Effects of After-School Education Vouchers on Children's Academic and Behavioral Outcomes: Evidence from a Randomized Experiment

By

Hideo Akabayashi (Keio University)
Hiroko Araki (Kindai University)
Ryuichi Tanaka (University of Tokyo)

December 2018

CREPE DISCUSSION PAPER NO. 48

CENTER FOR RESEARCH AND EDUCATION FOR POLICY EVALUATION (CREPE)
THE UNIVERSITY OF TOKYO
http://www.crepe.e.u-tokyo.ac.jp/

Effects of After-School Education Vouchers on Children's Academic and Behavioral Outcomes: Evidence from a Randomized Experiment

Hideo Akabayashi (Keio University), Hiroko Araki (Kindai University), Ryuichi Tanaka (University of Tokyo)[1]

First Draft: October 28, 2016; This Draft: December 4, 2018

Abstract

We estimated the causal impact of vouchers for after-school education programs on children's academic and behavioral outcomes using experimental data of middle and high school students. Our identification strategy relied on the random assignment of after-school education vouchers provided to children who suffered as a result of the Great East Japan Earthquake. We estimated the value-added models of various outcomes such as academic test scores (mathematics and Japanese language art) and non-cognitive skill measures (self-esteem and quality of life) of the children. We carefully treated potential biases due to sample attrition and the small sample property by employing the inverse probability weight for regression analyses. Our estimation results revealed that the assignment of vouchers had a positive and significant effect on the increase of mathematics test scores of high school students immediately, and of language art test scores of middle and high school students in one year. These results were robust to the fully non-parametric permutation tests. We found that the assignment of after-school education vouchers negatively affected the self-esteem scores of middle school students immediately, but this relationship became weak and insignificant in one year. We found that the assignment of vouchers was positively related to the quality of life measure, but these relationships were insignificant. We also estimated the effect of vouchers on the children's study time, finding that the assignment of vouchers had a positive and statistically significant effect on study time, mainly at home. However, these positive effects were insignificant in the fully non-parametric permutation tests. This indicates that the assignment of vouchers improved the quality of the study environment of the children without increasing their study time, thus resulting in better academic outcomes.

JEL classification: I21, I22
Keywords: Educational Voucher, Causality, Randomized Control Trial, Shadow Education, the Great East Japan Earthquake

---

[1]Corresponding author: Hideo Akabayashi, email: hakab@econ.keio.ac.jp. The authors gratefully acknowledge the financial support from Grants-in-Aid for Scientific Research (B) 24330090 and Grant-in-Aid for Challenging Exploratory Research 16K13372. The authors would like to thank the Chance for Children for providing us with detailed administrative data, and Withus Co. and Fujii Shobou for generously allowing us to use their original standardized tests. We are also grateful to Stacey Chen, Susan Dynarski, Daiji Kawaguchi, Yasuyuki Sawada, and participants at the 2013 JEA spring meeting, seminars at Hitotsubashi, Keio, U-Tokyo, and Kansai Labor Workshop. Special thanks go to Makiko Nakamura who contributed to this study in an early stage. The views and opinions expressed here are those of the authors and not necessarily those of the Chance for Children or any of its stakeholders. This research was approved by the IRB at the Institute for Economic Studies, Keio University. All remaining errors are solely ours. There are no financial conflicts of interest to disclose.

1. Introduction

In most advanced countries, public schooling is common at the primary and secondary levels of education. Governments support public schooling, because education in schools fosters the nation's human capital. Public schooling plays an important role in mitigating educational inequality stemming from household income inequality.

After-school activities are another important form of human capital investment before the tertiary level of education. An after-school activity is any organized program that invites youth to participate outside the traditional school day. For example, sports, performing arts, creative arts, and academic enrichment are typical forms of such activities. After-school activities are common; for example, in the US, 40 of the largest national youth organizations have a total membership of about 40 million youths. Some programs are run by externally funded non-profit or commercial organizations such as 4-H clubs and the YMCA, while others receive state or federal funds. While the 21^{st}-Century Community Learning Centers and No Child Left Behind are federally funded, 21^{st} century High School ASSETS (After School Safety and Enrichment for Teens) is financed by the state of California.

After-school activities are also very popular in East Asian countries, especially in Japan. Among various types of activities, elementary and secondary school students commonly participate in academic enrichment activities such as cramming school and private tutoring (called "shadow education"). According to a report by the Ministry of Education, Culture, Sports, Science and Technology (2015), 47.3 percent of 6^{th}-grade elementary school students and 61.1 percent of 3^{rd}-grade junior high school students participate in after-school cramming education. Similarly, 24.3 percent of high school students participate in academic enrichment after school (The Fourth Basic Survey of Learning in Japan by Benesse, 2006) to ensure they are prepared for college entrance exams.

There are arguments on both sides regarding the pros and cons of cram schooling and private tutoring. In support of such activities, Dang (2007) pointed out that cram schooling and private tutoring are useful in attaining higher academic achievement (higher test scores, graduating on time, fewer dropouts, etc.), which is important for success in later life. Matiashvili and Kutateladze (2006) argued that these out-of-school activities might remedy and compensate for the poor quality of schooling. On the other hand, in a report noting the downside of such activities, the National Plan Outline for Medium and Long-Term Education Reform and Development of China (2010) stated that heavy schoolwork (through shadow education such as cramming school and private tutoring) is harmful to the mental and physical wellbeing of youngsters and health of the nation. Bray et al. (2014) pointed out that tutoring imposes excessive burdens on students. Bray and Kwok (2003) argued that too much focus on academic cramming could interfere with educational processes in mainstream classes. Dang and Rogers (2008) highlighted the heavy cost on households. As a result, Bray and Kwok (2003) suggested that heavy reliance on private tutoring and out-of-school academic enrichment activities reduces intergenerational educational and income mobility. Hence, it is imperative to examine the effect of after-school academic activities not only on academic outcomes, but also on non-cognitive and behavioral outcomes to evaluate the total impact of such activities.

In this study, we estimated the causal impact of vouchers for after-school education programs on children's academic and behavioral outcomes using experimental data on middle and high school students. We designed a random assignment experiment on a sample of participants regarding after-school education vouchers provided to children who suffered as a result of the Great East Japan Earthquake. We estimated the value-added models of various outcomes such as academic standardized test scores (mathematics and Japanese language art) and non-cognitive skill measures (self-esteem and quality of life) of the children. Furthermore, we carefully treated potential biases due to sample attrition and the small sample property by employing an inverse probability weight for regression analyses.

Our estimation results revealed that the assignment of vouchers for after-school education programs had a positive and statistically significant effect on increasing the mathematics test scores of high school students immediately, and of language art test scores of both middle and high school students in one year. These results were robust to the fully non-parametric permutation tests. We found that the assignment of after-school education vouchers negatively affected the self-esteem score of middle school students immediately, but this relationship became weak and insignificant in one year. Furthermore, we found that the assignment of vouchers tended to be positively related to the quality of life measure, but these relationships were insignificant.

To uncover the mechanism by which these effects manifest in children's outcomes, we also estimated the effect of vouchers on the children's study time. We found that the assignment of vouchers has a positive and statistically significant effect on study time, mainly at home. However, these positive effects were insignificant in the fully non-parametric permutation tests. This finding indicates that the assignment of vouchers improved the quality of the children's study environment without increasing their study time, resulting in better academic outcomes.

This study is related to the strand of literature on remedial education. Banerjee, Cole, Duflo, and Linden (2007) estimated the effect of remedial education on the literacy and numeracy skills of elementary school students using randomized experiments in a school in urban India. They found positive effects on the test scores of students in the 3rd and 4th grades. Lavy and Schlosser (2005) evaluated the effects of remedial education programs for high school students in Israel, and found improvements in the mean pass rates in the baccalaureate exams. Briggs (2001) revealed that commercial private tutoring courses increased SAT mathematics scores in the US. Jacob and Lefgren (2004) used a regression discontinuity design to investigate the reform of the school system in Chicago, finding that remedial education increased 3rd-grade students' academic achievement in public elementary schools. One important factor differentiating our current analysis from these studies is that we estimate academic as well as non-cognitive and behavioral outcomes using a unique experimental sample. This enabled us to thoroughly evaluate the effectiveness of the voucher program.

Numerous papers attempted to evaluate the effects of cram schooling and private tutoring. Among many others, Lee (2013) used Korean junior high school students' data to estimate the effects of cram schooling on academic test scores and study behavior through propensity score matching, IV, and quantile regression. Dang (2007) found that private tutoring positively affected academic achievement among students in Vietnam. Akabayashi and Araki (2011) evaluated the effects of educational vouchers for students in private high schools using panel IV regression. Despite the number of papers, the literature is far from conclusive. Bray et al. (2014) stated, "The literature has not delivered clear findings about the relationships between tutoring and academic achievement." Other studies by Byun (2014); Liu (2012); Primont and Domazlicky (2006); Zhang (2013); and Zimmer, Hamilton, and Christina (2010) reported similar findings. We focus on the effect of the assignment of education vouchers on academic and behavioral outcomes using an experimental design. However, we attempt to provide indirect evidence of the effectiveness of after-school education programs.

A recent paper by Kobayashi (2018) is most closely related to ours. Similar to our paper, Kobayashi estimated the effects of shadow education vouchers provided in an area affected by the Great East Japan Earthquake on children's own subjective rating of academic rank at school using a regression discontinuity design. Our study differs in that we obtained both standardized academic test scores and behavioral outcomes measures to evaluate the value-added effect of the vouchers, and used experimental data generated from our randomized control design. In our empirical findings, unexpected results regarding non-cognitive outcomes are key to understanding the overall effects of after-school education vouchers for children.

The remainder of this paper is structured as follows. Section 2 provides a detailed institutional background of the program that provides private education vouchers. Section 3 explains our econometric specification and identification strategy. Section 4 reports the descriptive statistics of the data set used for the analysis. Section 5 reports our main findings, and Section 6 concludes the paper.

2. Institutional Background

2.1 Chance for Children's After-School Education Voucher Program

On March 11, 2011, a large area in the Tohoku region in East Japan was devastated by a powerful earthquake measuring magnitude 9 on the Richter scale. This earthquake was followed by a large tsunami that engulfed the Pacific coast along the area where the prefectures of Iwate, Miyagi, and Fukushima are located. The tsunami resulted in the severe malfunction of the Tokyo Electric Power Company's Fukushima No. 1 nuclear power plant located in the south of Fukushima prefecture, causing a major leak of radioactive material. This unprecedented event, known as the Great East Japan Disaster, took the lives of 19,667 people, and left 2,566 missing and 6,231 injured.[2] By September 2018, 7 and a half years after the earthquake, more than 37,000 people are still living in temporary shelters.[3] Thousands of people from the affected areas, many of them school-aged children, are still struggling to recover from the devastating losses caused by the disaster.

In total, 891 people aged less than 20 years lost their lives in the disaster, and 1,724 children aged less than 18 years lost at least one parent, among whom 241 were orphaned. The infrastructure of schools in the area was severely damaged by the earthquake and the tsunami that followed. More than 7,988 public and private schools suffered some damage.[4] By September 1, 2011, half a

---

[2] According to the Fire and Disaster Management Agency, as of September 2018.
[3] According to the Reconstruction Agency, as of September 2018.
[4] MEXT (September 2012).

year after the Great Disaster, 25,751 students from primary to upper secondary levels were forced to relocate to different schools after losing their families, homes, and schools. Among these children, 24,092 came from the tsunami-stricken prefectures of Miyagi and Iwate, and from neighborhoods in Fukushima prefecture forced to evacuate because of the nuclear power plant accident.

In this context, in September 2011, the non-profit organization (a public interest incorporated association) *Chance for Children* (CFC) announced and initiated the acceptance of applications for the distribution of education vouchers to guarantee the education of children affected by the Great East Japan Disaster, in this way contributing to the recovery of the devastated areas. Each voucher was worth JPY 250,000 and could be exchanged for registered after-school educational services such as cram schooling, distance learning, and private tutoring.

The eligible population included children enrolled in primary to upper secondary education at the time of the earthquake who could be classified into any of the following seven categories as a result of the disaster: (i) Children beginning to receive welfare assistance; (ii) Children for whom at least one household member had died; (iii) Children for whom at least one member of the household is still missing; (iv) Children for whom at least one household member, including the applicant, was injured or disabled; (v) Children for whom at least one member of the household became unemployed; (vi) Children whose house suffered damages; (vii) Other children recognized as victims by CFC.[5] When applying, the children had to submit official evidentiary documentation to prove they could be classified in one of the abovementioned categories.

CFC selected voucher recipients based on their eligibility index value, which captures the degree of applicants' physical and material damage due to the disaster, as well as the level of their need to attend after-school activities. Among other aspects, the index considered criteria such as the grade of enrollment as a measure of the time remaining before the student had to take a high school or university admission exam, and whether the student lost the opportunity to receive after-school education because of the disaster. Children with high eligibility index values were prioritized for the distribution of education vouchers (a detailed description of the selection mechanism and sample of analysis in this study is provided in Section 2.3).

[5] Students applying for the Certificate for Students Achieving the Proficiency Level of Upper Secondary School Graduates (koutougakkou sotsugyouteido nintei shiken in Japanese) were counted as students applying for upper secondary education, and therefore considered eligible for the program.

In November 2011, 150 recipients were selected among 1,700 candidates during the first round of the distribution of vouchers. The recipients received the vouchers in December 2011, and were allowed to use them until the end of March 2013 in registered after-school educational institutions.[6] Since CFC had additional funds, it announced a second round of voucher distribution in April 2012, selecting 50 additional recipients among those unsuccessful in the first round. The vouchers for this second round were distributed in May 2012. They were similar in amount and had the same validity period as those distributed in the previous round (i.e., JPY 250,000 per person, valid until the end of March 2013).

Voucher recipients were also offered counseling services by volunteers (called "Big Brothers and Sisters"), who were university students from faculties related to the fields of education and welfare. They provided academic assistance by telephone and Skype to voucher recipients.

2.2 Surveys of CFC Voucher Program Applicants

In September 2011, along with the implementation of the voucher program, CFC called on one of the authors of this paper to conduct an independent and external evaluation of the program. For this purpose, the authors conducted three surveys according to the schedule presented in Figure 1.

[Figure 1]

The first survey was conducted at the beginning of November 2011, right after the end of the application period for the first selection round of the program. The survey was posted to 809 applicants enrolled in the lower and upper secondary education levels. Responses were received through the same means after two weeks.[7] Since the responses were collected before the selection results were announced, the respondents had no information on whether their applications had succeeded or not. In total, 590 valid responses were collected for a response rate of 72.93 percent.[8]

[6]As of November 2011, the educative vouchers distributed through this program were accepted by 36.03 percent of the after-school educational institutions located in the regions inhabited by the eligible group. By April 2012, 66.90 percent of the institutions accepted the vouchers.

[7] Surveys were not collected for primary school students because of the small number of applicants in this group.

[8] A response was considered valid when respondents returned a survey with at least one of the following parts completed: Japanese Language test, Mathematics test, or the Questionnaire.

[Table 1]

The first survey focused on test scores in mathematics and Japanese language art, and included a separate questionnaire. Problem sets of the academic tests differed based on the grade in the case of lower secondary education students, while upper secondary students took the same set of questions regardless of grade.[9] The time limit for the test was 40 minutes per subject for the lower secondary students, and 50 minutes per subject for upper secondary students. The tests were self-administered in the students' places of residence. The questionnaire was divided into two sections: one oriented to the applicant and the other to the child's parents or guardians. The main questions for the parents centered on their lifestyle, household income, age, education, and employment, and their views and ideas on the future of their child, access to after-school education for the child, and so on. Questions for the child focused on study times, reading habits, academic aspirations, and a measure of self-esteem using Rosenberg's Self Esteem Scale (Rosenberg, 1965).

The second survey was performed in April 2012, four months after the distribution of the first round of vouchers. The survey was distributed and collected via post, and targeted 754 children enrolled from the first year of lower secondary education (middle schools) to the second year of upper secondary education (high schools) by September 2011.[10,11] In total, 371 valid responses were obtained for a response rate of 49.20 percent. The contents of the survey were similar to those in the first one, focusing on mathematics and Japanese language arts test scores, and included a questionnaire for the applicant. The questionnaire included questions on access to after-school education, Rosenberg's Self Esteem Scale, and the health and lifestyle of the applicant (quality of life). For students in the third year of middle school by September 2011, questions on their academic or labor market status were added.

---

[9] From the first to the third surveys, the authors employed a practice test for the high school entrance examination as the problem set for lower secondary students. Similarly, a practice test for the high school equivalence examination was used for upper secondary students.

[10] The authors also collected information on the after-graduation career plans of students enrolled in the third-year higher secondary level by September 2011. The analysis of this data is left for future research.

[11] Initially, the length of the response collection period for the second survey was supposed to be two weeks, as for the first survey. However, as the response rate was lower than expected, the period was extended by one month. Because of the deadline extension, several responses arrived after the second round of the provision of vouchers was announced. This may have resulted in a bias in questionnaire responses among these people, but since our randomized sample contained few such respondents, we considered their influence minimal.

Finally, a third survey was conducted in November 2012, which targeted 754 students enrolled from the first year of middle school to the second year of high school by September 2011. It employed the same definition of the target group and data collection methodology. In total, 397 responses were obtained for a response rate of 52.65 percent. The contents of the survey were the same as those of the second survey: a set of academic test scores and questionnaire.

2.3 Selection Process, Eligibility Index, and Random Sample for the Voucher Program

This subsection describes the process through which voucher program recipients were selected. It also provides details on the sample for the estimations in this study.

To select program recipients, CFC created an eligibility index, which was calculated as the sum of scores in six areas capturing the necessity of after-school education for the child and magnitude of the damage to the candidate's household due to the disaster.[12] In all, 150 recipients were selected during the first selection round in the following manner: 106 applicants were selected according to their eligibility index value, starting from the highest value. The next position in terms of the eligibility index was shared as a tie among 115 applicants who were similar in terms of the damage suffered and educational opportunities sought.[13] From this group, 44 applicants were randomly selected to receive the benefits of the program, while the remaining 71 applicants were excluded from the voucher program. In this study, we evaluated the impact of the vouchers for after-school education by estimating the change in academic performance before and after the implementation of the program for this randomized sample.

As explained in Section 2.2, we conducted three surveys targeting students enrolled in middle and high school at the time of the application period (September 2011). However, from the second survey in April 2012, students in the third year of high school were not included. Therefore, this study concentrated only on students enrolled from the first year of middle school to the second year of high school at the time of the application. From the 115 candidates selected through randomized sampling, 100 satisfied this condition, among which 37 were selected as recipients

[12] The eligibility index was calculated as the sum of the scores in the following six areas: (1) Whether the applicant receives welfare assistance, (2) Human damage suffered because of the disaster, (3) House damage suffered because of the disaster, (4) Labor market situation of the family after the disaster, (5) Classification of the school and grade of enrollment, and (6) History of use of after-school education and current utilization status.
[13] Two primary school students, 81 lower secondary school students, and 32 upper secondary school students (among whom 13 were enrolled in the third year).

(the treatment group of our analysis) and 63 failed the selection process (therefore becoming the control group) in the first round of voucher distribution.

In the second round of selections, 50 candidates were chosen as recipients starting from the highest values of the eligibility index, which was recalculated based on each candidate's grade of enrollment in April 2012. From the 63 candidates in the control group in the first round, 30 were selected as recipients during the second round, while 33 were not. The selection process during the second round cannot be considered random sampling. Thus, for the purposes of the analysis in this study, we primarily evaluated the impact of the program on the academic and behavioral performance observed during the second survey, which was conducted before the second round of vouchers were distributed and corresponds to the randomized sample. However, we also analyzed the effects of vouchers on the outcomes during the third survey to supplement the analysis using the second survey.

The response rates of the surveys for the randomized sample are provided separately in Table 1 for those selected to receive the voucher and those not selected. The response rates for the second and third surveys were calculated as a ratio of the number of original participants and number of respondents to the first survey. In general, the response rate for the second survey of those not selected was considerably lower than that of the selected group. Possibly, the characteristics of those who participated in the survey and those who dropped out were different. As explained in Section 3.2, our estimations were performed by employing an inverse probability weighted regression to control for the effect of selective attrition.

3. Estimation Strategy

In this section, we first explain our econometric model, and then discuss our identification strategy and method to correct biases due to sample selection.

3.1 Econometric Model

To estimate the effect of educational vouchers on students' outcomes, we used a simple linear regression model:

$$y_i = \alpha + \rho D_i + X_i'\gamma + \varepsilon_i \quad (1)$$

where y_i is an outcome variable such as test scores, D_i is the treatment status that takes 1 if individual i is a recipient of an educational voucher and 0 otherwise, X_i is the vector of other covariates including individual characteristics, and ε_i is the error term. The coefficient parameter ρ is our parameter of interest capturing the causal effect of educational vouchers on students' outcomes. This parameter was consistently estimated by regressing y_i on D_i and X_i provided that the error term was orthogonal to these regressors (D_i, X_i).

In our setup, we assumed that treatment status D_i would be exogenously determined and uncorrelated with the unobserved characteristics of applicants, because voucher recipients were selected based on how seriously they were affected by the earthquake and how soon they had to take the entrance examination around November 2011. However, since the eligibility index included information on parents' economic status and previous use of out-of-school education services, it may be correlated to unobserved characteristics in the error term of equation (1). Moreover, because the tsunami exacerbated the severe damage wrought by the Great East Japan Earthquake, the treatment status may capture unobserved characteristics of people who lived in coastal areas. Hence, we confirmed that the differences in student achievements measured by the deviation value (math and reading scores) prior to the intervention between recipients and non-recipients were not statistically significant, as shown in Table 2 (for the sample of middle school students, high school students, and the pooled sample). Furthermore, we controlled the observable household characteristics to reduce standard errors caused by the differences in the effect of vouchers due to the differences in these characteristics.

We estimated regression equation (1) using the randomized sample generated during the assignment process for the first voucher in December 2011. The coefficient of interest can be interpreted as the average treatment effect on the treated children by using the limited size of the randomized sample during the period between the first and second surveys. The estimate of this coefficient answers the question in a way closer to the original purpose of the external evaluation of CFC's activities, and thus evaluates the impact and effectiveness of the CFC voucher project.

3.2 Correcting Estimation Bias Due to Attrition

As described in the previous section, the data were highly incomplete, since the project design was part of a relief effort. There was no pre-requirement for voucher applicants to comply with our request to respond to a series of surveys, and the attrition rate was inevitably very high from the first through the third surveys. Limiting the observations to the randomized sample, only 53 of the 100 eligible families returned the second survey, and only 52 of the 100 eligible families returned the third survey.

Given the severe attrition of applicants and small sample size, especially of the randomized sample, it was important to correct for a potential bias due to selective attrition. We did this using the standard method based on a weighted regression (Wooldridge, 2010, Campbell et al., 2014, Fitzgerald et al., 1999).

We assumed that attrition was fully explained by the finite number of observable variables in the data set, similar to the standard matching estimation method. Then, we estimated the attrition for two cases: attrition from the applicants' pool or the first survey respondents to the second survey for an analysis of the random sample. For analyses using the random sample, weighting using the variables in the applicants' pool was attractive, because it would cover the largest sample. However, when we analyzed the data based on the value-added form, it was necessary to include information from the first survey where possible. The choice of weighting depends on the trade-off between the importance of information from the initial application form and that from the first survey.

Similar to the matching estimation, a practical issue was the choice of covariates to estimate attrition. Following Campbell et al. (2014), we used the Akaike Information Criterion (AIC) to assist the optimal choice of the model. We chose the model with a set of covariates that minimized AIC among all combinations of the control variables in addition to the treatment variable, gender of the child, and initial test scores (for the full sample case), which were always included. Then, we constructed the inverse probability weighting (IPW) to reweight the observations in each case.

4. Data

All variables used in the estimations are summarized in Table 2. The outcome variables were classified into academic and non-cognitive outcomes. As academic outcomes, students' achievements were measured according to their standardized test scores in mathematics and Japanese language art tests, which are closely aligned to the national school curriculum. These test scores were converted into the deviation values for the estimation with mean 50 and standard deviation 10. As non-cognitive indicators, (i) quality of life (*QOL Index*) and (ii) self-esteem (*Self Esteem Index*) were employed. Quality of life was derived from KINDLR (http://kindl.org/english/) based on Ravens-Sieberer and Bullinger (1998), which is a set of questionnaires to assess the health-related quality of life in children aged three years and older. This variable ranges from 0 to 100, where 100 indicates the highest score possible. Self-esteem was derived from the "Rosenberg Self-Esteem Scale" by Rosenberg (1965), which ranges from 0 to 100.

The key independent variable of interest was voucher recipient status, which was coded 1 if an individual received the voucher and 0 otherwise. Other variables including control variables are also listed in Table 2.

[Table 2]

The descriptive statistics are summarized in Table 2. The top panel is for the sample pooling the middle and high school students. The middle panel is for high school students, and the bottom panel for middle school students.

To test the balance of the baseline characteristics (i.e., mathematics and Japanese language test scores and self-esteem index in the first survey) between the treatment group with vouchers and control group without vouchers, we tested whether the difference in the mean between the treatment and control groups was statistically significant. For middle school students, there was no statistical difference in mean characteristics between the treatment and control groups. However, we found a statistically significant difference at the 10 percent significance level in the Japanese language test scores in the first survey for high school students. This difference in Japanese language art scores in the baseline survey remained significant for the pooled sample of middle and high school students. In the regression analysis with value-added specifications, we included this baseline test score as a control variable. Moreover, we found no statistically significant differences in household characteristics, namely parents' education levels, household income, and use of after-school education programs before treatment assignment. Thus, we think that this difference in the baseline scores for the Japanese language art test does not severely affect our main results and conclusions.

Finally, CFC did not have any direct control over where or how much of the voucher recipients spent on after-school activities and shadow education. This means that some students did not use the whole amount of the voucher, even though they were in the treatment group.

Furthermore, CFC allows students to purchase lessons at driving schools, swimming schools, or for piano, etc. Thus, our estimate should be interpreted as the intention-to-treat-effect estimator.

5. Results

5.1 Academic Outcomes

Table 3 summarizes the estimation results of the effects of receiving educational vouchers on children's academic outcomes.

[Table 3]

The first three columns report the estimation results of the model with the mathematics test scores in the second survey as the dependent variables. The first column reports the results of the specification with only the mathematics test scores in the baseline survey. The top panel (called "Pooled") reports the coefficient of the treatment dummy (i.e., the dummy variable indicating the assignment of vouchers). Although the sign of the coefficient was negative, it was insignificant (standard error robust to heteroskedasticity is reported in parenthesis). The insignificant coefficient was confirmed by a fully non-parametric test.[14] The third row of the first column reports the number of coefficients of 200 replicated estimates larger than the original coefficient (-0.1645). This number indicates that 97 of 200 coefficient estimates are larger than the original one, implying that the p-value for the two-sided test is 0.98.

In the second column, we report the estimation results of the regression model with the female dummy as control variable. Similarly, the third column reports the estimation results of the regression model with grade dummies in addition to the female dummy as control variables. We observed similar results as those in the first column: the sign of the coefficient is negative, but statistically insignificant.

The middle panel (labeled "High School") reports the same set of results using only the high school student sample. The coefficients are positive and statistically significant in all three specifications. The coefficients are from 3.15 to 3.49, indicating that the assignment of the voucher improved mathematics test scores in the standard deviation unit by an amount ranging from 0.315

[14] The details of the fully non-parametric permutation test are presented in the Appendix.

to 0.349. These results were robust in the non-parametric test: the coefficients are statistically significant at least at the 10 percent significance level.

The bottom panel (labeled "Middle School") reports the same set of results using only the middle school student sample. The coefficients were negative, but insignificant in all three specifications.

The fourth to sixth columns report the results of the model using the mathematics test scores in the third survey as the dependent variables. Regardless of the specification and education level, the coefficients were all negative and not statistically significant.

The seventh to ninth columns report the results of the model using the test scores of Japanese language art in the second survey as the dependent variables. In the top panel, the sign of the coefficients was positive, but insignificant. For the sub-sample of high school students reported in the middle panel, the coefficients were larger than those for the pooled sample. In the second specification, the coefficient was statistically significant at the 10 percent level. However, the results of the non-parametric permutation test indicate that this coefficient was insignificant. For middle school students, the coefficients were all negative, but insignificant.

The tenth to twelfth columns report the results of the model using the Japanese language test score in the third survey as the dependent variable. In all specifications, the coefficients were positive and statistically significant. With the pooled sample, the coefficients ranged from 7.6 to 7.7, indicating that the assignment of vouchers improved Japanese language art test scores in the standard deviation unit by an amount ranging from 0.76 to 0.77, conditional on the language test scores in the baseline survey. These results were supported by the non-parametric permutation tests.

Looking at the estimated effects by school level, the coefficients with middle school students are larger than those with high school students. The effects for high school students range from 5.5 to 6.6 and were statistically significant at least at the 10 percent significance level. However, these coefficients were insignificant in the permutation tests. In contrast, the coefficients for middle school students range from 8.4 to 8.6 and were statistically significant both in the asymptotic t-test and non-parametric permutation tests.

In summary, we found a positive effect of the assignment of vouchers on the mathematics test scores of high school students immediately (in the second survey), but this effect disappeared in one year. We also found positive and statistically significant effects of vouchers on Japanese language art test scores in one year, and a strong positive effect for middle school students.

5.2 Non-Cognitive Outcomes

Table 4 summarizes the estimation results of the effects of the assignment of educational vouchers on children's non-cognitive skill outcomes (quality of life score and self-esteem score).

[Table 4]

The first three columns report the effect of the assignment of vouchers on the self-esteem score in the second survey, conditional on the self-esteem score in the baseline survey. The top panel ("Pooled") reports the results of the middle and high school student samples. The coefficients were all negative and statistically significant at the 10 percent level in the first and third specifications. The results of the non-parametric permutation tests indicate that these coefficients were statistically significant at least at the 10 percent level.

The results using the high school student sample reported in the middle panel show that the treatment effect was negative, but insignificant. However, the results with middle school students reported in the bottom panel show that the assignment of vouchers had a negative and statistically significant effect on middle school students. These results were strongly supported by the non-parametric permutation tests.

The fourth to sixth columns report the results of the model using self-esteem scores in the third survey as the dependent variable. Though the sign of the coefficients was negative everywhere, the magnitude was much smaller than those in the model for responses in the second survey reported in the first three columns. As a result, none of these negative coefficients were statistically significant at the conventional significance level. These results combined with the previous ones indicate that the immediate negative impact of the assignment of vouchers on self-esteem scores became weak in one year.

The seventh to ninth columns report the effect of the assignment of vouchers on the quality of life score conditional on the quality of life score in the second survey. The interpretation of the coefficients for the quality of life equations differs from that for the self-esteem equations, because the regression equations for the quality of life score include the quality of life score in the second survey, which was potentially affected by the assignment of vouchers. The sign of all coefficients was positive, but they were not statistically significant. This may be because the quality of life score in the second survey was positively affected and thus, the value added from the second to third surveys was not large enough to ensure statistical significance. Alternatively, if the assignment of vouchers adversely affected the quality of life score in the second survey, these positive coefficients imply gradual recovery of the quality of life measure.

In summary, the assignment of vouchers negatively affected the self-esteem scores of middle school students immediately, but this relationship became small and insignificant in one year. We found that the assignment of vouchers was positively related to the quality of life measure, but these relationships were insignificant.

5.3 Study Time

Table 5 summarizes the estimation results of the effects of receiving educational vouchers on the children's study time.

[Table 5]

The first three columns report the effect of the assignment of vouchers on the total study time reported in the third survey. The top panel ("Pooled") reports the results of the middle and high school student sample. All coefficients were positive and statistically significant at the 10 percent level in the second specification. However, the results of the non-parametric permutation test indicated that this coefficient was insignificant.

The results for the high school student sample reported in the middle panel show that the coefficient was positive and much larger than that for the pooled sample, but none were statistically significant. The results with middle school students are reported in the bottom panel. The coefficient was positive, but much smaller than for the pooled sample. Again, none of these estimates were statistically significant.

The fourth to sixth columns report the results of the model with study time at home in the third survey as the dependent variable. The coefficients were positive and statistically significant in the asymptotic t-test, but these results were not supported by the non-parametric permutation test. We obtained similar conclusions for the results for high school students in the middle panel and middle school students in the bottom panel.

Finally, the seventh to ninth columns report the effect of the assignment of vouchers on study time out of home. We did not find any statistically significant effect of the assignment of vouchers on study time out of home.

In summary, the assignment of vouchers had no statistically significant effect on the study time of middle and high school students. These results imply that receiving vouchers improved the quality of the study environment, resulting in better academic outcomes.

6. Conclusions

In this study, we estimated the causal impact of vouchers for after-school education programs on children's academic and behavioral outcomes using experimental data on middle and high school students. We designed an experiment in which after-school education vouchers were randomly assigned to a portion of the participating children who suffered from the Great East Japan Earthquake. We estimated the value-added models of various outcomes of children such as academic standardized test scores (mathematics and Japanese language art) and non-cognitive skill measures (self-esteem and quality of life).

We found that the assignment of vouchers had a positive and significant effect on the increase of the mathematics test scores of high school students immediately, and language art test scores of both middle and high school students in one year. These results were robust in the fully non-parametric permutation tests. Unexpectedly, we found that the assignment of after-school education vouchers negatively affected the self-esteem scores of middle school students immediately, but this relationship became small and insignificant in one year. We found that the assignment of vouchers was positively related to the quality of life measure, but these relationships were insignificant.

To uncover the mechanism of these effects on children's outcomes, we further estimated the effect of vouchers on the study time of children. The assignment of vouchers had a positive and statistically significant effect on study time, mainly at home. This finding indicated that the assignment of vouchers improved the study environment of these children, resulting in better academic outcomes.

The provision of after-school education vouchers may negatively influence the self-esteem of recipients through at least two possible pathways. The first is the stigma of receiving external assistance. Previous research found that receiving public assistance tends to reduce the level of self-esteem of recipients (Sutton et al., 2014, Elliott, 1996). However, the CFC education voucher was provided to children affected by a natural disaster for which they were not responsible, and none were on welfare before the Great East Japan Earthquake. Therefore, it is unlikely that this explanation can explain the negative effect of the voucher on self-esteem.

The second is a change in the peer group after starting the after-school program. Previous socio-psychological studies found that the level of individuals' self-esteem tended to decrease if they felt they were at a lower rank in the hierarchy among their peers, especially when the comparison was externally forced.[15] In our study, all the children were forced into an extremely difficult situation by the earthquake and tsunami, and some of the treated children felt they were forcedly exposed to new peers with better educational achievement, because of the education voucher they did not expect to receive. If this negative effect overweighed other potentially positive effects, the self-esteem level of the treatment group could be lower than that of the control group.

The negative effect of education vouchers on self-esteem may explain why we failed to find consistent and significantly positive effects of the voucher on academic outcomes. Researchers including Heckman (2006) and Cunha et al. (2010) increasingly recognize the importance of developing non-cognitive skills at an early stage of life to foster the development of cognitive skills. Possibly, the negative effect of the after-school education voucher on non-cognitive outcomes hindered the potential short-term positive effect on cognitive outcomes in the estimates reported in Section 5.

One important policy implication from our interpretation of the evidence here is that the effectiveness of out-of-school educational activities for middle and high school students on academic outcomes can depend on the appropriate control of non-cognitive skills, especially in the short term. Policies that try to support after-school educational programs should be carefully

---

[15] Brown and Lohr (1987) correlated the self-esteem of 7th and 12th graders with their position in the peer group status hierarchy. Vogel et al. (2014) found that when using the social network sites, self-esteem decreased when exposure to others' profiles with higher statuses was forced.

designed so that participation therein may not immediately damage non-cognitive outcomes including self-esteem.

Certainly, more analyses using an experimental research design with a larger sample size and in different settings are necessary to generalize our findings for definitive policy suggestions, which are left for future research.

Appendix

In the appendix, we explain the fully non-parametric permutation test used in the robustness check of the estimates. We used a non-parametric test for the difference in statistics between two groups (Good, 2006, Davison and Hinkley, 1997). This test is a type of resampling method that is robust to extremely small samples, and does not need to rely on the underlying distribution or large sample properties of test statistics, unlike the traditional t-test. This test has recently become increasingly popular in evaluating randomized social experiments with small sample sizes (Campbell et al., 2014). We followed the procedure in Tang et al. (2009), which proposed the permutation method in the presence of complex longitudinal structures such as covariate adjustment and attrition weights.

The null hypothesis is that there is no treatment effect. Suppose we have n_1 observations in the treatment group and n_2 observations in the control group, where $n_1 + n_2$ corresponds to the original sample size. The procedure by Tang et al. (2009) consists of the following three steps:

Step 1: Fit the regression model using the original sample with appropriate attrition weights computed by a model of attrition under certain model selection criteria.

Step 2: Randomly re-assign n_1 observations to the treatment group and n_2 observations to the control group, and re-derive the attrition weights computed by the same method and criteria as in Step 1. Re-fit the same regression model as in Step 1 using the permuted sample with the new attrition weights. Repeat this step N times.

Step 3: Construct the empirical distribution of the statistics (i.e., treatment effect) based on the results with N-times permutated samples. Under the null hypothesis of no effect, the original test statistics should not be an extreme value in the distribution of test statistics with permutated samples. Hence, the empirical p-value determines whether the null hypothesis is rejected.

In our analysis, using the original randomized samples, we computed the inverse probability weight for attrition bias correction based on the Akaike model selection criteria. Then, we estimated the treatment effect (call it β_o) using the IPW method. For the estimates with the permutated samples, we first computed the R sets of IPW for attrition bias correction with the permutated samples, where R was the number of permuted samples. We denoted the IPW computed with the i-th permutated sample by IPW_i. We then estimated the treatment effect (called β_i) using each set of IPW_i. Under the null hypothesis with no treatment effect, β_i should have a zero mean. To test whether the null hypothesis was supported, we counted the number of permutation samples, called M, where $\beta_i \geq \beta_o$. Finally, we tested the null hypothesis that there was no treatment effect using the empirical p-value obtained by $p = (M+1)/(R+1)$ (Davison and Hinkley, 1997). In our application, R=200.

References

Akabayashi, H. and H. Araki (2011) "Do education vouchers prevent dropout at private high schools? Evidence from Japanese policy changes," Journal of The Japanese and International Economies, 25: 183–198

Banerjee, A. V., B. Cole, E. Duflo and L. Linden (2007) "Remedying Education: Evidence from Two Randomized Experiments in India," The Quarterly Journal of Economics, 122 (3): 1235-1264.

Bray, M., and Kwok, P. (2003) "Demand for private supplementary tutoring: Conceptual considerations, and socio-economic patterns in Hong Kong," Economics of Education Review, 22(6), 611–620.

Bray, M., Zhan, S., Lykins, C., Wang, D. and O. Kwo (2014) "Differentiated demand for private supplementary tutoring: Patterns and implications in Hong Kong secondary education," Economics of Education Review, 38, 24-37.

Briggs, D. C. (2001) "The Effect of Admissions Test Preparation: Evidence from NELS: 88." Chance 14(1):10-18.

Brown, B. B., and Lohr, M. J. (1987). Peer-group affiliation and adolescent self-esteem: An integration of ego-identity and symbolic-interaction theories. Journal of Personality and Social Psychology, 52(1), 47-55.

Byun, S. Y. (2014) "Shadow education and academic success in Republic of Korea," In H. Park & K. K. Kim (Eds.), Korean education in changing economic and demographic contexts. Dordrecht: Springer.

Campbell, F., Conti, G., Heckman, J., Moon, S. H., Pinto, R., Pungello, E. and Y. Pan, (2014), "Early Childhood Investments Substantially Boost Adult Health," Science 343, 1478-1485.

Cunha, F., Heckman, J. J., and Schennach, S. M. (2010). "Estimating the technology of cognitive and noncognitive skill formation." Econometrica, 78(3), 883-931.

Dang, H. (2007) "The Determinants and Impact of Private Tutoring Classes in Vietnam," Economics of Education Review, 26(6), 683-698.

Dang, H. and H. Rogers (2008) "The growing phenomenon of private tutoring: does it deepen human capital, widen inequalities, or waste resources?" The World Bank research observer, 23(2), 161-200.

Davison, A.C., and D.V. Hinkley (1997) Bootstrap methods and their application, Cambridge University Press.

Elliott, M. (1996). “Impact of Work, Family, and Welfare Receipt on Women's Self-Esteem in Young Adulthood.” Social Psychology Quarterly 59(1), 80-95.

Fitzgerald, J., Gottschalk, P. and R. Moffitt (1998) “An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics,” Journal of Human Resources, 33(2), 251-299.

Good, P. (2006) Resampling Methods: A Practical Guide to Data Analysis, Birkhauser Basel.

Heckman, J. J. (2006). “Skill Formation and the Economics of Investing in Disadvantaged Children.” Science 312, 1900.

Jacob, B. and L. Lefgren (2004) “Remedial Education and Student Achievement: A Regression Discontinuity Analysis.” Review of Economics and Statistics. LXXXVI (1): 226-244.

Kobayashi, Y. (2018) “The Effect of Shadow Education Vouchers after the Great East Japan Earthquake: Evidence from regression discontinuity design,” RIETI Discussion Paper Series 18-E-031.

Lavy, V. and A. Schlosser (2005) “Targeted Remedial Education for Underperforming Teenagers: Costs and Benefits,” Journal of Labor Economics, 23(4), 839-874.

Lee, Y. J. (2013) “Private tutoring and its impact on students' academic achievement, formal schooling, and educational inequality in Korea, (Unpublished doctoral dissertation)” Columbia University.

Liu, J. (2012) “Does cram schooling matter? Who goes to cram schools? Evidence from Taiwan,” International Journal of Educational Development, 32(1), 46–52.

Matiashvili, A., and Kutateladze, N. (2006). Georgia. In I. Silova, V. Bu¯diene, & M. Bray (Eds.), Education in a hidden marketplace: Monitoring of private tutoring (pp. 191–210). New York: Open Society Institute.

Primont, D., and B. Domazlicky (2006) “Student achievement and efficiency in Missouri schools and the No Child Left Behind Act,” Economics of Education Review, 25(1), 77–90.

Ravens-Sieberer, U. and M. Bullinger (1998). Assessing health related quality of life in chronically ill children with the German KINDL: first psychometric and content-analytical results, Quality of Life Research, 4(7).

Rosenberg, M. (1965) Society and the adolescent self-image. Princeton, NJ: Princeton University Press.

Sutton, E., Pemberton, S., Fahmy, E., and Y. Tamiya (2014) "Stigma, Shame and the Experience of Poverty in Japan and the United Kingdom," Social Policy and Society, 13(1), 143-154.

Tang, L., Duan, N., Klap, R., Asarnow, J. R. and Belin, T. R. (2009) "Applying permutation tests with adjustment for covariates and attrition weights to randomized trials of health-services interventions," Statist. Med., 28: 65-74.

Vogel, E. A., Rose, J. P., Roberts, L. R., & Eckles, K. (2014). "Social comparison, social media, and self-esteem." Psychology of Popular Media Culture, 3(4), 206-222.

Wooldridge, J. (2010). Econometric Analysis of Cross Section and Panel Data. 2nd ed. MIT Press.

Zhang, Y. (2013) "Does private tutoring improve students' national college entrance exam performance? A case study from Jinan, China." Economics of Education Review, 32(1), 1–28.

Zimmer, R., Hamilton, L., and R. Christina (2010) "After-school tutoring in the context of no child left Behind: Effectiveness of two programs in the Pittsburgh Public Schools," Economics of Education Review, 29(1), 18–28.

Figure 1: Timeline of CFC Voucher Program and Surveys

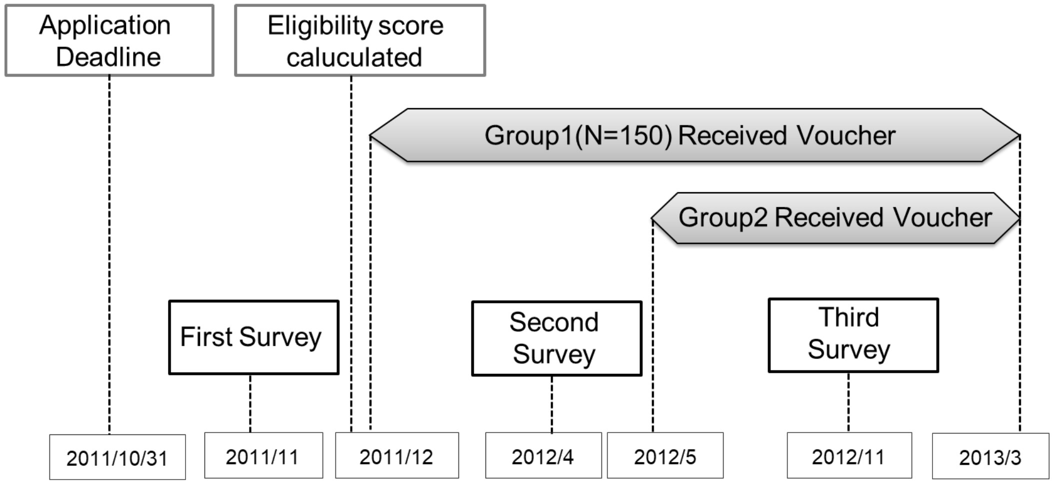

Table 1: Response Rate of Survey among the Randomized Sample

	Total	Recipients	Non-Recipients
First Survey : 7th-12th Grade Students	78/113 : 69.91%		
Second Survey : 7th-11th Grade Students	53/100 : 54.00%	25/37 : 67.57%	28/63 : 44.44%
Among 1st Survey Respondents	43/70 : 61.42%	21/27 : 77.78%	22/43 : 51.16%
Third Survey : 7th-11th Grade Students	52/100 : 52.00%	42/67 : 62.69%	10/33 : 30.30%
Among 1st Survey Respondents	40/70 : 57.14%	34/50 : 68.00%	6/20 : 30.00%

Table 2: Summary Statistics

Pooled	All			Recipients			Non-Recipients			Mean difference test	
Variable	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	P-value	
Math score 1st	41	50.91	11.14	20	49.43	11.38	21	52.31	10.99	0.41	
Math score 2nd	41	49.46	11.63	20	48.34	11.65	21	50.53	11.78	0.55	
Math score 3rd	33	51.33	10.70	15	49.81	11.03	18	52.60	10.56	0.46	
Japanese score 1st	41	49.98	9.67	20	46.89	8.30	21	52.91	10.16	0.05	**
Japanese score 2nd	41	48.63	11.32	20	46.07	9.71	21	51.07	12.41	0.16	
Japanese score 3rd	33	51.24	10.75	15	46.01	8.20	18	55.60	10.86	0.01	***
Self esteem Index 1st	41	16.73	4.60	20	16.60	3.75	21	16.86	5.38	0.86	
Self esteem Index 2nd	41	16.29	5.26	20	17.40	5.39	21	15.24	5.04	0.19	
Self esteem Index 3rd	33	16.61	3.66	15	17.27	3.45	18	16.06	3.83	0.35	
QOL Index 2nd	39	62.17	10.13	19	63.84	10.74	20	60.57	9.51	0.32	
QOL Index 3rd	33	63.35	10.37	15	65.83	10.48	18	61.28	10.09	0.21	
Female	41	0.44	0.50	20	0.40	0.50	21	0.48	0.51	0.63	
Mother's Education	34	12.56	1.42	16	12.63	1.41	18	12.50	1.47	0.80	
Father's Education	33	13.27	1.97	15	13.73	1.83	18	12.89	2.05	0.23	
Household Income	37	531.08	257.15	17	547.06	306.43	20	517.50	214.00	0.73	
No Prior Tutoring	41	0.39	0.49	20	0.40	0.50	21	0.38	0.50	0.90	

High School	All			Recipients			Non-Recipients			Mean difference test	
Variable	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	P-value	
Math score 1st	12	47.55	13.46	7	46.28	12.18	5	49.33	16.41	0.58	
Math score 2nd	12	46.61	12.76	7	44.17	11.37	5	50.02	15.13	0.98	
Math score 3rd	10	50.21	11.75	6	48.70	10.74	4	52.48	14.50	0.65	
Japanese score 1st	12	49.14	10.79	7	47.40	9.07	5	51.58	13.56	0.05	*
Japanese score 2nd	12	49.04	12.63	7	44.63	9.05	5	55.22	15.31	0.48	
Japanese score 3rd	10	49.53	10.55	6	46.85	8.61	4	53.56	13.21	0.02	**
Self esteem Index 1st	12	15.83	5.47	7	17.14	4.67	5	14.00	6.52	0.37	
Self esteem Index 2nd	12	12.92	7.46	7	14.29	7.95	5	11.00	7.11	0.04	**
Self esteem Index 3rd	10	16.50	4.62	6	17.00	4.77	4	15.75	4.99	0.36	
QOL Index 2nd	11	60.70	11.26	7	61.90	14.00	4	58.59	4.61	0.31	
QOL Index 3rd	10	62.08	9.10	6	64.06	9.54	4	59.11	8.77	0.29	
Female	12	0.50	0.52	7	0.57	0.53	5	0.40	0.55	0.31	
Mother's Education	8	12.25	0.71	4	12.00	0.00	4	12.50	1.00	0.60	
Father's Education	8	13.75	1.67	4	14.50	1.91	4	13.00	1.15	0.49	
Household Income	10	632.50	199.32	5	655.00	245.20	5	610.00	167.33	0.89	
No Prior Tutoring	12	0.17	0.39	7	0.29	0.49	5	0.00	0.00	0.84	

Middle School	All			Recipients			Non-Recipients			Mean difference test	
Variable	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	P-value	
Math score 1st	29	52.30	9.96	13	51.13	11.05	16	53.25	9.24	0.72	
Math score 2nd	29	50.64	11.14	13	50.59	11.61	16	50.69	11.14	0.46	
Math score 3rd	23	51.82	10.45	9	50.55	11.80	14	52.64	9.86	0.65	
Japanese score 1st	29	50.32	9.36	13	46.62	8.23	16	53.33	9.37	0.53	
Japanese score 2nd	29	48.46	10.97	13	46.84	10.32	16	49.77	11.64	0.16	
Japanese score 3rd	23	51.98	10.98	9	45.46	8.39	14	56.18	10.60	0.35	
Self esteem Index 1st	29	17.10	4.24	13	16.31	3.33	16	17.75	4.86	0.35	
Self esteem Index 2nd	29	17.69	3.29	13	19.08	2.40	16	16.56	3.54	0.48	
Self esteem Index 3rd	23	16.65	3.27	9	17.44	2.55	14	16.14	3.66	0.70	
QOL Index 2nd	28	62.74	9.81	12	64.98	8.84	16	61.07	10.44	0.66	
QOL Index 3rd	23	63.90	11.02	9	67.01	11.46	14	61.90	10.66	0.43	
Female	29	0.41	0.50	13	0.31	0.48	16	0.50	0.52	0.60	
Mother's Education	26	12.65	1.57	12	12.83	1.59	14	12.50	1.61	0.36	
Father's Education	25	13.12	2.07	11	13.45	1.81	14	12.86	2.28	0.23	
Household Income	27	493.52	269.04	12	502.08	327.43	15	486.67	223.78	0.74	
No Prior Tutoring	29	0.48	0.51	13	0.46	0.52	16	0.50	0.52	0.23	

Table3: Effects on Test Scores

	Mathematics (2)			Mathematics (3)			Japanese (2)			Japanese (3)		
Pooled												
Treatment Effect	-0.1645	-0.2054	-0.2504	-2.8780	-3.2305	-3.4020	0.7364	0.6245	0.2561	7.6895***	7.6978***	7.5805**
(Robust Std. Error)	(2.2399)	(2.2742)	(2.0238)	(2.2711)	(2.2245)	(2.3567)	(4.2370)	(4.4012)	(4.5952)	(2.5412)	(2.6055)	(2.9620)
beta*>=beta (200 replications)	97	99	103	175	180	179	76	81	86	0	0	4
Two-sided p-value	0.98	1.00	0.96	0.24	0.19	0.2	0.77	0.82	0.87	0.01	0.01	0.05
Observations	41	41	41	39	39	39	41	41	41	39	39	39
High School												
Treatment Effect	3.1516***	3.3798***	3.4912***	-1.0508	-0.9258	-1.0287	7.2303	8.2984*	7.9193	6.6071*	6.1616**	5.5019**
(Robust Std. Error)	(0.7957)	(0.8164)	(0.9279)	(2.5431)	(2.5821)	(2.8324)	(4.9526)	(4.3453)	(4.7683)	(3.3868)	(1.9243)	(1.7665)
beta*>=beta (200 replications)	4	4	7	135	123	125	34	24	31	21	11	10
Two-sided p-value	0.05	0.05	0.08	0.64	0.76	0.74	0.35	0.25	0.32	0.22	0.12	0.11
Observations	12	12	12	11	11	11	12	12	12	11	11	11
Middle School												
Treatment Effect	-1.2101	-1.3378	-1.4607	-3.6789	-4.2249	-4.0755	-1.1213	-1.2370	-1.5794	8.4134**	8.3568**	8.5759**
(Robust Std. Error)	(3.0010)	(3.1966)	(2.6958)	(2.9192)	(2.8307)	(2.9556)	(5.7631)	(6.1559)	(6.1315)	(3.3090)	(3.2864)	(3.6730)
beta*>=beta (200 replications)	118	119	139	164	175	175	108	108	118	2	5	5
Two-sided p-value	0.81	0.8	0.6	0.35	0.24	0.24	0.91	0.91	0.81	0.03	0.06	0.06
Observations	29	29	29	28	28	28	29	29	29	28	28	28
Specifications												
Test score (1)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Female dummy	no	yes	yes	no	yes	yes	no	yes	yes	no	yes	yes
Grade dummy	no	no	yes	no	no	yes	no	no	yes	no	no	yes

* p<0.1, ** p<0.05, *** p<0.001. Standard errors robust to heterosckedasticity are reported in parenthesis.
All specifications include test score of the same subject before provision of vouchers.
Coefficients are estimated using the inverse probability of sample attrition as weights.

Table4: Effects on Non-Cognitive Outcomes

	Self Esteem (2)			Self Esteem (3)			Quality of Life		
Pooled									
Treatment Effect	-2.6798*	-2.3956	-2.6904*	-1.2078	-0.7797	-0.5822	2.8745	3.3695	3.2913
(Robust Std. Error)	(1.3674)	(1.6147)	(1.4607)	(1.3183)	(1.2009)	(1.2056)	(3.6352)	(3.2643)	(3.6354)
beta*>=beta (200 replications)	191	187	194	163	133	129	43	40	44
Two-sided p-value	0.08	0.12	0.05	0.36	0.66	0.70	0.44	0.41	0.45
Observations	41	41	41	38	38	38	32	32	32
High School									
Treatment Effect	-1.4723	-2.3078	-1.8974	-0.8605	-0.7997	-0.4598	3.7094	3.5639	3.3275
(Robust Std. Error)	(3.9063)	(4.1076)	(4.5671)	(3.5049)	(3.6042)	(4.1121)	(3.8676)	(2.9700)	(3.6358)
beta*>=beta (200 replications)	119	134	124	101	101	95	38	36	45
Two-sided p-value	0.80	0.65	0.75	0.98	0.98	0.96	0.39	0.37	0.46
Observations	12	12	12	10	10	10	10	10	10
Middle School									
Treatment Effect	-3.1850***	-2.9250**	-2.9276**	-1.7980	-1.3551	-1.2671	2.2545	4.2411	4.1905
(Robust Std. Error)	(0.9774)	(1.1969)	(1.2245)	(1.3551)	(1.2099)	(1.2083)	(5.3714)	(5.0519)	(4.8433)
beta*>=beta (200 replications)	199	199	199	176	169	168	73	42	51
Two-sided p-value	0.00	0.00	0.00	0.23	0.30	0.31	0.74	0.43	0.52
Observations	29	29	29	28	28	28	22	22	22
Specifications									
Female dummy	no	yes	yes	no	yes	yes	no	yes	yes
Grade dummy	no	no	yes	no	no	yes	no	no	yes

* p<0.1, ** p<0.05, *** p<0.001. Standard errors robust to heterosckedasticity are reported in parenthesis.
All specifications for self esteem include self-esteem score before the provision of vouchers.
All specifications for Quality of Life include the score of quality of life from the second survey.
Coefficients are estimated using the inverse probability of sample attrition as weights.

Table5: Effects on Behavioral Outcomes

	Total Study Hour			Home Study Hour			Out-of-Home Study Hour		
Pooled									
Treatment Effect	3.7885	4.1293*	4.3336	3.3444**	3.4497**	3.6555**	0.4442	0.6796	0.6781
(Robust Std. Error)	(2.7730)	(2.3059)	(3.1209)	(1.3228)	(1.2583)	(1.6426)	(1.7064)	(1.3501)	(1.7105)
beta*>=beta (200 replications)	45	43	39	22	21	16	73	69	70
Two-sided p-value	0.46	0.44	0.4	0.23	0.22	0.17	0.74	0.7	0.71
Observations	39	39	39	39	39	39	39	39	39
High School									
Treatment Effect	15.8435	14.1027	12.5324	8.2658	7.1386	5.8592	7.5777	6.9641	6.6732
(Robust Std. Error)	(10.9669)	(8.3576)	(8.3931)	(5.8488)	(4.1881)	(4.3501)	(5.6271)	(4.8386)	(4.9352)
beta*>=beta (200 replications)	22	15	30	25	13	22	27	24	39
Two-sided p-value	0.23	0.16	0.31	0.26	0.14	0.23	0.28	0.25	0.4
Observations	11	11	11	11	11	11	11	11	11
Middle School									
Treatment Effect	0.1933	0.9323	1.2186	2.1326	2.4754	2.5924	-1.9393	-1.5431	-1.3738
(Robust Std. Error)	(5.6252)	(3.7144)	(4.0961)	(2.5615)	(1.8265)	(1.9045)	(3.1680)	(2.0604)	(2.3692)
beta*>=beta (200 replications)	85	76	71	37	33	25	126	117	115
Two-sided p-value	0.86	0.77	0.72	0.38	0.34	0.26	0.73	0.82	0.84
Observations	28	28	28	28	28	28	28	28	28
Specifications									
Test score (1)	yes	yes	yes	yes	yes	yes	yes	yes	yes
Female dummy	no	yes	yes	no	yes	yes	no	yes	yes
Grade dummy	no	no	yes	no	no	yes	no	no	yes

* p<0.1, ** p<0.05, *** p<0.001. Standard errors clustered at new residential location are reported in parenthesis.
Coefficients are estimated using the inverse probability of sample attrition as weights.