends entirely on whether the degree of standardization construction *xzhachrate$_i$* exceeds 80%. According to this, the school are randomly grouped in the smallest neighborhood [80% - ε, 80% + ε] near the cutoff to obtain quasi-experimental design , So, the local average treatment effect (LATE) near *xzhachrate* = 80% can be estimated:

$$\begin{aligned} \text{LATE} &= E(y_{1i} - y_{0i}|\text{xzhachrate} = 80\%) \\ &= E(y_{1i}|\text{xzhachrate} = 80\%) - E(y_{0i}|\text{xzhachrate} = 80\%) \\ &= \lim_{\text{xzhachrate}\downarrow 80\%} E(y_{1i}|\text{xzhachrate}) \\ &\quad - \lim_{\text{xzhachrate}\uparrow 80\%} E(y_{0i}|\text{xzhachrate}) \end{aligned} \tag{4}$$

Set the Regression Discontinuity model as follows:

$$\text{xxmean}_i = \alpha + \beta_1(\text{xzhachrate}_i - x_0) + \beta_2 T_i + \beta_3 T_i \cdot (\text{xzhachrate}_i - x_0) + \beta_i Z_i + \varepsilon_i$$
$$(\text{-h} \leq \text{xzhachrate}_i - x_0 \leq h) \tag{5}$$

Among them: *xxmean$_i$* represents the average score of the test results of students in i school; *xzhachrate$_i$* indicates the degree of standardized construction of i school; *x_0* indicates the cutoff value, which is 80%; *xzhachrate$_i$-x_0* is central treatment for the degree of standardization construction; T_i indicates whether the standardized construction of i school has reached the standard with 80% as the cutoff, 1 means that achieve standard, and 0 means that not up to standard; Z_i is the control variable, which includes the area where i school belongs, the per capita GDP of the city in which the school is located, education needs, and urban-rural categories; The interaction term *$\beta_3 T_i$ (xzhachratei-x_0)* is to allow regressions on both sides of the cutoff to have different slopes.

Data Source and Variable Description

The data used in this study mainly came from the monitoring data of standardization construction of compulsory education schools in Province A, the test data of academic

quality of compulsory education students, and the statistical yearbooks of various cities in Province A.

Among them, the monitoring indicators of standardization construction of province A in 2017 were composed of 8 first level indicators (school setting, campus construction, educational equipment, teacher team, education and teaching, school management, quality evaluation, and funding guarantee), 40 second level indicators and 69 monitoring points. The elementary schools (including teaching center), middle schools and nine-year schools with compulsory education in 13 cities and 116 counties (cities and districts) in Province A were estimated.

The academic quality monitoring of compulsory education students in Province A was an independent test of the academic quality of fourth and ninth grade students in Province A. The test content was the academic level of Chinese and mathematics of the third grade students, and the academic level of Chinese, mathematics, English and science of the eighth grade students. In 2016, the Province A adopted a two-stage stratified sampling method and sampled 1,771 elementary schools. Among them, the number of students sampled was 141,516 in Chinese, and 141,460 in math. In middle school, Chinese and Mathematics had 992 schools were sampled, and 82,358 and 82,319 students were chosen respectively; English and Science had 883 schools were sampled, 78,246 and 78,685 students were chosen, respectively.

The statistical yearbook data of 13 cities in Province A were from the websites of statistical bureau of each city. It was for obtaining the annual data of economic development and education demand of each district and city.

On this basis, we conducted data processing according to the research design. Given the research focus was on the relationship between standardized construction and students' academic performance, so we matched the school standardized construction data and student academic performance data. In the end, the data of 1,619 elementary schools and 880 middle schools were used.

Among them, the academic performance of elementary school was measured by the average of Chinese and math performance of each school; the academic performance of middle school was measured by the average of Chinese, math, English and science performance of each school. The calculation of the degree of the school's standardized construction was divided by the number of compliance monitoring points divided by the actual number of monitoring points. To make the estimation more accurate, control variables were added. Under the decentralized and county-based education management system in China, the development of local education was related to the local socio-economic level, so the per capita GDP index of the city where the school was located represented the level of local economic development. In addition, the demand for education was also one of the factors affecting the standardization of schools. Therefore, we divided the number of students in school in each prefecture-level city in the *Statistical Yearbook of Province A* by the number of permanent residents (calculated separately for elementary and middle schools) to represent the educational needs of the city where the school is located. In addition, urban-rural disparities and regional differences (central, southern, and northern province A) might also affect student perfor-

mance, so the urban-rural categories and regions to which schools belong were used as control variables in the model.

The specific variable settings are shown in **Table 1**. There was a large difference in the test results of elementary and middle schools in different schools. The minimum test score for elementary school was only 342 points, and the highest score is 663 points. The degree of the standardization construction of elementary schools was between 58.1%-100.0%, and the average rate was 82.4%. The level of middle school standardization was between 58.1%-98.4%, and the average rate was 84.8%. In addition, there were large differences in the level of economic development and education demand of different regions.

Results

Correlation analysis based on OLS

First, the relationship between school standardization and student performance was analyzed by OLS. Through the White test, we found that there was a problem of heteroscedasticity in the regression equation. Therefore, weighted least squares regression (WLS) was used for heteroscedasticity correction.

Models (1) and (2) showed that after controlling the school's area and urban-rural category, the education demand of the city, and the per capita GDP of the region where the school is located, the degree of standardization of elementary school construction and whether it meets the standard or not was negatively related to the average student's performance, but not show statistical significance. The regression results of model (1) showed that under the control of related variables, for every 1% increase in the level of standardization, the grade of elementary school students decreased by 0.113 points, but it was not statistically significant. The regression results of model (2) showed that the difference between achieve the standard or not of standardization construction in elementary schools and the achievement of students was 1.802 points, but it was not significant.

Models (3) and (4) showed that the degree of standardization construction in the middle school and whether the standardization construction meets the standards or not was significantly negatively related to student performance. Model (3) found that the degree of standardization construction increased by 1%, and the score of middle school students reduced by 1.225 points. Model (4) showed that compared with below the standard schools, the average score of students in middle schools that reached the standard was 15.779 points lower.

In terms of controlling variables, the economic level of the city where the elementary school is located had a significant positive impact on student performance, and there was a positive but not significant impact on the educational demands and student performance in the middle school stage. The coefficient of the region to which the school belongs indicated that the performance of the Central province A was higher than those of the Northern Province A. Students in urban areas were significantly better than rural students.

Variable	Variable Meaning		Mean	SD	Min	Max
xxmean	Test score	Elementary School	503.524	57.054	341.750	663.014
		Middle School	495.080	47.461	301.453	649.278
T_i	Reach the standard or not of standardization construction	Elementary School	0.676	0.468	0	1
		Middle School	0.773	0.419	0	1
xzhachrate	The degree of standardization construction (%)	Elementary School	82.442	6.898	58.060	100
		Middle School	84.800	6.922	58.060	98.390
cityGDP	Per capita GDP by city (10,000 CNY)	Based on Elementary School sample	9.932	3.227	4.831	14.556
		Based on Middle School sample	9.867	3.238	4.831	14.556
qy1	Central Province A	Elementary School	0.200	0.400	0	1
		Middle School	0.206	0.404		
qy2	Southern Province A	Elementary School	0.460	0.499	0	1
		Middle School	0.444	0.497		
cx1	City	Elementary School	0.525	0.500	0	1
		Middle School	0.557	0.497		
cx2	County town	Elementary School	0.343	0.475	0	1
		Middle School	0.362	0.481		
edudemand	Education demand of 13 cities in province A (%)	Elementary School	6.216	2.021	4.480	10.390
		Middle School	2.383	0.460	1.800	3.460

Table 1. Variable Descriptive Statistics.

Causal Inference Based on RDD

Based on the previous design, we used the standard of 80% as cutoff to analyze whether there is a causal relationship between standardization construction and student academic performance.

Figures 1 and 2 are the RDD data of the degrees of standardization construction and students' academic performance after centralized processing. The academic performance of elementary and middle school students had a slight fluctuation at 80% of the standardization construction indicating that no significant causal relationship exists between the construction of standardization and the differences in student performance at various schools.

In order to accurately estimate the causal relationship between them, we performed further analysis. For RDD, the choice of bandwidth is critical. We chose the optimal bandwidth based on minimizing the Mean Square Error of the two regression functions at the cutoff proposed by Imbens & Kalyanaraman (2012). We used a rectangular kernel and half and twice the optimal bandwidth for causality tests.

Table 3 shows that under different bandwidth settings, there is no significant causal relationship between the standardized construction of elementary and middle schools and student performance. In order to ensure the robustness of the discontinuity regression results, we also tested the continuity of the conditional density of the group-
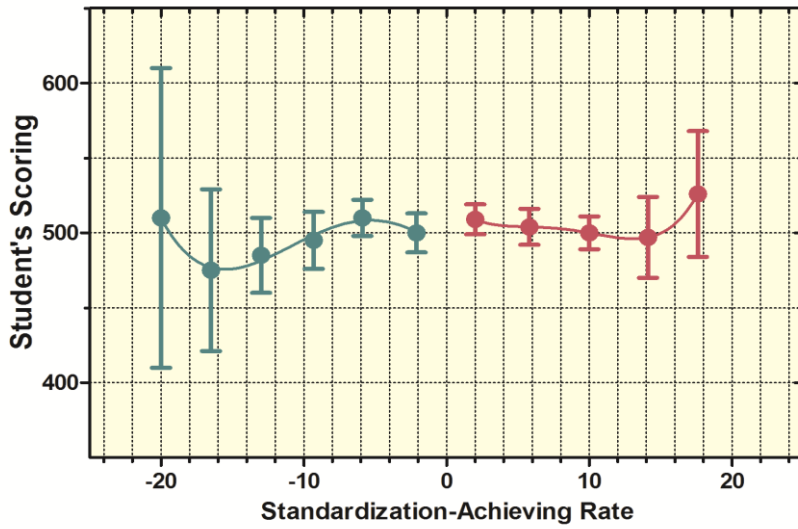
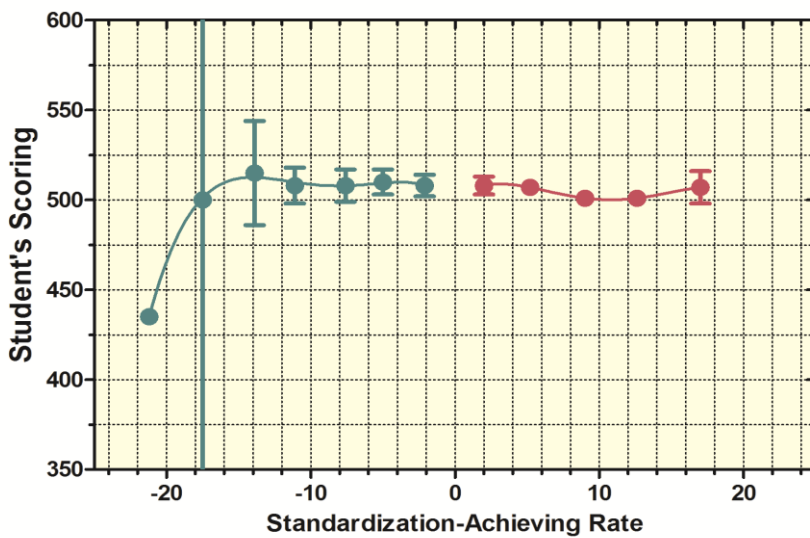**Figure 1. Regression Discontinuity (Elementary School).**

Figure 2. Regression Discontinuity (Middle School).

Table 2. OLS Regression Results of Compulsory Education School Students' Performance and the Degree of Standardization.

	Elementary School		Middle School	
	(1)	(2)	(3)	(4)
The degree of standardization construction (%)	-0.113		-1.225***	
	(-0.636)		(-5.807)	
Reach the standard or not of standardization construction		-1.802		-15.779***
		(-0.724)		(-4.374)
Per capita GDP by city (10,000 CNY)	2.928*	6.129***	0.562	1.737
	(1.792)	(3.134)	(0.291)	(0.813)
Educational demand of the city where the school is located (%)	-0.364	-0.471	8.789	8.527
	(-0.307)	(-0.383)	(1.518)	(1.432)
School Location — Central Province A	44.796***	41.883***	43.286***	41.001***
	(5.175)	(4.865)	(5.388)	(4.990)
School Location — Southern Province A	-12.517	-34.746**	35.986**	24.875
	(-0.967)	(-2.535)	(2.464)	(1.621)
Type of School Area — City	27.568***	29.380***	17.730***	15.556***
	(6.275)	(7.562)	(2.942)	(2.728)
Type of School Area — County/Town	-1.204	-1.568	9.087	7.543
	(-0.270)	(-0.389)	(1.537)	(1.365)
Constant Term	467.430***	439.001***	534.434***	438.892***
	(21.346)	(26.616)	(18.670)	(19.640)
Number of Samples	1,619	1,619	880	880
F	43.667	53.652	23.708	22.813
Adj. R^2	0.156	0.186	0.153	0.148

*Note: The value of t in parentheses; ***p , **p , and *p indicate that the levels of significance of 1%, 5%, and 10%, respectively.*

Table 3. Regression Discontinuity Results of Compulsory Education School Student Performance and Reach the Standard or not of Standardization Construction.

Achieve the standard or not of standardization construction	Elementary School	±3.541		±7.083		±14.165	
		-6.146	-11.164	8.661	7.619	8.158	5.845
		(-0.551)	(-0.963)	(1.190)	(1.005)	(1.583)	(1.083)
	Middle School	±3.76		±7.53		±15.06	
		12.216	11.222	4.217	5.549	2.026	1.815
		(0.754)	(0.670)	(0.456)	(0.571)	(0.278)	(0.238)
	Add control variables or not	Y	N	Y	N	Y	N

*Note: Z value in parentheses; ***p, **p, and *p indicate that the levels of significance of 1%, 5%, and 10%, respectively.*

ing variables and control variables at the cutoff point. The results showed that there is no obvious change at the cutoff between the assignment variable and the control variable, which was in accordance to the continuity assumption.

In summary, through RDD analysis, we did not find a clear causal relationship between standardization construction and student performance, that is, the level of standardization construction is not the cause of differences in student performance.

Discussion and Reflection

Based on our analyses, from the monitoring indicators of the standardization construction of compulsory education schools, in the dimensions involved in the indicators, especially in terms of hardware security, Province A had basically completed the task of balanced development, and the standardization construction work had reached a higher level. In other words, from the input side, the effectiveness of the standard construction was satisfactory.

However, many previous studies found that the improvement of school running conditions may not necessarily lead to the improvement of education quality. For example, the Coleman Report found that investment in educational resources was not the biggest cause of differences in student performance (Coleman et al., 1966). In addition, Hanushek at Stanford University analyzed 377 related studies and found that no strong correlation exists between student performance and school resources. Simple resource policies had little effect on improving student performance (Hanushek, 1997). Similarly, studies from China also reached these conclusions. For example, Hu (2007) used the data of rural public elementary and middle schools in western China through regression analysis and found that most of the physical and financial resources invested in elementary and middle schools did not significantly improve student performance. Hou and Shen (2014) used the 2009 Shanghai PISA test data for middle school students and found that there was no significant correlation between the quality of school education resources and student performance.

As thus, these findings show further that when we do policy assessment, we cannot look at the "inputs" only, but also the "outputs". Although we do not think that output-oriented evaluation is a better performance evaluation method than input-oriented one; on the contrary, we firmly believe that in the process of standardization, improvement and balance of guarantee conditions are the top priority of this work. However, we also believe that while most of the current similar evaluations focus on input performance, the review of output performance is a necessary link to ensure that the evaluation conclusions are comprehensive, scientific and robust.

In fact, our analysis also showed that when we conducted output-oriented policy performance evaluation, we oftentimes got different results from the input-oriented evaluation. In this study, although the construction results on the input side was impressive; many problems were revealed on the output side. OLS regression results showed that no significant correlation exists between standardized construction policies and elementary school students 'performance; but there was a negative correlation with

middle school students' performance, and the average score of school that was 15.78 points lower than those who were below the standard. The RDD analysis also found that there was insufficient evidence that standardization construction is the cause of differences in student performance.

Our conclusions are obviously different from the results of input-oriented performance evaluation. We are more concerned about what exactly caused this difference. So, we further analyzed the monitoring data of standardized construction, and performed variance analysis on 62 quantifiable detection points according to achieve the standard group and below the standard group to determine whether the students' performance in each group was significantly different under different indicators.[1] The results showed a significant difference between the two groups in the number of school tracks, the area per student, sports area per student, the provision of network multimedia classrooms per student, the setting of reading rooms, electronic reading rooms, and the number of stored books in libraries, etc. Taking the "number of school tracks" as an example, the ANOVA results show that the average score of the students in the schools that meet the standards is 13.01 points lower. According to the requirements in the standardized monitoring indicators, "complete elementary school is below 6 tracks". Only schools with less than or equal to 6 classes in one grade can be included in the compliance group. However, in the course of our investigation, we found that most of the rural elementary schools in Province A meet this condition, and in urban elementary schools, especially many "quality elementary schools" are often difficult to find. Therefore, it is impossible to reach the target on this indicator. However, the quality of teaching in such schools is often much higher than the average level, which in general raises the student performance of schools that do not meet this target.

At the level of middle school, indicators such as the number of school tracks, the average building area per student, the number of books per student, the number of new books per student per year, the student-computer ratio, and the ratio of middle school teachers with undergraduate or higher education levels were significantly different between achieve and the below the standard. Taking the "student-computer ratio" as an example, those who were achieve-the-standard scored 14.31 points lower than the contrast ones. According to the requirements of the monitoring indicators, "the ratio of the number of students in the school to the number of students using computers in the network environment must reach 8:1 in middle schools." High-quality schools often have difficulty in meeting this requirement due to the large scale. The low scores of resource indicators per student are often the main reason why the "high-quality schools" with higher grades have difficulty meeting the standards.

From our analyses, in many indicators, the phenomenon of "below the standard" group has better performance, which reveals exactly everyone's thirst for quality education resources. The relative shortage of quality education resources has led to a shortage of "good schools". The expansion of the scale of running a school will reduce the supply of resources per student, which will inevitably lead to lower-than-expected results at the output end. This divergence between the actual quality of running a school and the evaluation results at the input side reminds us once again that we need to reflect on the

setting of such education policy evaluation indicators, so as to use a more comprehensive perspective to define and evaluate the implementation of education policies.

Conclusions and Suggestions

We can draw the conclusions based on our analyses: from the perspective of input, the standardized construction of compulsory education schools in Province A had achieved gratifying results, but from the perspective of output, such kind of input had not significantly affected student performance. No sufficient evidence indicates that standardization construction is the cause of differences in student performance.

We need to point out that this conclusion does not mean we deny the value and significance of resource input. After all, resources are the bottom line guarantee, and our conclusion shows that the current standardization construction still has much room for improvement in terms of how to achieve balance between input and sufficient output. In particular, we suggest:

First, in the direction of standardization construction, we must achieve a transition from equal input to sufficient results. As pointed out by existing research, "Quantity and hardware standards are the appearance of standardized school construction, and quality and connotation standards are the essence of standardized school construction; therefore, the quality of school running must always be the first place for standardized evaluation" (Zhang & Zhang, 2017). Currently, the desire for higher quality education and the contradiction between the imbalance and inadequate supply of quality education resources are the main conflict in the process of educational reform and development. As mentioned earlier, since the 1980s in the United States, the principle of education supply had gradually changed from fairness to adequateness, and its core idea was to allow each child to achieve the maximum possible development under the conditions of certain resources (Xue, 2011). This is also of value to China's current standardization construction.

Second, in the focus of construction and the selection of performance evaluation indicators, to avoid "seeing things but not seeing people," we must coordinate the relationship between the unified requirements and the development of the school's characteristics. This requires that standardized construction is to increase the evaluation of output quality while ensuring the necessary facilities and equipment, teachers, curriculum resources and funding for the development of each school. In order to promote hardware investment and condition improvement to truly serve the development of students, and thus effectively improve the quality of education, strive to achieve the growth and development of each student. In this process, it is necessary to avoid setting all indicators in a balanced and stable manner. It is necessary to allow the development of school characteristics and the substitution and coordination between different evaluation dimensions. It is not necessary to be confined to the full realization of the indicators, as long as the construction of the school can better achieve the development of students, it should be regarded as completing the task of "standard construction".

Third, strengthen scientific research and set scientific and rational evaluation indicators for standardization construction. An important finding of our study is that

simply focusing on the standardization of the input side does not promote the improvement of student output. The reason for this is largely due to the fact that in the setting of standardized construction indicators, the relationship between "input" and "output" has not been thoroughly explored. Many indicators are often designed based on experience rather than scientific evidence. In the process of evidence-based reform, European and the USA require solid empirical evidence as the basis for making major policy decisions. In the "Every Child Success Act", the United States not only continues the requirements of evidence-based decision-making in the "No One Child Left Behind" act, but also makes clear the strength of various types of evidence (Slavin, 2017). From this point, China has a lot of room for improvement. In the process of future standardization, more emphases should be placed on the rationality and scientificity of the index setting, and each index should be fully demonstrated to make the evaluation play a better role in finding problems and guiding reform.

Note:

1. *Due to space limitations, we did not report the results of variance analysis of 62 indicators. For complete information, please contact Yao, the corresponding author of this article. Contact email: yaojijun_ njnu@163.com.*

References

Chen, J.L., Ding, Y.Q., Ye, X.Y. (2014). Causes and effects of school choice of the students of senior high school in China. Edu Res Mon, (05):36-45.

Chu, S., Meng, F.Q. (2017) College enrollment expansion in China and the difference return of education between urban and rural: Based on regression discontinuity design. South Chin J Econ, (10):16-35.

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., York, R.L. (1966). Equality of educational opportunity. Washington: Government Printing Office.

Education Department of A Province. (2016). Measures for monitoring the standardized construction of compulsory education schools in A Province. http://www.ec.js.edu.cn/art/2016/9/30/art_ 4267_198203.html.

Hanushek, E.A. (1997). Assessing the effects of school resources on student performance: An update. Edu Eva Policy Anal, 19(2):141-164.

Hou, Y.N. Shen, A.X. (2014). The impact of school resources on quality & equality of basic education in Shanghai: Empirical study based on PISA2009. Edu Res Mon, (09):38-45.

Hu, Y.M. (2007). The relationship between school resources allocation and student's attainment: Empirical research based on rural school survey of five provinces in west China. Beijing Normal Uni.

Imbens, G., Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. The review of economic studies, 79(3):933-959.

Lee, D.S., Lemieux, T. (2010). Regression discontinuity designs in economics. J Econ Lit, 48(2):281-355.

Lee, J. (2018). Does merit-based aid promote degree attainment? KEDI J Edu Policy, 15(1):99-118.

Li, W.L., Zeng, M.L. (2002)."New" finance of American basic education. Edu Res, (05):84-89.

Li, X.Y., Tao, X. (2015). From average to equity: Concept's evolution of American educational equality-Based on the American financial litigation about basic education since the1970s. Edu & Econ, (03):58-64.

Li,W., Konstantopoulos, S. (2016) Class size effects on fourth-grade mathematics achievement: Evidence from Times 2011. J Res Edu Effect, 9(4):503-530.

Liu, S.L., Zhou, S.J., Hu, A.G. (2016). Compulsory education law and economic return to education in urban China: Based on regression discontinuity design. Eco Res J, (02):154-167.

Outline of the national medium-and long-term education reform and development plan. (2010). Retrieved from http://old.moe.gov.cn/publicfiles/business /htmlfiles/moe/info_list/201407/xxgk_171 904.html

Provincial education evaluation institute of A. (2018). 2017 A provincial compulsory education school standardization construction monitoring report. Nanjing.

Provincial government of A. (2015). Standards for provincial A compulsory education schools (Trial). http://www.ec.js.edu.cn/art/2016/1/19/art_ 10338_187394.html.

Slavin, R.E. (2017). Evidence-based reform in education. J Edu Stu Place Risk, 22(3):178-184.

Thistlethwaite, D.L, Campbell, D.T. (1960). Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment. J Edu Psychol, 51(6):309-317.

Wang, J., Sun, Z.J. (2015). Does key senior high schools improve student's academic performance? A regression discontinuity study in county F. Peking Univ Edu Rev, 13(04):82-109.

Wei, X. (2012). Does NCLB improve the achievement of students with disabilities? A regression discontinuity design. J Res Edu Effect, 5(1):18-42.

Xue, E.Y. (2011). Policy innovation of American education equitable development. Res Edu Dev, 31(19):80-84.

Yao, J.J. (2018). The establishment and influence of education adequacy standard in the USA and its enlightenment. Pri & Sec Sch Abroad, (7):10-16.

Zhang, X.P., Zhang, R. (2017). Reflection on construction of standardized school in compulsory education: Status, problems and choices. Res Edu Dev, 37(18):1-7.