Teacher Professional Development, Cooperative Learning and Student Performance: Evidence from Two Interventions in Thailand

A Dissertation

Submitted to the National Graduate Institute for Policy Studies (GRIPS)

in Partial Fulfillment of the Requirements for the Degree of

Ph.D. in Development Economics

by

Somsiriwong Piyakul

August 2022

4 QUALITY EDUCATION

Abstract

This dissertation examines at-scale government teacher training programs and student performance. Our sample is drawn from officers of Thailand's Ministry of Education, the Institution for the Promotion of Teaching Science and Technology (IPST) and National Institute of Educational Testing Service (NIETS). This dissertation consists of two chapters. The first chapter evaluates the effects of the Active Learning School (ALS) Project in three provinces in the south of Thailand on national test scores during the period 2017–2019. Instruction in the project combines two interventions, the cooperative learning approach (CL) and professional development (PD), to improve the academic performance of students. Schools in the district were ranked and selected for the project on the basis of their grade six students' scores on the 2016 Ordinary National Education Test scores. The study provides regression discontinuity (RD) results on student performance pooled across years, grades and subjects. The results estimated from subsamples pooled across either subject and grade or subject and year are consistent with the main results. All results suggest that the project did not substantially improve student performance as measured by national test scores, since the effect sizes are very small with the 95 percent confidence interval estimates between -0.1 and 0.1 test score standard deviation. The major factors hindering the success of district-wide interventions were found to be inadequate teacher training, imperfect classroom implementation, and lack of administrative support and follow-up.

In the second chapter, evaluations of the effects of Science, Mathematics and Technology (SMT) promotion on national test scores of students in Thailand during 2018 – 2020 are discussed. In order to mitigate education inequality and improve academic performance of Thai students, the project aims to enhance the skills and teaching style of SMT teachers in participating primary and secondary schools. This study provides difference-in-difference results pooled across grades on science and mathematics performance of 6^{th} and 9^{th} grade students. Overall results suggest that the project did not substantially improve student performance as measured by national test scores. The effect sizes are smaller than 0.05 standard deviation with the 95 percent confidence interval estimates between -0.1 and 0.1 standard deviation in both grades. Nevertheless, the effect sizes of students in top performance schools are larger than 0.05 (0.1) standard deviation on science (mathematics) scores of 6^{th} grade students. According to survey data, the major hindrances preventing the success of the intervention are ineffective online training, difficulty in classroom implementation, and lack of school resources.

Dedication

To my parents, with love and gratitude.

Acknowledgements

This dissertation would have been impossible without the aid and support of many people. First and foremost, I am extremely grateful to my advisor, Professor Stephan Litschig, who supported my interests and gave me the room and freedom to try out different research questions. I could not have undertaken and overcome this journey without his invaluable advice, continuous support, and patience with me and my learning process. I would like to extend my sincere thanks to my committee members, Professor Dainn Wie and Professor Yutaro Izumi, who offered constructive comments and support.

I am also grateful to the Japanese Government and the National Graduate Institute for Policy Studies (GRIPS) for awarding me Monbukagakusho (MEXT) scholarship, providing me with financial means to complete this dissertation and doctoral degree.

Many thanks to Mr. Kitti Suthijitpaisarn, Ms. Waraporn Kaewsikao, and officers of Thailand's Ministry of Education, the Institution for the Promotion of Teaching Science and Technology (IPST) officials and National Institute of Educational Testing Service (NIETS) officials, who provided and supported the access to most of the data analyzed in this dissertation. Special thanks also go to Ms. Chomanard Wongwech, who supported my interests and established a connection with the educational officers from the outset. I would like to acknowledge the assistance of Mr. Lawrence M. Hunter on his constructive comment on the manuscript.

Last but not the least, I am eternally grateful to my parents and siblings for their unconditional love and support. Not to forget my best friends, Kwantawan, Reynalyn, and Pataranard, who always check up on and comfort me. Without their encouragement and support in the past few years, it would have been impossible for me to complete my study.

Table of contents

List of Tables

List of Figures

Appendix Table

Appendix Figures

CHAPTER ONE

Introduction

Many countries attempt to improve quality of education by implementing academic intervention various types, including after-school programs, coaching/ mentoring of students, coaching/mentoring of personnel, computer-assisted instruction, content changes, cooperative learning, summer programs, tutoring, and personnel development (Dietrichson and Jorgenson 2017). Two most common academic interventions are cooperative learning (CL) and teachers' professional development (PD) programs. Cooperative learning encourages students to cooperate and work together in small groups in order to maximize their own and each other's learning (Johnson and Johnson 1999). Teacher professional development usually offers in-service teacher training in the form of workshops or short-term courses, to provide teachers with new information and enhance their experience (Villegas-Reimers 2003).

Many developing countries implement CL and teacher PD programs, aiming to improve student performance. Over the past two decades, a growing literature on impacts of the programs on student academic achievement provides mixed evidence on large-scale interventions from many countries, ranging from successful (Albornoz et al. 2020; Sun and Du 2021; Cilliers et al. 2020) to ineffective (Mbiti I 2016; Loyalka et al. 2019; Abbiati et al. 2021; Schaffner et al. 2021; Carneiro et al. 2022). For instance, Tan et al. (2007) and Demie et al. (2019) find no statistically significant impact of CL intervention in two schools, Thurston et al. (2019) in 10 high schools with 95% confidence interval between -0.15 and 0.10 standard deviation, Tracey et al. (2010) in 34 primary schools, and Loyalka et al. (2019) in 300 schools with 95% confidence interval between -0.053

and 0.074 standard deviation. Other ineffective large scale interventions are, for example, studies, by Schaffner et al. (2021) in Nepal with 95% confidence interval between -0.239 and 0.019 standard deviation, and by Carneiro et al. (2022) in Ecuador with 95% confidence interval between -0.135 and 0.001 standard deviation. On the other hand, Albornoz et al. (2020), for example, reported positive impacts on performance of 7^{th}-grade students in Argentina with 95% confidence interval between 0.305 and 0.795 standard deviation. Cilliers et al. (2020) found positive impacts on performance of primary school students in South Africa with 95% confidence interval between 0.061 and 0.407 standard deviation.

This dissertation examines effects of two at-scale government teacher training programs on student performance. Those two programs are Active Learning School (ALS) project and Science, Mathematics and Technology (SMT) promotion. ALS project combines CL concept with teacher PD program, while SMT promotion mainly applies teacher PD program. Our sample is drawn from officers of Thailand's Ministry of Education, the Institution for the Promotion of Teaching Science and Technology (IPST) and National Institute of Educational Testing Service (NIETS).

There are two main chapters in this dissertation. The first chapter evaluates the effects of the Active Learning School (ALS) Project in three provinces in the south of Thailand on national test scores during the period 2017–2019. Schools in the district were ranked and selected for the project on the basis of their grade six students' scores on the 2016 Ordinary National Education Test scores. The study provides regression discontinuity (RD) results on student performance pooled across years, grades and subjects. The results estimated from subsamples pooled across either subject and grade or subject and year are consistent with the main results. All results suggest that the project did not substantially improve student performance as measured by national test scores, since the effect sizes are very small with the 95 percent confidence interval estimates between -0.1 and 0.1 test score standard deviation. The major factors hindering the success of district-wide interventions were found to be inadequate teacher training, imperfect classroom implementation, and lack of administrative support and follow-up.

In the second chapter, evaluation of the effects of Science, Mathematics and Technology (SMT) promotion on national test scores of students in Thailand during 2018 – 2020 is discussed. This study provides difference-in-difference results pooled across grades on science and mathematics performance of 6^{th} and 9^{th} grade students. Overall results suggest that the project did not substantially improve student performance as measured by national test scores. The effect sizes are smaller than 0.05 standard deviation with the 95 percent confidence interval estimates between -0.1 and 0.1 standard deviation in both grades. Nevertheless, the effect sizes of students in top performance schools are larger than 0.05 (0.1) standard deviation on science (mathematics) scores of 6^{th} grade students. According to survey data, the major hindrances preventing the success of the intervention are ineffective online training, difficulty in classroom implementation, and lack of school resources.

This dissertation makes two major contributions. First, this study provides evidence of a general lack of success from a case involving a large-scale professional development and cooperative learning intervention in Southeast Asia. Therefore, it provides some useful information for countries with similar background and culture, i.e.

for Southeast Asian countries. In addition, an online survey conducted by a third-party provides evidence and possible explanations for ineffectiveness of a project.

This paper is organized as follows. Chapter 2 discusses Effects of Cooperative Learning on Student Performance: Evidence from Southern Thailand introduces the Active Learning School Project in the South of Thailand. Chapter 3 discusses Science, Mathematics and Technology Promotion and Student Performance: Evidence from Thailand. Chapter 4 concludes the research and provides policy implications.

CHAPTER TWO

Effects of Cooperative Learning on Student Performance: Evidence from Southern Thailand

2.1 Introduction

Many countries attempt to carry out progress in the field of education by implementing educational policies in the form of projects. Such project style academic intervention can be of various types, including after-school programs, coaching/mentoring of students, coaching/mentoring of personnel, computer-assisted instruction, content changes, cooperative learning, summer programs, tutoring, and personnel development (Dietrichson and Jorgenson 2017). One of the most common academic interventions is cooperative learning (CL), in which the students are encouraged to cooperate and work together in small groups in order to maximize their own and each other's learning (Johnson and Johnson 1999). This learning method facilitates student learning by providing a group interaction environment and applying active learning teaching techniques. In addition to cooperative learning, both developed and developing countries have made massive educational investments in teachers' professional development (PD) programs, which usually offer in-service teacher training in the form of workshops or short-term courses, to provide teachers with new information and enhance their experience (Villegas-Reimers 2003).

Johnson and Johnson (2018) characterize cooperative learning, in which students work together in small groups to maximize their own and each other's learning, as the foundation of active learning procedures. In 2016, the Active Learning School (ALS) project was launched in the three southmost provinces of Thailand, using a combination of cooperative learning and personnel development approaches to improve the academic achievement of students in schools in the region. Since the region has language, religious, and cultural differences from the rest of Thailand, compounded by local political turmoil in Pattani, Yala and Narathiwas provinces, schools in the south find it difficult to design and deliver classes and activities which are appropriate for improving education outcomes. Despite recent interest in the effect of academic intervention on academic achievement in Thailand, few empirical studies have quantitatively examined the effects of education projects on student performance.

This study examines the effect of the ALS project on the performance of students in the three southern provinces of Thailand. Schools were selected for the project on the basis of their 2016 scores on the grade six Ordinary National Education Test (O-NET). Schools were ranked in ascending order within district and those schools with relatively low average test scores were selected for the project. In other words, schools with average test scores below the cutoffs were designated as the treatment group and those above the cutoffs as the control group. In order to evaluate the effectiveness of the project, individual O-NET scores of grade six students in 2017, 2018 and 2019 were used as main dependent variable of this study. Additionally, Reading Test (RT) scores of grade one students and National Test (NT) scores of grade three students were analyzed to confirm robustness. I compared the educational achievement of students in grades one, three and six in treatment and control schools by computing the mean of individual test scores at school level. For internal validity check, I conducted a balance test of pretreatment covariates. The results provide no evidence of discontinuity in 2016 covariates including number of teachers, number of students, number of classrooms, and student-teacher ratio. Consequently, schools in the same neighborhood are plausibly comparable.

The results provide regression discontinuity (RD) estimates of impact of the project on student performance using individual test scores as a proxy. The total effects estimated from the pooled test scores are insignificantly different from zero with a 95% confidence interval between minus 0.1 and 0.1 test score standard deviation. The results suggest that the project did not significantly improve the performance of students as measured by national test scores, since the effects on the scores were neither highly positive nor negative. I estimated the effect of the program on subsample test scores pooled across either years or grades. The estimates obtained from subsamples are similar to the original estimates, i.e. the total effects; however, I detected some heterogeneity across grades but none across years. The consistency of those subsample estimations allows us to confidently conclude that the impact of the ALS project was not sufficient to warrant the claim that the project improved academic achievement of students in the south of Thailand.

There was limited success of the use of the CL approach, but most of the instances of success were in small scale implementations, i.e. at the classroom or school level, for example, six-classroom intervention (Kramarski et al. 2001), 12-classroom intervention (Kramarski and Mevarech 2003), and single school intervention (Genc 2016). Those studies provide RCT evidence demonstrating positive effects of the CL approach on the academic achievement of elementary and junior high school students. On the flip side, some RCT studies failed to find evidence that cooperative learning gives rise to significant improvement when implemented on a large scale, i.e. at the multiple-school to district level. For instance, Tan et al. (2007) and Demie et al. (2019) find no statistically significant impact of CL intervention in two schools, Thurston et al. (2019) in 10 high schools with 95% confidence interval between -0.15 and 0.10 standard deviation, Tracey et al. (2010) in 34 primary schools, and Loyalka et al. (2019) in 300 schools with 95% confidence interval between -0.053 and 0.074 standard deviation. The explanation for the success of small scale interventions could be that the changes required for these interventions were made perfectly in the classrooms, and therefore increased the probability of improvement in students' academic achievement.

Regarding the effect of teacher PD on student academic achievement, Jacob and Lefgren (2004) provide fuzzy regression discontinuity evidence suggesting that staff development did not significantly enhance student performance in high-poverty schools. Likewise, using a randomized control trial (RCT), Loyalka et al. (2019) observe that neither teacher PD alone nor teacher PD with follow-up and/or evaluation significantly improved academic achievement of Chinese students both in short term and a longer term. Since they investigated the effects of PD intervention in a large number of schools (461 and 300, respectively), the studies lend support to the notion that large scale intervention is ineffective. Recent studies provide mixed evidence on interventions from many countries, ranging from successful to ineffective. Albornoz et al. (2020), for example, reported positive impacts on performance of 7th-grade students in Argentina with 95% confidence interval between 0.305 and 0.795 standard deviation. Cilliers et al. (2020) found positive impacts on performance of primary school students in South Africa with 95% confidence interval between 0.061 and 0.407 standard deviation. Some ineffective large scale interventions are, for example, studies of Loyalka et al. (2019) in China with 95% confidence interval between -0.053 and 0.074 standard deviation, Schaffner et al. (2021) in Nepal with 95% confidence interval between -0.239 and 0.019 standard deviation, and Carneiro et al. (2022) in Ecuador with 95% confidence interval between -0.135 and 0.001 standard deviation.

This study makes three major contributions. First, many studies have identified a general lack of success of large-scale academic intervention in South Africa (Carneiro et al, 2022), Central Asia (Loyalka et al., 2019), and in South Asia (Schaffner et al, 2021), this study provides RD evidence of a general lack of success from a case of a large scale professional development and cooperative learning intervention in Southeast Asia, involving about 171,000 students from 662 schools in 3 provinces in Thailand. Countries in Southeast Asia have implemented similar academic intervention. By studying teacher professional development and cooperative learning intervention in Thailand, we can provide policy implication for countries with similar background and culture, i.e. for Southeast Asian countries. Second, this study reports results from multiple-grade test scores, i.e. RT (grade one), NT (grade three), and O-NET (grade six), while other similar studies reported only academic performance on either one or two subjects of students in the same grade. For example, Genc (2016) reports science achievement of sixth-grade students, Demie et al. (2019) report mathematics achievement of ninth-grade students, and Tan et al. (2007) report geography achievement of eighth-grade students. Some large scale interventions affect performance of students of more than a specific grade, for example, the implement of ALS project aimed to improve the performance of all students. Therefore, examining multiple grade test scores could capture more effects and provide solid evidence on whether the project is effective. Finally, this study reports effects, for periods of up to three years, of a large scale professional development and cooperative learning intervention.

Large scale education interventions can be problematic for education policymakers due to the difficulty in implementation. Banerjee et al. (2007) report a significant improvement of 0.28 standard deviation resulting from an intervention by an

NGO in 77 schools in urban India. However, replicating the study on a larger scale (192 schools in eight provinces in Kenya), Bold et al. (2018) find a significant positive effect of 0.19 standard deviations in schools randomly assigned to NGO implementation, and zero effect in schools receiving contract teachers from the Ministry of Education. Banerjee et al. (2017) identify market equilibrium, spillover effects, political reactions, context dependence, randomization or site-selection bias, and piloting bias or implementation as challenges for large scale policy implementation. Regarding the ALS project, it is possible that information spillover from treatment to control schools and imperfect classroom implementation are two main obstacles to successful intervention.

Studies related to CL intervention suggest that, apart from imperfect classroom implementation, possible obstacles could be: inadequate training and practice; lack of familiarity with CL methodology; and unfavorable teacher attitude towards CL (Demie et al. 2019; Thanh et al. 2008). Tracey et al. (2010) suggest that more training and follow-up are required to make CL intervention effective. Although in-service training is a widely used approach in teacher PD, PD is considered a long-term process (Villegas-Reimers 2003). These findings point to a need for education policymakers to give careful consideration to the details of the implementation of large scale policy intervention. In case of Thai ALS project, the in-service training lasted only three days and the follow-up was performed infrequently, at most twice a month. Hence, policymakers might reconsider the training duration necessary to ensure the effectiveness of the teacher training process.

This chapter is organized as follows. Section 2.2 introduces the Active Learning School Project in the South of Thailand. Section 2.3 provides the data sources and some descriptive statistics. Section 2.4 mentions estimation approach. Section 2.5 shows the internal validity and robustness checks. Section 2.6 gives the estimation results. Section 2.7 explains why the program was not effective. The details of the robustness check are in section 2.8. Conclusions and directions for further study are in section 2.9.

2.2 Background

2.2.1 Target Schools

Ministry of Education (MoE) Thailand intended to increase education opportunities nationwide by providing nine years of compulsory education in line with National Education Act B.E. 2542. Any child below age 12 who wishes to study can enroll in a school near their home, since Thailand has nine years of compulsory education, i.e. elementary and junior high school. In order to increase education opportunities, the total public expenditure on education as a percentage of GDP and of total government expenditure has been generous, at an average of 4 and 20 percent respectively, since 1999 (Office of the Education Council, 2017). According to National Statistical Office (NSO) Thailand, MoE efforts have increased average years of schooling for the population aged above 15 from 7.4 years in 2002 to 8.6 years in 2018, whereas for those in the labor force, average years of schooling increased from 8.1 years to 9.6 years in the same period. There has been extensive debate as to whether the quality of education did in fact improve. Since 1999, Thailand has been working to increase educational opportunity nationwide; however, there is no broad agreement that academic achievement in the country has improved. Even though schools evaluate their students' achievement by means of midterm and final exams, those exams are created within school or within district, so there is no uniform evaluation standard across schools.

In order to achieve the same evaluation standard for all students, National Institute of Educational Testing Service (NIETS), Thailand, was established in 2005 as a public organization, nationalizing Thai education measurement and evaluation. The institution annually conducts the Ordinary National Education Test which evaluates students' achievement in Thai, English, mathematics, science and social studies for all students in grades six, nine and twelve. The MoE's Office of Basic Education Commission (OBEC) received consistently low average national test scores with a decrease observed in the southern provinces of Pattani, Yala, and Narathiwas. In 2016, Active Learning School (ALS) project, launched by MoE, encouraged teachers to let students work together on both individual and group assignments during class hours, rather than giving traditional lectures. The objectives were to create a more interactive environment and to promote self-learning. Students in the three southern provinces have performed consistently below the national average in terms of national test scores, mainly due to differences in language, religion, and culture, compounded by political turmoil. The language used in that region is Bahasa Melayu, reflecting the influence of neighboring Malaysia, even though Thai is the official language. Additionally, according to the 2000 census, NSO Thailand reported that 93.83% of the national population practices Buddhism, followed by 4.56% that practices Islam, 0.8% Christianity, and 0.81% others. However, Islam is the predominant religion in the three southern provinces, where roughly 80% of the population practices Islam. Students in those provinces face considerable barriers to educational achievement since everybody in their community speaks Bahasa Melayu although Thai is the language of instruction in schools. Moreover, political turmoil in the region affects study timetables: classes are often cancelled due to the severity of the turmoil.

There are two ALS projects in the country; one for schools in the south and one for the rest of the country. In both ALS projects, participating schools are provided with materials to support teachers in creating group- and self-learning activities. In the south, ALS staff also provide training sessions for school principals and teachers of ALS schools. The training emphasizes teaching methods for group- and self-learning activities in classrooms. The training is practical; it provides various activities which emphasize critical thinking skills. Practical group activities are encouraged since they allow students to learn from their experience.

The national tests are conducted annually in March by NIETS. In April 2017, the Ordinary National Education Test (O-NET) scores of grade six students in the academic year 2016 were used to select 158 schools for the project based on their performances. Each district was allocated a certain number of schools that could be selected for the ALS project. Schools were ranked by their 2016 average test score in ascending order within the district, and local authorities selected schools with relatively low scores for the project. Although schools knew about the program before their students took the 2016 O-NET test, they could not know the location of the cutoffs as these were determined based on the test itself. In April 2018, the 2016 O-NET scores were used to select another 192 schools into the project using the same approach. The project finally has 350 participant schools from both phases. This selection process created a group of schools just to the right of a given cutoff that had a very similar performance but were not selected for the ALS project. I exploit the fact that treatment and control schools were separated based on cutoffs in each district and estimate the effects of the ALS Project using Regression Discontinuity Design (RDD)

As can be seen in Figure 2.1, about 33% of the schools included in the project were divided into two phases, 45.14% in phase I and 54.86% in phase II. The project information reveals that only 56% of treatment schools registered for the second training sessions, i.e. only 819 teachers from 196 schools attended the sessions. Non-trained schools accounted for 46% of treatment schools; these schools only had access to active learning materials. Trained teachers, representing less than 5% of the teachers in the south, i.e. 819 out of 17,594 teachers, were expected to transmit the content to their colleagues so that the concept of active learning is applied widely in schools. In practice, project supervisors in each ESAO monitor and encourage teachers to implement ALS teaching methods. However, there is no clear data on either how effective the knowledge transfers among teachers in the same school were or how the trained teachers adjusted their teaching techniques. Regarding teacher incentives, teachers are likely to be promoted to higher academic positions by school principals if they can provide evidence of classroom implementation of active learning concept and its effect on improvement of student performance. The higher academic position ensures a higher base salary; nevertheless, the floor and ceiling of the salary are based on both academic position and seniority. This study seeks to determine whether Active Learning Schools project did in fact substantially contribute to the improvement of student performance as measured by the national test scores.

2.2.2 Project Timeline and Contents

The project was operated in the following steps. First, address the problem and identify teacher teaching requirement. Second, develop teacher professional development program to meet pre-standard. Third, implement the project. Finally, evaluate and follow up the project. The project timeline is presented in Figure 2.2. In January 2017, the ALS authorities did a baseline survey to collect information about schools in the south. The survey focused on problems reported by teachers and their requests. From a total of 17,594 teachers from nine Education Service Area Offices (ESAO), 392 teachers were randomly selected to respond to the survey. To ensure that the information reflects the situation in that area accurately, the number of teachers selected from each ESAO is in proportion to the total number of teachers in each ESAO. The teachers were asked to complete a questionnaire, responding to each question on a 5-point Likert scale, where five points indicated a serious problem and one point a least problematic issue. The teachers reported some difficulty with learning management processes including preparation, creation of learning activities, and evaluation, and average responses (standard deviation) were 4.52 (0.55), 4.53 (0.53) and 4.52 (0.58), respectively.

After the authorities analyzed the problems and requests, they designed an intensive course for teachers in the ALS schools. The concept of active learning, identified as a solution to the reported problems, was introduced and developed as a teaching model for the schools. The pilot training program was implemented in June, the beginning of the academic year 2017, with a small group of 10 teachers who gave feedback and comments. Development of the program continued until the program met the pre-set standard.

The contents of the training program include King's Philosophy [1], brain development concept, active learning concept, graphic organizer, and academic assessment. In order to improve academic achievement of students, the program requested teacher's cooperation to utilize technology, prioritize practice lessons, encourage idea sharing and question asking in class, assign group work, use graphic organizer, and do formative assessment. Moreover, the highlights of the program were to guide students to think systematically and to apply classroom knowledge to their real lives. Regarding the learning models, the project emphasized experiential learning, problem based learning, project based learning, and thinking based learning. The program design aimed to identify problems and intensify active learning and brain-based learning concepts to Thai, English, mathematics, and science teachers (Figure 2A-3 in Appendix). Therefore, the program introduced some brain-based learning tools that teachers can use in classrooms, including mind mapping, concept map, web, Venn diagram, circle diagram, fish bone diagram, table, ranking ladder, pie chart, bar chart, cycle map and diagram (Figure 2A-4 in Appendix).

The first training sessions were held in an intensive 1-day training in November 2017. Unfortunately, the conduct of the project was transferred to another team from a different department, and the training information including the number of schools and trained teachers was not transmitted. The new team proceeded with the plan and held the second training session in July 2018, after the training curriculum had been developed and finalized. A total of 1,036 professionals, consisting of 21 project supervisors, 196

[1] King's Philosophy emphasizes the concept of "process comprehension and development". The concept encourages us to research and analyze data before we take actions to create new innovation. (Tanguthai, 2018)

principals and 819 teachers from 196 schools in 9 ESAOs, attended the sessions. Schools were requested to send at least one teacher from each subject, namely Thai, English, mathematics and science; those teachers were mostly selected by school principals.

The teachers were divided into four batches and each batch attended the sessions for 3 consecutive days; the training timetable is shown in Table 2.1. A pretest on active learning knowledge and activities was administered before the sessions started. After the testing, the participants were divided into small groups of 5-7 persons and engaged in practical group activities to develop their active-learning-concept teaching skills. On the last day of the sessions, the subjects took the same test again. The test results indicate that the teachers acquired the knowledge and gained necessary experience from the sessions: their average scores (standard deviation) on the pretest and post-test were 16.2 (1.92) and 25.82 (1.56), respectively, out of 30 points in total.

After the training sessions, 65 supervisors from nine ESAOs were assigned to coach the participants and monitor the project. To evaluate the sessions, a few months later, local supervisors conducted a survey and distributed the questionnaire, asking about their satisfaction of the sessions, to all 819 trained teachers. The supervisors were able to collect 598 completed questionnaires which reported an average of 3.48 (0.35), 3.77 (0.42), and 3.19 (0.61) points (standard deviation) out of five for lessons on teaching preparation, active learning curriculum and after class evaluation, respectively.

2.3 Data

This study analyzes data from educational data sources for the period 2016 to 2019. The main data analyzed here are the national test scores from the National Institute of Educational Testing Service (NIETS), Thailand, i.e. the Ordinary National Education Test (O-NET) scores for grade six students from 2016 to 2019, the National Test (NT) scores for grade three students from 2017 to 2019 and the Reading Test (RT) scores for grade one students from 2018 to 2019. Every academic year, the O-NET tests includes Thai, English, mathematics and science tests whereas the RT tests speaking and reading. The NT tested literacy, numeracy and logic in 2017 and 2018; however, the tests were modified and only literacy and numeracy were tested from 2019 onward. The descriptive statistics can be found in Table 2.3.

The 2016 O-NET scores were used to select schools for inclusion in the project. After the scores were announced, the local authorities in each district ranked the schools in ascending order according to their average test scores and then set the cutoffs. The schools whose averages were below the cutoffs were to be included in the project; those above the cutoffs were not to be included. The project involved 38 districts in the south of Thailand; however, only data for 26 districts is analyzed in this study due to some limitations in terms of the cutoffs. Some schools below the cutoffs were not included in the project, while some schools above the cutoffs were included. I used individual scores to compute school level average test scores, then normalized the scores by subtracting the district cutoffs. The normalized scores were used as the selection criteria for pooling 662 schools from 26 districts; the threshold value is zero.

Figure 2.3 shows the sample selection process. There are 350 ALS schools out of 1,056 schools in 38 districts in the south (left panel in Figure 2.3). The first batch of 158 schools was selected in April 2017; while the second batch of 192 schools was added in April 2018. I dropped all schools in 10 districts where the district cutoffs were not reported, and additionally dropped schools in two districts where the selection process was not clear. The sample in this study, therefore, includes only 210 ALS schools out of 662 schools in 26 districts in the south (right panel in Figure 2.3).

The probabilities of treatment calculated from the pooled observations reveal the difference in the probability of being an ALS between the groups of schools below and above the cutoff. Figure 2.4 suggests that schools with relatively poor 2016 O-NET scores have a higher probability of being selected as ALS. In the academic year 2017, the probability of being ALS for the lowest cluster equals one, whereas schools above the cutoff have a probability of less than 0.1 of being ALS, on an average. The information visualized in panel A, based on the OLS estimation, indicates that the probability of being ALS schools was significantly higher in schools just below the cutoff when compared to those schools just above the cutoff, 0.432 and 0.065, respectively. When the second batch ALS schools were included in the treatment group, the lowest cluster was definitely included in the project with a probability of one. Nevertheless, the probabilities of being ALS schools, presented in panel B, were higher for both groups of schools, i.e. just above and just below the cutoff. There are evident jumps in both panels; as a result, I exploit the selection cutoffs and use a fuzzy regression discontinuity design to evaluate the effect of ALS on student performance as measured by the national test scores.

The national test scores from 2017 to 2019 are outcome variables in this study. In order to compare the scores across years, I computed the z-scores of students in grades one, three and six from 2017 to 2019. There are 24 tests: 12 tests for grade six – Thai, English, mathematics and science from 2017 to 2019; eight tests for grade three – literacy, and numeracy from 2017 to 2019 and logic for 2017 and 2018; and four tests for grade one – speaking and reading in 2018 and 2019. Then, I computed school average z-scores by subject by year from individual z-scores and created a large panel data of school average z-scores of 24 tests from 2017 to 2019. The panel data were used to estimate the results, both overall and subgroup analysis of the project.[2] Those subgroup analyses were for the groups pooled across years – 2017, 2018, and 2019 – and those pooled across grades – one, three, and six. In addition to the national test scores, the Ministry of Education provided important school information including the number of students, number of teachers and number of classrooms. That information was used to verify that schools just below and above the cutoff do not have any significant difference in terms of school characteristics.

2.4 Estimation Approach

To identify the causal effect of the ALS project on student performance as measured by the national test score, I exploit the project selection cutoffs, which were set separately in each district. These cutoffs create a discontinuity in the probability of treatment as a function of normalized 2016 O-NET scores after the samples were pooled across districts. The discontinuity allows me to apply a regression discontinuity design

[2] There is no data on specific trained teachers and classrooms; therefore, this study examined the project's impact on student academic performance at the aggregate level, i.e. school level.

(Lee and Lemieux 2010; Van der Klaauw 2008). To bypass the question of model specification in RDD estimation, Hahn, Todd, and Van der Klaauw (2001) and Imbens and Lemieux (2008) propose local linear regression techniques, whereas Lee and Lemieux (2010) introduce polynomial regression techniques. Those techniques locally identify treatment effects in the neighborhood of the cutoff where the probability of receiving treatment changes discontinuously. For the ALS project, the schools just above and just below the cutoffs had similar performance in 2016 but the probability of being treated jumps discontinuously.

As discussed above, the ALS selection process segregated schools into two groups, including the groups of schools whose average 2016 O-NET scores were above and below the district cutoffs. In order to analyze the pooled data across districts, I compute the normalized 2016 O-NET scores (X_{ij}) which is the difference between school's average scores and the district cutoffs. This normalization sets each district cutoff to zero where the normalized scores represent the score deviation from the cutoffs. The ideal selection should include all schools below the cutoffs and exclude all schools above the cutoffs. However, some schools below the cutoffs were not selected in the project, while some schools above the cutoffs were included. As a result, the jump in the probability of treatment schools at the normalized cutoff (c=0) is greater than zero but less than one, thus a fuzzy regression discontinuity design is applied in this analysis.

Given a normalized 2016 O-NET score as an assignment variable (X_{ij}), the model specification outcome (Y_{ijk}) can be written as

$$Y_{ijk} = \tau D_{ijk} + f\left(X_{ij}\right) + S_j + T_k + U_{ijk} \qquad (1)$$

$$D_{ijk} = \pi I\{X_{ij} \geq 0\} + g(X_{ij}) + S_j + T_k + V_{ijk} \qquad (2)$$

where Y_{ijk} denotes an average test score of school i in district j for test k; D_{ijk} is a dummy for ALS status, one for AL schools and zero otherwise; X_{ij} is the school i average test score normalized with respect to district j cutoff (c_j); $I\{X_{ij} \geq 0\}$ is an indicator for the school i above the cutoff in district j; τ is the effect of ALS Project on the test score; π is the effect on the ALS probability after crossing the cutoff; $f(X_{ij})$ and $g(X_{ij})$ are polynomial functions of the normalized score; S_j is a set of district dummies; T_k is a set of test dummies; and U_{ijk} and V_{ijk} capture the error terms.

I use a rectangular kernel which gives an equal weight for all observations in the estimation sample, as suggested by Imbens and Lemieux (2008). For the bandwidth selection, Imbens and Kalyanaraman (2012) propose the Mean Squared Error (MSE) optimal bandwidth, which was developed from the Mean Integrated Squared Error criterion (MISE). I specify the bandwidth selection by allowing two different MSE-optimal bandwidth selectors, below and above the cutoff, for the RD treatment effect estimator. The standard errors provided in the estimation is the robust standard errors clustered at district level (Calonico et al. 2017).

In this fuzzy RD design, an instrument for the treatment dummy D_{ij} is $I\{X_{ij} \geq 0\}$. As mentioned in Hahn, Todd, and Van der Klaauw (2001), $E[U_{ijk}|X_{ij} = x]$ continuous in x is a crucial assumption for the instrumental variable framework. Lee and Lemieux (2010) suggested an alternative assumption, the so-called imprecise control, which implies that the continuity of the density of the assignment variable is a sufficient condition for the continuity of unobservable (U_{ijk}). In the direct and intuitive test of the imprecise control of the assignment variable, introduced by McCrary (2008), a jump in density at the threshold implies evidence of some degree of manipulation. Hence, the smoothness of the density of the assignment variable can validate the appropriateness of RD design. I report the result of McCrary's test in the validity and robustness check section.

The second key identification of the fuzzy RD design is a discontinuity in the probability of treatment at the cutoff, the first stage assumption. In order to verify the assumption, I regress the linear specification

$$D_{ij} = \theta I\{X_{ij} \geq 0\} + \alpha_0 + \alpha_1 X_{ij} + \alpha_2 X_{ij} I\{X_{ij} \geq 0\} + W_{ij} \quad (3)$$

where D_{ij} is a dummy for ALS status, one for AL schools and zero otherwise; X_{ij} and $I\{X_{ij} \geq 0\}$ are as discussed earlier; θ is the discontinuity effect of the ALS probability at the cutoff; and W_{ijk} captures the error terms. Equation (3) allows a different slope for the regression function in the neighborhood of the cutoff.

Importantly, the third assumption is the exclusion restriction saying that an indicator for school being above or below the cutoff only affects outcomes (Y_{ijk}) through the treatment status (D_{ijk}) but not any other channels. I use the same specification in equation (3) to prove that there is no evident discontinuity in any pretreatment covariate.

Last but not the least, in the case of imperfect compliance, the estimates are average causal effects for different populations, depending on whether the samples are one-sided or two-sided non-compliance. When the first batch of schools were selected as treatment schools in the academic year 2017, it was a one-sided non-compliance case.

There were almost no cross-overs but about half of no-shows (Panel A in Figure 2.4). Specifically, the probability of ALS status of schools just above the cutoff was 0.065 while the probability for those just below the cutoff increased to 0.432. Therefore, the estimation results for the academic year 2017 are average treatment effects on the treated (ATT), i.e. average treatment effects (ATE) for the treatment schools (ALS). When the second batch of schools began participating in May 2018 (the beginning of academic year 2018), it was a two-sided non-compliance case. There were both cross-overs and no-shows (Panel B in Figure 2.4). The probability of ALS status of schools just above the cutoff was 0.3 while the probability for those just below the cutoff increased to 0.62. However, it is unlikely that schools below the cutoff that were designated as treatment refused to participate in the project but would have participated if they had been above the cutoff. As a result, it may be safe to conclude that the monotonicity assumption holds in this study, i.e. there were no defiers, and the estimation results for the academic year 2018 are local average treatment effects (LATE) of the complier (Angrist and Imbens 1994).

2.5 Internal Validity and Robustness Checks

This section provides evidence ensuring that the key assumptions for RDD hold. First, to show that the normalized 2016 O-NET scores were not manipulated, Figure 2.5 plots the histogram of the scores where the cutoff is zero. The bin width for this histogram is five and the lower limit of the first bin is set to minus 40 to ensure that the bins do not overlap at the cutoff. The histogram visually reveals neither a discontinuity nor a significant jump at the cutoff, thus the densities of the running variable are smooth at the cutoff.

In addition to the histogram, I also provide in Figure 2.5 the density test suggested by McCrary (2008). Even though the null hypothesis of a smooth density at the cutoff cannot be rejected at 10 percent significance level, we can see a jump at the cutoff in Figure 2.5. Nevertheless, the balance test of the raw 2016 O-NET scores presented in Table 2.2 verifies that the scores were not manipulated since there is no significant change in scores at the cutoff. In addition, the cutoff was exogenously determined by local authorities when they selected schools for the project. When they set the cutoffs, they did not look at any specific schools to ensure that certain schools would be included in the project. Moreover, schools did not know the exact values of the cutoffs because the cutoffs were determined after the 2016 O-NET scores were released. Another concern is whether teachers copied the exams and gave hints to their students a day before the tests were taken. In the case of Thai national test, the test papers are not distributed beforehand. Most of the time students take the tests on the exact same day and no papers can be taken out of the examination rooms. NEITS do not reuse the questions; they create new set of questions every year. Additionally, teachers are rotated on the examination date and not allowed to proctor their students. Lastly, there is no evidence of discontinuities in other relevant factors including the number of students, number of teachers, number of classrooms, student-teacher ratio, and raw O-NET score in 2016, as numerically shown in Table 2.2 and graphically shown in Figure 2.6. Therefore, the treatment and control groups did not have any different unique characteristics before the beginning of the project.

2.6 Estimation Results

The main estimation results presented in this session are the effects of ALS Project on national test scores pooled across years, grades and subjects; and the effects on subsamples pooled either across years and subjects or across grades and subjects. The main results, reported in this section, were estimated without test takers weights. The effect sizes shown in Figure 2.7, which are generated from Table 2.4, report both linear and quadratic estimates of the project on the test scores, where the small solid squares indicate the effect estimate and the horizontal lines represent the 95% confidence intervals. The first two lines are the total effects, which are the linear, and the quadratic effect on the pooled test scores across years, grades and subjects. The next (last) six lines are the subsample effects pooled across grades (years) and subjects.

The linear and quadratic RD point estimates of the total effects are 0.0121 and -0.0093 test score standard deviation, respectively. Those estimates for the first subsample pooled across grades and subjects in 2017 are 0.0293 and -0.0997 test score standard deviation. Those estimates for the subsample pooled across grades and subjects in 2018 are -0.0074 and 0.0294 test score standard deviation. Those estimates for the subsample pooled across grades and subjects in 2019 are -0.0074 and 0.0078 test score standard deviation. Those estimates for the subsample pooled across years and subjects in grade one are -0.0823 and -0.065 test score standard deviation. Those estimates for the subsample pooled across years and subjects in grade three are 0.0344 and 0.0214 test score standard deviation. The estimates of the last subsample pooled across years and subjects in grade three are 0.0192 and 0.0063 test score standard deviation. All the effect sizes are very small. Moreover, most of the 95% confidence intervals reported in Figure

2.7 allow us to rule out effect sizes whose absolute values are greater than 0.1 test score standard deviation except for grade one and grade 3 subsamples pooled across years and subjects. The total effects are small and very similar to those from subsamples.

In academic year 2017, before the second batch of schools joined the project, the first stage regression indicates a significant difference in probability of ALS status between the two groups, i.e. group of schools just below and above the cutoff. Specifically, the probability for schools just above the cutoff is 0.065, while the probability for those just below the cutoff is 0.432. Therefore, analysis of the 2017 national test scores suggests that the average treatment effects (ATE) for the treatment schools are insignificantly different from zero. The linear and quadratic estimations for about 50 percent of the schools near the cutoff indicate the effect sizes of 0.0293 and -0.0045 test score standard deviation, respectively.

Results of the analysis of the 2018 national test scores suggest that the local average treatment effects (LATE) for the complier schools (which account for one-third of all samples) are also insignificantly different from zero. The effect sizes of the linear and quadratic specification are -0.0074 and 0.0294 test score standard deviation, respectively. Similarly, the 2019 national test scores show no significant improvement in student performance, with effect sizes of -0.0074 and 0.0078.

The above results are presented graphically in Figure 2.8. Panel A presents the results for the total effects of the ALS Project on the pooled test scores. According to the project timeline, the authorities mainly did a baseline survey, one pilot training and the first training session near the end of the academic year 2017. The results in panel B are reasonable because they suggest no discontinuity in the test scores. In 2018 and 2019, after the second training session and the coaching/monitoring process, there is no clear jump at the cutoff (Panels C and D). Panels E, F and G also exhibit no jump at the cutoff. In other words, the project did not improve the test scores for grade one, three and six students. These graphical presentations of the results confirm that the ALS project did not substantially improve student performance as measured by national test scores.

In summary, the ALS Project, which is intended to improve student performance, has been implemented since the beginning of the academic year 2017. The authorities analyzed the problem and developed an intervention, the active learning concept teaching model. The first training session was conducted in November 2017, near the end of academic year 2017. As expected, we did not find any improvement of student performance on 2017 test scores. The second training session was conducted in July 2018, at the beginning of academic year 2018, followed by coaching and monitoring activities. Similarly, no significant effect was found in the 2018 and 2019 test scores. Schools near the cutoff performed very similarly and there was no noticeable jump at the cutoff. The RD estimates of test scores are insignificantly different from zero, with the 95 percent confidence interval estimates ranging from -0.1 to 0.1 test score standard deviation. I also evaluated the effects of the project on student literacy and found no statistically significant difference between first and third grade reading and literacy test scores for the treatment schools and those for the control schools. Nevertheless, the results in this section only explain effects of ALS project on student performance at aggregate level, i.e. at school level. It is possible that gain by some students is cancelled out by loss of some others within the same schools; therefore, we found null effects at school level. Further study, addressing the heterogeneity within school issue, is discussed in section 1.9.

2.7 Why was the program not effective?

One question may arise whether the number of trained teachers is proportional to the school size since the project implementation detail can be one of the possible reasons for the null effects. In order to answer the question, the ALS trained teacher list, as a piece of evidence, shows that even though the number of trained teachers in each school are not proportional to the school size, the program was implemented as intended. The original plan was to have one teacher from each subject from each participant school participate in the training session. According to the name list, the majority of trained schools, about 95%, have at least 4 teachers who participated in the training sessions as the project was designed. Hence, the doubt on whether the training sessions were properly implemented is clarified. The name list is a solid evidence to prove that trained teachers are not concentrated in a small group of schools since the majority of trained schools got the same portion of training slots, where the non-weighted average proportion of trained teachers from 196 schools is 24%. Another important information to be clarified is the fraction of trained to total teachers; the total number of trained teachers account for only about 16 percent of the total number of teachers in treatment schools. Ideally, these teachers were expected to share the content obtained from the sessions; however, there is no evidence regarding the knowledge transfer among teachers.

There are a few possible explanations, in term of human resources, as to why the program have null effects. First, the number of trained teachers is relatively small compared to the number of total teachers; therefore, they may struggle to widely transfer the knowledge to their colleagues. Second, trained teachers may not successfully apply the ALS concept in classes, as a result, the concept did not reach students. Moreover, it could be challenging when each project supervisor was in charge of 3 – 5 schools due to the fact that the project has only 65 supervisors but 196 trained schools, if not 350 ALS schools.

The project design could be one of the hindrances. According to the project timeline, the ALS training sessions were held only for three days. The duration of the project in-service training was limited, and may not have been sufficient to fully train and provide the teachers with a sufficient number of activities. In addition, the training content was not provided by subject and none of ready-to-implement teaching plans were given. Most tools introduced in the training sessions are graphic organizers. The tools help stimulate students' thinking skills but may not be applicable to a specific subject. Moreover, the report on development of teacher training curricula for the ALS project (south), prepared by the ALS project supervisors in cooperation with senior and local education authorities (2020), addressed the following feedback from trained teachers. They suggested that the training sessions should have been held regularly, i.e. once a semester before the semester starts, and extended to newly assigned teachers for more effective outcomes. In addition, they requested for variety and effective content, practical activities and demonstration, and example based guidelines for classroom implementation. Lastly, they recommended that the number of project supervisors should be proportional to the number of participants to effectively provide advice and acquaint teachers with active learning techniques.

In terms of classroom intervention, there is heterogeneity in implementation according to the report (Tanguthaisuk 2020). Project supervisors encouraged teachers in a majority of the ALSs to replace top-down lectures with group discussion and activities;

ask questions that allow students to think critically, and prepare supporting materials and equipment in advance. The supervisors reported that teachers used positive reinforcement, allowed knowledge exchange, asked more questions in classes, and designed and provided a variety of fun activities. However, small evidence of practical and fun activities, observed in a few Active Learning Schools (ALSs), may not be enough to inductively conclude that the project led to effective classroom intervention.

Last but not the least, effects at school level may be net effects of ALS project on student performance. It could be that gain from some students is cancelled out by loss from others in the same school. In addition, effects may be heterogeneous depending on student-to-teacher and supervisor-to-teacher ratios. The project could possibly increase performance of students in some schools but negatively affect performance of students in other schools. However, the estimates report null effects when all treatment schools, with both high and low student-to-teacher and supervisor-to-teacher ratios, are pooled together.

In conclusion, possible explanations for the null effects are described as follows. First, the limited duration of in-service training may not sufficiently and effectively prepare teachers to comprehend and implement active learning concept in classes. Second, trained teachers only accounted for less than one-fourth in each school and there was no clear evidence on knowledge transfer among teachers. Third, the content provided in the sessions were not specific and did not include ready-to-implement lesson plans; therefore, teachers could spend some time to create teaching plans that fit the contents. Forth, the intensity of classroom implementation mostly depends on individual teachers and I cannot assume specific-to-general conclusion in this context. Finally, supervisor-to-teacher ratio is not large enough to ensure sufficient follow-ups and monitoring. As a result, it is difficult to verify whether the active learning concept was in fact implemented in treatment schools.

2.8 Robustness Check

As a robustness check, I also estimated the effects of ALS Project in a manner similar to that for estimation of the main results; the effects were weighted using the number of test takers in each school with robust standard errors clustered at district level . The linear and quadratic estimates of the effects of the project on the test scores are presented in Figure 2A-1 in Appendix, generated from Table 2A-1; the small solid squares indicate the effect estimates and the horizontal lines represent the 95% confidence intervals. The first two lines are the total effects, which are the linear and quadratic effects on the pooled test scores across years, grades and subjects. The final six lines are the subsample effects pooled across grades (years) and subjects.

The linear and quadratic RD point estimates of the total effects are 0.027 and 0.005 test score standard deviations, respectively. The estimates for the first subsample pooled across grades and subjects in 2017 are 0.078 and 0.004 test score standard deviations. The estimates for the subsample pooled across grades and subjects in 2018 are 0.022 and 0.021 test score standard deviations. The estimates for the subsample pooled across grades and subjects in 2019 are -0.005 and -0.027 test score standard deviations. The estimates for the subsample pooled across years and subjects in grade one are -0.035 and 0.133 test score standard deviations. The estimates for the subsample pooled across years and subjects in grade three are 0.026 and 0.005 test score standard deviations. The estimates for the last subsample pooled across years and subjects in grade three are 0.030 and 0.007

test score standard deviations. Most of the effect sizes are small, with 95% confidence interval between -0.2 and 0.2 test score standard deviations, except for grade one subsamples pooled across years and subjects. The effects are small and correspond to the main results reported in the previous section.

Due to the fact that there was very little crossover, and since about half of the schools just below the cutoff were no-shows, the analysis of the 2017 test scores suggests that the average treatment effects (ATE) for the treatment schools are insignificantly different from zero. The linear and quadratic estimations for about 50 percent of the schools near the cutoff indicate effect sizes of 0.078 and 0.004 test score standard deviations, respectively. Although there were both crossovers and no-shows after the second batch of schools began participating in May 2018 (the beginning of academic year 2018), it is reasonably safe to conclude that the monotonicity assumption holds for this study, i.e. there were no defiers, because it is unlikely that schools below the cutoff that were designated as treatment refused to participate in the project; rather, they would have participated if they had been above the cutoff. As a result, the analysis of the 2018 national test scores suggests that the local average treatment effects (LATE) for the complier schools (which account for one-third of all samples) are insignificantly different from zero. The effect sizes of the linear and quadratic specification are 0.022 and 0.021 test score standard deviations, respectively. Similarly, the 2019 national test scores show no significant improvement in student performance, with effect sizes of -0.005 and -0.027.

The above results are presented graphically in Figure 2A-2 in Appendix. Panel A presents the results for the total effect of the ALS Project on the pooled test scores. The results in panel B, i.e. no discontinuity in test scores, are compatible with the project timeline since the first training was held near the end of academic year 2017. Even after the second training session and the completion of the coaching/monitoring process, there is no clear jump at the cutoff for 2018 and 2019 test score results, presented in panels C and D. Considering student performance by grade, panels E, F and G also exhibit no jump at the cutoff. In other words, the project did not improve the test scores for grade one, three and six students. These graphical presentations of the results coincide with the graphs shown in the estimation result section; this provides further evidence that the ALS project did not substantially improve student performance as measured by national test scores.

2.9 Conclusion

The ALS project, which was intended to improve academic achievement of students in the south of Thailand, has been implemented since April 2017, the beginning of the academic year 2017, by Thailand's Ministry of Education. This study uses grades one, three and six national test scores from the National Institute of Educational Testing Service (NIETS) during the academic years 2016-2019 to provide evidence of the effect of Active Learning School (ALS) Project on student performance in the most southern Thailand. Schools with relatively low average test scores were designated as treatment schools and those with relatively high average test scores as control schools. I exploit the project selection criteria and estimated the effects of the ALS Project by using a fuzzy regression discontinuity design. The RD estimates on the test scores pooled across years, grades and subjects are consistent with the results estimated from subsamples pooled across either subjects and grades or subjects and years. All results suggest that the project did not substantially improve student performance as measured by national test scores, since the effect sizes are insignificantly different from zero. Moreover, the 95 percent confidence interval allows us to eliminate the effect sizes whose absolute values are greater than 0.1 test score standard deviation. Our findings regarding the effects of the ALS project on student performance are consistent with those of studies on the effects of CL and teacher PD on academic achievement (Jacob and Lefgren 2004; Tan et al. 2007; Tracey et al. 2010; Thurston et al. 2019; Demie et al. 2019; Loyalka et al. 2019).

The primary limitations of the generalization of these results are limited access to and availability of data. The effect estimates in the RDD model are based on accessible and available data and might not perfectly represent the effects of the project on the total student population in the south of Thailand. However, the available data covers all provinces where the project was operated and could be reasonable representations of the population. Thus, it is likely that the findings of this study are reliable and valid despite the limitations. Moreover, this study only examined the effects at an aggregate level, i.e. school level, since specific data on trained teachers and their classes are not available. Although the test scores used in this study might not be the best indicator to capture the effects of ALS project on student performance, the national test scores guarantee the same evaluation standard countrywide. Since there are no other indicators available to be assessed and evaluated, the national test scores are the best available data for this study. In addition to limitations in terms of data, another limitation regarding the estimation approach should be pointed out: the RDD approach only affords an explanation of the local average treatment effect (LATE) of the complier schools around the cutoff. Therefore, the findings of this study do not explain the effect of the ALS project on noncomplier schools, i.e. schools that would have participated or not participated regardless of their 2016 average performance relative to the cutoff.

The followings are recommendations for further research work. The null effects of ALS project on student performance at school level does not necessarily imply no effects since the effects could represent the net gain from different groups of students. Therefore, further study could explore the effects of the project on higher moment, for example, variance of the test scores within school. In order to statistically provide evidence for ineffectiveness, subgroup analysis grouped by share of trained teachers, supervisor-to-teacher, and student-to-teacher ratios should be considered.

CHAPTER THREE

Science, Mathematics and Technology Promotion and Student Performance: Evidence from Thailand

3.1 Introduction

Many developing countries implement teacher professional development (PD) programs, aiming to improve student performance. A growing literature on impacts of teacher PD programs on student academic achievement over the past two decades provides mixed evidence on large-scale interventions from many countries, ranging from successful (Albornoz et al. 2020; Sun and Du 2021; Cilliers et al. 2020) to ineffective (Mbiti I 2016; Loyalka et al. 2019; Abbiati et al. 2021; Schaffner et al. 2021; Carneiro et al. 2022). Characteristics of effective training include, for example, practical lesson plans with supported materials and long-term follow-up and monitoring, while possible causes of ineffective interventions include weak accountability and motivation, inadequate central-level efforts, and lack of trainer expertise (Popova et al. 2022; Schaffner et al. 2021).

This study examines the effect of a teacher professional development program in Thailand on student performance as measured by the national test scores of grade six and nine students. The program was launched in 2018 by Thailand's Ministry of Education to improve the quality of science, mathematics and technology teaching in sub-provincial schools across Thailand. Schools voluntarily applied for and were selected for the project based on their readiness and total numbers of students. The project provided online learning resources and materials starting in early 2018, followed by a series of online training sessions for science, mathematics and technology teachers of participant schools starting in 2020. This study focuses on up to three-year effects of the project using 2015 – 2017 test scores as pre-treatment scores and 2018 – 2020 test scores as post-treatment scores.[3]

I employ a difference-in-difference (DID) design to estimate the impacts of the project on mathematics and science test scores for grade six and nine students of 1,640 participant schools in 77 provinces of Thailand. The common trends assumption seems to hold during the pre-treatment period for grade six test scores but not for grade nine test scores. Impact estimates suggest that the project had, if anything, only a small positive effect on grade six test scores. Effect sizes are smaller than 0.05 standard deviation for both mathematics and science and the 95% confidence intervals allow us to rule out the effect sizes that are greater than 0.07 and 0.05 standard deviation for mathematics and science, respectively. Subgroup analyses show that the project appears to have larger impacts on test scores of students in high-performance schools but insignificant impacts on the scores of those in low-performance schools. The effect sizes of ninth-grade students are also small; however, these results are more tentative since the common trends assumption seems to fail for grade nine students. The results are robust to estimations with or without test taker weights and to the choice of base years.

These results are in line with Loyalka et al. (2019) and Schaffner et al. (2021) who report that at-scale government teacher PD programs had only little or no impact on student learning in China and Nepal: effects larger than 0.074 and 0.1 standard deviations

[3] The project is ongoing and an additional booster of teacher training and material distribution happened during June – August 2021. However, taking the national tests is no longer compulsory from academic year 2020, which makes it difficult to evaluate the project because of potential sample selection.

were ruled out. Some PD programs in developing countries only improve the quality of teacher-student interaction; however, they do not translate into higher student achievement, or even lower their achievement in some cases (Carneiro et al. 2022; Mbiti I. 2016). Another large scale teacher PD for lower secondary math teachers in southern Italy had no impact on math achievement of grade six students in the first wave; however, they found a significantly positive impact in the second wave (Abbiati et al. 2021). On the other hand, some studies in South Africa, Argentina, ad China found significantly positive effects of teacher PD program on student performance; however, the effects are heterogeneous (Albornoz et al. 2020; Sun and Du 2021; Cilliers et al. 2020). High performance and urban students relatively benefit more from the program compared to low performance and rural students.

There are several explanations for ineffective PD program such as weak accountability and motivation, inadequate central-level efforts, lack of trainer expertise, low training participation rates, and insufficient complementary materials (Popova et al. 2022; Schaffner et al. 2021). It is possible that the newly implemented program can be unfamiliar to both teachers and students, as a result, it could take some time for them to open up and adapt to the new pedagogical approach (Abbiati et al. 2021; Carneiro et al. 2022). In addition to the DID analysis, I conducted an online survey by randomly selecting 150 out of 575 SMT schools, then randomly selecting three teachers from the lists of trained teachers in each school. I obtained 80 responses from SMT teachers. Based on this online survey and some information acquired via telephone interviews, the insignificant effects of the Thai SMT project can be explained by the following possible reasons. First, SMT teachers have low motivation in implementing the newly trained content in classrooms. Second, the training participation rates were disappointingly low, with less than 60% of eligible teachers attending training sessions. Third, the training contents do not provide ready-to-implement teaching plans, thus teachers have to spend some time designing lesson plans themselves. Forth, even though teachers found some of the training content useful and practical, lack of supporting material distribution in the beginning of the program became an obstacle for integrating the training into practice. Fifth, there is a lack of human resources to monitor and follow-up whether there is real implementation in treatment schools after central training was arranged. Finally, it could be challenging for teachers to fully apply the new approach via online instruction amid the COVID-19 situation since the training contents were not designed for online classes.

This chapter is organized as follows. Section 3.2 introduces the Science Mathematics and Technology Project (IPST) Thailand. Section 3.3 provides the data sources and some descriptive statistics. Section 3.4 describes the estimation approach. Section 3.5 and Section 3.6 provide estimation results and robustness checks. The survey results and possible reasons for ineffectiveness of the intervention are detailed in Section 3.7. Section 3.8 concludes the research.

3.2 Background

Thailand's Ministry of Education has implemented a number of projects to strengthen underprivileged schools, to create high potential schools, and to support the development of talented students. However, the mid-level schools, which rank between underprivileged schools and high potential schools, were not the first priority target group of those projects. The mid-level schools, located in every municipality, are the majority of schools; more than half of Thai students are under their care, hence there is a need to promote these schools to the qualified school level. The Institution for the Promotion of Teaching Science and Technology (IPST) under Thailand's Ministry of Education, together with Ministry of Interior, Bangkok Metropolitan and Pattaya City, launched the Science Mathematics and Technology (SMT) school project to improve the quality of science, mathematics and technology instruction in sub-provincial schools. The objectives of the project are to mitigate education inequality, to increase educational opportunity, to enhance knowledge and develop skills of SMT teachers and principals, to develop students' analytical, critical and innovative thinking skills, and to create a better network for science, mathematics and technology teachers. The ultimate goal is to support and help students improve their academic performance and apply their knowledge in real life.

The SMT project has three phases consisting of school selection, project implementation, and annual evaluation. The selection of 1,640 primary and secondary schools included four announcements, where 760 first batch schools were announced in December 2017 and July 2018, and 880 second batch schools were announced in May 2019 and February 2020, respectively (Figure 3.1). IPST provides materials to participant schools: curricula manual; textbooks (e-books); teacher manual; instructional media; equipment to support science/computer laboratory (not including any construction, building upgrade or maintenance, and computer hardware); and academic training for principal and teachers. Meanwhile, SMT schools designed school development plan, in terms of learning management; computer and science laboratories; teacher network; and science, mathematics and technology materials, and report back to IPST.

In order to achieve the objectives, IPST introduced a learning process called *coding*, which is aimed at the development of students' analytical, critical and creative thinking skills, as well as reading, writing and problem solving. Teachers' roles are generally thought to be the most important factor in a *coding* learning system; hence, the IPST, in cooperation with the Office of the Basic Education Commission (OBEC), developed the online *Coding for Teacher* (C4T) curriculum to develop teachers' potential in terms of both content authoring and teaching techniques. C4T courses are mostly online training sessions for SMT teachers. In the early stages, training topics focus on computing science; programming (KB-IDE, Micro Python, Python[4], and Scratch[5]); unplugged 1 and 2 (non-PC courses); data science; and AI for school level 1–4.[6] Later in the training series, SMT also provides online training for analytical, critical, creative thinking and problem solving, mathematics, biology, physics, earth science and astronomy, science and technology, design and technology, and summative assessment. Starting in the academic year 2018, computing science became a new subject in science curriculum for grade one, four, seven, and ten, followed by grade two (three), five (six), eight (nine), and eleven (twelve) in academic year 2019 (2020). Computing science aims at computational thinking with fundamental knowledge of digital technology and media and information

[4] Python Language is an open source computer language, an interpreter computer program, for application development in any available platform.

[5] Scratch is a coding community for children. The coding language includes a sample visual interface that allows children to create digital stories, games, and animations. It promotes computational thinking and problem solving skills, creative teaching and learning, self-expression and collaboration, and equity in computing.

[6] AI for schools is a four-level Artificial Intelligence (AI) program. The first level, awareness, provides basic information of AI and how it is used in our daily lives. The second level, AI components and basic concepts, tells about the concept, components of AI and also factors that affect AI agent's decision making. The third level, model, introduces how to use AI with computer, ask a computer to identify the categories of input information and then accurately forecast quantitative data. Lastly, integration constitutes lessons on how to develop practical AI application and use it in our routine activities.

literacy to create systematical thinking process, which leads to good problem solving abilities in students. In addition to C4T, there are specific training sessions (*SMT Principal Training* and *Coding for School Directors*) for SMT school principals and directors. Those training sessions have been in place since May 2020, the beginning of academic year 2020.

In addition to the training sessions, IPST also provides online learning resources and materials via the IPST learning space and knowledge repository *SciMath*, which is a science, mathematics and technology knowledge library. The *SciMath* collection of widely varied teaching materials include educational videos, photo galleries, articles, projects, lessons and lesson plans, online applications, and E-books. All materials can be accessed freely except E-books, for which registration and login are required. The content is academically correct and consistent with the learning standards set out in Thai academic curricula. Teachers can review content or develop their own mastery of both academic content and teaching techniques.

The project timeline in Figure 3.1 shows that the curriculum development and supporting materials were available from spring 2018 onwards, while a series of online training sessions for science, mathematics and technology teachers of participant schools started to be conducted from spring 2020 onwards. The project does not target only a specific grade; it treats all grades equally. Thai educational system mainly consists of primary schools (grades one through six) and secondary schools (grades seven through twelve); however, we also have primary-middle schools (grade one through nine), primary-secondary schools (grade one through twelve), and a few high schools (grade 10 through 12). All schools in Thailand are included in the analysis of this study regardless of the type of institution they are. The total number of trained teachers was about 6,000 as of early 2021. During June – August 2021, there was a massive additional online training including about 32,000 mathematics and science teachers of SMT schools. Unfortunately, we do not have data from 2022 onward since national tests were no longer compulsory. Detailed information on training timetable and some training contents can be found in Table 3A-1 in Appendix. After the project was implemented, IPST annually evaluates the effectiveness of the project for three consecutive academic years starting from the academic year 2020, i.e. the first report of academic year 2020 will be done during March-May 2021.

3.3 Data

3.3.1 Test scores and Science, Mathematics and Technology participant schools.

The main data analyzed in this study are individual national test scores of grade six and nine students from academic year 2015 to 2020 from the National Institute of Educational Testing Service (NIETS), Thailand, i.e. the Ordinary National Education Test (O-NET) scores. The tests include mathematics and science tests, where the descriptive statistics of grade 6 and grade 9 students can be found in Table 3.1 and Table 3.2, respectively. In order to compare the scores across years, I computed the individual standardized scores of the two tests of grade six and nine students. Then, I calculated school average standardized scores by subject by year from individual standardized scores and created a large panel data of school average standardized scores of the two tests from 2015 to 2020.

The participant school lists are from The Institute for the Promotion of Teaching Science and Technology (IPST), Thailand. The project involves over 1,600 schools from all provinces of Thailand. Figure 3.2 is the population diagram. Chart A shows that the total of about 33,000 schools in 77 provinces of Thailand includes 1,640 treatment schools and over 31,000 control schools. There are approximately 35,000 teachers in treatment schools, shown in Chart B. Only about one-sixth of them, roughly 6,000 teachers, participated in the online training sessions during April 2020 – March 2021, i.e., academic year 2020. Due to data availability, this study covers only about 90% of schools in Thailand[7], about 30,000 schools under Office of the Basic Education Commission (OBEC) and about 300 schools under Bangkok affiliation. Students in schools under OBEC and Bangkok affiliation are accounted for about 75% of Thai students; therefore, test scores analyzed in this study are from the majority of Thai students. As a result, this study only explained the effects of the project on performance of students in schools under OBEC and Bangkok affiliation; it did not include, if any, the effects of students in schools under Office of the Private Education Commission (OPEC), schools under Department of Local Administration (DLA), and homeschooling.

The school selection process was not randomized and one might expect participating schools to be better performing even in the absence of any intervention. However, the common trend assumption, which says that the unobserved differences between the treatment and control groups are constant over time might still hold. There are clear differences in average performance between treatment and control groups before the intervention and those differences are time invariant (Figure 3.3). Therefore, the common trend assumption holds in pre-treatment period.

[7] There are about 33,000 schools in Thailand; however, the numbers of schools keep decreasing each year due to the shortage of academic professionals, especially in rural areas.

In the academic year 2020, there was a change in the national test policy, in which the test became voluntary and schools ceased to be in charge of students' test registration. From academic years 2015 – 2019, numbers of schools and test takers were quite constant, about 800,000 grade six students and 700,000 nine grade students. Numbers of schools registering for the test may vary due to their current 6^{th}- and 9^{th}-grade students. However, number of schools decreased dramatically, about 18%, in academic year 2020 according to the policy change.[8] As a result, only students who were willing to take the exams registered for the test. I will address the policy change issue and deal with the sample selection in the robustness check section.

3.3.2 Survey Data

In addition to the DID analysis, I conducted an online survey, during October – November 2021, to assess whether or not the project was effectively implemented as planned. The survey data were collected from 80 SMT teachers, 22 responses via contact persons and 58 responses from official letter invitation. The list of names of teachers who received the letters was selected randomly from the list of names of teachers who were eligible for the training sessions. The list included 6,195 teachers from 575 schools. This survey randomly selected 150 schools from the list, then randomly selected 3 teachers from each school. The letters were finally distributed to 397 teachers[9] from 150 schools in 57 provinces; however, only 58 teachers from 52 schools in 35 provinces responded. The survey respondents consist of 31% male teachers and 69% female teachers, these

[8] Prior to academic year 2020, schools were in charge of O-NET application of their students, i.e. schools needed to submit grade six and nine student name list to NIETS as for O-NET registration. From academic year 2020 onward, students apply for the test by themselves.
[9] There are some schools with less than 3 eligible teachers in the name list.

numbers are compatible with the total number of Thai education professionals, i.e. 34% and 66% male and female professionals, respectively. The average participation rate from the survey data is about 57%: the participation rate of 22 teachers from contact persons is 50%, 10 participants and 12 non-participants, while the participation rate of 58 teachers from the invitation letters is about 62%, 36 participants and 22 non-participants. Teachers who answered the survey are in charge of all subjects related to the project, i.e. mathematics; science; and technology. The average age of respondents is 38.59 years.

3.4 Estimation Approach

3.4.1 Difference-in-difference design for overall and subgroup analysis

I employ a difference-in-difference (DID) design to estimate the impacts of the project on mathematics and science test scores for grade six and nine students of 1,640 participant schools in 77 provinces of Thailand. In order to estimate the causal effect of the SMT project on student performance I use the following equation:

$$TS_{st} = \alpha_0 + \alpha_1 SMT_s + \sum_t \beta_t t + \sum_t \delta_t SMT_s t + \varepsilon_{st} \qquad (1)$$

where TS_{st} denotes an average test score of school s in academic year t; SMT_s is a dummy for treatment variable, one for treatment schools and zero otherwise; t is a set of dummy variables for academic year 2016 – 2020, one if the test scores were taken in that academic year and zero otherwise (i.e. 2016 is a dummy for academic year 2016, etc.); $SMT_s t$ is a set of interaction terms between treatment and academic year dummies; α_0 represents the average test score of control group in a base year which is academic year 2015 in this study; α_1 implies how better or worse the performance of the treatment group is when compared to the control group in the base year; δ_t are coefficients which indicate whether the common trends assumption held in 2017 and 2018, and whether there are effects of the SMT Project on test scores in the academic years 2018, 2019 and 2020. ε_{st} captures the influence of unobserved determinants of test scores.

I use OLS regressions with weights equal to the number of test takers in each school and cluster standard errors at school level to estimate the effects of the SMT project on school level test scores for all estimates except for unweighted estimates (section 2.6.1). The estimates were produced separately by subject, i.e. mathematics and science, and grade, i.e. grade six and nine.

The SMT school project targets mathematics, science and technology, thus it is reasonable to compare the effect of the project between groups of schools with high and low science and mathematics performance. As a result, I created subgroup using the national test scores prior to the implementation of the project, i.e. academic years 2015-2017, as a proxy of their students' performance. High (low) performance schools are schools with average mathematics or science scores above (below) national average for 3 consecutive years.[10] For grade six schools, the total of 22,143 schools consist of 7,273 (6,248) low performance schools, 5,112 (5,058) high performance schools and 9,758 (10,837) inconsistent schools when grouped by mathematics (science) scores. For grade nine schools, the total of 7,517 schools consist of 2,741 (2,395) low performance schools, 1,410 (1,557) high performance schools, and 3,366 (3,565) inconsistent schools when grouped by mathematics (science) score. I ran regression using equation (1) on mathematics and science test scores separately for all subgroups in both grades. As a

[10] Schools with inconsistent performance, i.e. fall into high performance in 2015 but low performance in 2016 and 2017, were dropped out of the sample.

result, the estimates presented the impacts of the project on the following four subgroups of schools: high mathematics performance, low mathematics performance, high science performance, low science performance.

As a robustness check, I also estimate equation (1) without test taker weights. In addition, I check whether the estimates are robust when I use different base years, i.e. 2016 and 2017 instead of 2015 in the main results. I also provide heterogeneity check by estimating the following equation (2) so as to do joint hypothesis test of coefficients between β_t and δ_t, to statistically test whether the effects on performance of students in top and bottom SMT schools are significantly different.

$$TS_{st} = \alpha_0 TM_s + \alpha_1 SMT_s TM_s + \sum_t \gamma_t TM_s t + \sum_t \beta_t SMT_s TM_s t +$$

$$\theta_0 BM_s + \theta_1 SMT_s BM_s + \sum_t \rho_t BM_s t + \sum_t \delta_t SMT_s BM_s t + \varepsilon_{st} \qquad (2)$$

where TM_s is a dummy for top mathematics performance schools, one for schools with school avarage mathematics scores above national average test scores for academic year 2015 – 2017 and zero otherwise; BM_s is a dummy for bottom mathematics performance schools, one for schools with school avarage mathematics scores below national average test scores for academic years 2015 – 2017 and zero otherwise; SMT_s is a dummy for treatment variable, one for treatment schools and zero otherwise; t is a set of dummy variables for academic years 2016 – 2020, one if the test scores were taken in that academic year and zero otherwise.

3.4.2 Lee Bounds

There was a change in the national test policy in academic year 2020 where the test became voluntary for students; therefore, the numbers of schools registering for the test fell sharply by about 18%. Only students who were willing to take the exams registered for the test. As a result, this brought in different participation rates between treatment and control groups and raised a sample selection issue in academic year 2020. In the presense of nonrandom sample selection, Lee (2009) proposed a bounds estimator to determine an interval for the true value of the treatment effect. The estimator requires two assumptions which are random assignment of treatment and monotonicity. The bounds estimator trims observations either in treatment or in control group so as to get the equal share of obsevations for both groups. Therefore, the trimming process corresponds to two extreme assumptions about missing information, i.e either trim the bottom or trim the top. In this study, the practical estimates for the bounds, stemmed from Tauchman H. (2014), are computed as follows.

$$q_T = \frac{\sum_i 1(SMT_i = 1, S_i = 1)}{\sum_i 1(SMT_i = 1)}$$

$$q_C = \frac{\sum_i 1(SMT_i = 0, S_i = 1)}{\sum_i 1(SMT_i = 0)}$$

Let Y_i denote student test score, SMT_i is a binary treatment indicator with SMT_i=1 for treatment group, S_i is a binary selection indicator with S_i=1 when student test score can be observed, and $1(\cdot)$ denotes the indicator function. Therefore, q_T represents the participation rate of the treatment group, and q_C represents the participation rate of the control group. In this study, participant rates of treatment and control groups are 83% and

80%, respectively. Therefore, the trimming proportion in a case where participation rate of treatment group is higher than participation rate of control group, $q_T > q_C$, can be calculated as follows.

$$q = \frac{q_T - q_C}{q_T}$$

Here, q and $1 - q$ determine the position at which student test scores in the treatment group are trimmed from the bottom and the top of the distribution. Therefore, y_q^T and y_{1-q}^T are bottom and top marginal values of test scores at the trimming points, where G_Y^{-1} denotes an inverse distribution of student test scores.

$$y_q^T = G_{Y|T=1,S=1}^{-1}(q)$$

$$y_{1-q}^T = G_{Y|T=1,S=1}^{-1}(1 - q)$$

The upper and lowwer bounds can be calculated using y_q^T and y_{1-q}^T as follows.

$$\hat{\theta}^{upper} = \frac{\sum_i l(T_i = 1, S_i = 1, Y_i \geq y_q^T)\, Y_i}{\sum_i l(T_i = 1, S_i = 1, Y_i \geq y_q^T)} - \frac{\sum_i l(T_i = 0, S_i = 1) Y_i}{\sum_i l(T_i = 0, S_i = 1)}$$

$$\hat{\theta}^{lower} = \frac{\sum_i l(T_i = 1, S_i = 1, Y_i \leq y_{1-q}^T)\, Y_i}{\sum_i l(T_i = 1, S_i = 1, Y_i \leq y_{1-q}^T)} - \frac{\sum_i l(T_i = 0, S_i = 1) Y_i}{\sum_i l(T_i = 0, S_i = 1)}$$

3.5 Main Results

3.5.1 Overall

Average test scores at school level grouped by treatment status were reported in Figure 3A-2 in Appendix, where open (solid) dots represent average scores of control (treatment) group. For both grade six and nine test scores, there are clear differences in average performance between treatment and control groups, better performance from students in treatment group. Those differences in grade six test scores are time invariant before the intervention and the gaps were slightly larger after the intervention. In other words, the common trend assumption holds in pre-treatment period (Table 3.3) and there were small improvements in their performance after the intervention. The average test score gaps of grade nine students fluctuated during 2015 – 2020; hence, both the common trend assumption (Table 3.4) and the impact of the project might be inconclusive.

The main results in this session are the effects of SMT project on national test scores, i.e. mathematics and science scores, of grade six and nine students. The results estimated using Difference-in-Difference (DID), with test taker weights and standard errors clustered at school level, present the effects from the beginning of the project to April 2021. In Figure 3.4[11], each dot represents a coefficient of an interaction term between a treatment school dummy (SMT) and an academic year dummy (2016 – 2020) while the horizontal lines represent 95% confidence intervals. According to the project timeline shown in Figure 3.1 (Section 2), the implementation of the project has started in academic year 2018; therefore, O-NET scores of academic years 2016 – 2017 (2018 – 2020) are pre- (post-) treatment outcomes. The coefficients of interaction terms between

[11] Figure 2.4 is generated using the results from column (1) in Table 3.5, 3.6, 3.11, and 3.12.

SMT and post treatment academic years, i.e. 2018, 2019 and 2020, represent the first, the second, and the third year effects of teacher professional development program on mathematics and science performance of grade six and nine students. The results suggest that the project does not significantly improve student performance as measured by national test scores.

As shown in Panel A and B of Figure 3.4, the project contributes to less than 0.05 standard deviation improvement in both mathematics and science performance of grade six students. Specifically, the project increases mathematics performance of grade six students by 0.043, 0.011, and 0.027 standard deviation, on average, after one, two, and three-year implementation, respectively. The project's impacts on science performance of grade six students are -0.009, 0.024, and 0.016 standard deviation, on average, at the end of the first, second, and third year of the project. It is highly likely that the common trend assumption, implied by the first two lines, holds for both mathematics and science; therefore, the estimated results are promising. I can rule out the effect sizes that are greater than 0.07 (0.05) standard deviation for mathematics (science) performance of grade six students.

The estimates for grade nine students, presented in Panel C and D of Figure 3.4, suggest that the project either has null or negative effects on science performance of grade nine students after the first and the second year implementation. The project seems to positively affect their science performance after the project has been implemented for three years; however, effect sizes for academic year 2020 are not greater than 0.05 standard deviation. Unfortunately, according to the first two lines of the graph, the average test score differences between treatment and control groups are not constant. In other words, the common trend assumption does not hold for grade nine test scores. As a result, the estimates could not explain much about the effect of the project on the mathematics and science performance of grade nine students.

3.5.2 High vs Low performance schools

In this subsection, I will assess whether high performance schools benefited more from the PD program compared to low performance schools. To be consistent with the main results, subgroup analysis results presented in this section use the Difference-in-Difference (DID) estimates with test taker weights and cluster standard errors at school level. Figure 3.5 presents the effects of the project on mathematics and science scores of students in high and low performance schools grouped by pre-treatment science scores. Effects on mathematics of students in top and bottom science schools are shown in Panel A and Panel C, while effects on science performance of students in top and bottom science schools are shown in Panel B and Panel D. The graphs show that the project significantly improves both mathematics and science performance of grade six students in high performance schools, Panel A and Panel B. On the contrary, the project fails to improve both mathematics and science performance of students in low performance schools, Panel C and Panel D. For grade nine students, it seems that the project takes three years to show the impacts on their mathematics and science performance.

The effects on both mathematics and science scores of grade six students in top performance schools are mostly greater than 0.1 standard deviation, except for science scores of these students in academic year 2018. Specifically, the effect sizes on mathematics (science) scores in academic years 2018, 2019 and 2020 are 0.119 (0.075), 0.105 (0.123), and 0.136 (0.119) standard deviation, respectively. Similar to the main results, the first two lines of Panel B imply that the common trend assumption hold for top science schools. As a result, it is convincing that the project significantly improves science performance of grade six students in top science schools. On the flip side, the project seems to have insignificant effects, relatively close to zero, on both mathematics and science performance of grade six students in low performance schools. The project effects on mathematics (science) performance of grade six students in bottom science schools are 0.021 (-0.001), 0.011 (0.006), and 0.004 (0.019) standard deviation, respectively. The results support recent findings where high performance and urban students benefit relatively more from the program compared to low performance and rural students (Albornoz et al. 2020; Sun and Du 2021; Cilliers et al. 2020).

The project's impacts on science performance of grade nine students are different from those of grade six students. The impacts of the project in the academic years 2018 and 2019 are not as obvious as those in the academic year 2020, the project took relatively longer to make impacts on performance of grade nine students. In contrast to the main results, the common trend assumption can be assumed in the scores of 9^{th}-grade students. As a result, it is compelling that the project has significantly positive impacts on mathematics and science performance of students in top performance schools after three years. However, the impacts are not as large as those on grade six students. For the academic year 2020, the effects on mathematics and science performance are both 0.06 standard deviation. Similar to grade six results, the project has null or even negative effects on science performance of grade nine students in low performance schools.

Similar results of subgroup analysis on mathematics and science scores of grade six and nine students in high and low performance schools, grouped by pre-treatment mathematics scores, i.e. academic years 2015 – 2017, are discussed in robustness check section (Figure 3.6 and Figure 3A-4).

3.6 Robustness Check

3.6.1 Unweighted estimates

Similar results without test taker weights obtained from DID estimation are presented in Figure 3A-1 in Appendix. The project improves mathematics (science) performance of grade six students by 0.049 (0.02), 0.021 (0.036), and 0.06 (0.038) standard deviation in academic year 2018, 2019, and 2020, respectively. The impacts on mathematics (science) performance that are greater than 0.1 (0.08) standard deviation can be ruled out. The differences in average test scores grouped by treatment status, shown in Figure 3A-3 in Appendix, are much smaller, especially those of grade six students, when I take average without test taker weights. The common trend assumption can be assumed according to the graphs; as a result, it is convincing to say that the results are robust.

3.6.2 High vs Low Performance Schools

This subsection reports results from two different subgroups to check whether the results in subsection 5.2 are robust when the schools were grouped by pre-treatment mathematics scores instead of science scores. Results for grade six students shown in Figure 3.6 suggest that the project successfully improves mathematics and science scores of students in top performance schools grouped by mathematics score; it, however, fails to improve the scores of students in bottom performance schools. To be more accurate, the project significantly increases mathematics scores of students in top mathematics schools by at least 0.1 standard deviation (Panel A), since the first year of the implementation, while it increases their science scores only by 0.01 – 0.12 standard deviation (Panel B). Conversely, the project has null or even slightly negative impacts on both mathematics and science performance of students in bottom performance schools (Panel C and Panel D). These results are consistent with those reported in subsection 5.2. As a result, it is convincing to conclude that the project significantly improves academic performance of grade six students in top performance schools while it yields either no or slightly negative impacts on academic performance of grade six students in bottom performance schools.

Results on test scores of grade nine students, shown in Figure 3A-4 in Appendix, are compatible with those of grade nine students reported in Figure 3.7 in subsection 5.2, with only a small difference in their effect sizes. The project does not significantly affect academic performance of grade nine students both in top and bottom performance schools.

3.6.3 Change of base year

One common issue that should be considered is whether using different base years alters the results; therefore, I run the same regression with test take weight average using different base years. The project impacts on mathematics and science performance of grade six and nine students, using academic year 2016 and 2017 as base line, are discussed as follows. Focusing on base year 2016, Figure 3.8 presents overall effects on mathematics and science performance of students in both grades. The results are very similar to those results reported in subsection 5.1. Overall, the project has small impacts on performance of grade six students, the effects that are greater than 0.08 standard deviation can be ruled out (Panel A and Panel B in Figure 3.8). The impacts on performance of grade nine students cannot be concluded since the common trend assumption cannot be assumed (Panel C and Panel B in Figure 3.8). Figure 3A-5, in Appendix, demonstrates the robustness of the results when using academic year 2017 as base line.

The heterogeneity of project impacts on grade six students in top and bottom science schools is reported in Figure 3A-5. Students in top science schools significantly benefit from the project, their performance is, on average, 0.1 standard deviation higher both in mathematics and science scores when compared to students in control schools (Panel A and Panel B in Figure 3A-5). On the other hand, the project has null effects on mathematics and science performance of grade six students in bottom science schools (Panel C and Panel D in Figure 3A-5). Similar results using academic year 2017 as base line are shown in Figure 3A-6 in Appendix.

The effects of SMT project on performance of grade nine students in top and bottom science schools, using academic year 2016 and 2017 as base line, are presented in Appendix (Figure 3A-7 and Figure 3A-8). Regardless of group the school students are in, the project does not significantly boost either their mathematics or their science performance. Specifically, students in top science schools benefit slightly, about 0.07 standard deviation, in academic year 2020, which is the third year of the implementation. Students in bottom science schools obviously has no gain from the implementation.

The effects afterwards are robust even with a different base line. To conclude, the project positively affects performance of grade six students; however, it does not significantly improve performance of grade nine students. The impacts on grade six students in top and bottom science schools are heterogeneous, students in top science schools successfully increase their mathematics and science scores while students in bottom science schools fail to improve their performance.

3.6.4 Lee bounds

The national test policy change in academic year 2020 stated that tests have become voluntary for Thai students, including grade six and nine students. As a result, number of schools and students participating the test dropped significantly in the academic year 2020 – the tests were taken around March 2021. The policy change affected the test participation rates[12] directly; overall participation rate of grade six students[13] decreases by 18%, on average. The average participation rates of students in treatment and control groups are 83% and 80%, respectively. To deal with sample selection problem, we can assume two extreme assumptions – bottom or top performance students did not participate in the 2020 national tests – and correspondingly trim either bottom or top performance students in treatment group by the number, $q = (q_T - q_C)/q_T$, suggested by Lee (2009) and Tauchmann H. (2014).

In this study, by using the participation rates from the two groups, the suggested number, q, becomes 3.65%. We can assume that additional 3.65% of examinees in the treatment group are either the weakest or the strongest students, and trim 3.65% of students in either the bottom or the top of the distribution at individual level. Under those circumstances, the DID estimates for the academic year 2020 explain the effects of the project as if treatment and control groups are subjected to the same participant ratio. The

[12] Ratios of grade six test takers to grade six students.
[13] Lee Bounds in this study focuses on analysis of grade six students because the project had impacts on performance of grade six students but null effects on performance of grade nine students.

results, shown in Figure 3.10, compare the two new estimates to the original estimate of academic year 2020. The first line in each panel represents the original estimate where test scores of neither top nor bottom performance students in treatment group were dropped from the sample. The second (third) line in each panel represents the new estimate where test scores of students in top (bottom) of the distribution were dropped from the sample. Specifically, the trimming processes were done separately by subject, for example, Bottom Math (or BM) is a trimmed sample where mathematics scores of students in the lowest 3.65% of the distribution were dropped from the sample, and Top Science (or TS) is a trimmed sample where science scores of students in the highest 3.65% of the distribution were dropped from the sample.

The results, before and after test scores were trimmed, are heterogeneous. In general, dropping test scores of the weakest students, i.e. trimming the scores of students in the lowest 3.65% of the distribution, in treatment group strengthens the impact of SMT project on student performance. On the contrary, dropping test scores of the strongest students, i.e. trimming the scores of students in the highest 3.65% of the distribution, in the treatment group generally reverts the impact of the project on student performance. For example, the project improved student's mathematics performance slightly by less than 0.05 standard deviation in Panel A in the original result, without trimming. In comparison, the project boosted their performance significantly, over 0.1 standard deviation, when mathematics scores of students in the bottom of the distribution were trimmed. However, it worsens their performance, by at least 0.1 standard deviation, when the scores of students in the top of the distribution were trimmed. Similar effects are found in overall mathematics, overall science, mathematics of top science schools, and science of top science schools. Results from bottom science schools are slightly incompatible, especially when test scores of students in the top of the distribution were trimmed. Even though mathematics/science scores of students in the top of the distribution were dropped from the sample, the project did not negatively affect students' performance, i.e. the results are consistent with original results, without trimming.

3.7 Why was the project not more successful?

The SMT project only slightly improved performance of grade six students if at all, with the possible exception of students from top performance schools. According to Schaffner et al. (2021), at-scale government teacher training in Nepal had little or no impact on student learning due to the following factors: training session and central-level efforts, disappointingly low participation rates, weak accountability and motivation, deficit in post-training support, and weakness in teachers' and students' knowledge. In order to gather informative evidence from SMT participants, the online survey was conducted from 80 SMT teachers in 35 provinces during October – November 2021. In this section, I will discuss those issues in context of SMT project.

The training sessions, first of all, have been conducted since spring 2020 where lists of training courses were available for registration. Nevertheless, most of the training sessions were held online amid the Covid-19 pandemic and a series of online sessions, during May 2020 – February 2021, only included about 6,000 teachers from treatment schools. Cooperation among central educational institutes aimed to provide great impacts through the sessions; however, it is questionable whether a series of sessions with a limit of about 6,000 teachers is effectively extensive. Moreover, the majority, about 83%, of the participants reported that they attended online sessions with top-down lecture method, which is mostly one-way communication. The communication does not generously allow participants to raise questions during the sessions; therefore, it is up to them to research and clarify their inquiries themselves. Only about 20 – 30% of the participants experienced group discussion and group activities during the sessions.

Secondly, according to the survey, the average participation rate of the eligible SMT teachers is only about 60% while the ideal participation rate should be 100%. Participants revealed that they only attended either one training session, 70% of them, or two training sessions, 30% of them. The majority of participants attended the training sessions during May – August 2020, with 14 training hours, on average. This brings in another question- whether an average of 14 training hours adequately and significantly enhances teachers' knowledge and skills.

Speaking of accountability and motivation, SMT teachers may not be satisfyingly responsible for and enthusiastic about the newly introduced contents. Although about 76% of participants said the contents were applicable and were adapted accordingly, about 72% of them revealed that implementing SMT contents in schools was an obligation, depending on each school policy. As a result, about half of the participants did not have any incentive to apply the new contents in classes even though some of them would apply the content expecting academic and salary promotion. The major hindrances are not only motivation issues but also heterogeneity in applying the contents at school level, i.e. some schools applied the contents in all classrooms while the others only applied in a small number of classrooms in specific grades. After the online session, participants were asked to do and submit some specific assignments given by the lecturers; however, we do not have information regarding the assignments.

Accountability and motivation issues in Nepal's project also seem to have developed in the Thai SMT project.

In addition to the above three factors, insufficient materials and support are problems in SMT project as reported in the survey: 54% of trained teachers received training materials; only 11% of them received teaching materials; only 28% of them received laboratory equipment; and no other supporting materials were provided. Moreover, contents provided in the sessions do not supplement ready-to-implement teaching plans. SMT teachers would take some time designing their lesson plans before they can apply the new contents to the classes.

Last but not the least, the group of SMT authorities who take care of the project is relatively small compared to the size of the project. As a result, monitoring and follow-up process can be a challenge both before and after the Covid-19 outbreak. Tracking real implementation in SMT schools can; therefore, become problematic.

3.8 Conclusion

In 2018, the Institution for the Promotion of Teaching Science and Technology (IPST) together with Thai educational organizations provided supporting materials and a series of training sessions to improve quality of science, mathematics, and technology taught in SMT schools. The ultimate goals of the project were to create a better understanding in the subjects and to boost student performance. This study aims to evaluate the effects of the Science, Mathematics and Technology (SMT) school project on student performance as measured by national test scores. Science and mathematics scores of grade six and nine students, from the National Institute of Educational Testing Service (NIETS) during academic years 2015 – 2020 were used to compare performance of students in SMT and non-SMT schools. I adopted difference-in-difference design to estimate the effects of the project on students' test scores.

The overall and subgroup estimates suggest that the SMT project only improved science and mathematics scores of grade six students in the top performance schools by 0.1 standard deviation, on an average. Students in grade nine or in the bottom performance schools did not benefit much from the project. These estimates are robust to change of base lines; only grade six students in top performance schools benefited from the project. A sample selection issue arose in 2020 due to the policy change in national test, i.e. the test was no longer compulsory. I used the method suggested by Lee (2009) and Tauchmann H. (2014) to obtain DID estimates for academic year 2020 when treatment and control groups were subjected to the same participating ratio; however, the last year effects were not very informative. The effects from at-scale teacher PD intervention in Thailand on student performance are similar to other at-scale projects in China, Nepal, and Ecuador (Loyalka, et al. 2019, Sun and Du 2021, Schaffner, et al. 2021, Carneiro, et al. 2022).

This study conducted, apart from quantitative evidence, an online survey to gather information from SMT teachers who participated in training session(s). The survey revealed a few possible explanations why the project was not more effective, for instance, informative but one-way communication training sessions; unexpectedly low participation rate; weak motivation and accountability; inadequate material distribution; insufficient supports and follow-ups. The insufficiency might be due to the fact that a group of SMT authorities who take care of the project is relatively small compared to the size of the project. The problems in policy design, which arose in SMT project, are not unique; these issues also happened in projects in African and South Asian countries (Schaffner, et al. 2021, Popova, et al. 2022). In order to fill the voids and develop more effective policies, policymakers may consider the issues mentioned above.

This study has some limitations due to data availability. First of all, this study only explained the effects of the program on performance of students in schools under Office of the Basic Education Commission (OBEC) and Bangkok affiliation. In other words, it did not include effects on performance of student in schools under Office of the Private Education Commission (OPEC). In addition, this study only presents the effects up to three years, from academic years 2018 to 2020, even though this is an ongoing project with a continuous series of training sessions. Lastly, the results in this study only explains effects of SMT project on student performance at aggregate level, i.e. at school level. It is possible that the null effects at school level are net effects in which gain by some students is cancelled out by loss of some others within the same schools. Therefore, further study may explore effects of SMT project on higher moments to confirm that the project has no effect, or only little effect if any, on student performance assessed at school level.

CHAPTER FOUR

Conclusion and Policy Implications

4.1 Conclusion

In 2017 and 2018, ALS and SMT projects have been implemented by Thailand's Ministry of Education, to improve academic achievement of students. The projects combine concepts of cooperative learning and teacher professional development, two popular academic interventions. This study evaluates the effects of the projects on student performance as measured by national test scores. The effects, estimated using a fuzzy regression discontinuity design (for ALS project) and difference-in-difference (for SMT project), suggest that overall, the projects did not substantially improve student performance when considering the effects at school level. Our findings regarding the effects of large scale academic intervention in Southeast Asia are consistent with studies conducted by, for example, Loyalka et al. (2019) and Thurston et al. (2019). However, the SMT project improved science and mathematics scores of grade six students in the top performance schools by 0.1 standard deviation, on an average.

In addition to quantitative evidence, this study conducted an online survey to gather information from SMT teachers who participated in training session(s). The survey revealed a few possible explanations why the project was not more effective, for instance, informative but one-way communication training sessions; unexpectedly low participation rate; weak motivation and accountability; inadequate material distribution; insufficient support and follow-ups. The problems in policy design, which arose in the SMT project, are not unique; these issues also existed in projects in African and South Asian countries (Schaffner, et al. 2021, Popova, et al. 2022).

4.2 Policy Implications

The findings in this study contribute some useful policy implications to policymakers and project implementers, including teachers, supervisors, and education professionals. In order to fill the voids and develop more effective policies, policymakers may consider the following details for future policy designs. Firstly, it might be helpful to include more practical lessons and ready-to-implement teaching plans in in-service training sessions, as the lessons and the plans could facilitate teachers' classroom implementation. Secondly, project supervisors may want to consider whether the number of trained teachers in each school should be proportional to the school size. When the number of teachers is relatively small compared to the number of total teachers in the same school, trained teachers may struggle to effectively transfer the knowledge to their colleagues. Another important factor is the duration of in-service training which can be one of the most challenging questions to be answered. Duration and frequency of training sessions might be limited and subjected to resources allocation both in terms of budget and human resources; nevertheless, the duration and frequency could strongly determine the effectiveness of the implementation.

In addition to the project design, the online survey reveals that the actual participation rate of eligible teachers from treatment schools (60%) is far less than the ideal participation rate (100%). The difference between a list of registration and a list of participation could be another important issue that prevents the success of the projects. Last but not the least, weak accountability and motivation diminish the training quality; and should therefore be prioritized and taken into consideration.

4.3 Limitation and future research

This study has some limitations due to data availability. First of all, this study only explained the effects of the program on performance of students in schools under Office of the Basic Education Commission (OBEC) and Bangkok affiliation. Moreover, this study only examined the effects at an aggregate level, i.e. school level, since specific data on trained teachers and their classes are not available. The null effects of the projects at school level do not necessarily imply no effects since the effects could represent the net effects in which gain by some students is cancelled out by loss of some others within the same schools. Lastly, other scores and/or indicators are not available to be accessed and evaluated except the national test scores, which guarantee the same evaluation standard countrywide. The scores used in this study are the best available option, yet they might not be the perfect indicator to capture effects of the projects on student performance.

The followings are recommendations for future research work. Firstly, the effects of the projects on higher moments, for example, variance of the test scores, should be examined in order to conclude that the projects had null effects on student performance at school level. To statistically provide evidence for ineffectiveness, share of trained teachers, supervisor-to-teacher, and student-to-teacher ratios should be taken into consideration. Alternatively, subgroup analysis, grouped by variables mentioned earlier, should be examined.

References

Abbiati, G., Argentin, G., Caputo, A., & Pennisi, A. (2021). Repetita Iuvant? A Repeated Randomized Controlled Trial on the Effectiveness of an At-Scale Teacher Professional Development Program. *Evaluation Review*, 0193841X211055354.

Albornoz, F., Anauati, M. V., Furman, M., Luzuriaga, M., Podestá, M. E., & Taylor, I. (2020). Training to teach science: experimental evidence from Argentina. *The World Bank Economic Review*, 34(2), 393-417.

Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics*, 122(3), 1235-1264.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., ... & Walton, M. (2017). From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives*, 31(4), 73-102.

Bold, T., Kimenyi, M., Mwabu, G., & Sandefur, J. (2018). Experimental evidence on scaling up education reforms in Kenya. *Journal of Public Economics*, 168, 1-20.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, 17(2), 372-404

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, 23(2), 192-210.

Carneiro, P., Cruz-Aguayo, Y., Intriago, R., Ponce, J., Schady, N., & Schodt, S. (2022). When promising interventions fail: Personalized coaching for teachers in a middle-income country. *Journal of Public Economics Plus*, 100012.

Cilliers, J., Fleisch, B., Prinsloo, C., & Taylor, S. (2020). How to improve teaching practice? An experimental comparison of centralized training and in-classroom coaching. *Journal of Human Resources*, 55(3), 926-962.

Demie, S., Seyoum, Y., & Tsehayu, W. (2019). Cooperative Learning as a Window of Opportunity to Transact Mathematics Instruction in Alamata and Korem Secondary Schools of Tigray, Ethiopia. *International Journal of Education and Literacy Studies*, 7(4), 136-143.

Dietrichson, J., Bøg, M., Filges, T., & Klint Jørgensen, A. M. (2017). Academic interventions for elementary and middle school students with low socioeconomic status: A systematic review and meta-analysis. *Review of Educational Research*, 87(2), 243-282.

Genç, M. (2016). An evaluation of the cooperative learning process by sixth-grade students. *Research in Education*, 95(1), 19–32.

Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1), 201-209.

Imbens, G. and J. D. Angrist. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica,* 62 (2), 467-475.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of econometrics*, 142(2), 615-635.

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

Jacob, B. A., & Lefgren, L. (2004). The impact of teacher training on student achievement quasi-experimental evidence from school reform efforts in Chicago. *Journal of Human Resources*, 39(1), 50-79.

Johnson, D. W., & Johnson, R. T. (1999). Making cooperative learning work. *Theory into practice*, 38(2), 67-73.

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational researcher*, 38(5), 365-379.

Johnson, D. W., & Johnson, R. T. (2018). Cooperative learning: The foundation for active learning. *Active Learning—Beyond the Future*.

Kramarski, B., Mevarech, Z. R., & Lieberman, A. (2001). Effects of multilevel versus unilevel metacognitive training on mathematical reasoning. *The Journal of Educational Research*, 94(5), 292-300

Kramarski, B., & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal*, 40(1), 281-310.

Kyndt, E., Raes, E., Lismont, B., Timmers, F., Cascallar, E., & Dochy, F. (2013). A meta-analysis of the effects of face-to-face cooperative learning. Do recent studies falsify or verify earlier findings?. *Educational research review*, 10, 133-149.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3), 1071-1102.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of economic literature*, 48(2), 281-355.

Loyalka, P., Popova, A., Li, G., & Shi, Z. (2019). Does teacher training actually work? Evidence from a large-scale randomized evaluation of a national teacher training program. *American Economic Journal: Applied Economics*, 11(3), 128-54.

Mbiti, I. M. (2016). The need for accountability in education in developing countries. *Journal of Economic Perspectives*, 30(3), 109-32.

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

Popova, A., Evans, D. K., Breeding, M. E., & Arancibia, V. (2022). Teacher professional development around the world: The gap between evidence and practice. *The World Bank Research Observer*, 37(1), 107-136.

Schaffner, J., Glewwe, P., & Sharma, U. (2021). Why Programs Fail: Lessons for Improving Public Service Quality from a Mixed-Methods Evaluation of an Unsuccessful Teacher Training Program in Nepal (No. 1701-2021-3437).

Sun, R., & Du, P. (2021). Does teacher training narrow the educational gap between urban and rural students? Empirical evidence from ceps baseline data. *Best Evidence in Chinese Education*, 9(2), 1243-1261.

Tan, I. G. C., Sharan, S., & Lee, C. K. E. (2007). Group investigation effects on achievement, motivation, and perceptions of students in Singapore. *The Journal of Educational Research*, 100(3), 142-154.

Tanguthaisuk, B. (2020). King's Philosophy training for schools in the southmost provinces of Thailand. Ministry of Education Report.

Tauchmann, H. (2014). Lee (2009) treatment-effect bounds for nonrandom sample selection. *The Stata Journal*, 14(4), 884-894.

Thanh, P. T. H., Gillies, R., & Renshaw, P. (2008). Cooperative Learning (CL) and Academic Achievement of Asian Students: A True Story. *International education studies*, 1(3), 82-88.

Thurston, A., Cockerill, M., & Craig, N. (2019). Using cooperative learning to close the reading attainment gap for students with low literacy levels for grade 8/year 9 students. *International Journal of Educational Research*, 94, 1-10.

Tracey, L., Madden, N. A., & Slavin, R. E. (2010). Effects of co-operative learning on the mathematics achievement of Years 4 and 5 pupils in Britain: a randomized control trial. *Effective Education,* 2(1), 85–97.

Tymms, P., Merrell, C., Thurston, A., Andor, J., Topping, K., & Miller, D. (2011). Improving attainment across a whole district: school reform through peer tutoring in a randomized controlled trial. *School effectiveness and school improvement*, 22(3), 265-289.

Van der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: A regression–discontinuity approach. *International Economic Review*, 43(4), 1249-1287.

Van der Klaauw, W. (2008). Regression–discontinuity analysis: a survey of recent developments in economics. *Labour*, 22(2), 219-245.

Villegas-Reimers, E. (2003). Teacher professional development: an international review of the literature. *Paris: International Institute for Educational Planning.*

Table 2. 1: Training Session Timetable

Day 1	Day 2	Day 3
Registration, Orientation	Active Learning Concept	Design Teaching Plan
Pre-test, King's Philosophy	Graphic Organizers	Academic Assessment
Brain Development	Design Teaching Plan	Post-test
Active Learning Concept		

Source: the report on development of teacher training curricula for the ALS project (south)

Table 2. 2: The Balance Test

	Constant estimate	Discontinuity estimate
Number of teachers	11.69***	-0.496
	(0.769)	(1.170)
Number of students	206.9***	-27.72
	(18.21)	(29.28)
Number of classrooms	10.06***	-0.751
	(0.423)	(0.743)
Student-teacher ratio	17.23***	-0.731
	(0.781)	(0.661)
Raw O-NET score	29.10***	0.0254
	(0.576)	(0.253)

Note: Linear OLS estimations of 2016 covariates following specification in equation (3) shown in estimation approach.
Standard errors in parentheses are clustered at district level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2. 3: Descriptive Statistics

	Variables	Control			Treatment		
		Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.
Grade 1							
2018	Speaking	50.731	29.229	20533	43.113	27.437	3395
	Reading	63.797	20.211	20533	58.894	20.717	3395
2019	Speaking	50.398	30.003	19989	44.844	27.811	3056
	Reading	64.377	21.478	19989	60.277	21.579	3056
Grade 3							
2017	Literacy	38.669	16.827	22246	35.041	14.722	3628
	Numeracy	29.194	14.506	22246	27.243	12.649	3628
	Logic	33.453	16.015	22246	30.299	13.989	3628
2018	Literacy	37.26	18.001	21455	32.465	14.628	3475
	Numeracy	33.972	16.952	21455	30.96	14.555	3475
	Logic	33.486	16.408	21455	29.511	13.807	3475
2019	Literacy	31.937	16.359	21514	28.961	14.037	3678
	Numeracy	32.174	15.676	21514	28.434	13.01	3678
Grade 6							
2016	Thai	41.652	15.37	23634	35.496	13.509	4055
	English	30.416	12.577	23634	27.613	9.413	4055
	Math	31.42	15.595	23634	26.443	12.504	4055
	Science	35.418	10.753	23634	32.292	9.382	4055
2017	Thai	36.13	13.752	23336	32.151	11.981	3960
	English	29.812	12.171	23336	28.66	10.726	3960
	Math	28.647	13.509	23336	25.917	12.009	3960
	Science	32.36	10.652	23336	30.384	9.848	3960
2018	Thai	41.85	16.524	23057	37.178	14.702	3762
	English	31.215	12.086	23057	29.468	9.844	3762
	Math	26.173	15.026	23057	23.32	12.079	3762
	Science	32.441	10.908	23057	30.35	9.808	3762
2019	Thai	36.824	14.315	22749	32.622	12.439	3827
	English	27.818	10.481	22749	26.109	7.755	3827
	Math	24.655	12.058	22749	22.371	10.523	3827
	Science	28.758	10.479	22749	27.478	9.121	3827

Note: Average scores and standard deviations computed from individual test scores from the National Institute of Educational Testing Service (NIETS), Thailand.

Table 2. 4: Effects of ALS project on student performance (without test taker weight)

	Coef.	Std.Err.	[95%Conf.	Interval]	Order	Dist. FE	Test FE	Obs.
All grade,	0.013	0.045	-0.076	0.101	1	Yes	No	5287
All year	0.012	0.046	-0.077	0.101	1	Yes	Yes	5239
	-0.015	0.055	-0.123	0.094	2	Yes	No	8019
	-0.009	0.053	-0.113	0.095	2	Yes	Yes	7971
All grade	0.029	0.051	-0.071	0.129	1	Yes	No	1558
in 2017	0.029	0.051	-0.071	0.129	1	Yes	Yes	1558
	-0.005	0.053	-0.110	0.099	2	Yes	No	2846
	-0.004	0.053	-0.109	0.100	2	Yes	Yes	2846
All grade	-0.004	0.058	-0.118	0.109	1	Yes	No	2166
in 2018	-0.007	0.057	-0.119	0.104	1	Yes	Yes	2175
	0.029	0.072	-0.113	0.170	2	Yes	No	2994
	0.029	0.072	-0.112	0.171	2	Yes	Yes	2994
All grade	0.009	0.067	-0.121	0.140	1	Yes	No	1676
in 2019	-0.007	0.068	-0.141	0.126	1	Yes	Yes	1692
	0.011	0.072	-0.130	0.151	2	Yes	No	2544
	0.008	0.072	-0.132	0.148	2	Yes	Yes	2560
Grade 1	-0.082	0.143	-0.362	0.198	1	Yes	No	1256
	-0.082	0.143	-0.362	0.197	1	Yes	Yes	1256
	-0.065	0.132	-0.323	0.193	2	Yes	No	1640
	-0.065	0.132	-0.323	0.193	2	Yes	Yes	1640
Grade 3	0.034	0.098	-0.159	0.227	1	Yes	No	1927
	0.034	0.099	-0.160	0.228	1	Yes	Yes	1935
	0.016	0.109	-0.197	0.229	2	Yes	No	3007
	0.021	0.110	-0.195	0.237	2	Yes	Yes	3015
Grade 6	0.019	0.031	-0.041	0.080	1	Yes	No	2896
	0.019	0.031	-0.041	0.079	1	Yes	Yes	2896
	0.007	0.040	-0.072	0.085	2	Yes	No	3864
	0.006	0.040	-0.072	0.085	2	Yes	Yes	3864

Note: Linear and quadratic RD estimates on equation (1). All grades in all years use all test scores from 24 tests during 2017-2019. All grades in 2017 use the test scores from the 7 tests in 2017. All grades in 2018 use the test scores from the 9 tests in 2018. All grades in 2019 use the test scores from 8 tests in 2019. Grade 1 all years use the test scores from the 4 tests during 2018-2019. Grade 3 all years use the test scores from the 8 tests during 2017-2019. Grade 6 all years use the test scores from the 12 tests during 2017-2019. Order refers to polynomial order and Dist. FE refers to district fixed effect.

Table 3. 1: Descriptive Statistics of Grade 6 Students

		Control			Treatment		
Year	Variables	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.
2015	Math	39.729	18.606	503731	46.82	21.512	53662
	Science	40.113	12.482	503731	44.798	14.221	53662
2016	Math	36.68	17.107	512039	43.507	20.297	55654
	Science	38.971	11.272	512039	43.092	13.175	55654
2017	Math	33.662	14.398	497121	39.545	18.243	55784
	Science	36.723	11.104	497121	40.663	12.589	55784
2018	Math	33.282	17.968	489077	41.03	22.483	55932
	Science	37.293	11.753	489077	41.601	13.417	55932
2019	Math	29.862	13.965	482214	35.413	17.484	55684
	Science	32.853	12.336	482214	37.867	15.495	55684
2020	Math	27.304	12.741	338954	32.699	17.561	44771
	Science	36.152	12.609	338954	41.147	15.003	44771

Note: Average scores and standard deviations computed from individual test scores from the National Institute of Educational Testing Service (NIETS), Thailand.

Table 3. 2: Descriptive Statistics of Grade 9 Students

		Control			Treatment		
Year	Variables	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.
2015	Math	31.299	14.222	381188	34.276	15.87	133080
	Science	36.824	12.874	381188	39.48	14.152	133080
2016	Math	28.211	14.868	370780	31.826	16.801	130103
	Science	34.356	9.97	370780	36.298	10.867	130103
2017	Math	25.393	15.524	369792	28.518	17.38	131070
	Science	31.784	9.535	369792	33.444	10.191	131070
2018	Math	29.049	15.187	370283	32.173	17.097	132374
	Science	35.469	10.728	370283	37.648	11.376	132374
2019	Math	25.871	15.044	379788	28.757	16.934	137220
	Science	29.764	8.421	379788	30.839	8.889	137220
2020	Math	24.636	14.184	197241	28.279	16.505	67722
	Science	29.45	9.317	197241	31.47	10.316	67722

Note: Average scores and standard deviations computed from individual test scores from the National Institute of Educational Testing Service (NIETS), Thailand.

Table 3. 3: Common trend assumption check for grade 6 students

	(1) Mathematics	(2) Science	(3) Mathematics	(4) Science
SMT*2015			-0.0231*	0.0175
			(0.0140)	(0.0130)
SMT*2016	0.0130	-0.0106	-0.0101	0.00699
	(0.0118)	(0.0115)	(0.0128)	(0.0124)
SMT*2017	0.0231*	-0.0175		
	(0.0140)	(0.0130)		
N	2805869	2805869	2805869	2805869
R-squared	0.0658	0.0687	0.0658	0.0687

Note: School-level DID estimates based on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2015, 2016, 2017} are academic year dummies. All columns used test scores of grade six students during 2015-2020. Columns (1) and (2) used 2015 as base year while columns (3) and (4) used 2017 as base year. Standard errors in parentheses are clustered at school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3. 4: Common trend assumption check for grade 9

	(1) Mathematics	(2) Science	(3) Mathematics	(4) Science
SMT*2015			0.00718	0.0304***
			(0.00798)	(0.00884)
SMT*2016	0.0289***	-0.0135*	0.0361***	0.0168**
	(0.00688)	(0.00756)	(0.00627)	(0.00728)
SMT*2017	-0.00718	-0.0304***		
	(0.00798)	(0.00884)		
N	2561872	2561872	2561872	2561872
R-squared	0.0246	0.0276	0.0246	0.0276

Note: School-level DID estimates based on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2015, 2016, 2017} are academic year dummies. All columns used test scores of grade nine students during 2015-2020. Columns (1) and (2) used 2015 as base year while columns (3) and (4) used 2017 as base year. Standard errors in parentheses are clustered at school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3. 5: Effects of SMT project on mathematics performance of grade 6 students (base year: 2015)

	(1) Overall	(2) Top Math Schools	(3) Bottom Math Schools	(4) Top Science Schools	(5) Bottom Science Schools
SMT	0.379***	0.123***	0.0979***	0.162***	0.0633**
	(0.0333)	(0.0366)	(0.0214)	(0.0380)	(0.0252)
2016	-0.00186	-0.0129**	0.0132***	-0.0135**	0.0191***
	(0.00302)	(0.00610)	(0.00322)	(0.00635)	(0.00404)
2017	-0.00566*	-0.0475***	0.0413***	-0.0365***	0.0429***
	(0.00331)	(0.00727)	(0.00335)	(0.00740)	(0.00437)
2018	-0.00780**	-0.122***	0.109***	-0.0774***	0.0816***
	(0.00356)	(0.00885)	(0.00378)	(0.00841)	(0.00462)
2019	-0.00672*	-0.176***	0.145***	-0.118***	0.115***
	(0.00363)	(0.00888)	(0.00403)	(0.00854)	(0.00485)
2020	-0.0252***	-0.258***	0.196***	-0.185***	0.167***
	(0.00442)	(0.0120)	(0.00399)	(0.0113)	(0.00502)
SMT*2016	0.0130	0.0334**	-0.0131	0.0351**	0.000908
	(0.0118)	(0.0160)	(0.0161)	(0.0156)	(0.0201)
SMT*2017	0.0231*	0.0616***	-0.0132	0.0551***	0.0265
	(0.0140)	(0.0196)	(0.0175)	(0.0190)	(0.0225)
SMT*2018	0.0446***	0.156***	-0.0335	0.120***	0.0211
	(0.0136)	(0.0195)	(0.0213)	(0.0185)	(0.0232)
SMT*2019	0.00997	0.143***	-0.0364	0.105***	0.0105
	(0.0146)	(0.0213)	(0.0241)	(0.0201)	(0.0254)
SMT*2020	0.0207	0.187***	-0.0196	0.136***	0.00396
	(0.0196)	(0.0286)	(0.0223)	(0.0272)	(0.0239)
cons	-0.0205***	0.503***	-0.444***	0.403***	-0.412***
	(0.00715)	(0.0131)	(0.00363)	(0.0145)	(0.00506)
N	2805869	941108	878067	985218	742425
R-squared	0.0658	0.0701	0.0846	0.0546	0.0413

Note: School-level DID estimates based on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2016, 2017, 2018, 2019, 2020} are academic year dummies. All columns used mathematics test scores of grade six students during 2015-2020. Standard errors in parentheses are clustered at school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3. 6: Effects of SMT project on science performance of grade 6 students (base year: 2015)

	(1) Overall	(2) Top Math Schools	(3) Bottom Math Schools	(4) Top Science Schools	(5) Bottom Science Schools
SMT	0.373***	0.210***	0.118***	0.141***	0.0699***
	(0.0307)	(0.0366)	(0.0291)	(0.0343)	(0.0217)
2016	0.000176	-0.00293	0.00312	-0.00702	0.00366
	(0.00309)	(0.00638)	(0.00419)	(0.00604)	(0.00369)
2017	-0.00162	-0.0136**	0.0168***	-0.0225***	0.0116***
	(0.00323)	(0.00669)	(0.00443)	(0.00612)	(0.00382)
2018	-0.00143	-0.0534***	0.0573***	-0.0961***	0.0943***
	(0.00339)	(0.00761)	(0.00455)	(0.00748)	(0.00433)
2019	-0.00499	-0.0729***	0.0797***	-0.116***	0.132***
	(0.00359)	(0.00864)	(0.00451)	(0.00876)	(0.00429)
2020	-0.0212***	-0.118***	0.0744***	-0.162***	0.132***
	(0.00396)	(0.00925)	(0.00510)	(0.00910)	(0.00496)
SMT*2016	-0.0106	0.0121	-0.0493***	0.0143	0.000971
	(0.0115)	(0.0154)	(0.0178)	(0.0146)	(0.0167)
SMT*2017	-0.0175	-0.00612	-0.0348*	-0.00120	-0.00145
	(0.0130)	(0.0180)	(0.0208)	(0.0170)	(0.0172)
SMT*2018	-0.0105	0.0410**	-0.0447*	0.0746***	-0.00983
	(0.0130)	(0.0177)	(0.0229)	(0.0171)	(0.0225)
SMT*2019	0.0219	0.0925***	-0.0472*	0.123***	0.00589
	(0.0147)	(0.0207)	(0.0276)	(0.0204)	(0.0215)
SMT*2020	0.0105	0.0755***	-0.00244	0.119***	0.0195
	(0.0160)	(0.0221)	(0.0292)	(0.0213)	(0.0289)
cons	-0.0224***	0.346***	-0.333***	0.403***	-0.418***
	(0.00605)	(0.0121)	(0.00479)	(0.0106)	(0.00415)
N	2805869	941108	878067	985218	742425
R-squared	0.0687	0.0674	0.0160	0.0626	0.0519

Note: School-level DID estimates based on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2016, 2017, 2018, 2019, 2020} are academic year dummies. All columns used science test scores of grade six students during 2015-2020. Standard errors in parentheses are clustered at school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3. 7: Effects of SMT project on mathematics performance of grade 6 students (base year: 2016)

	(1) Overall	(2) Top Math Schools	(3) Bottom Math Schools	(4) Top Science Schools	(5) Bottom Science Schools
SMT	0.392^{***}	0.157^{***}	0.0847^{***}	0.197^{***}	0.0642^{**}
	(0.0359)	(0.0410)	(0.0205)	(0.0418)	(0.0259)
2017	-0.00380	-0.0346^{***}	0.0281^{***}	-0.0230^{***}	0.0238^{***}
	(0.00291)	(0.00596)	(0.00304)	(0.00636)	(0.00394)
2018	-0.00594^{*}	-0.109^{***}	0.0957^{***}	-0.0639^{***}	0.0625^{***}
	(0.00317)	(0.00779)	(0.00335)	(0.00740)	(0.00424)
2019	-0.00486	-0.163^{***}	0.132^{***}	-0.104^{***}	0.0963^{***}
	(0.00337)	(0.00802)	(0.00369)	(0.00777)	(0.00451)
2020	-0.0233^{***}	-0.245^{***}	0.183^{***}	-0.172^{***}	0.148^{***}
	(0.00421)	(0.0112)	(0.00377)	(0.0107)	(0.00476)
SMT*2017	0.0101	0.0282	-0.0000512	0.0200	0.0256
	(0.0128)	(0.0176)	(0.0146)	(0.0169)	(0.0198)
SMT*2018	0.0316^{***}	0.123^{***}	-0.0203	0.0845^{***}	0.0202
	(0.0117)	(0.0165)	(0.0182)	(0.0155)	(0.0204)
SMT*2019	-0.00304	0.110^{***}	-0.0233	0.0700^{***}	0.00955
	(0.0131)	(0.0187)	(0.0211)	(0.0179)	(0.0224)
SMT*2020	0.00769	0.153^{***}	-0.00648	0.101^{***}	0.00305
	(0.0185)	(0.0269)	(0.0218)	(0.0253)	(0.0243)
cons	-0.0224^{***}	0.490^{***}	-0.431^{***}	0.390^{***}	-0.393^{***}
	(0.00735)	(0.0139)	(0.00332)	(0.0153)	(0.00474)
N	2319261	782338	722074	819203	610092
R-squared	0.0692	0.0717	0.0755	0.0578	0.0367

Note: School-level DID estimates based on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2017, 2018, 2019, 2020} are academic year dummies. All columns used mathematics test scores of grade six students during 2016-2020. Standard errors in parentheses are clustered at school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3. 8: Effects of SMT project on science performance of grade 6 students (base year: 2016)

	(1) Overall	(2) Top Math Schools	(3) Bottom Math Schools	(4) Top Science Schools	(5) Bottom Science Schools
SMT	0.363***	0.222***	0.0684***	0.156***	0.0708***
	(0.0319)	(0.0376)	(0.0248)	(0.0356)	(0.0212)
2017	-0.00180	-0.0107	0.0137***	-0.0155**	0.00791**
	(0.00312)	(0.00662)	(0.00412)	(0.00643)	(0.00355)
2018	-0.00161	-0.0505***	0.0542***	-0.0891***	0.0906***
	(0.00330)	(0.00733)	(0.00430)	(0.00751)	(0.00416)
2019	-0.00517	-0.0700***	0.0766***	-0.109***	0.129***
	(0.00347)	(0.00837)	(0.00433)	(0.00862)	(0.00412)
2020	-0.0214***	-0.115***	0.0713***	-0.155***	0.128***
	(0.00387)	(0.00895)	(0.00500)	(0.00905)	(0.00482)
SMT*2017	-0.00699	-0.0182	0.0145	-0.0155	-0.00243
	(0.0124)	(0.0172)	(0.0190)	(0.0161)	(0.0169)
SMT*2018	0.0000197	0.0289*	0.00467	0.0602***	-0.0108
	(0.0115)	(0.0161)	(0.0201)	(0.0156)	(0.0208)
SMT*2019	0.0324**	0.0804***	0.00216	0.109***	0.00492
	(0.0142)	(0.0204)	(0.0242)	(0.0199)	(0.0196)
SMT*2020	0.0210	0.0634***	0.0469*	0.104***	0.0185
	(0.0160)	(0.0221)	(0.0269)	(0.0214)	(0.0290)
cons	-0.0222***	0.343***	-0.330***	0.396***	-0.414***
	(0.00621)	(0.0126)	(0.00460)	(0.0114)	(0.00387)
N	2319261	782338	722074	819203	610092
R-squared	0.0701	0.0698	0.0145	0.0640	0.0485

Note: School-level DID estimates based on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2017, 2018, 2019, 2020} are academic year dummies. All columns used science test scores of grade six students during 2016-2020. Standard errors in parentheses are clustered at school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3. 9: Effects of SMT project on mathematics performance of grade 6 students (base year: 2017)

	(1) Overall	(2) Top Math Schools	(3) Bottom Math Schools	(4) Top Science Schools	(5) Bottom Science Schools
SMT	0.402***	0.185***	0.0847***	0.217***	0.0898***
	(0.0379)	(0.0444)	(0.0171)	(0.0449)	(0.0240)
2015	0.00566*	0.0475***	-0.0413***	0.0365***	-0.0429***
	(0.00331)	(0.00727)	(0.00335)	(0.00740)	(0.00437)
2016	0.00380	0.0346***	-0.0281***	0.0230***	-0.0238***
	(0.00291)	(0.00596)	(0.00304)	(0.00636)	(0.00394)
2018	-0.00214	-0.0744***	0.0676***	-0.0409***	0.0387***
	(0.00280)	(0.00669)	(0.00334)	(0.00628)	(0.00404)
2019	-0.00106	-0.129***	0.104***	-0.0814***	0.0725***
	(0.00307)	(0.00718)	(0.00356)	(0.00687)	(0.00426)
2020	-0.0195***	-0.211***	0.155***	-0.149***	0.124***
	(0.00386)	(0.0101)	(0.00373)	(0.00952)	(0.00461)
SMT*2015	-0.0231*	-0.0616***	0.0132	-0.0551***	-0.0265
	(0.0140)	(0.0196)	(0.0175)	(0.0190)	(0.0225)
SMT*2016	-0.0101	-0.0282	0.0000512	-0.0200	-0.0256
	(0.0128)	(0.0176)	(0.0146)	(0.0169)	(0.0198)
SMT*2018	0.0215**	0.0946***	-0.0203	0.0645***	-0.00537
	(0.0103)	(0.0145)	(0.0150)	(0.0136)	(0.0200)
SMT*2019	-0.0132	0.0813***	-0.0232	0.0500***	-0.0161
	(0.0109)	(0.0149)	(0.0191)	(0.0142)	(0.0221)
SMT*2020	-0.00244	0.125***	-0.00643	0.0810***	-0.0226
	(0.0150)	(0.0213)	(0.0190)	(0.0202)	(0.0213)
cons	-0.0262***	0.455***	-0.403***	0.367***	-0.369***
	(0.00768)	(0.0158)	(0.00338)	(0.0167)	(0.00466)
N	2805869	941108	878067	985218	742425
R-squared	0.0658	0.0701	0.0846	0.0546	0.0413

Note: School-level DID estimates based on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2015, 2016, 2018, 2019, 2020} are academic year dummies. All columns used mathematics test scores of grade six students during 2015-2020. Standard errors in parentheses are clustered at school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3. 10: Effects of SMT project on science performance of grade 6 students (base year: 2017)

	(1) Overall	(2) Top Math Schools	(3) Bottom Math Schools	(4) Top Science Schools	(5) Bottom Science Schools
SMT	0.356***	0.204***	0.0829***	0.140***	0.0684***
	(0.0307)	(0.0364)	(0.0250)	(0.0344)	(0.0209)
2015	0.00162	0.0136**	-0.0168***	0.0225***	-0.0116***
	(0.00323)	(0.00669)	(0.00443)	(0.00612)	(0.00382)
2016	0.00180	0.0107	-0.0137***	0.0155**	-0.00791**
	(0.00312)	(0.00662)	(0.00412)	(0.00643)	(0.00355)
2018	0.000189	-0.0398***	0.0405***	-0.0735***	0.0827***
	(0.00293)	(0.00636)	(0.00418)	(0.00628)	(0.00411)
2019	-0.00337	-0.0593***	0.0629***	-0.0932***	0.121***
	(0.00306)	(0.00717)	(0.00413)	(0.00720)	(0.00397)
2020	-0.0196***	-0.104***	0.0576***	-0.139***	0.120***
	(0.00360)	(0.00816)	(0.00484)	(0.00803)	(0.00481)
SMT*2015	0.0175	0.00612	0.0348*	0.00120	0.00145
	(0.0130)	(0.0180)	(0.0208)	(0.0170)	(0.0172)
SMT*2016	0.00699	0.0182	-0.0145	0.0155	0.00243
	(0.0124)	(0.0172)	(0.0190)	(0.0161)	(0.0169)
SMT*2018	0.00701	0.0472***	-0.00988	0.0758***	-0.00838
	(0.0101)	(0.0139)	(0.0190)	(0.0133)	(0.0204)
SMT*2019	0.0394***	0.0986***	-0.0124	0.124***	0.00734
	(0.0127)	(0.0179)	(0.0246)	(0.0173)	(0.0184)
SMT*2020	0.0280*	0.0816***	0.0323	0.120***	0.0209
	(0.0144)	(0.0200)	(0.0270)	(0.0188)	(0.0282)
cons	-0.0240***	0.333***	-0.316***	0.381***	-0.406***
	(0.00619)	(0.0123)	(0.00457)	(0.0109)	(0.00389)
N	2805869	941108	878067	985218	742425
R-squared	0.0687	0.0674	0.0160	0.0626	0.0519

Note: School-level DID estimates based on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2015, 2016, 2018, 2019, 2020} are academic year dummies. All columns used science test scores of grade six students during 2015-2020. Standard errors in parentheses are clustered at school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3. 11: Effects of SMT project on mathematics performance of grade 9 students (base year: 2015)

	(1) Overall	(2) Top Math Schools	(3) Bottom Math Schools	(4) Top Science Schools	(5) Bottom Science Schools
SMT	0.190***	-0.0406	0.0410***	-0.00600	0.0430***
	(0.0366)	(0.0495)	(0.00909)	(0.0511)	(0.0135)
2016	-0.00516	0.00439	-0.00564*	0.00374	-0.000690
	(0.00335)	(0.00621)	(0.00292)	(0.00630)	(0.00398)
2017	-0.000854	0.00152	0.0178***	0.00300	0.0148***
	(0.00391)	(0.00772)	(0.00280)	(0.00787)	(0.00376)
2018	-0.00224	-0.0209**	0.0638***	-0.0131	0.0316***
	(0.00430)	(0.00886)	(0.00348)	(0.00888)	(0.00407)
2019	0.000147	-0.0267***	0.0763***	-0.0174*	0.0491***
	(0.00443)	(0.00929)	(0.00343)	(0.00930)	(0.00406)
2020	-0.0354***	-0.0135	0.0793***	-0.0154	0.0517***
	(0.0104)	(0.0227)	(0.00407)	(0.0228)	(0.00480)
SMT*2016	0.0289***	0.0226**	0.00103	0.0249**	0.00192
	(0.00688)	(0.00955)	(0.00806)	(0.00999)	(0.0116)
SMT*2017	-0.00718	-0.0126	-0.00391	-0.0108	0.00164
	(0.00798)	(0.0116)	(0.00929)	(0.0121)	(0.0114)
SMT*2018	-0.00148	0.0129	-0.0170	0.00664	-0.0188
	(0.00876)	(0.0130)	(0.0112)	(0.0135)	(0.0142)
SMT*2019	-0.0152	0.00349	-0.0146	-0.00440	-0.0238*
	(0.00945)	(0.0140)	(0.0144)	(0.0145)	(0.0133)
SMT*2020	0.0550***	0.0551*	0.00266	0.0597*	-0.0120
	(0.0196)	(0.0318)	(0.0149)	(0.0322)	(0.0174)
cons	-0.0264	0.333***	-0.422***	0.328***	-0.381***
	(0.0185)	(0.0325)	(0.00286)	(0.0331)	(0.00398)
N	2561872	1370117	510571	1313795	496871
R-squared	0.0246	0.0013	0.0804	0.0005	0.0262

Note: School-level DID estimates based on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2016, 2017, 2018, 2019, 2020} are academic year dummies. All columns used mathematics test scores of grade nine students during 2015-2020. Standard errors in parentheses are clustered at school level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 3. 12: Effects of SMT project on science performance of grade 9 students (base year: 2015)

	(1) Overall	(2) Top Math Schools	(3) Bottom Math Schools	(4) Top Science Schools	(5) Bottom Science Schools
SMT	0.192***	-0.0185	0.0541***	-0.00715	0.0626***
	(0.0340)	(0.0452)	(0.0166)	(0.0457)	(0.0124)
2016	0.00289	-0.0265***	0.0466***	-0.0257***	0.0463***
	(0.00359)	(0.00637)	(0.00457)	(0.00628)	(0.00372)
2017	0.00323	-0.0331***	0.0668***	-0.0348***	0.0642***
	(0.00407)	(0.00760)	(0.00468)	(0.00753)	(0.00380)
2018	-0.00203	-0.0334***	0.0513***	-0.0461***	0.0767***
	(0.00432)	(0.00836)	(0.00487)	(0.00841)	(0.00439)
2019	0.0112*	-0.113***	0.180***	-0.131***	0.212***
	(0.00607)	(0.0103)	(0.00515)	(0.0101)	(0.00480)
2020	-0.0240***	-0.0551***	0.109***	-0.0800***	0.152***
	(0.00903)	(0.0181)	(0.00563)	(0.0180)	(0.00501)
SMT*2016	-0.0135*	0.00293	-0.0187	-0.00412	-0.00727
	(0.00756)	(0.0102)	(0.0166)	(0.0103)	(0.0108)
SMT*2017	-0.0304***	-0.00880	-0.0428***	-0.0126	-0.0220*
	(0.00884)	(0.0122)	(0.0158)	(0.0122)	(0.0121)
SMT*2018	-0.0000625	0.0221	-0.0235	0.0253*	-0.0171
	(0.0100)	(0.0138)	(0.0167)	(0.0142)	(0.0131)
SMT*2019	-0.0740***	0.00515	-0.0386**	0.00398	-0.0351**
	(0.0131)	(0.0169)	(0.0173)	(0.0168)	(0.0142)
SMT*2020	0.0184	0.0476*	-0.0111	0.0600**	-0.0292*
	(0.0175)	(0.0265)	(0.0185)	(0.0267)	(0.0158)
cons	-0.0284*	0.305***	-0.397***	0.334***	-0.440***
	(0.0168)	(0.0290)	(0.00433)	(0.0288)	(0.00384)
N	2561872	1370117	510571	1313795	496871
R-squared	0.0276	0.0056	0.0910	0.0075	0.1811

Note: School-level DID estimates based on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2016, 2017, 2018, 2019, 2020} are academic year dummies. All columns used sciences test scores of grade nine students during 2015-2020. Standard errors in parentheses are clustered at school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3. 13: Effects of SMT project on mathematics performance of grade 9 students (base year: 2016)

	(1) Overall	(2) Top Math Schools	(3) Bottom Math Schools	(4) Top Science Schools	(5) Bottom Science Schools
SMT	0.219***	-0.0181	0.0420***	0.0189	0.0449***
	(0.0375)	(0.0504)	(0.00876)	(0.0520)	(0.0120)
2017	0.00431	-0.00287	0.0234***	-0.000736	0.0155***
	(0.00320)	(0.00612)	(0.00268)	(0.00632)	(0.00358)
2018	0.00292	-0.0253***	0.0694***	-0.0168**	0.0323***
	(0.00361)	(0.00738)	(0.00327)	(0.00744)	(0.00377)
2019	0.00531	-0.0311***	0.0819***	-0.0211***	0.0498***
	(0.00366)	(0.00762)	(0.00332)	(0.00763)	(0.00380)
2020	-0.0302***	-0.0179	0.0850***	-0.0191	0.0524***
	(0.0105)	(0.0226)	(0.00395)	(0.0227)	(0.00456)
SMT*2017	-0.0361***	-0.0352***	-0.00494	-0.0357***	-0.000278
	(0.00627)	(0.00888)	(0.00825)	(0.00942)	(0.0115)
SMT*2018	-0.0304***	-0.00973	-0.0180*	-0.0182	-0.0207*
	(0.00746)	(0.0110)	(0.00951)	(0.0114)	(0.0110)
SMT*2019	-0.0441***	-0.0191	-0.0157	-0.0293**	-0.0257**
	(0.00811)	(0.0118)	(0.0118)	(0.0123)	(0.0130)
SMT*2020	0.0261	0.0325	0.00163	0.0348	-0.0139
	(0.0191)	(0.0312)	(0.0126)	(0.0316)	(0.0159)
cons	-0.0315*	0.338***	-0.428***	0.332***	-0.382***
	(0.0192)	(0.0334)	(0.00290)	(0.0340)	(0.00396)
N	2098116	1119866	419081	1074160	407612
R-squared	0.0247	0.0013	0.0736	0.0006	0.0243

Note: School-level DID estimates based on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2017, 2018, 2019, 2020} are academic year dummies. All columns used mathematics test scores of grade nine students during 2016-2020. Standard errors in parentheses are clustered at school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3. 14: Effects of SMT project on science performance of grade 9 students (base year: 2016)

	(1) Overall	(2) Top Math Schools	(3) Bottom Math Schools	(4) Top Science Schools	(5) Bottom Science Schools
SMT	0.178***	-0.0156	0.0353***	-0.0113	0.0554***
	(0.0326)	(0.0439)	(0.0134)	(0.0447)	(0.0115)
2017	0.000337	-0.00665	0.0201***	-0.00909*	0.0180***
	(0.00314)	(0.00566)	(0.00443)	(0.00551)	(0.00340)
2018	-0.00492	-0.00687	0.00468	-0.0204***	0.0305***
	(0.00379)	(0.00723)	(0.00477)	(0.00725)	(0.00441)
2019	0.00830*	-0.0867***	0.133***	-0.105***	0.166***
	(0.00494)	(0.00814)	(0.00501)	(0.00797)	(0.00466)
2020	-0.0269***	-0.0286	0.0628***	-0.0543***	0.105***
	(0.00887)	(0.0178)	(0.00540)	(0.0180)	(0.00495)
SMT*2017	-0.0168**	-0.0117	-0.0240*	-0.00850	-0.0148
	(0.00728)	(0.00978)	(0.0141)	(0.00986)	(0.0116)
SMT*2018	0.0135	0.0191	-0.00480	0.0295**	-0.00981
	(0.00844)	(0.0117)	(0.0162)	(0.0121)	(0.0158)
SMT*2019	-0.0605***	0.00222	-0.0198	0.00810	-0.0278**
	(0.0109)	(0.0139)	(0.0147)	(0.0141)	(0.0141)
SMT*2020	0.0319*	0.0446*	0.00761	0.0641**	-0.0219
	(0.0166)	(0.0256)	(0.0166)	(0.0258)	(0.0160)
cons	-0.0255	0.278***	-0.351***	0.309***	-0.394***
	(0.0163)	(0.0288)	(0.00414)	(0.0286)	(0.00369)
N	2098116	1119866	419081	1074160	407612
R-squared	0.0275	0.0052	0.0759	0.0068	0.1477

Note: School-level DID estimates based on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2017, 2018, 2019, 2020} are academic year dummies. All columns used science test scores of grade nine students during 2016-2020. Standard errors in parentheses are clustered at school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3. 15: Effects of SMT project on mathematics performance of grade 9 students (base year: 2017)

	(1) Overall	(2) Top Math Schools	(3) Bottom Math Schools	(4) Top Science Schools	(5) Bottom Science Schools
SMT	0.183***	-0.0532	0.0371***	-0.0168	0.0446***
	(0.0383)	(0.0529)	(0.00854)	(0.0547)	(0.0107)
2015	0.000854	-0.00152	-0.0178***	-0.00300	-0.0148***
	(0.00391)	(0.00772)	(0.00280)	(0.00787)	(0.00376)
2016	-0.00431	0.00287	-0.0234***	0.000736	-0.0155***
	(0.00320)	(0.00612)	(0.00268)	(0.00632)	(0.00358)
2018	-0.00139	-0.0224***	0.0460***	-0.0161***	0.0168***
	(0.00279)	(0.00545)	(0.00317)	(0.00545)	(0.00343)
2019	0.00100	-0.0282***	0.0585***	-0.0204***	0.0343***
	(0.00308)	(0.00626)	(0.00327)	(0.00625)	(0.00357)
2020	-0.0345***	-0.0150	0.0616***	-0.0184	0.0369***
	(0.0105)	(0.0222)	(0.00393)	(0.0223)	(0.00456)
SMT*2015	0.00718	0.0126	0.00391	0.0108	-0.00164
	(0.00798)	(0.0116)	(0.00929)	(0.0121)	(0.0114)
SMT*2016	0.0361***	0.0352***	0.00494	0.0357***	0.000278
	(0.00627)	(0.00888)	(0.00825)	(0.00942)	(0.0115)
SMT*2018	0.00571	0.0254***	-0.0131	0.0175*	-0.0204*
	(0.00685)	(0.00944)	(0.0102)	(0.00988)	(0.0106)
SMT*2019	-0.00799	0.0161	-0.0107	0.00643	-0.0254**
	(0.00737)	(0.0104)	(0.0123)	(0.0108)	(0.0113)
SMT*2020	0.0621***	0.0677**	0.00657	0.0705**	-0.0137
	(0.0194)	(0.0311)	(0.0134)	(0.0317)	(0.0151)
cons	-0.0272	0.335***	-0.404***	0.331***	-0.367***
	(0.0199)	(0.0358)	(0.00254)	(0.0365)	(0.00357)
N	2561872	1370117	510571	1313795	496871
R-squared	0.0246	0.0013	0.0804	0.0005	0.0262

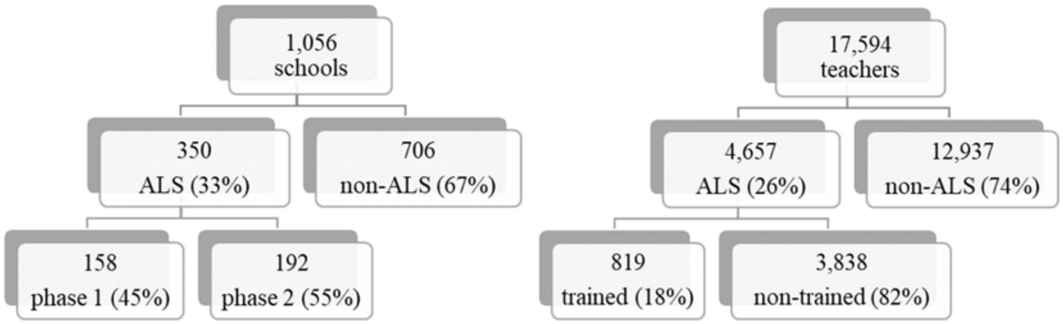
Note: School-level DID estimates based on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2015, 2016, 2018, 2019, 2020} are academic year dummies. All columns used mathematics test scores of grade nine students during 2015-2020. Standard errors in parentheses are clustered at school level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 3. 16: Effects of SMT project on science performance of grade 9 students (base year: 2017)

	(1) Overall	(2) Top Math Schools	(3) Bottom Math Schools	(4) Top Science Schools	(5) Bottom Science Schools
SMT	0.161***	-0.0273	0.0113	-0.0198	0.0406***
	(0.0312)	(0.0427)	(0.0111)	(0.0435)	(0.00891)
2015	-0.00323	0.0331***	-0.0668***	0.0348***	-0.0642***
	(0.00407)	(0.00760)	(0.00468)	(0.00753)	(0.00380)
2016	-0.000337	0.00665	-0.0201***	0.00909*	-0.0180***
	(0.00314)	(0.00566)	(0.00443)	(0.00551)	(0.00340)
2018	-0.00526	-0.000223	-0.0155***	-0.0113*	0.0125***
	(0.00335)	(0.00630)	(0.00459)	(0.00629)	(0.00421)
2019	0.00796*	-0.0801***	0.113***	-0.0960***	0.148***
	(0.00449)	(0.00735)	(0.00467)	(0.00715)	(0.00441)
2020	-0.0273***	-0.0219	0.0427***	-0.0452**	0.0873***
	(0.00880)	(0.0180)	(0.00511)	(0.0181)	(0.00475)
SMT*2015	0.0304***	0.00880	0.0428***	0.0126	0.0220*
	(0.00884)	(0.0122)	(0.0158)	(0.0122)	(0.0121)
SMT*2016	0.0168**	0.0117	0.0240*	0.00850	0.0148
	(0.00728)	(0.00978)	(0.0141)	(0.00986)	(0.0116)
SMT*2018	0.0303***	0.0309***	0.0192	0.0380***	0.00496
	(0.00768)	(0.0106)	(0.0149)	(0.0109)	(0.0141)
SMT*2019	-0.0437***	0.0140	0.00418	0.0166	-0.0131
	(0.00914)	(0.0118)	(0.0126)	(0.0117)	(0.0122)
SMT*2020	0.0487***	0.0564**	0.0316*	0.0726***	-0.00718
	(0.0156)	(0.0248)	(0.0164)	(0.0253)	(0.0132)
cons	-0.0252	0.271***	-0.330***	0.299***	-0.376***
	(0.0162)	(0.0289)	(0.00356)	(0.0288)	(0.00326)
N	2561872	1370117	510571	1313795	496871
R-squared	0.0276	0.0056	0.0910	0.0075	0.1811

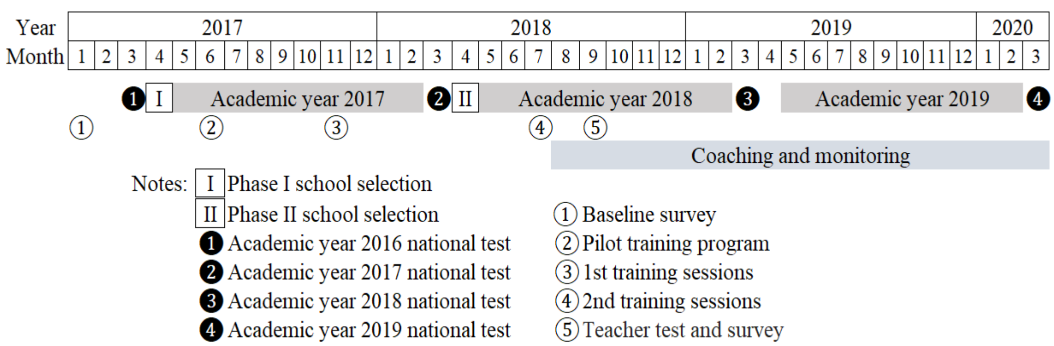
Note: School-level DID estimates based on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2015, 2016, 2018, 2019, 2020} are academic year dummies. All columns used science test scores of grade nine students during 2015-2020. Standard errors in parentheses are clustered at school level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Figure 2. 1: Number of schools and teachers in the south



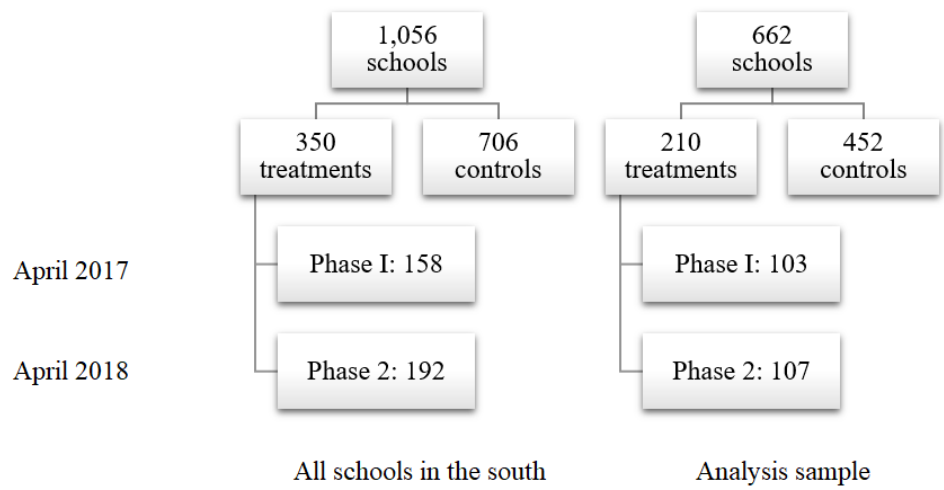
Source: ALS project report and correspondence with program officials.

Figure 2. 2: Active Learning School Project Timeline



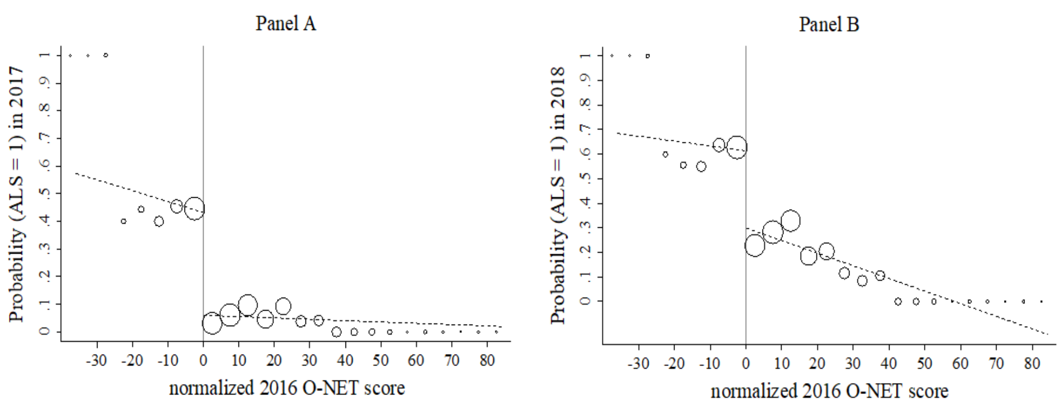
Source: ALS project report and correspondence with program officials.

Figure 2. 3: Sample selection diagram



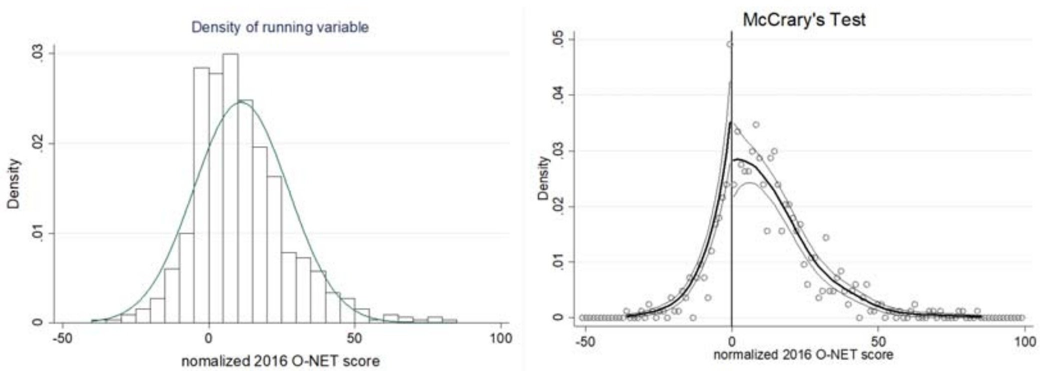
Source: ALS project report and correspondence with program officials.

Figure 2. 4: Probability of ALS status



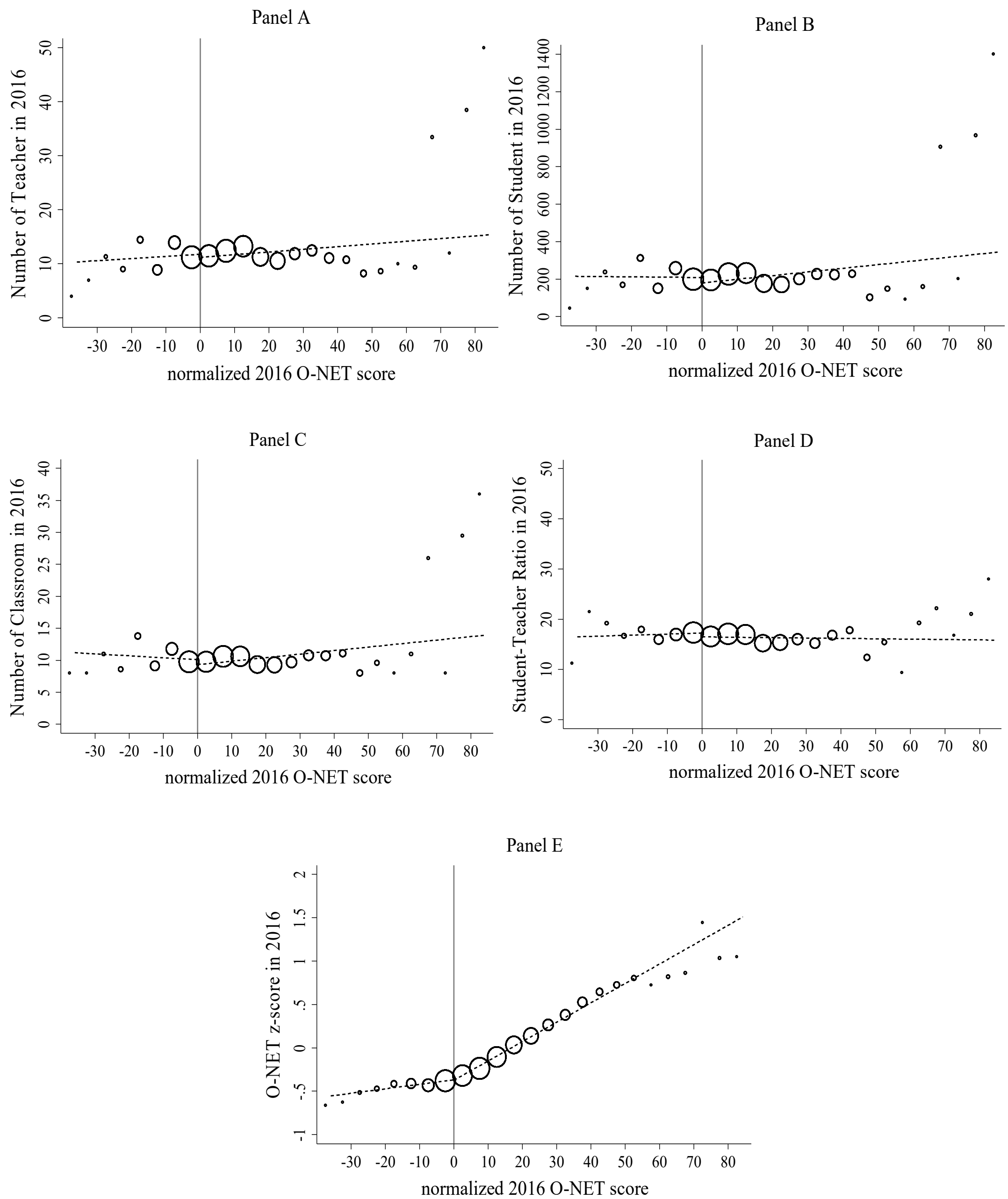
Note: Linear OLS regression on equation (3) in estimation approach session. Panel A and Panel B are ALS treatment status for academic year 2017 and 2018. Each dot represents the average probability of being ALS in a given bin. The bin-width is 5 normalized test score. The sizes of the dots are proportional to the number of observations in the bin.

Figure 2. 5: Histogram of normalized 2016 O-NET score and McCrary's Test



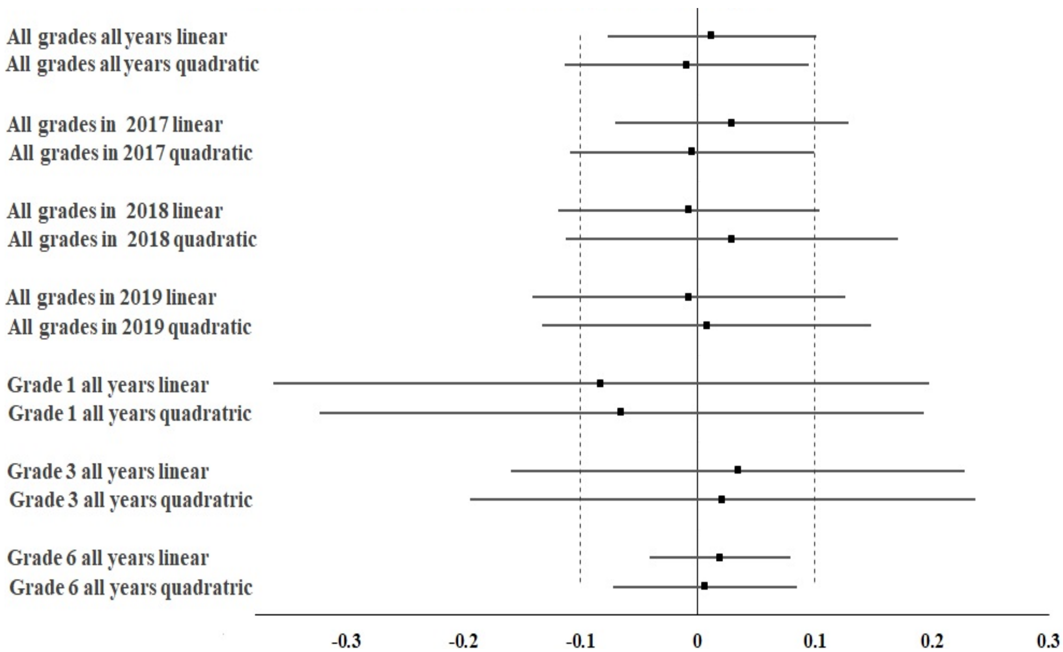
Note: The normalized 2016 O-NET scores are the school average scores subtracted by the district cutoffs

Figure 2. 6: Graphs of pre-treatment covariates



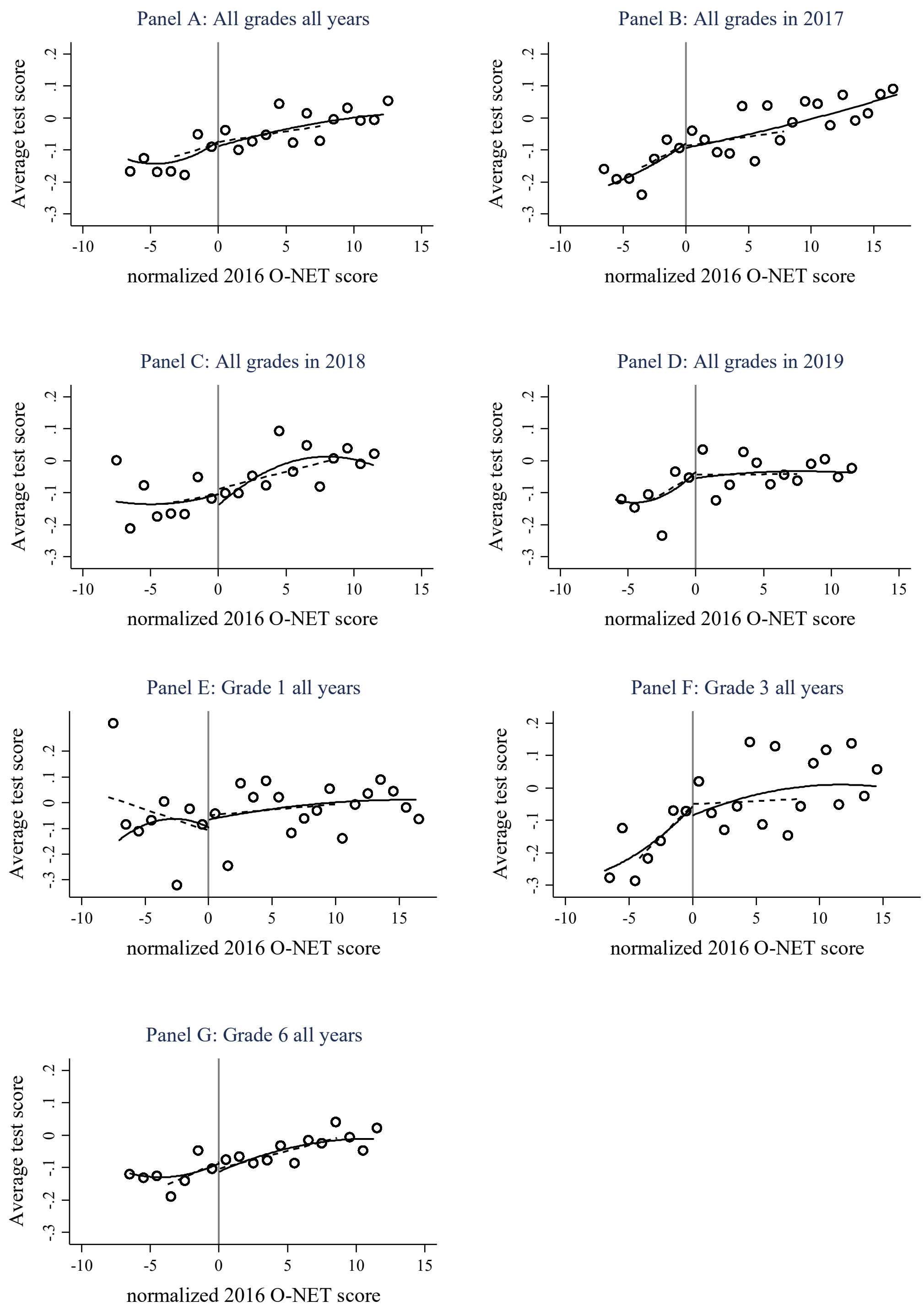
Note: Panel A, Panel B, Panel C, Panel D and Pane E represent number of teachers, number of students, number of classrooms, student-teacher ratio and standardized O-NET score for academic year 2016, respectively. Each dot represents the average value of the pre-treatment covariates in a given bin. The bin-width is 5 normalized test score. The sizes of the dots are proportional to the number of observations in the bin.

Figure 2. 7: Effect Sizes of Active Learning School Project



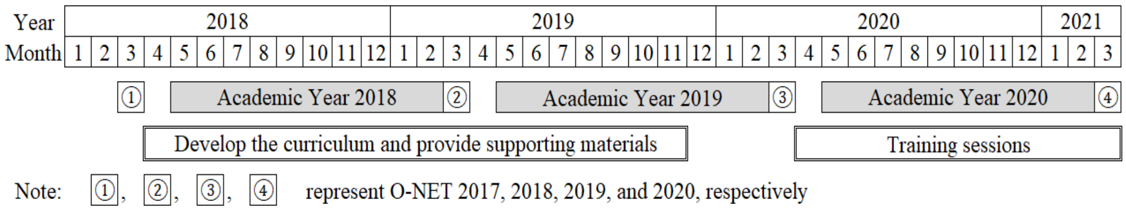
Note: Linear and quadratic RD estimates on equation (1). All grades in all years use all test scores from 24 tests during 2017-2019. All grades in 2017 use the test scores from the 7 tests in 2017. All grades in 2018 use the test scores from the 9 tests in 2018. All grades in 2019 use the test scores from 8 tests in 2019. Grade 1 all years use the test scores from the 4 tests during 2018-2019. Grade 3 all years use the test scores from the 8 tests during 2017-2019. Grade 6 all years use the test scores from the 12 tests during 2017-2019.

Figure 2. 8: Impacts on school average test score 2017-2019



Note: Bandwidths for these graphs are mean squared error (MSE) optimal bandwidths (Calonico et al. 2017) with district and test fixed effect and cluster district for standard error. Each dot represents the average test score in a given bin of bin width one. Dashed lines represent linear fit and solid lines are quadratic fit.

Figure 3. 1: SMT Project Timetable



Source: http://smt.ipst.ac.th/ and correspondence with program officials.

Figure 3. 2: Population diagram

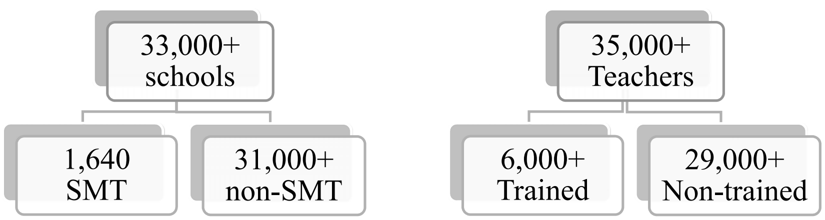
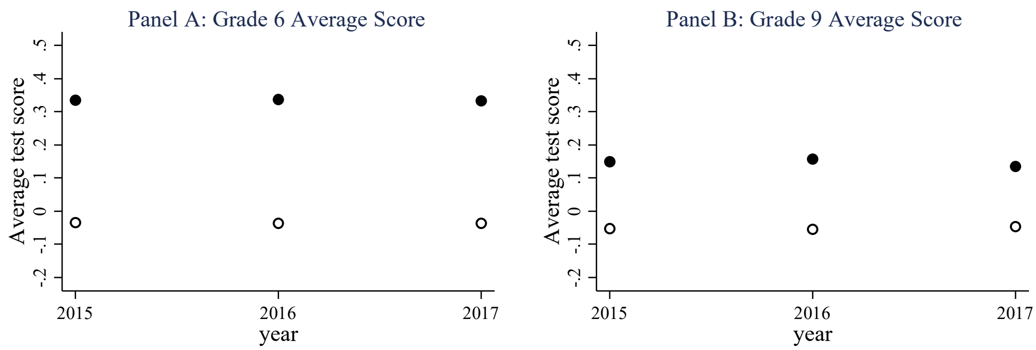


Chart A: Schools in Thailand Chart B: Teacher in Treatment Schools

Note: School statistic as of academic year 2019 when both batches were included in the project.
Teacher statistic from IPST as of early 2021.

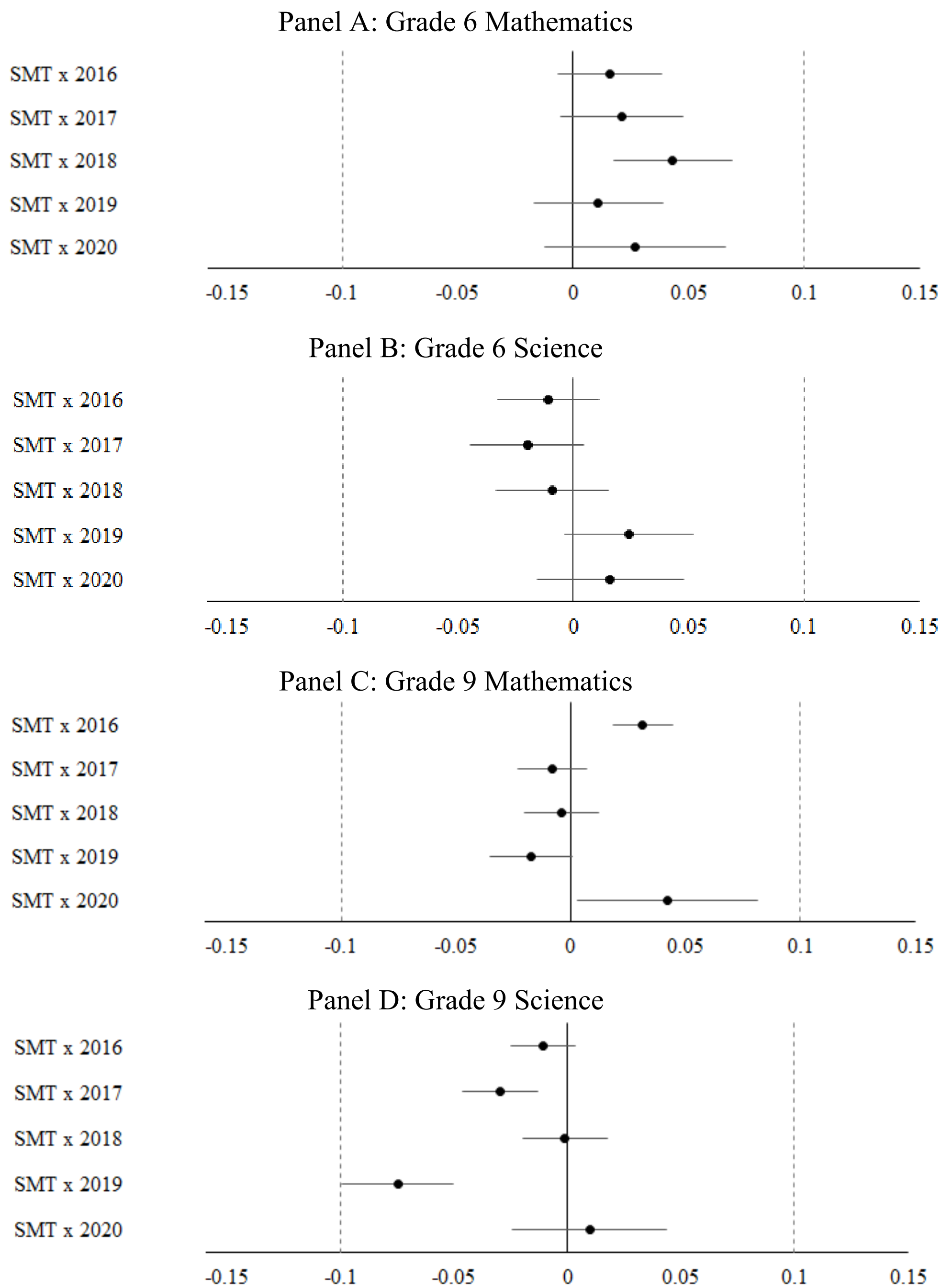
Figure 3. 3: Average test scores of SMT and comparison schools before the project started

.



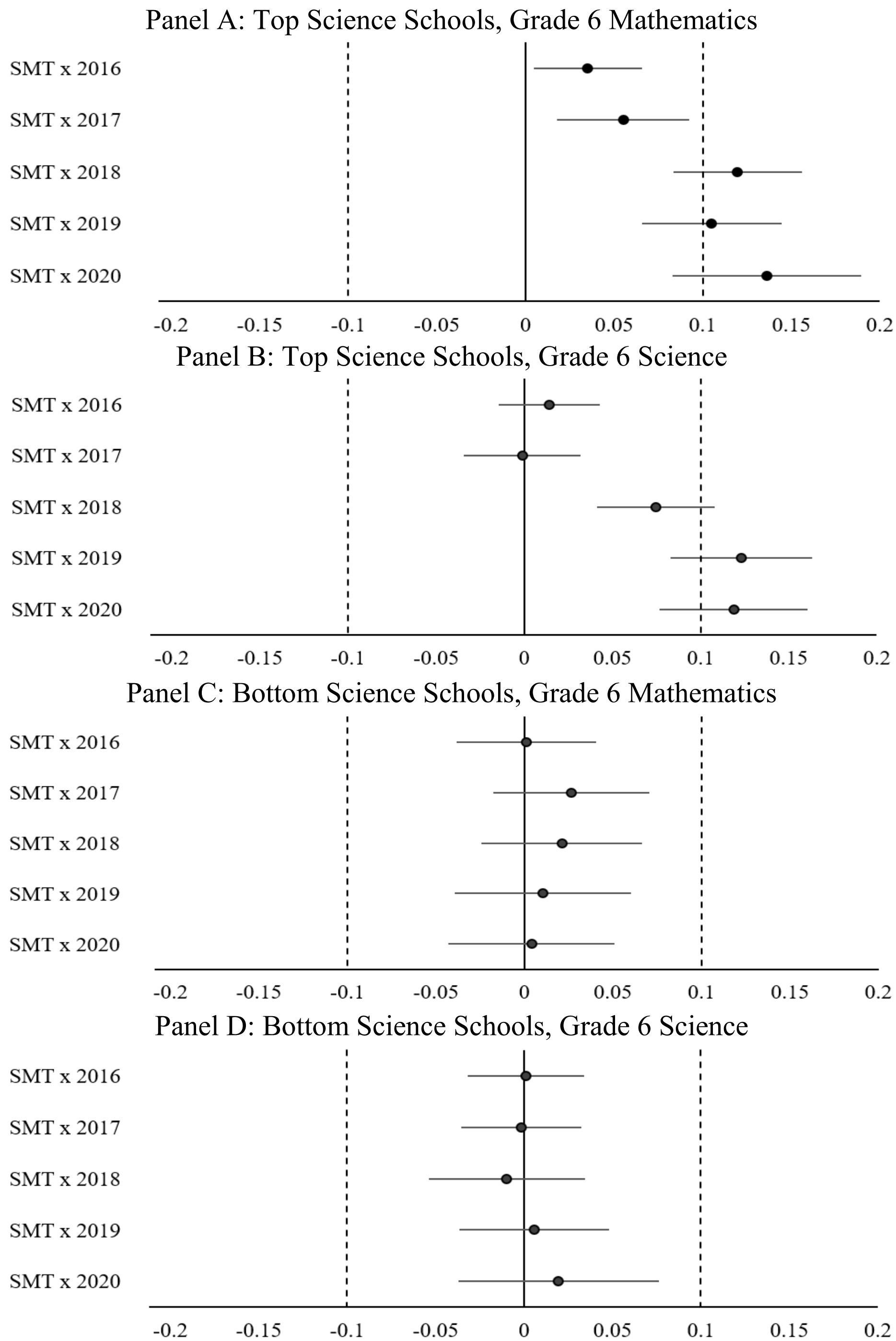
Note: Each dot represents the test-taker weighted average at school level of grade six and nine standardized scores. The scores were the average scores of mathematics and science. Solid and open dots represent the average scores of the treatment and control group, respectively.

Figure 3. 4: Effect sizes of SMT Project, base year 2015



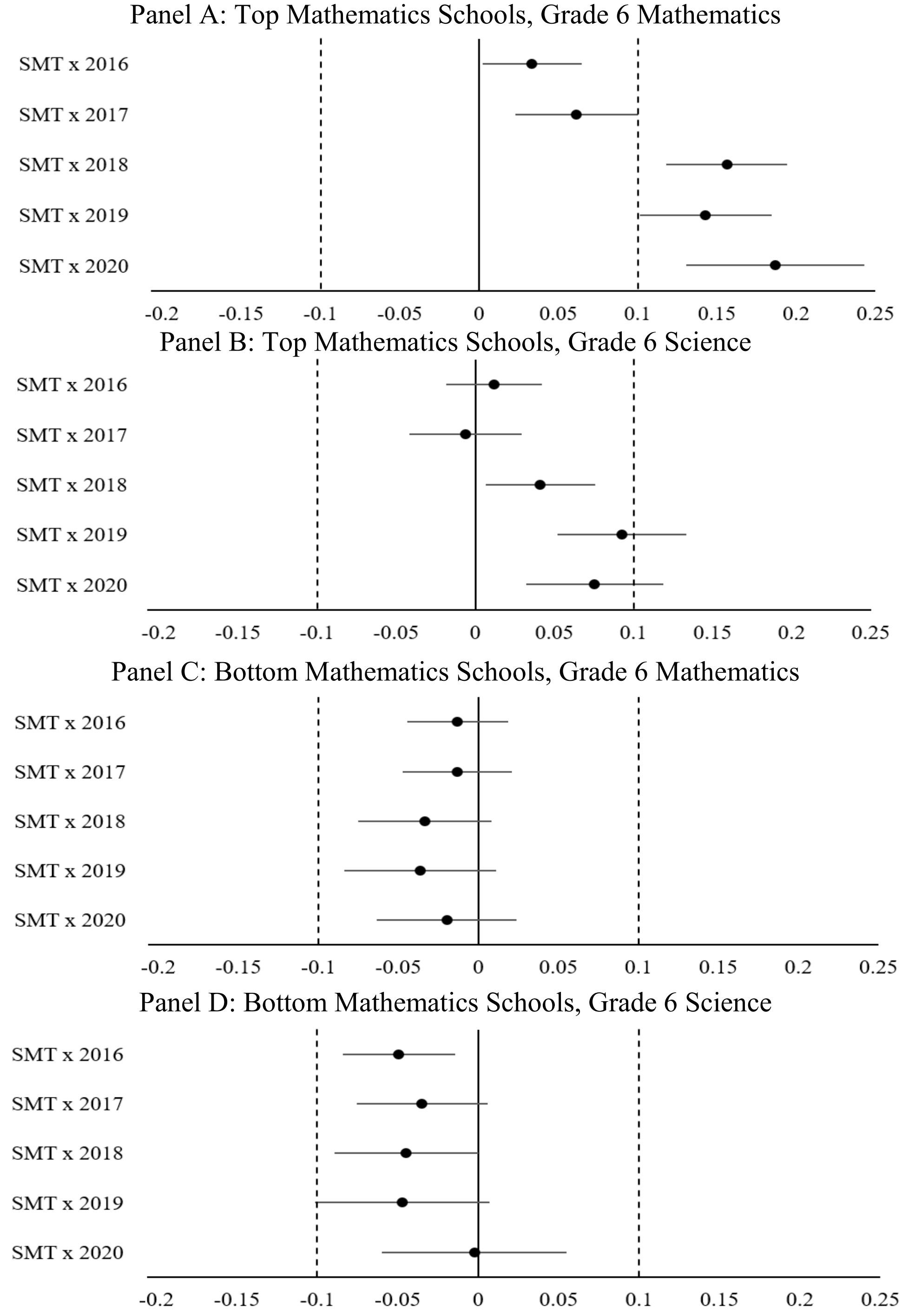
Note: School-level DID estimates based on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2016, 2017, 2018, 2019, 2020} are academic year dummies. Panel A and Panel B used mathematics and science test scores of grade six students during 2015-2020. Panel C and Panel D used mathematics and science test scores of grade nine students during 2015-2020. The solid black dots represent the effect sizes while the horizontal lines cover 95% confidence interval values.

Figure 3. 5: Effect sizes of SMT Project, top vs bottom science schools, base year 2015



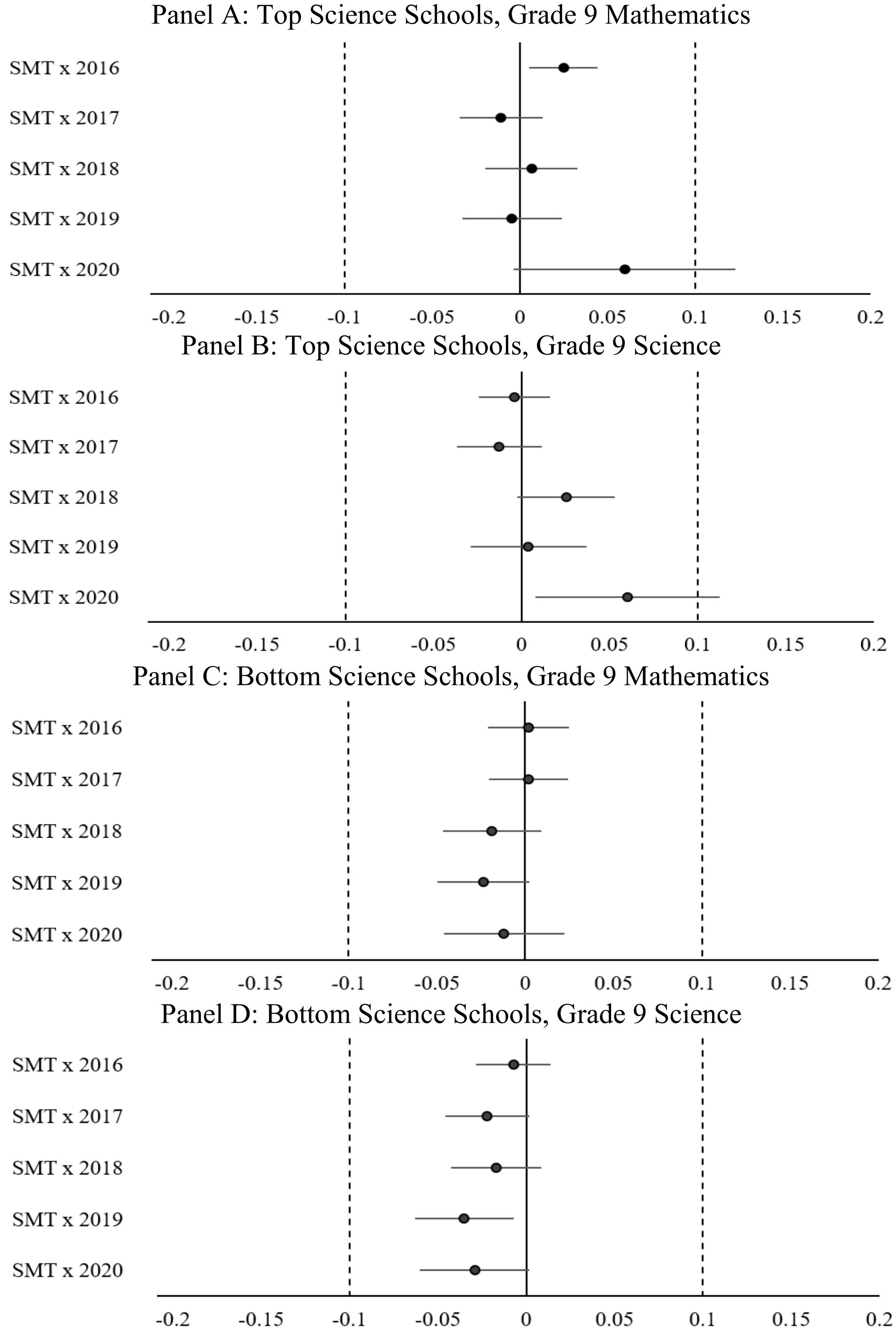
Note: DID estimates on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2016, 2017, 2018, 2019, 2020} are academic year dummies. Panel A and Panel B used mathematics and science test scores of grade six students in top science schools during 2015-2020. Panel C and Panel D used mathematics and science test scores of grade six students in bottom science schools during 2015-2020. The solid black dots represent the effect sizes while the horizontal lines cover 95% confidence interval values.

Figure 3. 6: Effect sizes of SMT Project, top vs bottom math schools, base year 2015



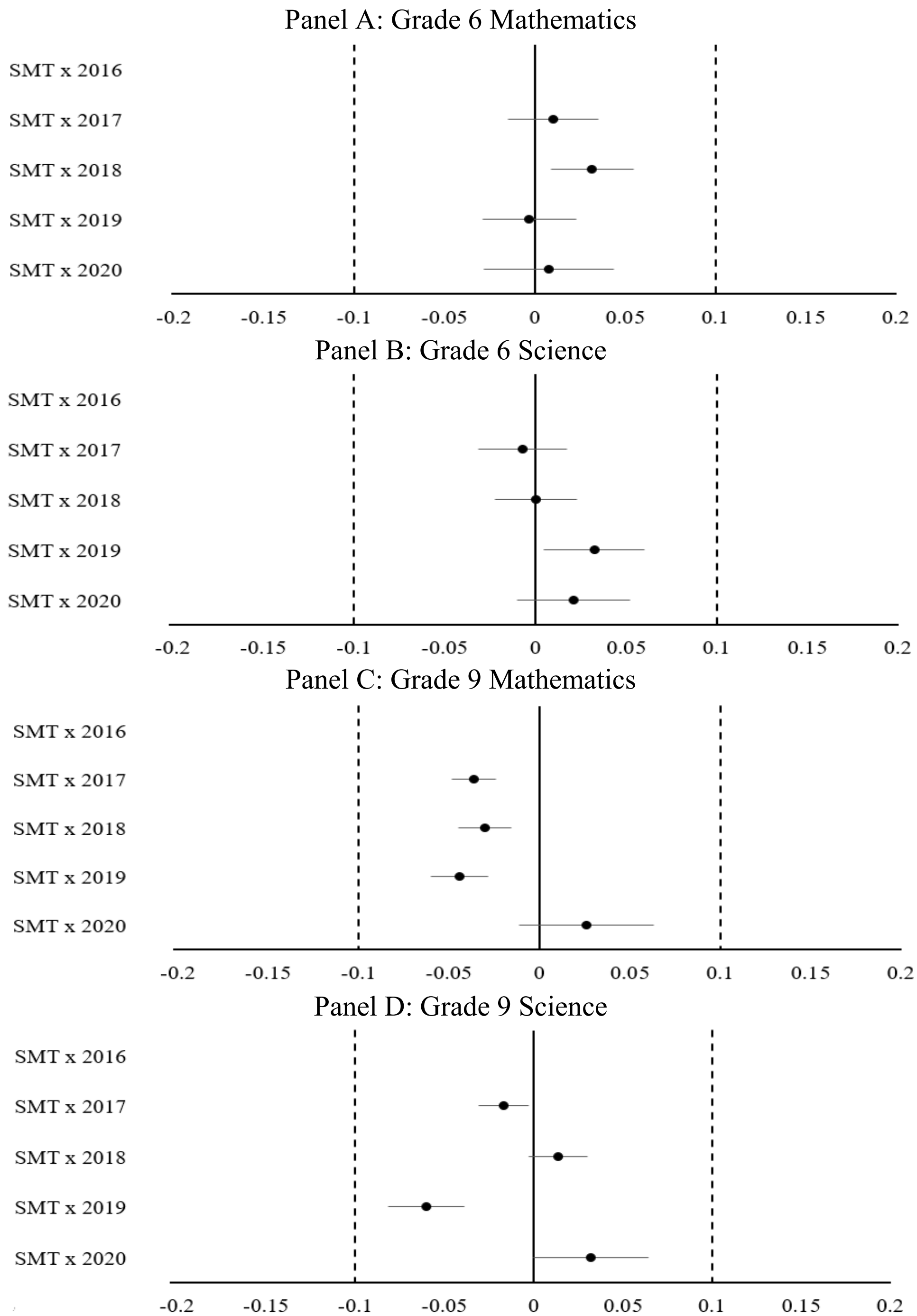
Note: DID estimates on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2016, 2017, 2018, 2019, 2020} are academic year dummies. Panel A and Panel B used mathematics and science test scores of grade six students in top mathematics schools during 2015-2020. Panel C and Panel D used mathematics and science test scores of grade six students in bottom mathematics schools during 2015-2020. The solid black dots represent the effect sizes while the horizontal lines cover 95% confidence interval values.

Figure 3. 7: Effect sizes of SMT Project, top vs bottom science schools, base year 2015



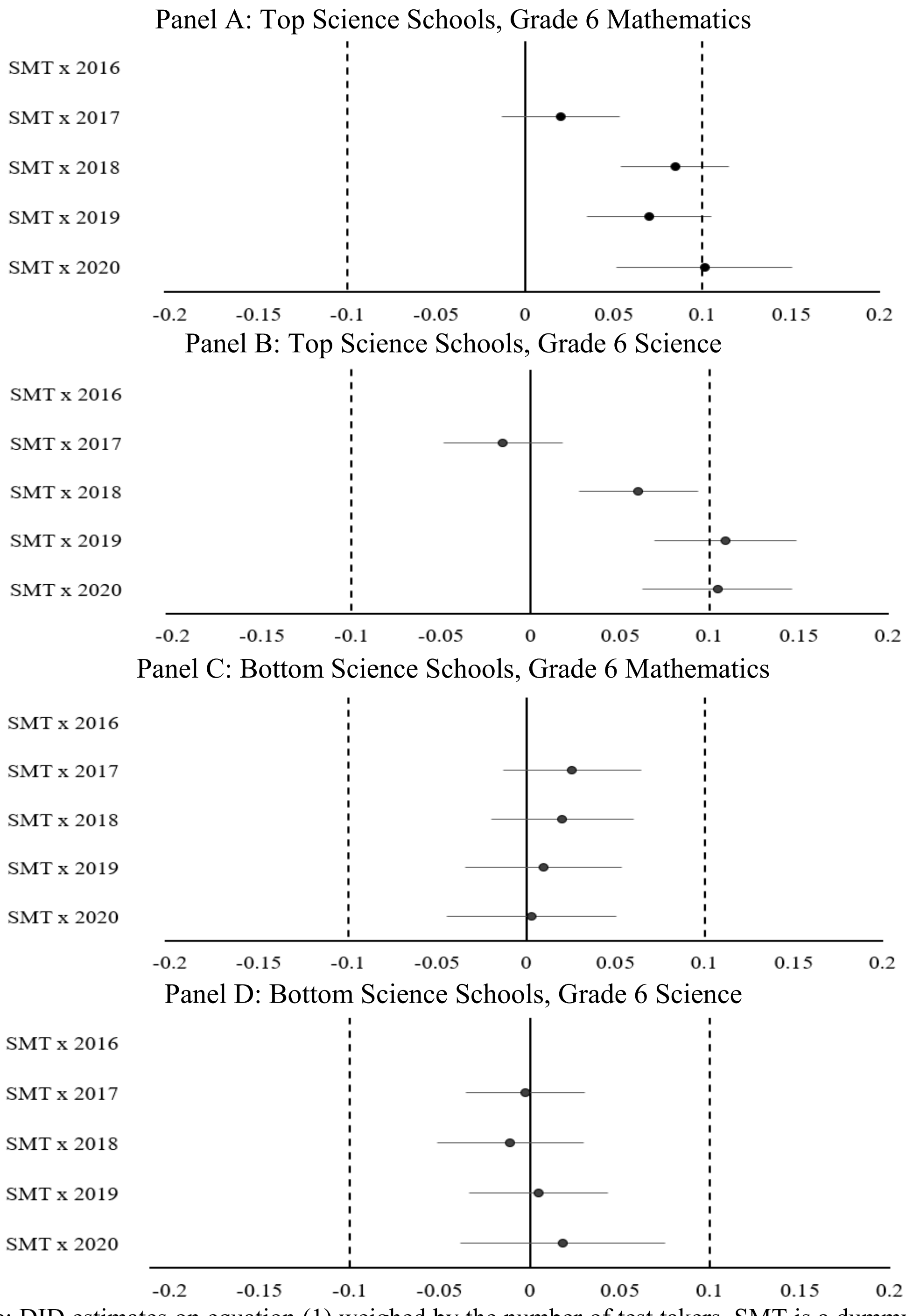
Note: DID estimates on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2016, 2017, 2018, 2019, 2020} are academic year dummies. Panel A and Panel B used mathematics and science test scores of grade nine students in top science schools during 2015-2020. Panel C and Panel D used mathematics and science test scores of grade nine students in bottom science schools during 2015-2020. The solid black dots represent the effect sizes while the horizontal lines cover 95% confidence interval values.

Figure 3. 8: Effect sizes of SMT Project, base year 2016



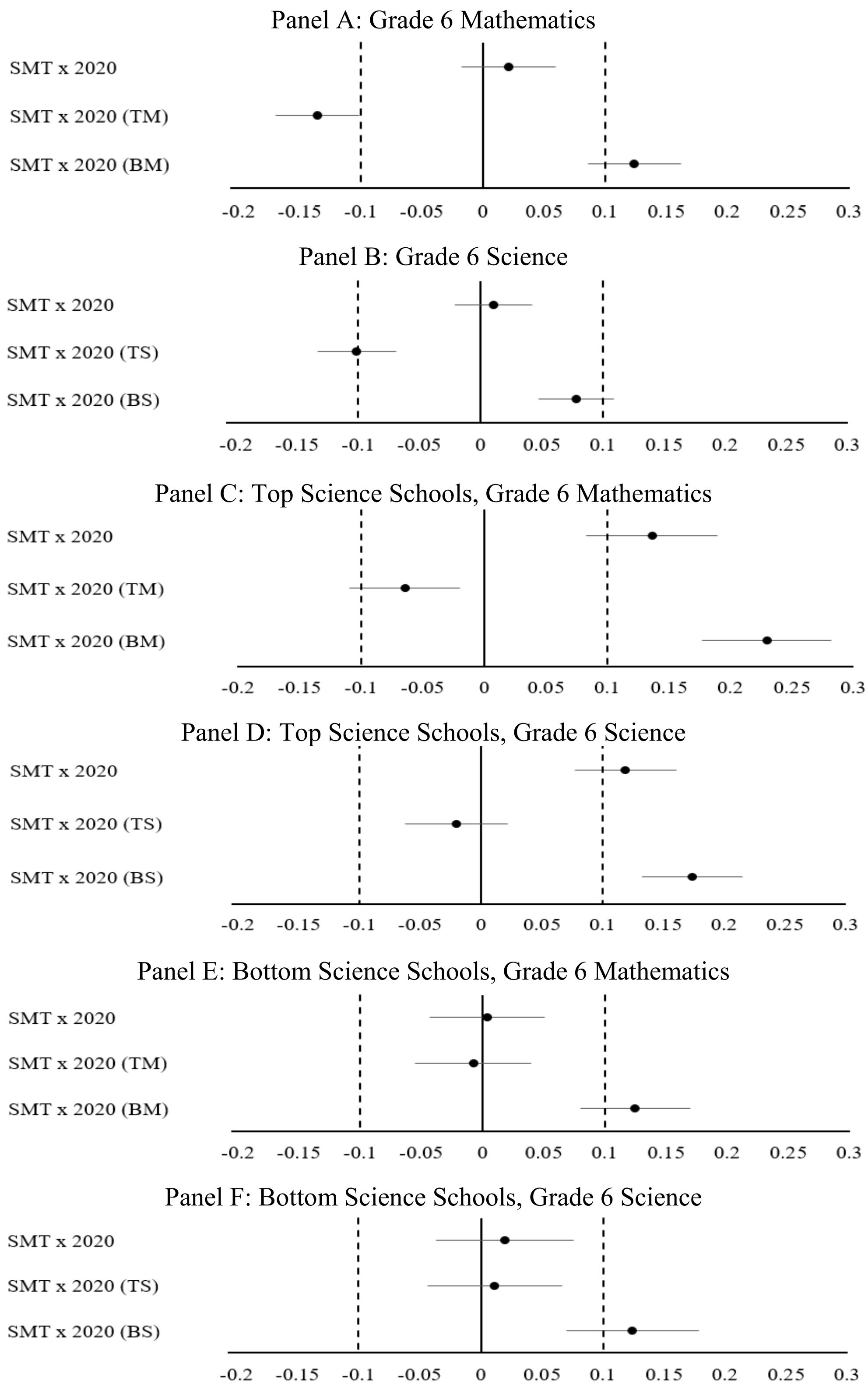
Note: DID estimates on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2015, 2016, 2017, 2018, 2019, 2020} are academic year dummies. Panel A and Panel B used mathematics and science test scores of grade six students during 2015-2020. Panel C and Panel D used mathematics and science test scores of grade nine students during 2015-2020. The solid black dots represent the effect sizes while the horizontal lines cover 95% confidence interval values.

Figure 3. 9: Effect sizes of SMT Project, base year 2016



Note: DID estimates on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2016, 2017, 2018, 2019, 2020} are academic year dummies. Panel A and Panel B used mathematics and science test scores of grade six students in top science schools during 2015-2020. Panel C and Panel D used mathematics and science test scores of grade six students in bottom science schools during 2015-2020. The solid black dots represent the effect sizes while the horizontal lines cover 95% confidence interval values.

Figure 3. 10: Effect sizes of SMT Project, original vs trimmed data



Appendix

Table 2A- 1: Effects of ALS project on student performance (with test taker weight)

	Coef.	Std.Err.	[95%Conf.	Interval]	Order	Dist. FE	Test FE	Obs.
All grade,	0.042	0.043	-0.042	0.126	1	Yes	No	6915
All year	0.027	0.041	-0.054	0.108	1	Yes	Yes	6891
	0.036	0.055	-0.072	0.143	2	Yes	No	6343
	0.005	0.047	-0.086	0.097	2	Yes	Yes	6967
All grade	0.078	0.052	-0.024	0.180	1	Yes	No	2027
in 2017	0.078	0.052	-0.025	0.180	1	Yes	Yes	2027
	0.004	0.050	-0.094	0.102	2	Yes	No	2188
	0.004	0.050	-0.093	0.102	2	Yes	Yes	2188
All grade	0.024	0.047	-0.069	0.117	1	Yes	No	2436
in 2018	0.022	0.048	-0.071	0.116	1	Yes	Yes	2436
	0.034	0.062	-0.087	0.155	2	Yes	No	2661
	0.021	0.064	-0.104	0.145	2	Yes	Yes	2769
All grade	-0.017	0.054	-0.122	0.088	1	Yes	No	2248
in 2019	-0.005	0.053	-0.109	0.098	1	Yes	Yes	2180
	-0.018	0.067	-0.130	0.151	2	Yes	No	2340
	-0.027	0.067	-0.158	0.104	2	Yes	Yes	2332
Grade 1	-0.039	0.127	-0.288	0.209	1	Yes	No	1180
	-0.035	0.129	-0.287	0.217	1	Yes	Yes	1164
	0.127	0.142	-0.151	0.406	2	Yes	No	1288
	0.133	0.143	-0.147	0.413	2	Yes	Yes	1284
Grade 3	0.026	0.077	-0.125	0.176	1	Yes	No	2239
	0.026	0.077	-0.124	0.177	1	Yes	Yes	2255
	0.009	0.090	-0.167	0.185	2	Yes	No	2543
	0.005	0.090	-0.171	0.182	2	Yes	Yes	2543
Grade 6	0.032	0.033	-0.031	0.096	1	Yes	No	2896
	0.030	0.033	-0.033	0.094	1	Yes	Yes	2896
	-0.008	0.038	-0.082	0.066	2	Yes	No	3900
	0.007	0.038	-0.067	0.080	2	Yes	Yes	3864

Note: Linear and quadratic RD estimates on equation (1). All grades in all years use all test scores from 24 tests during 2017-2019. All grades in 2017 use the test scores from the 7 tests in 2017. All grades in 2018 use the test scores from the 9 tests in 2018. All grades in 2019 use the test scores from 8 tests in 2019. Grade 1 all years use the test scores from the 4 tests during 2018-2019. Grade 3 all years use the test scores from the 8 tests during 2017-2019. Grade 6 all years use the test scores from the 12 tests during 2017-2019. Order refers to polynomial order and Dist. FE refers to district fixed effect.

Table 2A- 2: First stage results (without test taker weight)

	Coef.	Std.Err.	[95%Conf.	Interval]	Order	Dist. FE	Test FE	Obs.
All grade,	-0.834***	0.057	-0.945	-0.723	1	Yes	No	5287
All year	-0.833***	0.058	-0.946	-0.720	1	Yes	Yes	5239
	-0.909***	0.075	-1.056	-0.762	2	Yes	No	8019
	-0.930***	0.071	-1.069	-0.791	2	Yes	Yes	7971
All grade	-0.838***	0.055	-0.945	-0.730	1	Yes	No	1558
in 2017	-0.837***	0.055	-0.945	-0.730	1	Yes	Yes	1558
	-0.927***	0.072	-1.068	-0.787	2	Yes	No	2846
	-0.927***	0.072	-1.068	-0.787	2	Yes	Yes	2846
All grade	-0.787***	0.051	-0.888	-0.687	1	Yes	No	2166
in 2018	-0.789***	0.051	-0.889	-0.690	1	Yes	Yes	2175
	-0.884***	0.074	-1.029	-0.740	2	Yes	No	2994
	-0.885***	0.074	-1.029	-0.740	2	Yes	Yes	2994
All grade	-0.809***	0.057	-0.921	-0.698	1	Yes	No	1676
in 2019	-0.814***	0.055	-0.922	-0.706	1	Yes	Yes	1692
	-0.905***	0.070	-1.042	-0.767	2	Yes	No	2544
	-0.911***	0.070	-1.048	-0.773	2	Yes	Yes	2560
Grade 1	-0.665***	0.056	-0.774	-0.556	1	Yes	No	1256
	-0.665***	0.056	-0.774	-0.556	1	Yes	Yes	1256
	-0.885***	0.072	-1.025	-0.744	2	Yes	No	1640
	-0.885***	0.072	-1.025	-0.744	2	Yes	Yes	1640
Grade 3	-0.776***	0.052	-0.877	-0.675	1	Yes	No	1927
	-0.775***	0.051	-0.875	-0.675	1	Yes	Yes	1935
	-0.872***	0.073	-1.015	-0.730	2	Yes	No	3007
	-0.874***	0.073	-1.017	-0.732	2	Yes	Yes	3015
Grade 6	-0.822***	0.052	-0.925	-0.720	1	Yes	No	2896
	-0.822***	0.052	-0.924	-0.719	1	Yes	Yes	2896
	-0.914***	0.071	-1.052	-0.776	2	Yes	No	3864
	-0.914***	0.071	-1.052	-0.776	2	Yes	Yes	3864

Note: Linear and quadratic RD estimates on equation (2). All grades in all years use all test scores from 24 tests during 2017-2019. All grades in 2017 use the test scores from the 7 tests in 2017. All grades in 2018 use the test scores from the 9 tests in 2018. All grades in 2019 use the test scores from 8 tests in 2019. Grade 1 all years use the test scores from the 4 tests during 2018-2019. Grade 3 all years use the test scores from the 8 tests during 2017-2019. Grade 6 all years use the test scores from the 12 tests during 2017-2019. Order refers to polynomial order and Dist. FE refers to district fixed effect.

Table 2A- 3: First stage results (with test taker weight)

	Coef.	Std.Err.	[95%Conf.	Interval]	Order	Dist. FE	Test FE	Obs.
All grade, All year	-0.859***	0.053	-0.962	-0.756	1	Yes	No	6915
	-0.854***	0.053	-0.958	-0.750	1	Yes	Yes	6891
	-0.911***	0.072	-1.052	-0.769	2	Yes	No	6343
	-0.981***	0.073	-1.124	-0.839	2	Yes	Yes	6967
All grade in 2017	-0.864***	0.059	-0.980	-0.747	1	Yes	No	2027
	-0.864***	0.059	-0.980	-0.747	1	Yes	Yes	2027
	-0.894***	0.063	-1.017	-0.771	2	Yes	No	2188
	-0.894***	0.063	-1.017	-0.771	2	Yes	Yes	2188
All grade in 2018	-0.782***	0.054	-0.888	-0.675	1	Yes	No	2436
	-0.782***	0.054	-0.888	-0.675	1	Yes	Yes	2436
	-0.961***	0.071	-1.100	-0.822	2	Yes	No	2661
	-0.929***	0.075	-1.076	-0.783	2	Yes	Yes	2769
All grade in 2019	-0.863***	0.054	-0.969	-0.756	1	Yes	No	2248
	-0.857***	0.054	-0.963	-0.750	1	Yes	Yes	2180
	-0.929***	0.074	-1.073	-0.784	2	Yes	No	2340
	-0.953***	0.075	-1.101	-0.805	2	Yes	Yes	2332
Grade 1	-0.753***	0.056	-0.863	-0.643	1	Yes	No	1180
	-0.753***	0.056	-0.864	-0.643	1	Yes	Yes	1164
	-0.964***	0.068	-1.097	-0.830	2	Yes	No	1288
	-0.962***	0.068	-1.096	-0.829	2	Yes	Yes	1284
Grade 3	-0.820***	0.052	-0.923	-0.717	1	Yes	No	2239
	-0.819***	0.052	-0.922	-0.716	1	Yes	Yes	2255
	-0.894***	0.061	-1.014	-0.773	2	Yes	No	2543
	-0.894***	0.061	-1.014	-0.773	2	Yes	Yes	2543
Grade 6	-0.818***	0.067	-0.935	-0.700	1	Yes	No	2896
	-0.817***	0.060	-0.935	-0.700	1	Yes	Yes	2896
	-0.965***	0.067	-1.097	-0.834	2	Yes	No	3900
	-0.953***	0.067	-1.084	-0.822	2	Yes	Yes	3864

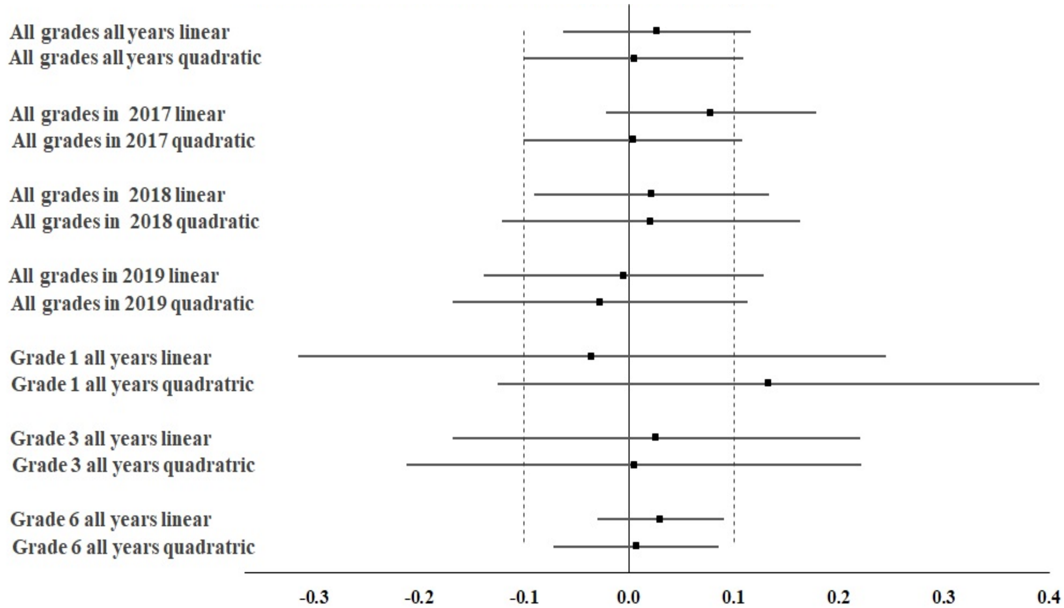
Note: Linear and quadratic RD estimates on equation (2). All grades in all years use all test scores from 24 tests during 2017-2019. All grades in 2017 use the test scores from the 7 tests in 2017. All grades in 2018 use the test scores from the 9 tests in 2018. All grades in 2019 use the test scores from 8 tests in 2019. Grade 1 all years use the test scores from the 4 tests during 2018-2019. Grade 3 all years use the test scores from the 8 tests during 2017-2019. Grade 6 all years use the test scores from the 12 tests during 2017-2019. Order refers to polynomial order and Dist. FE refers to district fixed effect.

Table 3A- 1: SMT Online Training Session Schedule

Start	End	Courses	Batch	Hours
1-May-20	1-Jul-20	C4T-Computational Science (G1-G12)	1	-
12-May-20	1-Jul-20	SMT Principal Training	1	-
6-Jul-20	13-Sep-20	SMT Online 4 Skills (G1-G12)	1(2020)	-
21-Jul-20	29-Aug-20	SMT Principal Training	2	-
1-Aug-20	31-Aug-20	Summative Assessment (Science)	1	-
24-Aug-20	16-Oct-20	Biology (G10-G12))	1	-
1-Sep-20	1-Oct-20	Summative Assessment (Science)	2	-
1-Sep-20	11-Nov-20	Earth Science and Astronomy (Fundamental)	1	20
1-Sep-20	2-Nov-20	C4T Plus-KB-IDE	1	16
		C4T Plus-MicroPython	1	16
		C4T Plus-Python	1	16
		C4T Plus-Scratch	1	16
8-Sep-20	9-Nov-20	C4T Plus-KB-IDE	2	16
		C4T Plus-MicroPython	2	16
9-Sep-20	8-Oct-20	Coding for School Director (C4S)	1	12
14-Sep-20	12-Oct-20	Biology (Grade12)	1	-
18-Sep-20	2-May-21	Earth Science and Astronomy (Advanced)	1	18
19-Sep-20	20-Nov-20	C4T Plus-KB-IDE	3	16
		C4T Plus-MicroPython	3	16
21-Sep-20	21-Nov-20	C4T Plus-Unplugged 2 (Secondary)	1	16
21-Sep-20	26-Nov-20	C4T Plus-Unplugged 1 (Primary)	1	16
26-Sep-20	19-May-21	Physics (Fundamental G10-G12)	1	-
26-Sep-20	21-May-21	Physics (Advanced G10-G12)	1	-
26-Sep-20	27-Nov-20	C4T Plus-KB-IDE	4	16
		C4T Plus-MicroPython	4	16
28-Sep-20	28-Nov-20	C4T Plus-Unplugged 2 (Secondary)	2	16
28-Sep-20	3-Dec-20	C4T Plus-Unplugged 1 (Primary)	2	16
1-Oct-20	2-Dec-20	C4T Plus-Data Science	1	16
5-Oct-20	6-Dec-20	C4T Plus-Unplugged 2 (Secondary)	3	16
		C4T Plus-Unplugged 1 (Primary)	3	16
5-Oct-20	1-Dec-20	AI for School Level 1	1	8
12-Oct-20	13-Dec-20	C4T Plus-Unplugged 2 (Secondary)	4	16
		C4T Plus-Unplugged 1 (Primary)	4	16
30-Oct-20	1-Jan-21	AI for School Level 2	1	8
30-Nov-20	21-Jan-21	AI for School Level 3	1	8
21-Dec-20	26-Feb-21	SMT Online 4 Skills (G1-G12)	1(2021)	-

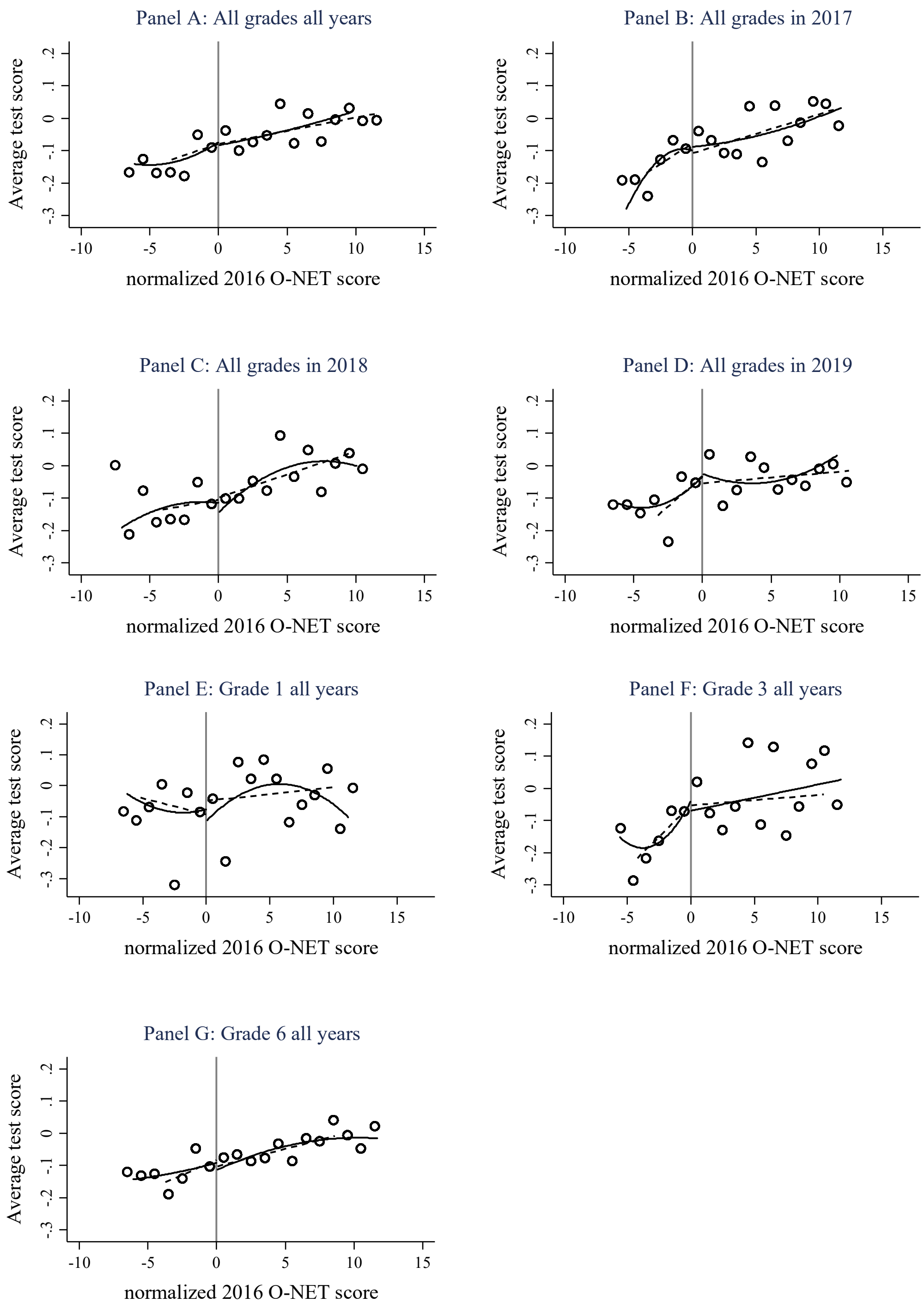
Source: Gathering information from https://learn.teacherpd.ipst.ac.th/courses

Figure 2A- 1: Effect Sizes of Active Learning School Project, test taker weights



Note: Linear and quadratic RD estimates on equation (1). All grades in all years use all test scores from 24 tests during 2017-2019. All grades in 2017 use the test scores from the 7 tests in 2017. All grades in 2018 use the test scores from the 9 tests in 2018. All grades in 2019 use the test scores from 8 tests in 2019. Grade 1 all years use the test scores from the 4 tests during 2018-2019. Grade 3 all years use the test scores from the 8 tests during 2017-2019. Grade 6 all years use the test scores from the 12 tests during 2017-2019.

Figure 2A- 2: Impacts on school average test score 2017-2019, test taker weights



Note: Bandwidths for these graphs are mean squared error (MSE) optimal bandwidths (Calonico et al. 2017) command with district and test fixed effect and cluster district for standard error. Each dot represents the average test score in a given bin of bin width one. Dashed lines represent linear fit and solid lines are quadratic fit.

Figure 2A- 3: The learning models/styles

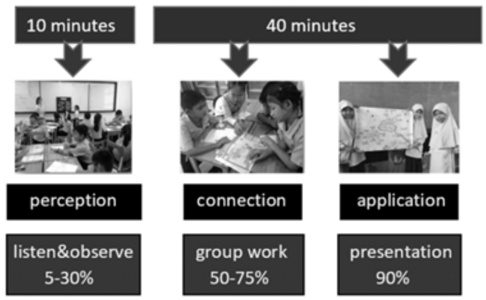


Figure A: Experiential

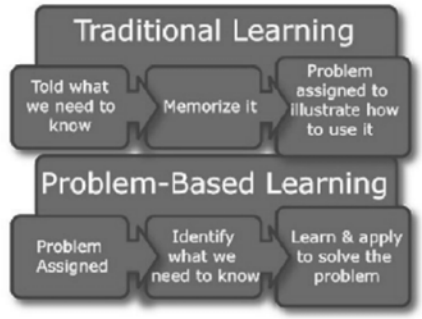


Figure B: Problem-based

Figure C: Problem-based

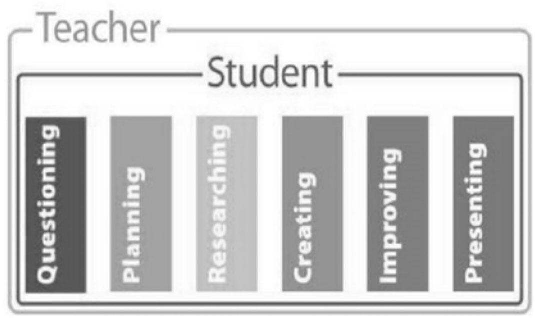
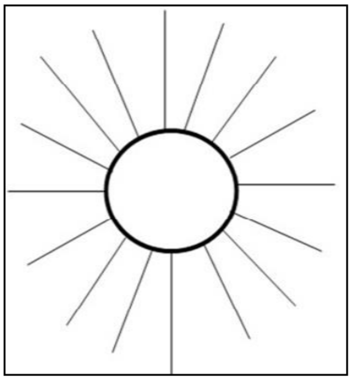


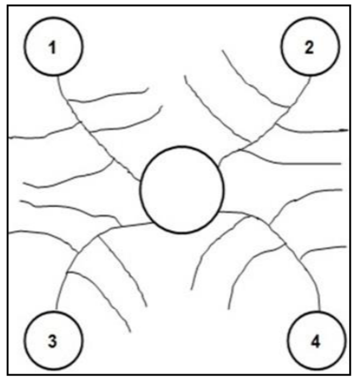
Figure D: Thinking-based

Source: Active learning guideline by an ALS supervisor team with author's translation

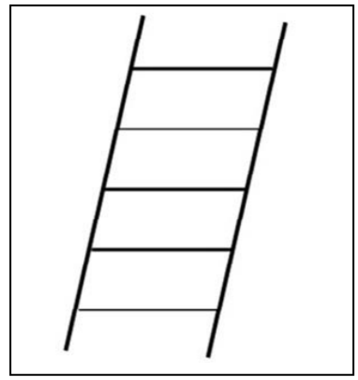
Figure 2A- 4: Example of graphic organizers



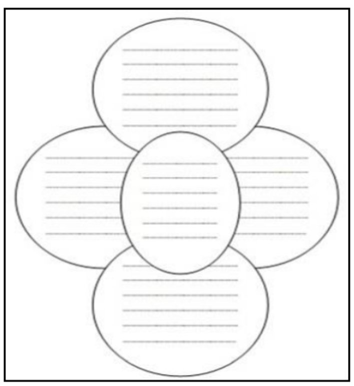
Web



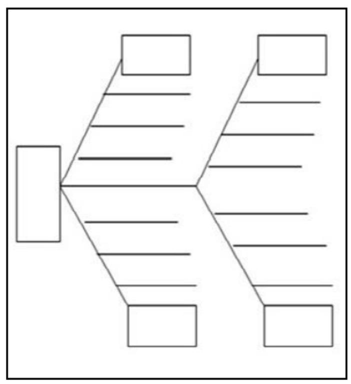
Mind Mapping



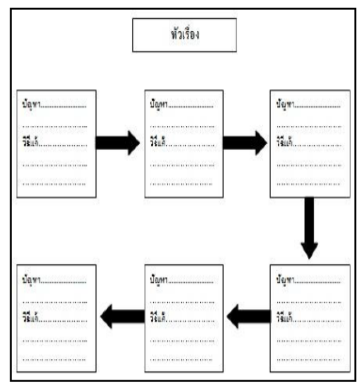
Ranking Ladder



Circle Diagram

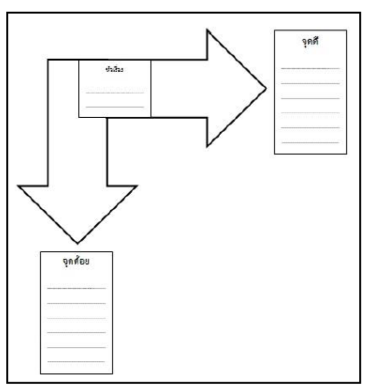


Fish Bone

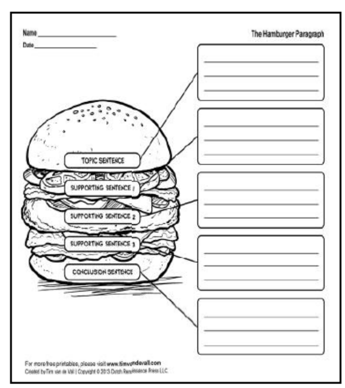


Diagram

Table

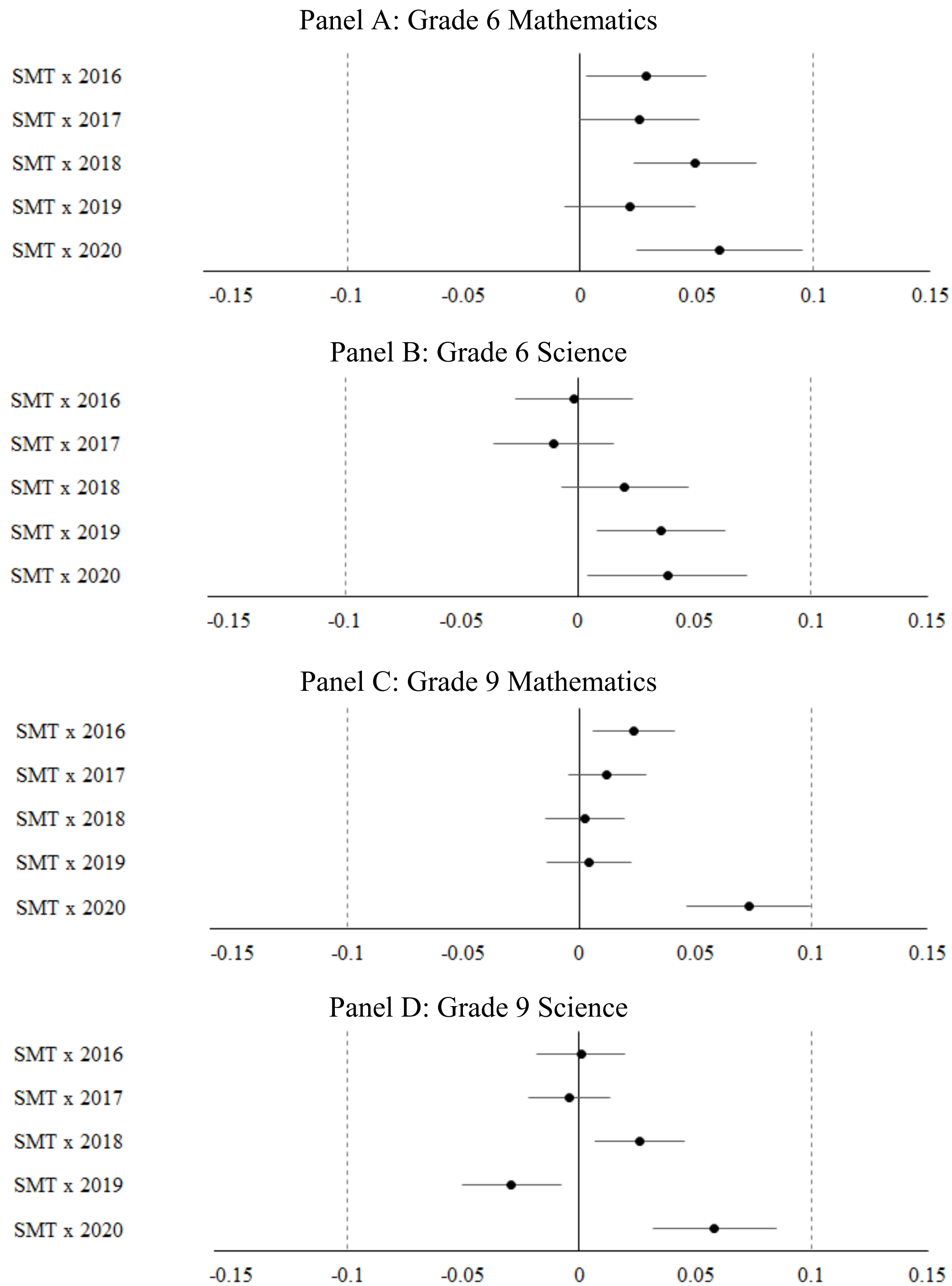


Diagram



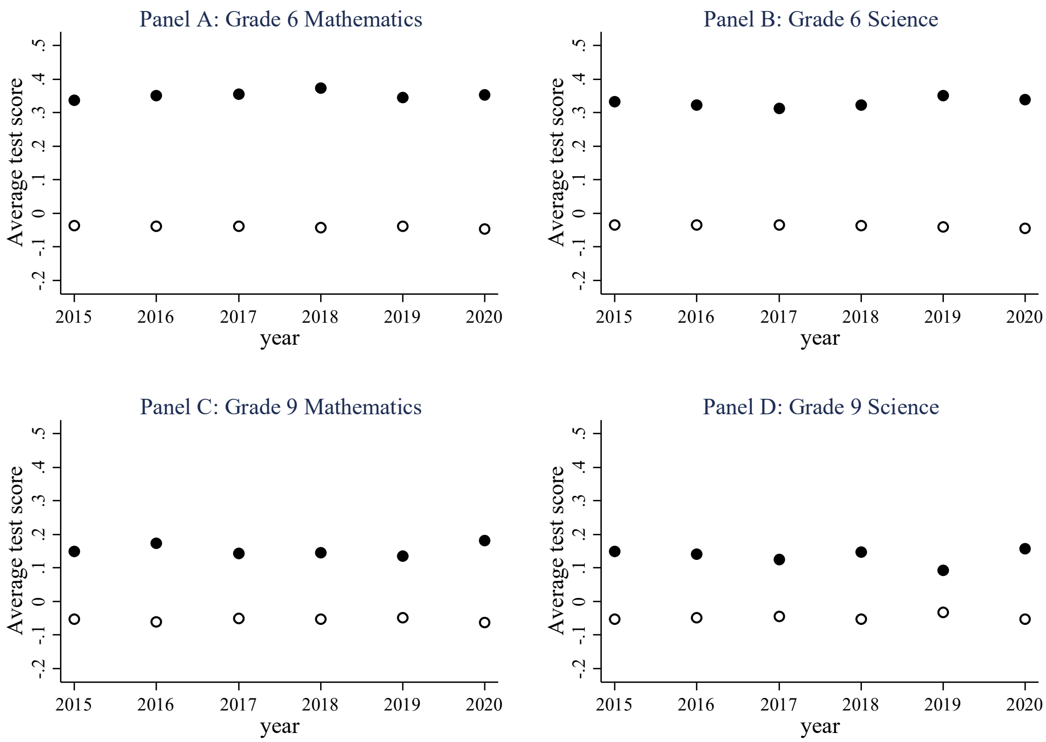
Hamburger Approach

Figure 3A- 1: Effect sizes of Science Mathematics and Technology School Project



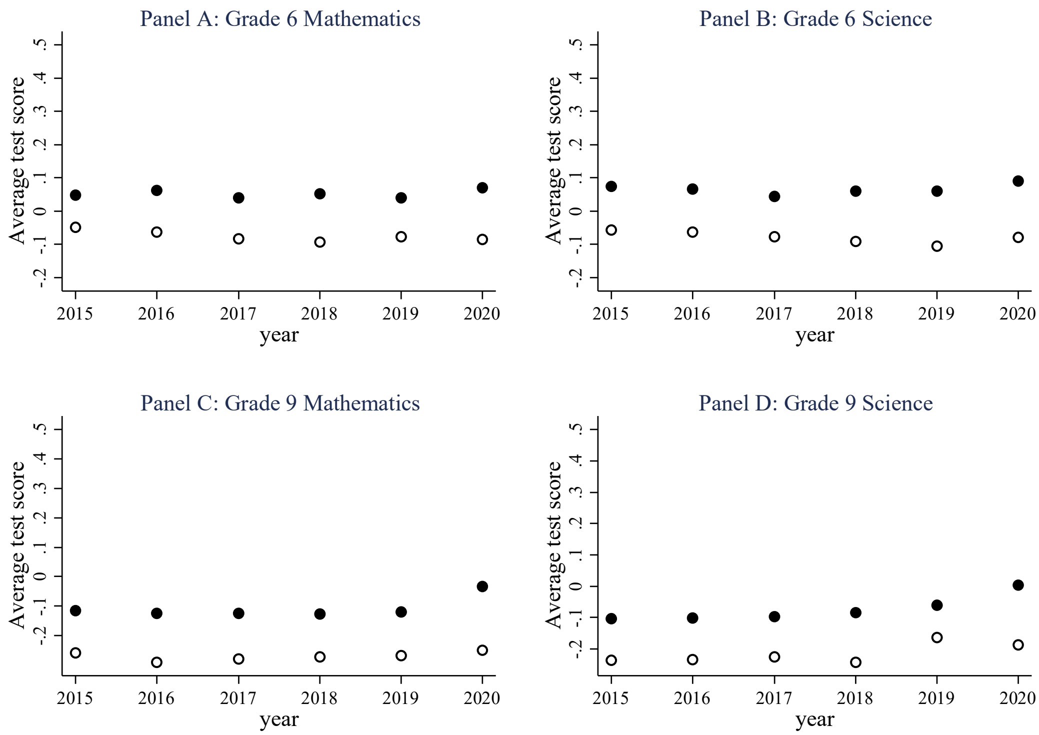
Note: DID non-weighed estimates on equation (1). SMT is a dummy for treatment school and a set of {2016, 2017, 2018, 2019, 2020} are academic year dummies. Panel A and Panel B used mathematics and science test scores of grade six students during 2015-2020. Panel C and Panel D used mathematics and science test scores of grade nine students during 2015-2020. The solid black dots represent the effect sizes while the horizontal lines cover 95% confidence interval values.

Figure 3A- 2: Average test scores grouped by treatment status



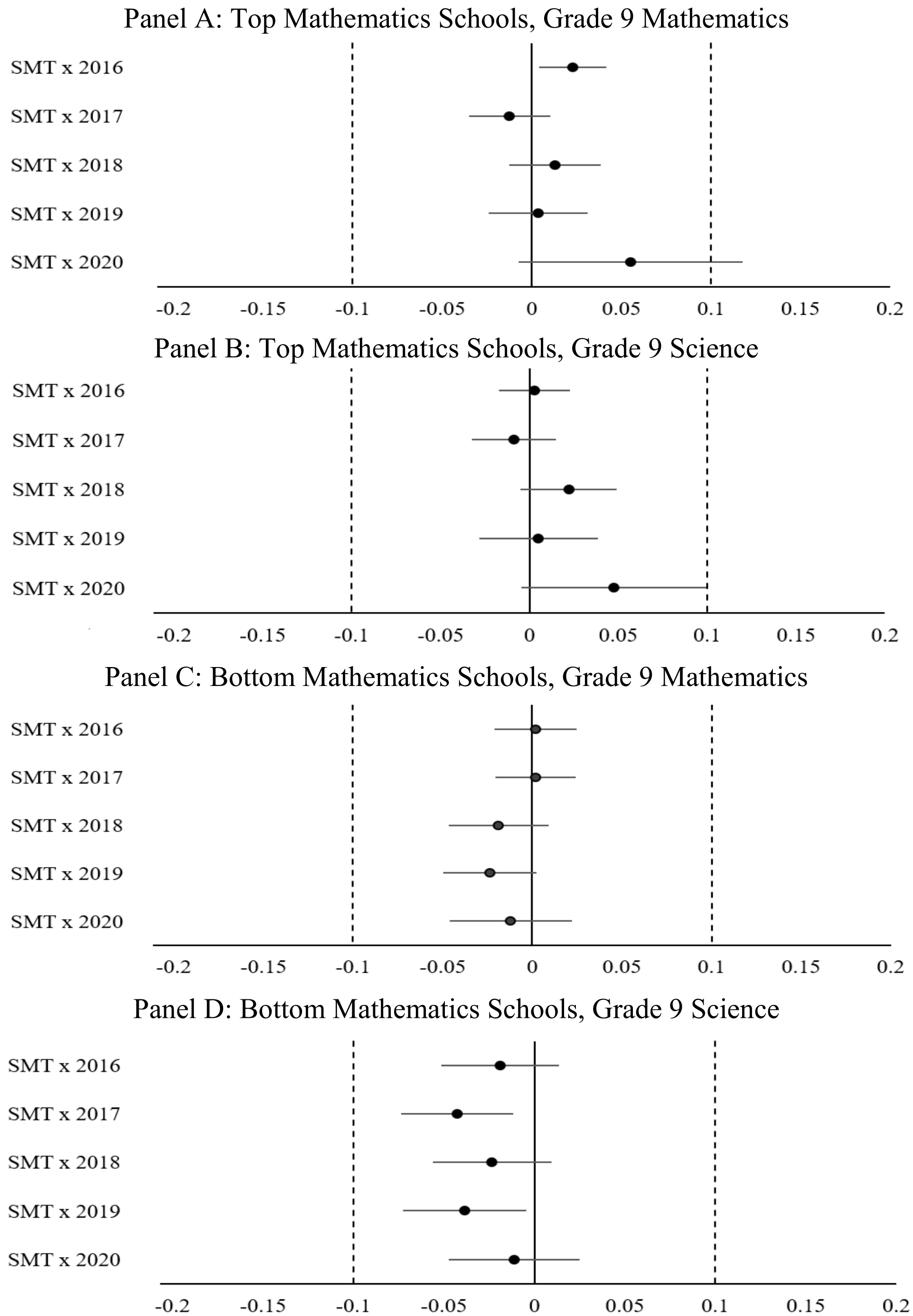
Note: Solid black dots represent average standardized test scores of the treatment group weighted using number of test takers in each school in each year. Open black dots represent average standardized test scores of control groups weighted using number of test takers in each school in each year. Panel A and Panel B show average standardized mathematics and science test scores of sixth-grade students. Panel C and Panel D show average standardized mathematics and science test scores of ninth-grade students.

Figure 3A- 3: Average test scores grouped by treatment status (without weight)



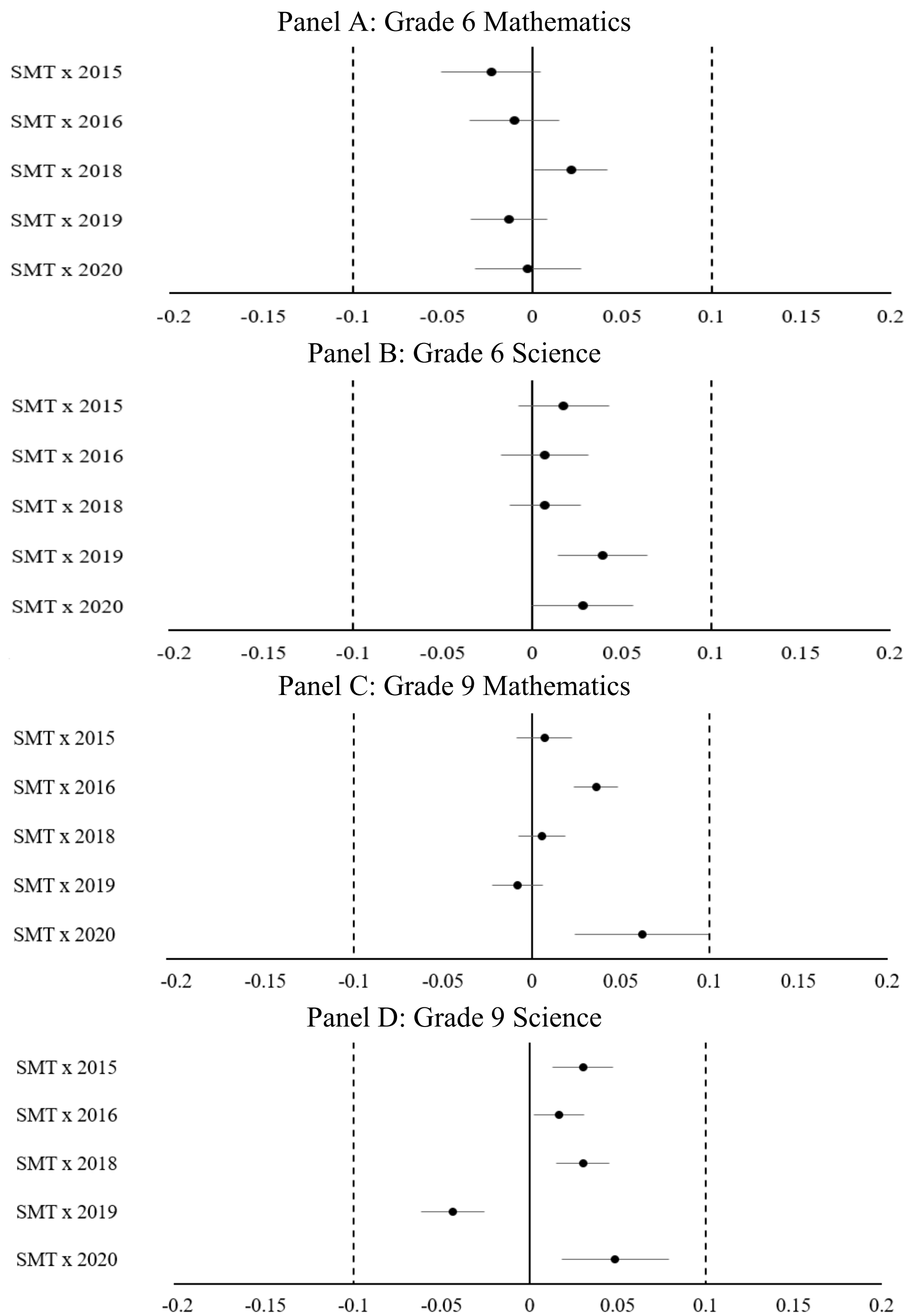
Note: Solid black dots represent average standardized test scores of the treatment group weighted using number of test takers in each school in each year. Open black dots represent average standardized test scores of control groups weighted using number of test takers in each school in each year. Panel A and Panel B show average standardized mathematics and science test scores of sixth-grade students. Panel C and Panel D show average standardized mathematics and science test scores of ninth-grade students.

Figure 3A- 4: Effect sizes of SMT Project, top vs bottom science schools, base year 2015



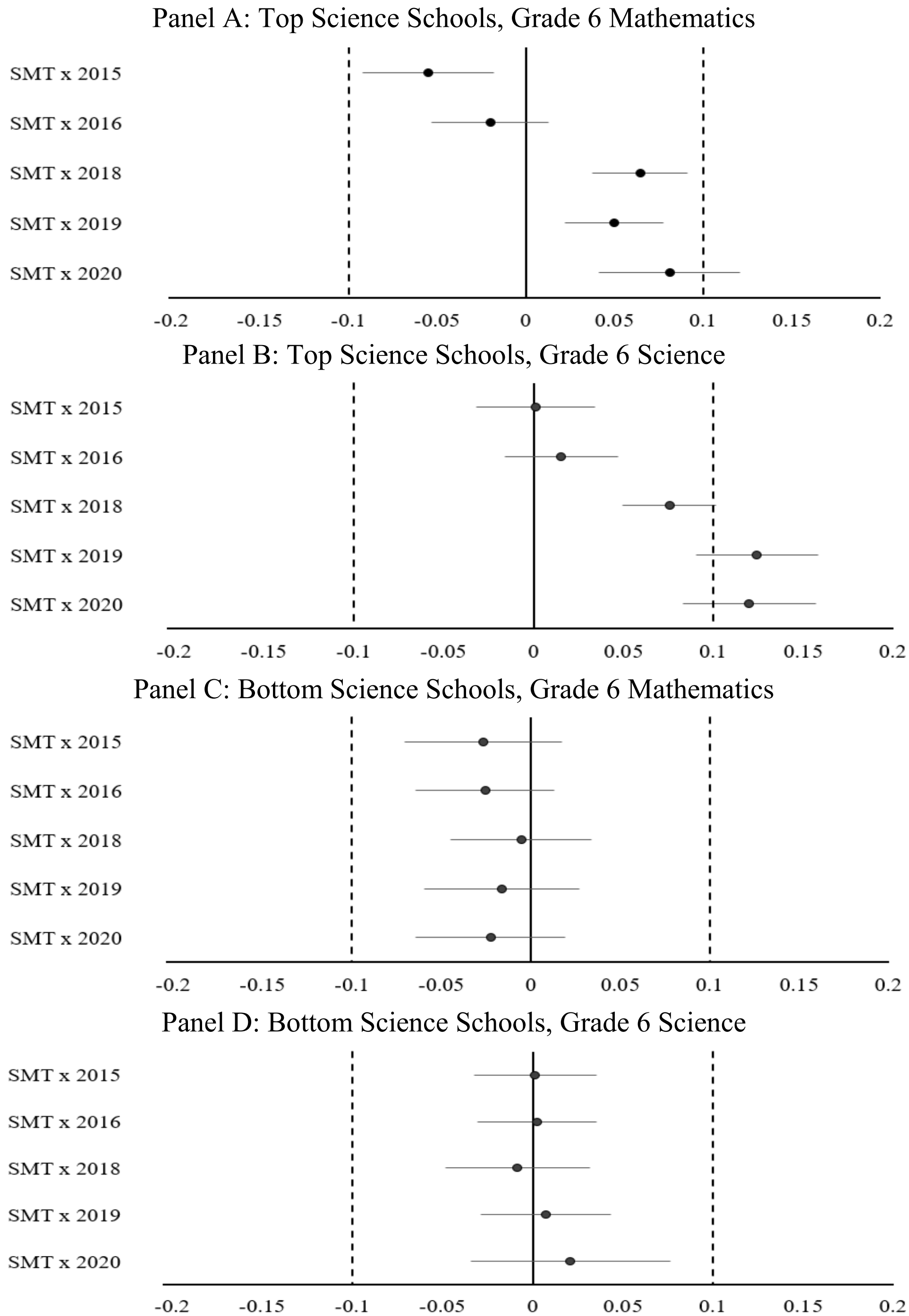
Note: DID estimates on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2016, 2017, 2018, 2019, 2020} are academic year dummies. Panel A and Panel B used mathematics and science test scores of grade nine students in top mathematics schools during 2015-2020. Panel C and Panel D used mathematics and science test scores of grade nine students in bottom mathematics schools during 2015-2020. The solid black dots represent the effect sizes while the horizontal lines cover 95% confidence interval values.

Figure 3A- 5: Effect sizes of SMT Project, base year 2017



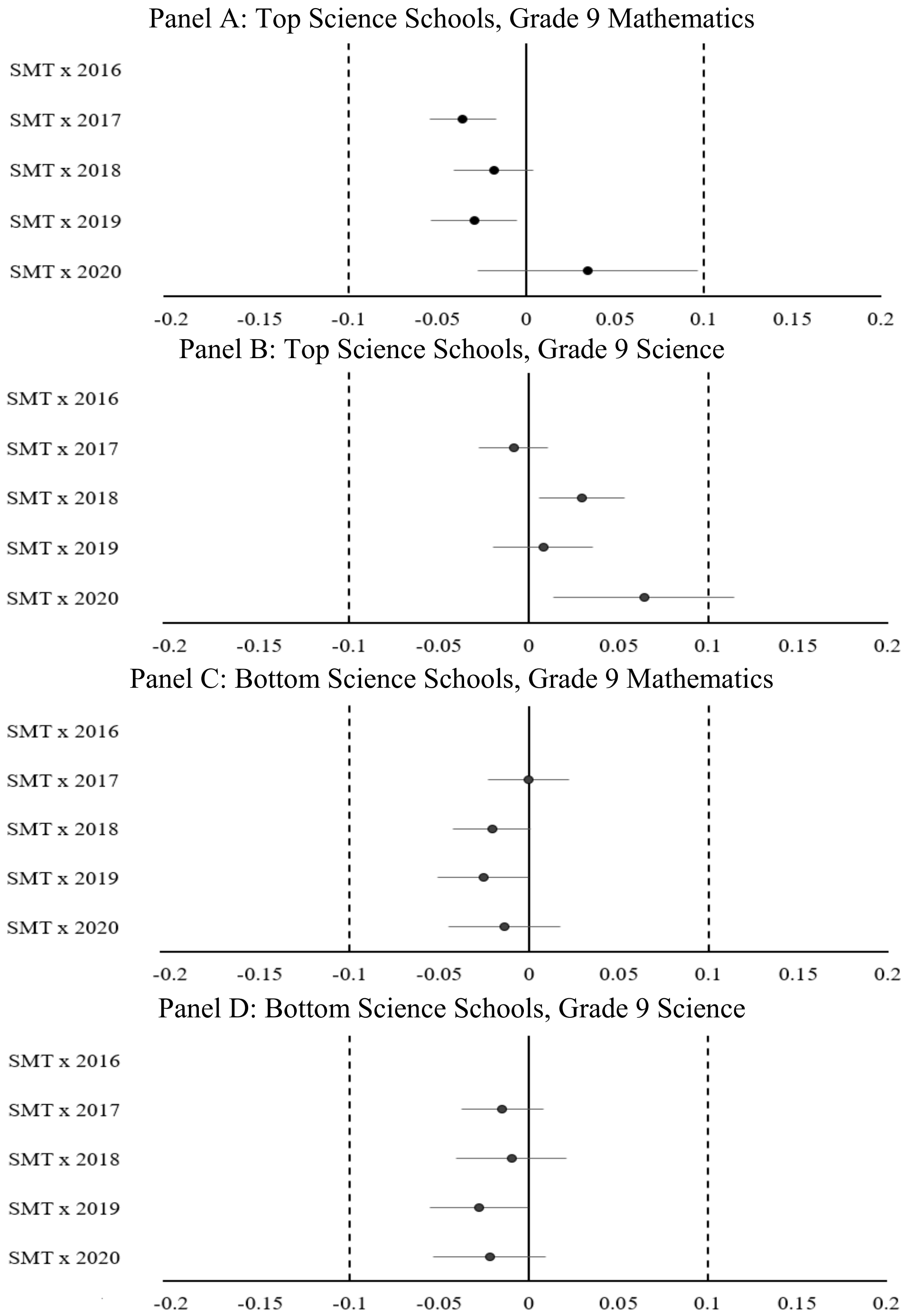
Note: DID estimates on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2015, 2016, 2017, 2018, 2019, 2020} are academic year dummies. Panel A and Pabel B used mathematics and science test scores of grade six students during 2015-2020. Panel C and Panel D used mathematics and science test scores of grade nine students during 2015-2020. The solid black dots represent the effect sizes while the horizontal lines cover 95% confidence interval values.

Figure 3A- 6: Effect sizes of SMT Project, top vs bottom science schools, base year 2017



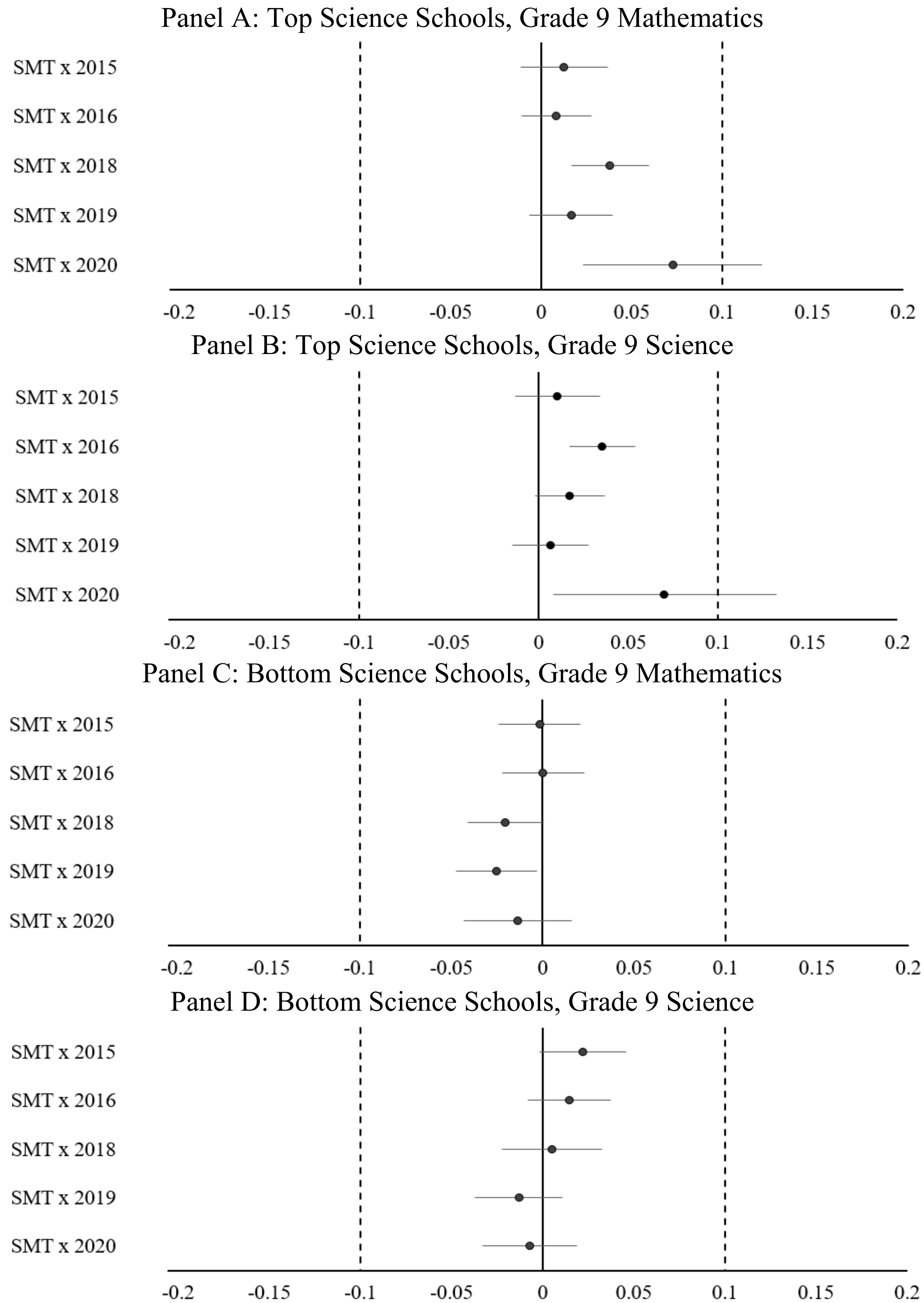
Note: DID estimates on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2016, 2017, 2018, 2019, 2020} are academic year dummies. Panel A and Panel B used mathematics and science test scores of grade six students in top science schools during 2015-2020. Panel C and Panel D used mathematics and science test scores of grade six students in bottom science schools during 2015-2020. The solid black dots represent the effect sizes while the horizontal lines cover 95% confidence interval values.

Figure 3A- 7: Effect sizes of SMT Project, top vs bottom science schools, base year 2016



Note: DID estimates on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2016, 2017, 2018, 2019, 2020} are academic year dummies. Panel A and Panel B used mathematics and science test scores of grade nine students in top science schools during 2015-2020. Panel C and Panel D used mathematics and science test scores of grade nine students in bottom science schools during 2015-2020. The solid black dots represent the effect sizes while the horizontal lines cover 95% confidence interval values.

Figure 3A- 8: Effect sizes of SMT Project, top vs bottom science schools, base year 2017



Note: DID estimates on equation (1) weighed by the number of test takers. SMT is a dummy for treatment school and a set of {2016, 2017, 2018, 2019, 2020} are academic year dummies. Panel A and Panel B used mathematics and science test scores of grade nine students in top science schools during 2015-2020. Panel C and Panel D used mathematics and science test scores of grade nine students in bottom science schools during 2015-2020. The solid black dots represent the effect sizes while the horizontal lines cover 95% confidence interval values.