2019

HARVARD-YENCHING INSTITUTE WORKING PAPER SERIES

HIGH SCHOOL ADMISSION REFORMS, EQUALITY OF EDUCATIONAL OPPORTUNITY, AND ACADEMIC PERFORMANCE IN TAIWAN

Huang Min-Hsiung | Academia Sinica

High School Admission Reforms, Equality of Educational Opportunity, and Academic Performance in Taiwan

Min-Hsiung Huang

Institute of European and American Studies, Academia Sinica,

Nangang District, Taipei 11529, Taiwan

mhhuang@sinica.edu.tw

TEL:886-2-3789-7246

January 3, 2019

High School Admission Reforms, Equality of Educational Opportunity, and Academic Performance in Taiwan

Abstract

Recent school admission reforms in Taiwan have been designed to avoid an unhealthy focus on academic competence and on preparation for high-stakes tests. In response, concerns have been raised about whether a more holistic assessment of student performance for school admission would dampen academic competence and equality of educational opportunities. Using nationally-representative trend data from different sources and a difference-in-differences methodology, findings point to a remarkable decline in the proportion of high-achieving students in mathematics, but a similar decline is not observed in science. The impact of family socioeconomic status on student academic performance has been stable over time, before and after major school admission reforms.

Keywords: school admission reform, high-stakes test, academic achievement, equality of educational opportunity, Taiwan

Introduction

Taiwan has earned a reputation as one of the top-performing countries in high school mathematics and science in international assessments such as the PISA (the Programme for International Student Assessment) and the TIMSS (Trends in International Mathematics and Science Study). According to recent research (Huang, 2017), Taiwanese students of two recent cohorts, fourth-graders in 2003 and 2007, consistently showed remarkable improvement in the percentage reaching the advanced international benchmark in mathematics, rising from about a quarter to nearly a half, as they progressed from grade 4 to grade 8. Such a dramatic improvement in just four years of schooling is not seen in other countries. Huang (2017) speculates that the improvement is likely a consequence of a national examination administered to nearly all Taiwanese ninth graders for admission to senior high schools. According to a theory proposed by Bishop (1996, 1997, 1999) and extended by Woessmann (2003, 2005), large-scale and high-stakes public examinations, such as the high-school entrance examination administered in Taiwan, hold teachers and schools accountable for student performance and provide incentives for students, teachers, and school administrators to focus on learning.

Nevertheless, recent educational reforms in Taiwan have been designed to reduce the impact of test performance on admission to the next level of schooling. These reforms raise concerns about whether a more holistic assessment of student performance for school admission would weaken academic competence. Whether educational opportunity would become more unequal among students of different socio-economic backgrounds is another concern. The present study aims to address these issues.

These issues are not specific to Taiwan, other societies share similar concerns. In the City of New York, for example, eight elite public high schools (specialized high schools) admit students on the sole basis of a single-day test, formally called "the Specialized High Schools Admissions Test." Much of the debate about this admission policy focuses on the bias towards high-SES and privileged ethnic groups (Monahan, Chapman, & Fermino, 2013; Taylor, 2018). Using test scores as the sole criterion for admission to the city's specialized high schools has been criticized for creating a "rich-get-richer phenomenon," because children of wealthy parents have the funds to attend expensive test-prep classes and tutoring. The presence of specialized high schools in New York City has been characterized as "undemocratic," and these elite high schools criticized for "not reflecting the city's diversity" (Taylor, 2018). It was the intention of the Mayor of the City of New York, Bill de Blasio, to overhaul the schools' admissions process, so multiple criteria, such as grades, portfolios and extracurricular activities, could be taken into account. However, that intention has not yet been realized (Taylor, 2018).

Nevertheless, some citizens support the use of high-stakes admission test as the sole criterion for admission to the specialized high schools in New York City. These supporters claim that introducing other criteria into admission decisions, such as extracurricular activities, would reduce the quality of students in the schools and may bring even more privileged children into these elite schools (Monahan et al., 2013).

While admission to these eight specialized high schools in the City of New York is based solely on performance in a high-stakes test, admission to senior high schools in Taiwan has become less dependent on test performance. Taiwan follows the 6-3-3-4 system of schooling in which a child spends six years in primary school, three years in junior secondary, and then another three years in the senior secondary before proceeding to higher education. At the senior secondary level, there is a distinction between academic and vocational programs (Jen, Lee, Chen, Lo, & Lin, 2016).

For decades, senior high schools in Taiwan have been highly stratified and subtly ranked according to academic performance. During the period between 1958 and 2000, when the high school entrance examination (henceforth, HSEE) was used for school admission, each senior high school has had a minimum required test score for admission, with places in the best senior high schools going to those who do extremely well on the examination (Hsu, 2002). Thus, high-performing Taiwanese students compete ferociously for places in the best public academic high schools. Those who fail to achieve the minimum required test scores for the best high schools subsequently compete to enter second- or third-tier high schools, and may devote substantial time to improve their scores to win a place in a relatively good high school.

The stratified system facilitated by the HSEE has been criticized for an unhealthy focus on academic competence and preparing for a high-stakes test (Wang, 1989). In 2001, school admission reforms were implemented at the national level to move toward a more holistic admission system, and the HSEE was replaced by the Basic Competence Test for Junior High School Students (henceforth, BCT). In 2014, 9-year compulsory education was extended to 12 years, and an attempt was made to further downplay the role of high-stakes testing for admission to senior high schools (Ministry of Education, 2015). In 2014, the BCT test was replaced by the Comprehensive Assessment Program for Junior High School Students (henceforth, CAP) which is a standards-based test.

Since the implementation of HSEE in 1958, there has been three different approaches to high school admissions and the role of assessment changed across these different approaches. Therefore, in the present study, I distinguish between three historical periods -- the HSEE period between 1958 and 2000, when test performance was the only criteria for admission to senior high schools; the BCT period between 2001 and 2013, when test was the only criteria for admission for only about half of the students; and the CAP period, which started in 2014, in which test performance is no longer the only criteria for admission. I examine whether or not academic performance has declined over time, and whether or not the effect of family socioeconomic status on student academic performance has increased over time, as test performance becomes less crucial for school admission across the HSEE, BCT, and CAP periods.

Many Taiwanese people think that standardized tests, when used as the sole admissions factor, create the fairest system. It is ironic, then, that Taiwan has been among those countries with the least equality of educational opportunity (Huang, 2008). Schuetz, Ursprung and Woessmann (2008) use international data from the TIMSS for more than 50 countries, and they consider the effects of family background on student performance in Grade-8 mathematics and science as an index of inequality of educational opportunity. According to Schuetz et al. (2008, Table 3, p. 291), England, Taiwan, Scotland, Hungary and Germany are the five countries with the largest estimated effects of family background in the sample of 54 countries.

To investigate trends in academic performance and inequality of educational opportunity among Taiwanese students, I use data from the TIMSS, the PISA, and the TEPS (Taiwan Education Panel Survey). These data, all together, cover student performance before and after educational reforms, across the HSEE, BCT, and CAP periods. In the next section, I introduce Taiwan's recent policy reforms in admission to senior high schools. The third section provides a summary on the data sets used for analysis. The fourth reports research findings, and the final section concludes.

Reforms in Admission to Taiwanese Senior High Schools

The HSEE Period between 1958 and 2000

Before the Chinese Kuomintang (KMT) administration moved to Taiwan in 1949, each Taiwanese senior high school enrolled students on an independent basis. From 1952 onwards, public senior high schools in Taipei began to enroll students in a collective fashion. Senior high schools in many other cities gradually adopted this mode of school enrollment. In 1958, the KMT administration began to implement regional examinations for senior high school admission. This policy implementation marked the dawn of the senior high school entrance exam era (C. P. Chang, 2002; Hsu, 2002).

The most important characteristics of the HSEE were that taking the examination was the only way to enter senior high school, and each student's test performance was the sole criteria for his or her subsequent admission to a specific senior high school. In addition to this peculiar characteristic, there are several other attributes such as "joint examination-enrollment", "one-examination-per-year", "collective construction of test items within examination quarters", "multiple forms of test items", "cross-subject heterogeneity of total test scores", and "raw score evaluation" (C. P. Chang, 2002).

"Joint examination-enrollment" refers to the fact that insofar as each student took part in the examination and obtained his or her test results, the Collective Enrollment Examination Committee would assign him or her to the corresponding senior high school in accordance with the rank order of school as well as the preference and the test performance of the student. "One-examination-per-year" refers to the fact that each student had only one shot at the senior high school entrance exam per year. Should the student's test results turn out to be subpar relative to his or her prior performance, perhaps due to unforeseen contingencies such as illness, he or she would need to wait an entire year to re-take the entrance examination (C. P. Chang, 2002).

“The collective construction of test items within examination quarters” refers to the convention that examination setters would enter examination quarters to construct test items about 10 days before the examination. While all examination setters were confined to these examination quarters, no external contact was allowed. In practice, there was no rigorous study with respect to the level of test item difficulty, and test items of differing levels of difficulty usually arise without specific sequence (C. P. Chang, 2002).

“Multiple forms of test items” pertains to the fact that there were different types of test items. Specifically, there were multiple-choice test items and test items that required each student to fill in the blank, to calculate and to solve arithmetic problems, or to derive mathematical proofs. Test items of different types did not carry the same weight, and there was no objective standard as to why specific types of test items counted more than others (C. P. Chang, 2002).

“Cross-subject heterogeneity of total test scores” refers to the fact that the total test scores were not equal across subjects. For instance, the total test score for the senior high school entrance exam was 700. This total test score could be decomposed into 200 for Chinese literature, 120 for mathematics, 140 for science, 140 for social science, and 100 for English. There was no objective standard as to why specific subjects had higher total test scores. “Raw score evaluation” refers to the fact that each student’s total test score on the examination was the sum of all of his or her subject test scores, and the subject test scores were the sum of his or her raw test scores based on the answers given to the test items within each subject (C. P. Chang, 2002).

During the HSEE regime, there were some proposals to reform admission to senior high schools. These premature proposals were initially experiments that co-existed in parallel with the entrance exam system. The first experiment with reform began in 1990 when the Ministry of Education proposed a “voluntary admission” mechanism to allow junior high school students to enter specific senior high schools without needing to take an entrance examination (Lin, 2004). This voluntary admission program was run on a small scale in Taipei, Kaohsiung, as well as some less populated areas, such as Yilan, Taitung, Penghu, and Kinmen (Lin, 2000). On the basis of overall academic performance in junior high school and test performance on a public examination set by a regional examining board, program participants were assigned to specific senior high schools, vocational schools, and polytechnics. In 1992, when the voluntary admission program was most popular, the sum of all participants was about one-tenth of total junior high school graduates. In 1993, despite the growing popularity of the new admission program, 84 legislators requested that the Ministry of Education terminated this school admission reform because, they claimed, it was unfavorable to graduates of higher academic caliber, and it may have dampened the motivation to learn (Chou, Hu, & Liu, 2013). In addition, there was no evidence, they argued, that the new admission program reduced the academic burden on junior high school students.

Apart from the voluntary school admission program, there were different admission options implemented on a small scale in several regions towards the end of the HSEE regime. All these regional and small-scale reforms shared the same goal: all aimed to reduce the impact of HSEE on admission to senior high schools (Ding, 1998; Executive Yuan’s Education Reform Audit Committee, 1996).

The BCT Period between 2001 and 2013

Due to the excessive stress placed on students under the HSEE regime, the Ministry of Education replaced the HSEE with the BCT and a diversified plan for admission to senior high schools in 2001 (Ministry of Education, 1998). This made it necessary for all graduating students to take the BCT for admission to senior high schools, but for about a half of the graduates, performance on the BCT is not the sole criteria for admission.

The BCT.

Before the Ministry of Education implemented the BCT in 2001, the Ministry began building a bank of test items. To compose an examination sheet for a subject, examiners would choose test items of commensurate levels of difficulty from the bank. In stark contrast to the HSEE, the BCT rested on more rigorous test item research (Wang & Yu, 2006). In the BCT, all test items were single-choice questions, and for each there was only one correct answer among the four multiple choices. All the test questions were ranked in the typical order from easiest to most difficult. Each student's final test result was composed of "standardized scores", that is scores converted from raw scores on the basis of test score normalization (Wang & Yu, 2006).

The full name of BCT is the Basic Competence Test for Junior High School Students. The word "Basic" does not imply that the BCT was easy. The word "Basic" means that the BCT emphasized the core, most important pieces of knowledge that junior high school students were required to master. The BCT usually avoided the inclusion of esoteric or arcane test questions, instead, emphasizing basic knowledge and targeting difficulty appropriate to a typical middle-school student's intellectual capacity (C. P. Chang, 2002; Song, Hsu, Zheng, Jiang, & Sun, 2007).

The BCT subjects included Chinese literature, English, mathematics, science, and social science. Each of these five subjects were worth 60 points for a full score of 300. After 2007, a Chinese analytical writing exam with six score levels became a part of the BCT (Ministry of Education, 2005). The overall exam result would be calculated from the total scores across all the subjects plus twice the Chinese analytical writing score level. As a consequence, the BCT would be out of a total of 312 points. After 2009, each BCT subject was worth 80 points with a BCT out of a total of 412 points (Ministry of Education, 2008). The BCT was held twice in May and July each year. Students whose BCT scores were the sole criteria for admission could take the BCT in May or July. Other students must take the BCT in May only.

Diversified Program for Admission.

The diverse admission program provides three different channels for admission to senior high schools. The first channel is for students who possess special skills and talents in areas such as mathematics, science, language, music, art, dance, drama, or sports. Students submitted their applications to particular schools, or to an enrollment committee organized by several schools. The criteria for selection included performance on the BCT, but the relative weight of the BCT scores should not have been more than 50% of the overall evaluation for this selection. No consideration should have been given to the student's academic records during junior high school. Each senior high school could take into account the student's nonacademic performance in junior high school, such as extracurricular skills and performances. For selection, each senior high school could also choose to conduct various forms of assessments, such as experiments, oral examinations, mini research projects, practical operations, and other artistic activities, but each school should not impose additional academic pressure on middle-school students in the form of academic written examinations (Ministry of Education, 1998).

The second channel was for students who wish to enroll at a nearby senior high school within the local catchment area. Each student could apply for admission to one academic high school, one vocational school, or both at the same time. Each senior high school would grant admission offers on the basis of the BCT test score, or a transformed score which placed greater weight on one or two BCT subjects. Each senior high school could also consider the student's prior performance in nonacademic subjects and extracurricular activities. Should the student apply for direct admission to an affiliated senior high school, its admission committee would consider his or her academic records within the affiliated junior high school (Ministry of Education, 1998).

The third and the final channel relied on the BCT scores as the sole criterion for admission to senior high schools. Each BCT subject carried equal weight for academic evaluation and admission. For academic senior high schools, each school could have a quota of up to 10% for enrolling students from the first channel. For the second and the third channels, the quotas were 30% and 60%, respectively. In the case that some of these academic high schools could not accommodate students from the first channel -- those who had special interests and talents -- their quota for enrolling students from the third channel -- in which the BCT scores were the sole criterion for admission -- could be extended to 70%. For vocational high schools, the percentage of total student intake for the first and the second channels could be as high as 60%, leaving a percentage of 40% for students enrolling through the third channel (Ministry of Education, 1998).

The CAP Period Starting 2014 and after

After compulsory education in Taiwan was extended from nine to 12 years of schooling in 2014, an attempt was made to further downplay the importance of test performance on admission to senior high schools. Two major reforms have been implemented to reach that goal. First, the BCT standardized test was replaced with a standards-based test, the CAP, in which subject-specific scores reported to students range down to three categories (A, B, and C) which are composed of 7 levels (A++, A+, A, B++, B+, B, and C) (National Elementary School and Junior High School Student Performance Evaluation Principles, 2015). Second, most students graduating from junior high schools are channeled to senior high schools via the "examination-free admission program" (Ministry of Education, 2015). A minority of students, about 10% of the total junior high school graduating cohort, are channeled to senior high schools via the "special examination admission program." Students following this channel for admission are mostly those who possessed special skills and talents in areas such as mathematics, science, language, music, art, dance, drama, and sports, etc.

The CAP.

Subsequent to the removal of the BCT, the CAP served as a device to evaluate and monitor the academic performance of all students in their final year of junior high school. It is necessary for all the final-year junior high school students to take the CAP. Performance on the CAP is consequential for students seeking admission to more prestigious senior high schools. The CAP functions as a means of objective academic selection as it is not possible for students to gain admission into elite senior high schools without excellent performance on the CAP. For this reason, the policy title "examination-free admission program" is misleading. In reality, performance on the CAP is the key criteria, though not the sole criteria, for admission to prestigious senior high schools.

Examination-Free Admission Program.

Under the examination-free admission program, all senior high schools should offer admissions to applicants as long as the number of applicants does not exceed school admission quota. In such cases, it was not permissible for senior high schools to specify revised admission criteria or erect new hurdles for admission. Neither were these schools allowed to consider the prior academic performance of junior high school students for admission. Alternatively, if the number of applicants from junior high school were to exceed the school's admission quota, these applicants would have to engage in a competition and selection based on three criterions: (a) test performance on the CAP, (b) whether or not the specific high school applied for is the student's most preferred, and (c) non-academic and behavioral records in junior high schools (Keelung-Taipei Senior High School Examination-Free Admission Committee, 2018).

In practice, there are more applicants to prestigious senior high schools, which are usually located in more populated urban areas, than these schools can admit. In such cases, test performance on the CAP, though not the sole criteria, become crucial for admission. The percentage of junior high school graduates who gained admission to senior high schools via the so called exam-free admission program exceeded 90% after 2014 (K-12 Education Administration, 2016, 2017).

Special Examination Admission Program.

Another admission channel is via the special examination. The special examination admission serves two purposes. One is to provide an admission channel for students who possess special skills and talents in music, art, dance, drama, sports, etc.; the other is to meet the need of senior high schools that prefer admitting students independently via an entrance examination of their own design (Wu, 2016).

Students who possess special skills and talents can apply to senior high schools that provide curriculum suitable for their special interests. Such applications need to go through the official audit review process for music, art, dance, drama, sport, and other practical subjects. Senior high schools grant admissions based on students' performance on these special skills. In most cases, they also consider students' academic performance in certain subjects, as indicated by their CAP test scores.

Some senior high schools prefer admitting students independently and designing their own entrance-examinations. These schools usually use students' CAP scores to set a minimum requirement for application. Once students have a CAP score high enough to meet the minimum requirement, they are eligible to attend an entrance examination designed by the schools. These schools may design their own test items or delegate this task to external institutions. Usually these entrance examinations cover only one to three academic subjects.

Many senior high schools became less interested in admitting students independently, probably because they did not attract better students in this way, or could not afford the administrative burden. In 2014, about 15,000 junior high school graduates entered senior high schools via independent school entrance-examinations. By 2015, however, only eight senior high schools continued to offer the option, with a total quota of merely 910 students (K-12 Education Administration, 2017).

Data

For trend analysis, I use data from the TIMSS, the PISA, and the TEPS. Taiwanese eighth-graders participated in the TIMSS in 1999, 2003, 2007, 2011, and 2015. Therefore, data from the TIMSS, all together, allow for a trend analysis for Taiwanese Grade 8 performance in mathematics and science in the HSEE, BCT, and CAP periods. The TIMSS 1999 provides Taiwanese data for student performance in the HSEE period because students of this cohort took the HSEE after they progressed to Grade 9, the last year of junior high school. Taiwanese data from the TIMSS 2003, 2007 and 2011, altogether, represent student performance in the BCT period. Finally, the TIMSS 2015 reports student performance in the CAP period. These Taiwanese TIMSS data, all together, cover student performance before and after policy reforms with respect to admissions to senior high schools.

15-year-old Taiwanese students participated in the 2006, 2009, 2012, and 2015 PISA. These Taiwanese students were in Grade 9 or 10 at the time of the PISA survey. The trend analysis uses data from Grade 9 alone because it is the grade level facing and experiencing the transition to senior high schools. Data from Grade 10 are not used for analysis because these students have already been admitted to senior high schools, and their test performance in PISA is likely affected by learning experiences after entering senior high schools. Taiwanese data from the PISA 2006, 2009 and 2012 covers student performance in mathematics, science and reading in the BCT period, while data from the PISA 2015 allow for an analysis of student performance in the CAP period. Data from the PISA, therefore, make possible a comparison between the BCT period and the CAP period in student performance.

The TEPS is a longitudinal survey in Taiwan, and the survey covers student performance in mathematics, curriculum-free problem solving tasks, and multiple subjects learned in school. The TEPS includes two nationally-representative samples, eleventh-graders in 2001 and 2005. Students in these two national samples participated in a follow-up survey in 2003 and 2007, respectively, when the eleventh-graders had moved up to Grade 12. The eleventh-graders of 2001 were admitted under the HSEE regime, while the eleventh-graders of 2005 were admitted under the BCT regime. Therefore, data from the TEPS are used for a comparison between the HSEE and the BCT periods in student performance.

TIMSS

The use of the TIMSS data in this study is confined to those collected in 1999, 2003, 2007, 2011, and 2015 when Taiwan was a participating country. The TIMSS targeted students of two populations: those at the end of the fourth year of formal schooling and those at the end of the eighth year—the equivalents of the fourth and eighth grades in most countries. The TIMSS adopted a two-stage clustered sample design within each country, with schools sampled in the first stage and classrooms sampled in the second stage (Foy & Joncas, 2000). In most countries, only one classroom was randomly selected from a sampled school, and when that classroom was sampled, all the students in the classroom were surveyed.

A "plausible values" method was used to produce proficiency scores in mathematics and science. The achievement scores, therefore, are available as a set of five plausible values for every individual student in each subject. Country-specific statistics used in this study, such as mean, standard deviation, and scores at various percentiles of the performance distribution, were estimated five times, each time using a different plausible value. The final statistic used for analysis is the average of the five estimates. In the base year of TIMSS, 1995, the international proficiency scores (math or science) for each grade level were scaled to have a mean of 500 and a standard deviation of 100. The metric of the 1995 scale has been preserved, so that test scores in later survey years can be compared with those from 1995. This facilitates a grade-specific trend analysis for countries participating in more than one TIMSS survey.

PISA

The first PISA survey was conducted in 2000 in 32 countries. Since 2000, the

PISA has been conducted every three years, in 2003, 2006, 2009, 2012, and 2015. It is expected that the PISA will continue to be administered every three years. In this study, I use PISA data from the 2006, 2009, 2012, and 2015 surveys which include Taiwanese students. Every three years, the PISA survey measures student performance in reading, mathematics and scientific literacy, but each survey year focuses more narrowly on a particular academic domain. The focus was on reading literacy in 2000, mathematical literacy in 2003, and scientific literacy in 2006. In the 2009 survey, reading literacy was again the focus. The aim of the PISA assessment is not to evaluate the mastery of a specific school curriculum, but to assess the ability to use knowledge and skills to meet real-life challenges (Organization for Economic Co-operation and Development [OECD], 2010). PISA test scores in reading, mathematics or science were designed to have an average score of 500 points and a standard deviation of 100 across OECD countries in PISA 2000, with the data weighted so that each OECD country contributed equally (OECD, 2005).

For every PISA participating country or jurisdiction, the PISA student sample is age-based, a random and representative sample of students aged 15 years old, when students in most countries are approaching the end of compulsory schooling. For most countries, the sample size range from 4,000 to 7,000 students in several hundred sampled schools. Fifteen-year-old students do not attend the same grade level across the board, as the typical grade level for this age cohort varies by country.

TEPS

Huang and Hauser (2010) provides a thorough introduction to the TEPS. In terms of survey content and design, the TEPS resembles the National Education Longitudinal Study of 1988 (NELS) in the United States. However, unlike the NELS, which followed a single cohort (eighth-graders in 1988), the TEPS followed two nationally representative samples: seventh-graders and eleventh-graders who were first surveyed in 2001 (L. Y. Chang, 2002). Grade 7 is equivalent to the first year of lower secondary schooling, and the students are about 13 years old. Grade 11 is one grade below the highest grade level in Taiwanese secondary schooling, and most eleventh-graders are 17 years old.

In the present analysis, I use TEPS data from Grade 11 only because they contain two different nationally-representative samples of Grade 11, one in 2001 and the other in 2005, which allow for a comparison between the HSEE and the BCT periods in student performance. The TEPS data of Grade 7 are confined to a single cohort, therefore, do not allow for a comparison between two different cohorts.

The baseline survey sampled 19,051 students in Grade 11, from 1,040 sampled classes in 286 sampled schools. The sample of students in Grade 11 were first surveyed in 2001 and then they were followed up in 2003 when they were about to graduate from high school. After 2003, the sample was no longer studied because the TEPS aimed to study student life longitudinally up to high school graduation. Therefore, the third wave of data collection contained only one cohort, a nationally representative sample of 20,136 eleventh-graders in 2005. A follow-up study was conducted in 2007, when these students progressed to Grade 12 and were about to graduate from high school.

A 90-minute cognitive test was administered to all students, not only in the baseline survey, but also in every later wave of data collection. The cognitive test was designed to measure student learning growth across the secondary schooling years (Yang, Tam, & Huang, 2003). The test aims more to measure the ability to learn and less to measure the accumulation of factual knowledge. The test includes two components: (a) a subject-specific component, which assesses students' math, science, and verbal achievements; and (b) a curriculum-free component, which assesses students' analytical, practical, and creative abilities. The subject-specific component includes 20 test items in math, 12 items in science, and 16 items in vocabulary knowledge (Chinese and English), for a total number of 48 items for Grade 11. The curriculum-free component includes three nine-item subtests, measuring analytical, practical, and creative abilities. Within each ability domain, there are three verbal, three graphical, and three numerical test items. Eleventh-graders were given 94 minutes to complete 75 test items. Both raw and IRT (item response theory) scores were made available for research (Yang et al., 2003). Data from the TEPS report IRT test scores in three domains: mathematics, curriculum-free component, and the overall performance which includes all test items. These IRT test scores are used to display trends in student performance.

Method

This study includes a descriptive analysis which reports trends in student performance and the effects of family background on student performance, covering the years before and after policy reforms in admission to senior high schools. In addition, the structure of the TIMSS data allows for a different methodology to evaluate the effects of policy changes. The TIMSS 2003, 2007, 2011, and 2015 targeted both the fourth and the eighth grades in Taiwan. Therefore, the TIMSS data contain students of three birth cohorts which had participated in the TIMSS surveys twice; first when each of these cohorts were in Grade 4, and again when reached Grade 8 four years later. These three cohorts include (a) TIMSS sampled students born around 1993, who were fourth graders in 2003 and eighth graders in 2007; (b) TIMSS sampled students born around 1997, who were fourth graders in 2007 and eighth graders in 2011; and (c) TIMSS sampled students born around 2001, who were fourth graders in 2011 and eighth graders in 2015. In Taiwan, the first two cohorts were subject to the BCT regime, while the last cohort was subject to the CAP regime.

As not all cohorts were subject to the same policy regime, it is possible to adopt a difference-in-differences (henceforth, D-in-D) analysis to evaluate the impact of policy changes introduced in 2014 by the CAP regime on student performance. This is done by comparing the level of performance at Grade 4 with that at Grade 8 across cohorts, before and after admission reform in 2014.

Difference-in-differences Methodology

To estimate the treatment effect when treatment is randomly assigned to some units, a simple method would be to compare the treated units before and after treatment. This way, however, tends to give biased results because the treatment effect is confounded with the effects of other factors that take place around the time of treatment. The D-in-D methodology deals with this problem by using a control group to difference out these confounding factors and isolate the treatment effect (Meyer, 1995). Specifically, the D-in-D methodology models the treatment effect by estimating the difference between outcome measures at two time points for both the treatment and the control groups and then calculating the difference between these two groups.

The D-in-D methodology requires a treatment group and a control group, with repeated measures of the same units from two time points: before and after treatment. A panel survey with data collected on the same individuals at different time points meets this requirement, but repeated cross-sections, such as two national random survey samples separately collected, can also be used in a D-in-D analysis (Meyer, 1995). For example, Hanushek and Woessmann (2006), taking advantage of the facts that no country tracked students before the fourth grade, and that some countries did track their students at a higher grade level, adopted a D-in-D methodology by comparing differences in achievement between younger (the fourth graders) and older students (eighth, ninth, or tenth graders) among tracking and non-tracking countries. Hanushek and Woessmann estimated the effect of tracking by comparing the average achievement gain from Grade 4 to Grade 8 (or higher) in tracked countries to that in untracked countries. The younger and older student samples used by Hanushek and Woessmann were not from a panel survey, but collected separately from two cross-sectional national surveys.

The present study adopts a methodological strategy similar with that used by Hanushek and Woessmann (2006). I assume that the fourth graders are too young to be affected by policy changes in admissions to senior high schools. For eighth graders, however, who are much closer to the transition from junior high to senior high schools, the impact of policy changes in admissions to senior high schools on student learning and equality of educational opportunity should have already taken place. Therefore, Grade 4 represents the time point before treatment while Grade 8 represents the time point after treatment. The two TIMSS cohorts under the regime of BCT are used as control groups, while the TIMSS cohort experiences the CAP regime characterized by the "examination-free admission", in which test performance is no longer the sole criterion for admission to senior high schools, is used as a treatment group.

Outcomes of interest are observed for control and treatment groups for two time periods which in this case are when students were in Grade 4, and four years later, when they were in Grade 8. The treatment group is exposed to a treatment in Grade 8 but not in Grade 4. The control group is not exposed to the treatment either in Grade 4 or Grade 8. The D-in-D estimate in the present study can be obtained by subtracting the difference between Grade 4 and Grade 8 in the control group from the difference between Grade 4 and Grade 8 in the treatment group. This removes biases in Grade 8 comparisons between the treatment and control group that could be the result from permanent differences between those groups, as well as biases from comparisons between Grade 4 and Grade 8 in the treatment group that could be the result of growth in outcomes as students move up to higher grade levels in school.

To evaluate the impact of policy changes between the BCT and CAP regime on equality of educational opportunity, I also conduct a D-in-D analysis using data from the TIMSS. A regression of test scores on family SES is conducted to evaluate the equality of educational opportunity. Because test scores are not directly comparable between Grades 4 and 8, in the regression analysis, test scores of either grade level are standardized to have a mean of 500 and a standard deviation of 100.

Measures of Family Socioeconomic Status

In TIMSS, family socioeconomic status is measured by the number of books at home, and there are five categories. The number of books at home was reported by the students according to the following question: "About how many books are there in your home? (Do not count magazines, newspapers, or your school books.)".

The five possible levels for answers were (1) none or very few (0–10 books), (2) enough to fill one shelf (11–25 books), (3) enough to fill one bookcase (26–100 books), (4) enough to fill two bookcases (101–200 books), and (5) enough to fill three or more bookcases (more than 200 books). The number of books at home used in the model is a continuous variable ranging from 1 to 5 along the lines of the five answer categories listed above. The reason is that the estimated coefficients of the four dummies of the number of books at home showed a highly linear pattern. Schuetz et al. (2008, 285–88) provide a detailed discussion on the use of the number of books at home in measuring student family background.

When analyzing data from the PISA, the index of economic, social and cultural status (ESCS) is used to measure family socioeconomic status. The ESCS index is a measure of family SES on the basis of the following variables: the International Socio-Economic Index of Occupational Status (ISEI); the highest level of parental education converted into years of schooling; the PISA index of family wealth; the PISA index of home educational resources; and the PISA index of possessions related to “classical” culture in the family home. When using data from the TEPS, family SES is measured by a dichotomous variable, with either parent having a college degree or higher coded as 1, otherwise, coded as 0.

Results

Admission Policy Changes and Student Performance

A descriptive trend analysis.

Table 1 presents trends in mathematics and science performance among Taiwanese eighth-graders, using data from the TIMSS. In mathematics, the average test score was 585 in 1999, the year when the HSEE was used for admission to senior high schools. In 2003, when the BCT administration was in its early stage, the average test score in mathematics was still 585, suggesting a stable trend during the transition from the HSEE to the BCT regime. Since 2003, there has been an upward trend in mathematics performance within the BCT regime, with an average score of 585 in 2003, 598 in 2007, and 609 in 2011. This demonstrates a significant growth of 24 points over a span of eight years. This growth in not only statistically significant, it is nearly one-fourth standard deviation in Taiwanese TIMSS mathematics test scores in 2003. Therefore, Taiwanese eighth-graders in 2011 performed one-fourth standard deviation higher in mathematics than their counterparts in 2003. However, a statistically significant decline of 10 points followed after 2011, due to a lower performance level among eighth-graders in 2015 when the CAP was in place, a regime when performance on high-stakes test is no longer the sole criterion for high school admission. In science, changes in average test scores over time are not as dramatic as those in mathematics, and they do not respond to policy changes over the HSEE, BCT, and CAP era. For example, changes in average science test scores between the two adjacent policy regimes, either between the HSEE (1999) and the BCT (2003) or between the BCT (2011) and the CAP (2015), are small and not reaching statistical significance.

Based on data from the TEPS, Table 2 reports test score trends between 2001 and 2005, covering the HSEE and the BCT era, among eleventh-graders in Taiwan. Findings suggest that students under the BCT regime performed better than students in the HSEE regime in which test performance is the sole criterion for high school admission. The difference is modest in mathematics performance or in the overall performance covered by the achievement test of the TEPS. However, the difference is more substantial in the curriculum-free component of the test. Taiwanese eleventh-graders in the BCT regime performed nearly one-fourth standard deviation higher in the curriculum-free component of the test than their counterparts in the HSEE regime.

Table 3 is based on data from the PISA, showing test performance of Taiwanese ninth-graders in mathematics, science, and reading in 2006, 2009, 2012, and 2015. The survey years of 2006, 2009, and 2012 took place during the BCT era, while those in the year of 2015 belong to the CAP regime. With respect to mathematics, the average test scores peak in 2012, the final survey year of the BCT era. Ninth-graders in 2015, whose performance on high-stakes test is no longer the sole criterion for high school admission, scored 7 points lower than those in 2012. The decline in mathematics performance is not large but statistically significant. Reading performance follows a similar pattern, as indicated in Table 3. The average test scores in reading in the BCT era peak in 2012, then a decline followed. The decline in reading performance is larger than that in mathematics performance. Ninth-graders in 2015 scored 18 points lower in reading than those in 2012, and the decline is statistically significant. Trends in science performance, however, follow a very different pattern. There has been a decline within the BCT era in science performance, and the average score reached its lowest point in 2009 and 2012, with a score of 515. However, a statistically significant increase of 16 points was introduced in 2015 when the CAP was in place.

Findings from the TIMSS are consistent with those from the PISA in showing a decline in mathematics performance, but not in science performance, after the CAP was implemented. Data from PISA show a notable increase in science performance after the CAP was in place. Over-time changes in reading performance follow a similar pattern with those in mathematics performance, also suggesting a drop in test performance after the CAP was introduced.

A difference-in-differences analysis.

In addition to a descriptive trend analysis presented above, I also evaluate the impacts of the new admission policies implemented in and after 2014 on student performance, using the D-in-D methodology. In TIMSS, four points on the achievement scales are identified as international benchmarks: Advanced (625), High (550), Intermediate (475), and Low (400). These benchmarks are used for mathematics and science performance, either in Grade 4 or Grade 8. I follow this definition and assess the impacts of the new policies on the percentages of Taiwanese students scoring 625 or higher, 550 or higher, 475 or lower, and 400 or lower.

Two cohorts under the BCT regime are used as control groups, while the youngest cohort which was admitted after admission reform in 2014 is used as a treatment group. Both control groups are under the same policy regime, so a D-in-D analysis between these two control groups is expected to show no significant differences in academic performance as students progressed from Grade 4 to Grade 8.

Table 4 presents results from the D-in-D analysis in three panels. The first and the second panels compare either control group with the treatment group, while the third panel presents findings for a comparison between the two control groups. The first panel of Table 4 compares the first control group with the treatment group. The last column shows the D-in-D estimates, indicating the estimated impacts of the new policies on the percentages of students scoring 625 or higher, 550 or higher, 475 or lower, and 400 or lower. Statistical significant D-in-D estimates at $p < 0.01$ level are underlined and bolded. The D-in-D estimates presented in the upper two rows of the last column suggest a remarkable decline in the percentage of high-performing students in mathematics after the reform in 2014. Under the CAP regime (the treatment group), as panel 1 of Table 4 indicates, only 44% of the Taiwanese eighth-graders scored 625 or higher in mathematics. However, it is expected that 63% of these Taiwanese students would score 625 or higher in mathematics, had students not experiencing the admission reform in 2014. The reform in 2014 is associated with a drop in the percentage of students scoring 625 or higher from an expected estimate of 63% to an actual estimate of 44%, and a decrease in the percentage of students scoring 550 or higher from an expected estimate of 84% to an actual estimate of 72%.

Compared to those in the first control group, students in the second control group are younger and they are closer to students in the treatment group in the year of birth. When the treatment group is compared with the second control group, as indicated in panel 2 of Table 4, there is still a remarkable decline in the percentage of high-performing students in mathematics, albeit to a lesser degree. The reform in 2014 is associated with a drop in the percentage of students scoring 625 or higher from 59% to 44%, and a decrease in the percentage of students scoring 550 or higher from 81% to 72%.

In 2015, 44% of Taiwanese eighth-graders reached advanced international benchmark in mathematics, and this percentage is not much lower than that in 2011 (49%) and that in 2007 (45%). However, findings from the D-in-D analysis suggest a much larger reduction in the percentage of high-performing eighth-graders in mathematics. This has to do with the fact that the share of fourth-graders reaching advanced international benchmark in mathematics increased dramatically from 16% in 2003, 24% in 2007, to 34% in 2011. The first and the second control groups in the D-in-D analysis started with a relatively low percentage of students reaching the advanced international benchmark in mathematics in Grade 4 (16% and 24% respectively), and this percentage increased remarkably as students progressed to grade 8 (45% and 49% respectively). Relative to these two control groups, the treatment group started with a higher percentage (34%) of students reaching the advanced international benchmark in mathematics in Grade 4. To have the same magnitude of growth between Grade 4 and Grade 8 as the control groups, it requires the treatment group to have a percentage of eighth-graders reaching the advanced international benchmark in mathematics up to 59% or 63%. However, in the treatment group, only 44% of eighth-graders reached that benchmark. The gap between 44% and 59%, or between 44% and 63%, is considered as the size of treatment effect which is negative.

In science performance, changes associated with the 2014 reform are much less dramatic, and findings are not consistent when the treatment group is compared to either control group. It is not certain that the 2014 reform is associated with a reduction in the proportion of low-performing students in either mathematics or science because panel 1 and panel 2 of Table 4 do not show consistent results. As to whether or not the 2014 reform is associated with an increase in the proportion of high-performing students in science, panel 1 and panel 2 do not show consistent findings either.

One concern is whether or not differences between the two control groups are insignificant in their academic progress in either mathematics or science. Findings from the third panel of Table 4 support the hypothesis that students in these two control groups, which are under the same policy regime, do not differ significantly in academic growth as they progress through the grades.

Admission Policy Changes and Equality of Educational Opportunity

A descriptive trend analysis.

In TIMSS, family socioeconomic status is measured by the number of books at home at five levels. Table 5 presents the effect of family socioeconomic status on student performance in mathematics and science across TIMSS survey years in 1999, 2003, 2007, 2011, and 2015. In mathematics, as indicated in Table 5, an increase in family SES from one level to a higher level is associated with an increase of about 30 points in test score, a magnitude of about 0.3 standard deviation. There have been little changes over time from 1999 to 2015, across all three policy regimes, in the effect of family socioeconomic status on student performance in mathematics, whether or not excluding students with an immigrant background. In science, over-time changes in the effect of family SES on test performance are not statistically significant between either two adjacent regimes, whether or not excluding students with an immigrant background. Table 6 shows findings from the TEPS which covers the transition from the HSEE to the BCT regime, and they suggest that the effect of family SES on test performance has declined during the transition from the HSEE to the BCT regime. While the decline is modest, it reaches statistical significance due to a very large sample size in the TEPS.

Table 7 presents findings from the PISA which covers the transition from the BCT to the CAP regime. An increase of a standard deviation in PISA's index of family SES is associated with an increase of 40 to 49 points in mathematics, 32 to 42 points in science, and 33 to 37 points in reading test scores. The effect of family SES on test performance is insensitive to the exclusion of students with an immigrant background. In mathematics, the effect of family SES remained stable during the BCT regime, but there was a statistically significant decrease after the CAP took place, from a coefficient of 48.99 in the BCT regime to 40.53 in the CAP administration. In science, on the other hand, the effect of family SES decreased during the BCT regime, but there was a statistically significant increase after the CAP took place. In reading performance, changes in the effect of family SES over time have been insignificant.

A difference-in-differences analysis.

Table 8 presents findings from the D-in-D analysis in which two cohorts under the BCT regime are used as control groups, while the youngest cohort which was admitted to senior high schools after admission reform in 2014 is used as a treatment group. Table 8 presents results in three panels. The first and the second panels compare either control group with the treatment group, while the third panel presents findings for a comparison between the two control groups. Findings from panels 1 and 2 suggest that the 2014 reform in admission to senior high schools did not introduce an increase or a significant change in the effect of family socioeconomic status on student performance in either mathematics or science. Findings from the third panel suggest that differences between the two control groups, which are under the same policy regime, are insignificant as well in the effect of family SES on student performance in either mathematics or science.

Discussion and Conclusions

Taiwan has experienced a series of educational reforms in the process of being admitted to senior high schools in the last two decades. These reforms share the same goal: all aim to reduce the importance of large-scale testing and move toward a more holistic assessment of student performance for school admission. School admission reforms in the transition from junior to senior high school are divided into three historical periods: the HSEE period (1958-2000) when test performance was the only criteria for admission to senior high schools, the BCT period (2001-2013) when test performance was the only criteria for admission for only about half of the students, and the CAP period after 2014 when test performance is no longer the only criteria for admission. The purpose of the present study is to examine whether or not academic performance has declined over time, and whether or not the effect of family socioeconomic status on student academic performance has increased over time, as test performance becomes less crucial for admission to senior high schools across the HSEE, BCT, and CAP periods.

With respect to student performance, a descriptive trend analysis suggests that the transition from the HSEE to the BCT regime did not weaken the academic performance of Taiwanese students. During the BCT regime, mathematics and reading performance improved over time and peaked at the end of the BCT regime, but then followed by a decline after the CAP took place. Findings from the D-in-D analysis also point out a decline in mathematics performance after the CAP was implemented in 2014, a decline attributable to a reduction in the share of high-performing students. Unlike mathematics and reading performance, the implementation of the CAP after 2014 did not lead to a decline in science performance. In fact, it may have introduced an improvement in science performance.

With regards to the impact of family SES on student academic performance, the descriptive trend analysis presents inconsistent findings, while the D-in-D analysis suggests a stable trend in the effect of family socioeconomic status on either mathematics or science performance across the BCT and the CAP periods. The finding of a stable condition could ease the concern that the admission reform of 2014 would increase the degree of inequality of educational opportunities. As mentioned in the introduction, Taiwan has been one of a handful countries with the least equality of educational opportunity among more than 50 countries. Therefore, there is still a concern whether or not high-stakes testing is indeed a fair mechanism for admissions to either the next higher level of schooling or elite schools.

The most consistent and significant finding of the present study is a decline in mathematics performance after the implementation of the CAP in 2014. One explanation is that the decline in mathematics performance is due to the fact that test performance is no longer the sole criterion for admission to senior high schools in the CAP regime, and students under the new regime no longer need to work so hard to prepare for a test. Such a speculation fails to explain why science performance did not experience a similar decline.

An alternative explanation is that the decline in mathematics performance may have to do with the switch from a standardized test (the BCT) to a standard-based test (the CAP) for admission to senior high schools in 2014. In the CAP regime, test performance in each subject was given in seven levels, with a score range from 1 to 7.

This is very different from the BCT regime in which test scores for each subject range from 1 to 80. For a better test performance to be seen in a narrow test score range from 1 to 7, students need to invest significant effort to overcome the hurdle from one level to the next higher level. In addition, not all upgrades from one level to the next higher level require the same amount of effort. Perhaps reaching a higher level from B to B+ is easier than the upgrade from A to A+, and some subjects maybe easier for students to upgrade to a higher level than others. In the CAP regime, admissions to senior high schools consider the total test scores across all five subjects, and each subject has a perfect score of 7 (A++). Only when two or more students have the same total test score do school admissions give priorities to those who score higher on an individual subject. Therefore, students can be very strategic in investing their time on test preparation. When mathematics is considered a more difficult subject to improve on, students may choose to invest their time on different subjects, in order to have a higher total test score across all five subjects. After all, the sum of test scores across all five subjects has a higher priority for admission than test scores on any individual subject. This is a possible reason for why the same policy change does not have the same impact on test performance of different subjects. When students consider mathematics a more difficult subject to move up to the next higher level, they may direct their time and efforts to a different subject, such as science. When this is the case, mathematics performance is expected to decline, while science performance would improve.

Taiwanese fourth-graders targeted by the TIMSS 2105 will be sampled by the TIMSS 2019 when they reach Grade 8. This will provide an additional cohort under the regime of the CAP, and will make possible a comparison between the two treatment groups. If these two treatment groups do not differ significantly in the progress that students make over time, it may strengthen the findings reported in this study regarding the impacts of the CAP regime on student performance and equality of educational opportunity. Since the TIMSS covers several dozen countries, the D-in-D methodology applied in this study for the case of Taiwan can potentially be used to evaluate the impacts of educational policies on student outcomes in other countries.

Reference

Bishop, J. H. (1996). The impact of curriculum-based external examinations on school priorities and student learning. *International Journal of Education Research*, *23*(8), 653-752.

Bishop, J. H. (1997). The effect of national standards and curriculum-based exams on achievement. *American Economic Review*, *87*(2), 260-264.

Bishop, J. H. (1999). Are national exit examinations important for educational efficiency? *Swedish Economic Policy Review*, *6*(2), 349-401.

Chang, C. P. (2002). *A primer on the junior high school basic proficiency test*. *Taipei: Linkin Publisher* (in Chinese).

Chang, L. Y. (2002). A brief introduction to the Taiwan Education Panel Survey (TEPS) database (in Chinese). *Empirical Inquiry*, *12*, 111-118.

Chou, J. Y., Hu, J. H., & Liu, Y. M. (2013). A comprehensive review of 12-year national education through the lens of Taiwan's middle-school voluntary matriculation support program. *Education Data Research*, *109*, 79-106 (in Chinese).

Ding, Y. W. (1998). A key change in the history of school exam admission: Senior high school diverse matriculation program. *Senior High School Education*, *2*, 22-27 (in Chinese).

Executive Yuan's Education Reform Audit Committee. (1996). *A comprehensive report on education reform*. Taipei: National Education Institute (in Chinese). Retrieved from http://www.naer.edu.tw/files/15-1000-7992,c1315-1.php?Lang=zh-tw

Foy, P., & Joncas, M. (2000). TIMSS sample design. In M. O. Martin, K. D. Gregory, & S. R. Stemler (Eds.), *TIMSS 1999 technical report* (pp. 29-46). Retrieved from https://timssandpirls.bc.edu/timss1999i/pdf/T99_TR.book.pdf

Hanushek, E. A., & Woessmann L. (2006). Does educational tracking affect performance and inequality? Differences- in-differences evidence across countries. *The Economic Journal*, *116*, C63–C76.

Hsu, T. Y. (2002). Recent evolution of senior high school matriculation regime. *National Junior High School Student Basic Proficiency Test Journal*, *13* (in Chinese).

Huang, M. H. (2008). Classroom homogeneity, family background, and math performance: A cross-national difference-in-differences analysis. *Taiwanese Journal of Sociology*, *40*(2), 1-44 (in Chinese).

Huang, M. H. (2017). Excellence without equity in student mathematics performance: The case of Taiwan from an international perspective. *Comparative Education Review*, *61*(2), 391-412.

Huang, M. H., & Hauser, T. S. (2010). Tracking persons from high school through adult life: Lessons from the Wisconsin Longitudinal Study. *EurAmerica*, *40*(2), 311-358.

Jen, T. H., Lee, C. D., Chen, K. M., Lo, P. H., & Lin, C. Y. (2016). Chinese Taipei. In I. V. S. Mullis, M. O. Martin, S. Goh, & K. Cotter (Eds.), *TIMSS 2015 encyclopedia: Education policy and curriculum in mathematics and science*. Retrieve from http://timssandpirls.bc.edu/timss2015/encyclopedia/download-center/

K-12 Education Administration. (2016, January 15). *Key exam-free admission application guide* (in Chinese). Retrieved from http://www.edu.tw/News_Content.aspx?n=9E7AC85F1954DDA8&s=52C8960CEB15D650

K-12 Education Administration. (2017, January 16). *Senior high school special admission student assignment* (in Chinese). Retrieved from http://www.edu.tw/News_Content.aspx?n=9E7AC85F1954DDA8&sms=169B8E91BB75571F&s=E20FE7871D499429

Keelung-Taipei Senior High School Examination-Free Admission Committee. (2018). *Keelung-Taipei senior high school exam-free admission application guide* (in Chinese). Retrieved from https://12basic.tp.edu.tw/attachment/more/brochures

Lin, J. C. (2000). Voluntary matriculation support program. In *National Education Institute Education Glossary* (in Chinese). Retrieved from http://terms.naer.edu.tw/detail/1305246/

Lin, S. C. (2004). A clairvoyant review of Taiwan's recent education reforms. *Education Academic Journal*, *23*, 1-36 (in Chinese).

Meyer, B. D. (1995). Natural and quasi-experiments in economics. *Journal of Business and Economic Statistics*, *13*(2), 151-161.

Ministry of Education. (1998). *Senior high school and vocational school diverse matriculation program* (in Chinese). Retrieved from http://www3.cnsh.mlc.edu.tw/~aao/ruels/edurules/hschpolyinto92.htm

Ministry of Education. (2005). *Junior high school student writing test program* (in Chinese). Taipei: Ministry of Education.

Ministry of Education. (2008). *Press release on the new master-scale for the junior high school student basic proficiency test* (in Chinese). Retrieved from https://www.edu.tw/default.aspx

Ministry of Education. (2015, June 23). *K-12 basic education implementation program* (in Chinese). Retrieved from http://12basic.edu.tw/Detail.php?LevelNo=8

Monahan, R., Chapman, B., & Fermino, J. (2013, October 10). Bill de Blasio: Tests shouldn't be only way kids get into NYC elite schools. *The New York Times*. Retrieved from http://www.nytimes.com

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016a). *TIMSS 2015 international results in mathematics*. Retrieved from http://timssandpirls.bc.edu/timss2015/international-results/download-center/

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016b). *TIMSS 2015 international results in science*. Retrieved from http://timssandpirls.bc.edu/timss2015/international-results/download-center/

National elementary school and junior high school student performance evaluation *principles*. (2015, January 7) (in Chinese). Retrieved from http://law.moj.gov.tw/LawClass/LawAll.aspx?PCode=H0070019

OECD. (2005). *PISA 2003 data analysis manual: SPSS*. Paris: OECD Publishing. Retrieved from https://doi.org/10.1787/9789264010666-en

OECD. (2010). *PISA 2009 results: What students know and can do–student performance in reading, mathematics and science (volume I)*. Paris: OECD Publishing. Retrieved from http://dx.doi.org/10.1787/9789264091450-en

Schuetz, G., Ursprung, H. W., & Woessmann, L. (2008). Education policy and equality of opportunity. *Kyklos*, *61* (2): 279–308.

Song, Y. T., Hsu, F. Y., Tzeng, F. L., Jiang, L. P., & Sun, W. M. (2007). National middle-school basic proficiency test review and development (in Chinese). *Education Reform and Development Journal*, *3*(4), 29-50.

Taylor, K. (2018, March 7). Elite schools make few offers to Black and Latino students. The New York Times. Retrieved from http://www.nytimes.com

Wang, P. L. (1989). Anticipatory examination stress on health of adolescence. *Journal of Taipei Municipal Teachers College*, *20*, 337-374 (in Chinese).

Wang, H. Y., & Yu, M. N. (2006). An alternative method for representing scaled scores of the basic competency test for junior high school students (in Chinese). *Psychological Testing*, *53*(2), 205-237.

Woessmann, L. (2003). Central exit exams and student achievement: International Evidence. In P. E. Peterson, & M. R. West (Eds.), *No child left behind? The politics and practice of school accountability*. Washington, DC: Brookings Institution Press.

Woessmann, L. (2005). The effect heterogeneity of central examinations: Evidence from TIMSS, TIMSS-repeat and PISA. *Education Economics*, *13*(2), 143-169.

Wu, C. S. (2016). Glossary of education: Special examination admission. *Pulse of Education*, *6* (in Chinese). Retrieved from https://pulse.naer.edu.tw/Home/Content/9fc7708d-d290-4cf3-8bf4-5b03aa6c3d8b?insId=40977899-d342-4f01-94a7-66d446c9d3bb

Yang, M. L., Tam, T., & Huang, M. H. (2003). *Taiwan Education Panel Survey (TEPS) psychometric report*. Taipei: Academia Sinica Empirical Research Center (in Chinese).

Table 1. Over-Time Changes in Mathematics and Science Performance among Taiwanese Students in Grade 8, TIMSS

	Year	N	Mean	Median	S.D.	% > 625	% < 400
Math	1999	5772	585	595	104	37%	5%
	2003	5379	585	596	100	38%	4%
	2007	4046	598	614	106	45%	5%
	2011	5042	609	623	106	49%	4%
	2015	5711	599	612	97	44%	3%
Science	1999	5772	569	574	89	27%	4%
	2003	5379	571	577	79	26%	2%
	2007	4046	561	571	89	25%	5%
	2011	5042	564	571	84	24%	4%
	2015	5711	569	579	83	27%	4%

Note. In mathematics, all between-year differences in average test scores are statistically significant, except for that between 1999 and 2003, between 2007 and 2011, and between 2007 and 2015. In science, all between-year differences in average test scores are statistically insignificant, except for that between 2003 and 2007 and that between 2007 and 2015.

Source: Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016a). *TIMSS 2015 international results in mathematics*; Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016b). *TIMSS 2015 international results in science*. Retrieved from http://timssandpirls.bc.edu/timss2015/international-results/download-center/

Table 2. Over-Time Changes in Test Performance among Taiwanese Students in Grade 11, TEPS 2001-2005

	Year	N	Mean	Median	S.D.
Math	2001	18956	1.37	1.41	1.35
	2005	19935	1.54	1.62	1.35
Curriculum-free	2001	18956	1.47	1.58	1.45
	2005	19935	1.82	2.02	1.42
Overall	2001	18956	1.34	1.43	1.30
	2005	19935	1.51	1.62	1.31

Note. In all three subjects, the difference in average test scores between 2001 and 2005 is statistically significant at the $p < 0.01$ level.

Table 3. Over-Time Changes in Mathematics, Science and Reading Performance among Taiwanese Students in Grade 9, PISA 2006-2015

	Year	N	Mean	Median	S.D.	% > Level 6	% < Level 1
Math	2006	2709	539	546	97	7.5%	3.7%
	2009	1870	535	541	103	8.8%	5.1%
	2012	1989	543	550	109	12.2%	5.3%
	2015	3386	536	541	98	7.9%	4.4%
Science	2006	2709	526	531	89	1.0%	1.7%
	2009	1870	515	522	87	0.5%	3.0%
	2012	1989	515	521	79	0.3%	1.8%
	2015	3386	531	538	96	1.9%	2.7%
Reading	2006	2709	487	494	81	0.1%	3.9%
	2009	1870	485	494	88	0.2%	5.8%
	2012	1989	513	521	89	0.9%	3.5%
	2015	3386	495	502	90	0.4%	5.1%

Note. In mathematics, all between-year differences in average test scores are statistically insignificant, except for that between 2009 and 2012 and that between 2012 and 2015. In science, all between-year differences in average test scores are statistically significant, except for that between 2009 and 2012 and that between 2006 and 2015. In reading, all between-year differences in average test scores are statistically significant, except for that between 2006 and 2009.

Table 4. Difference-in-differences (D-in-D) Analysis of the 2014 Reform's Impacts on Student Mathematics and Science Performance, Taiwan, Grades 4 and 8, TIMSS 2007, 2011, and 2015

	2003	2007	2011	2015	Control 1	Treatment	D-in-D
	Grade 4	Grade 8	Grade 4	Grade 8	Grd. Diff.	Grd. Diff.	T – C1
Math							
%>625	16	45	34	44	29	10	**-19**
%>550	61	71	74	72	10	-2	**-12**
%<475	8	14	7	12	6	5	-1
%<400	1	5	1	3	4	2	**-2**
Science							
%>625	14	25	15	27	11	12	1
%>550	52	60	53	63	8	10	2
%<475	13	17	15	14	4	-1	**-5**
%<400	2	5	3	4	3	1	-2
	2007	2011	2011	2015	Control 2	Treatment	D-in-D
	Grade 4	Grade 8	Grade 4	Grade 8	Grd. Diff.	Grd. Diff.	T – C2
Math							
%>625	24	49	34	44	25	10	**-15**
%>550	66	73	74	72	7	-2	**-9**
%<475	8	12	7	12	4	5	1
%<400	1	4	1	3	3	2	-1
Science							
%>625	19	24	15	27	5	12	**7**
%>550	55	60	53	63	5	10	5
%<475	14	15	15	14	1	-1	-2
%<400	3	4	3	4	1	1	0
	2003	2007	2007	2011	Control 1	Control 2	D-in-D
	Grade 4	Grade 8	Grade 4	Grade 8	Grd. Diff.	Grd. Diff.	C2-C1
Math							
%>625	16	45	24	49	29	25	-4
%>550	61	71	66	73	10	7	-3
%<475	8	14	8	12	6	4	-2
%<400	1	5	1	4	4	3	-1
Science							
%>625	14	25	19	24	11	5	-6
%>550	52	60	55	60	8	5	-3
%<475	13	17	14	15	4	1	-3
%<400	2	5	3	4	3	1	-2

Note. Statistical significant D-in-D estimates at the $p < 0.01$ level are underlined and bolded.

Table 5. Effect of Family Socioeconomic Status on Student Performance in Mathematics and Science, TIMSS, 1999, 2003, 2007, 2011, and 2015

		Grade 8 Math		Grade 8 Science	
	N	Coef.	t-value	Coef.	t-value
Include Immigrants					
1999	5750	31.27	21.81	26.42	20.74
2003	5370	31.07	20.34	25.26	21.44
2007	4033	31.64	19.54	29.15	21.65
2011	5039	30.65	21.95	26.78	27.82
2015	5706	31.32	25.69	29.00	29.49
Exclude Immigrants					
1999	5453	30.76	21.17	26.09	20.16
2003	5099	29.68	19.80	24.22	21.10
2007	3670	29.18	18.09	27.03	20.36
2011	4625	29.11	21.39	25.47	27.04
2015	4833	30.20	23.79	27.83	27.87

Note. All between-year differences in effect of family socioeconomic status on either mathematics or science performance are statistically insignificant, whether or not excluding students with an immigrant background.

Table 6. Effect of Family Socioeconomic Status on Test Performance among Taiwanese Students in Grade 11, TEPS 2001 and 2005

		Math		Curriculum-free		Overall	
Year	N	Coef.	t-value	Coef.	t-value	Coef.	t-value
2001	18029	0.82	22.91	0.85	23.17	0.90	25.97
2005	18996	0.77	23.47	0.75	22.07	0.84	26.71

Note. In all three subjects, the difference in the effect of family socioeconomic status on test performance between 2001 and 2005 is statistically significant at the $p < 0.01$ level.

Table 7. Effect of Family Socioeconomic Status on Student Performance in Mathematics, Science, and Reading, Grade 9, PISA, 2006, 2009, 2012, and 2015

	Math, include immigrants			Math, exclude immigrants		
Year	N	Coef.	t-value	N	Coef.	t-value
2006	2696	44.90	14.34	2635	43.25	14.23
2009	1856	43.03	16.27	1817	42.35	15.41
2012	1978	48.99	15.41	1931	48.24	14.53
2015	3376	40.53	14.76	3345	40.15	14.55
	Science, include immigrants			Science, exclude immigrants		
Year	N	Coef.	t-value	N	Coef.	t-value
2006	2696	41.35	15.24	2635	40.44	15.41
2009	1856	35.37	14.62	1817	34.70	14.28
2012	1978	32.85	13.97	1931	32.12	13.26
2015	3376	41.62	17.31	3345	41.38	17.03
	Reading, include immigrants			Reading, exclude immigrants		
Year	N	Coef.	t-value	N	Coef.	t-value
2006	2696	37.38	12.55	2635	36.26	12.43
2009	1856	34.14	14.53	1817	33.59	14.41
2012	1978	36.42	11.79	1931	35.77	11.42
2015	3376	35.51	14.13	3345	35.29	13.98

Note. In mathematics, all between-year differences in the effect of family socioeconomic status on test performance are statistically insignificant, except for that between 2012 and 2015. In science, all between-year differences in the effect of family socioeconomic status on test performance are statistically insignificant, except for that between 2006 and 2012 and that between 2012 and 2015. In reading, all between-year differences in the effect of family socioeconomic status on test performance are statistically insignificant. Whether or not excluding students with an immigrant background does not make a significant difference in results.

Table 8. Difference-in-differences Estimates of the 2014 Reform's Impacts on Equality of Educational Opportunity, Taiwan, Grades 4 and 8, TIMSS 2003, 2007, 2011, and 2015

	2003	2007	2011	2015	Control 1	Treatment	D-in-D
	Grade 4	Grade 8	Grade 4	Grade 8	Grd. Diff.	Grd. Diff.	T – C1
Math							
Coef.	25.7	30.8	28.7	33.2	5.2	4.5	-0.6
t-value	17.6	19.9	20.4	26.3	2.4	2.4	-0.2
Science							
Coef.	24.7	33.7	29.3	36.5	9.0	7.2	-1.8
t-value	19.0	21.7	21.8	30.5	4.4	3.9	-0.6
	2007	2011	2011	2015	Control 2	Treatment	D-in-D
	Grade 4	Grade 8	Grade 4	Grade 8	Grd. Diff.	Grd. Diff.	T – C2
Math							
Coef.	29.6	29.9	28.7	33.2	0.3	4.5	4.2
t-value	20.7	22.2	20.4	26.3	0.2	2.4	1.6
Science							
Coef.	29.1	33.4	29.3	36.5	4.3	7.2	2.9
t-value	22.3	28.5	21.8	30.5	2.4	3.9	1.2
	2003	2007	2007	2011	Control 1	Control 2	D-in-D
	Grade 4	Grade 8	Grade 4	Grade 8	Grd. Diff.	Grd. Diff.	C2-C1
Math							
Coef.	25.7	30.8	29.6	29.9	5.2	0.3	-4.9
t-value	17.6	19.9	20.7	22.2	2.4	0.2	-1.7
Science							
Coef.	24.7	33.7	29.1	33.4	9.0	4.3	-4.7
t-value	19.0	21.7	22.3	28.5	4.4	2.4	-1.8

Note. Statistical significant D-in-D estimates at the $p < 0.01$ level are underlined and bolded. None of the estimates reported in this table are statistically significant.