2013

HARVARD-YENCHING INSTITUTE WORKING PAPER SERIES

INTRA-COHORT GROWTH IN THE INEQUALITY OF MATHEMATICS PERFORMANCE:

TAIWAN, THE U.S. AND THE STATE OF MASSACHUSETTS FROM AN INTERNATIONAL PERSPECTIVE

Huang Min-Hsiung | Academia Sinica

Intra-cohort Growth in the Inequality of Mathematics Performance:
Taiwan, the U.S., and the State of Massachusetts from an International Perspective

Min-Hsiung Huang
Institute of European and American Studies
Academia Sinica, Taipei, Taiwan
Visiting Scholar (2012-13), Harvard-Yenching Institute,
Email: mhhuang@sinica.edu.tw

Abstract

As students in Taiwan progress from elementary to junior high school, a remarkable increase in the inequality of student achievement in mathematics has been observed, at a level significantly higher than that seen in other countries. This study investigates the widening-gap phenomenon in Taiwan with respect to (a) its presence in different studies; (b) its magnitude and pattern; (c) its coverage of students in different birth cohorts; and (d) its relevance to initial achievement levels, family background, gender, the rural-urban gap, as well as between- and within-classroom differences. Findings show that the widening-gap phenomenon occurring before Grade 9 accompanied a significant increase in average performance. High-achievers in mathematics generally learned faster than low-achievers over time. The widening-gap phenomenon in Taiwan was commonly seen in students of different birth cohorts, gender, and residential areas, and it occurred mainly within classrooms. Taiwanese students with more favorable family backgrounds improved more significantly in mathematics over time, and they displayed a less significant widening-gap phenomenon amongst themselves.

Key words: widening-gap phenomenon, Matthew effect, mathematics performance, performance inequality, TIMSS, TEPS, Taiwan

Intra-cohort Growth in the Inequality of Mathematics Performance:

Taiwan, the U.S., and the State of Massachusetts from an International Perspective

1. Introduction

From the time children enter schools, they show differences in learning outcomes. The initial differences may increase over time because higher-performing students may learn faster than their poorer-performing class mates. As children move successively through the grades, average academic performance improves; at the same time, however, dispersion among individual students may widen as well.

The widening gap in achievement between the advantaged and the disadvantaged is referred to as the “fan-spread effect” or the “Matthew effect.” The later term originates from the Biblical statement that: “For everyone who has will be given more, and he will have an abundance. Whoever does not have, even what he has will be taken from him” (Matthew 25:29, the New International Version). Research demonstrating such an effect extends to that concerned with scientific productivity (Merton, 1968) and student achievement (Coleman et al., 1966). Jensen (1991) considers the effect a ubiquitous phenomenon: “So consistently has been found that it could almost be called The First Law of Individual Differences, to wit: In achievement that do not have a low performance ceiling, instruction that succeeds in raising the group mean also increases the variance among individuals” (p. 178).

While the Matthew effect in student performance may be ubiquitous, it seems particularly evident in the performance of Taiwanese students in mathematics in primary and secondary schools. As a participating country in the Trends in International Mathematics and Science Study (TIMSS), Taiwan has earned a reputation as one of the top-performing countries over a number of survey years. However, a much less known

TIMSS finding peculiar to Taiwan concerns the remarkable contrast between students in Grade 4 and Grade 8 with regards to the inequality of achievement in mathematics. Taiwan stands out from other countries as its results reveal very narrow dispersion among fourth-graders, but extraordinarily wide dispersion among eighth graders (Mullis, Martin & Foy, 2008; Mullis et al., 2000; Mullis et al., 2004; Mullis et al., 2012). Taiwan, compared with more than three-dozen TIMSS-participating countries, consistently shows much larger dispersion in eighth-grade mathematics performance across TIMSS survey years (1999, 2003, 2007, 2011), trailing only South Africa, Qatar, Macedonia, and Turkey. Taiwanese fourth-graders, however, perform quite homogeneously in mathematics from a cross-national perspective. Such a narrow dispersion consistently appears across TIMSS survey years (2003, 2007, and 2011). The dramatic increase in the variation of mathematics performance in just four years, as students progress from Grade 4 to Grade 8, is not seen in other countries.

Longitudinal data from the Taiwan Education Panel Survey (TEPS) also supports the widening-gap phenomenon. For example, Huang (2007) reported a 40% increase in the standard deviation in IRT (Item Response Theory) scores in mathematics in just two years, as students progressed from Grade 7 in 2001 to Grade 9 in 2003 (Huang, 2007, p. 180).[1] As this inequality increased over this two-year period, there was a corresponding 43% growth in the effect of parental education on student performance in mathematics (Huang, 2008, p. 30).

Cross-national data from the Program for International Student Assessment (PISA) for 15-year-old students provide additional support with respect to the significance of the widening-gap phenomenon in Taiwan. Among 65 PISA participating countries in 2009, Taiwan exhibited the widest dispersion in mathematics performance. Therefore, the

[1] In Taiwan, Grade 7 is equivalent to the first year of lower secondary schooling, and the students are about 13 years old. Grade 9 is the highest grade level of junior high school, and most ninth-graders are 15 years old.

widening-gap phenomenon in Taiwan, according to these three data sets (TIMSS, PISA, and TEPS), may arise as early as Grade 4, when, from an international perspective, Taiwanese students exhibited very narrow dispersion in mathematics performance. By the time these Taiwanese students were approaching the end of compulsory schooling (Grade 9), they performed so unequally in mathematics that, internationally, they comprised one of the most heterogeneous student cohorts.

In this paper, I investigate the widening-gap phenomenon in mathematics performance among Taiwanese students. It is important to examine such a phenomenon because the goal for every country is to achieve both excellence and equity in student performance, not one or the other alone. A dramatic increase in the variation in student performance as students progress through the grades may due to a lack of progress early on by some students, who thereafter fall ever further behind. In addition, inequality in math performance in school may have implications for inequality in socioeconomic success later in adult life.

The focus on mathematics achievement has several advantages. Some studies suggest that mathematics skills are better predictors of future earnings than other academic skills learned in high school (Bishop, 1992; Murnane, Willett, and Levy, 1995). In addition, high school achievement in mathematics is more highly correlated with growth in the economic productivity of a nation than that in other subjects (Hanushek and Woessmann, 2012). Finally, performance in mathematics is more appropriate for rigorous comparisons across countries and cultures than performance in other academic subjects, such as science and reading (Hanushek, Peterson, and Woessmann, 2010). This is because of a relatively clear international consensus on the mathematics concepts and techniques that students need to master, and on the sequence in which mathematics knowledge should be taught.

The rest of the paper is organized as follows. First, I introduce the data sets employed in the analysis. Second, I report on the methods used to address several research questions. Third, I examine whether or not the widening-gap phenomenon is found among Taiwanese students of different birth cohorts. If the Matthew effect operates in every successive birth cohort, this would imply that some institutional force, not the characteristics of a specific birth cohort, is responsible for both reproducing and accelerating performance inequality between students of different educational and socioeconomic origins. Fourth, I estimate the magnitude, and identify the characteristics, of the widening-gap phenomenon. Such an analysis is important because the attributes of the widening-gap phenomenon may have implications for identifying its causes. Fifth, I investigate whether the widening gap phenomenon is associated with initial achievement levels, family background, gender, the rural-urban gap, as well as between- and within-classroom differences. Specifically, I address several questions in the fifth section: (a) Do those who perform better initially progress more rapidly over time? (b) Does the Matthew effect occur within or between levels of student socioeconomic background? (c) Does the Matthew effect operate within or between gender groups? (d) Does the Matthew effect broaden the urban-rural gap in mathematics performance? (e) Does the widening-gap phenomenon occur within or between classrooms? The sixth and final section concludes research findings. Some of the research questions in this study are addressed through international comparisons.

2. Data

The investigate the widening-gap phenomenon, it is necessary to have data from students who born around the same year and whose math performance were assessed in multiple grade levels. This study, therefore, uses data from the TIMSS and the TEPS. The TIMSS allows for an international comparison, so it is possible to examine the widening-gap phenomenon in Taiwan in an international context.

2.1. *TIMSS*

The use of the TIMSS data in this study is confined to those collected in 2003, 2007, and 2011 when Taiwan was a participating country. The TIMSS 2003, 2007, and 2011 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 of the students in the classroom were surveyed.

The TIMSS achievement test was designed to reflect as many curricula as possible in all participating countries. To ensure that test items were appropriate and reflected different curricula among countries, the construction of math and science tests were based on the TIMSS curriculum frameworks, which were set by consensus and endorsed by all participating countries.

The TIMSS test was composed of a series of multiple-choice test items and open-ended response questions requiring short, or more elaborate, explanations. Not all students were given the same test questions. To minimize the response burden on individual students, matrix-sampling techniques were used to divide the test item pool so that each sampled student responded to only a portion of the test questions, but a portion which nonetheless covered all subject areas (Garden, 2000). Therefore, students answered different test items depending upon which test booklets they received. On the basis of Item Response Theory, student responses were scaled to provide accurate estimates of achievement that could be compared across countries (Yamamoto, 2000). In addition, 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 against those from 1995. This facilitates a grade-specific trend analysis for countries participating in more than one TIMSS survey.

2.2. *TEPS*

In terms of survey content and design, the TEPS resembles the National Education Longitudinal Study of 1988 (NELS) in the United States.[2] 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 first surveyed in 2001 (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. .

2.2.1. *The first wave of data collection, 2001*

Taiwanese high schools run on a two-semester calendar, with Fall and Spring semesters. An academic year begins in early autumn and ends the following summer. The first (Fall) semester runs from September to January, while the second (Spring) semester runs from February until June. The TEPS baseline survey was carried out in late 2001 when the students were in the first semester of Grades 7 and 11. The baseline survey collected data from students, parents, teachers and schools. A 90-minute cognitive test was administered to all students, not only those participating in the baseline survey, but also those in every later wave of data collection.

The TEPS adopted a three-stage stratified cluster sample design, with schools as the first stage, classes as the second stage, and individual students as the third stage. For each grade

---

[2] The TEPS is designed and supported by Academia Sinica, National Science Council, Ministry of Education, and National Academy for Educational Research Preparatory Office. The principal investigator is Professor Ly-Yun Chang at the Institute of Sociology, Academia Sinica. TEPS data and documentation are available on the World Wide Web at http://www.teps.sinica.edu.tw/

level, the TEPS sample is nationally representative of the Taiwanese student population (Chang, 2009). Within each sampled school of a given grade level, three to six classes were randomly sampled. Within each sampled class, 15 or more students were randomly sampled. In the baseline survey of Grade 7 students, a total of 20,055 students were sampled from 1,244 sampled classes across 333 sampled schools. In Grade 11, the baseline survey sampled 19,051 students from 1,040 sampled classes in 286 sampled schools.

2.2.2. *Cognitive test*

The cognitive test was designed to measure growth in student learning across the secondary schooling years (Yang, Tam & Huang, 2003). The aim of the test is more to measure students' ability to learn, and less to measure the accumulation of factual knowledge. The test includes two components: (a) a subject-specific component, which assesses student achievement in math, science, and verbal; and (b) a curriculum-free component, which assesses analytical, practical, and creative abilities. The subject-specific component includes 20 math items, 10 to 12 science items, and 14 to 16 items of vocabulary knowledge (Chinese and English), for a total of 44 items in Grade 7, and 48 items for Grade 11 students. The curriculum-free component includes three 9-item subtests, measuring analytical, practical, and creative abilities. Within each ability domain, there are three verbal, three graphical, and three numerical test items.

Seventh-graders had 88 minutes to complete a test consisting of 71 items. Eleventh-graders were given 94 minutes to complete a 75-item test. Both raw and IRT (Item Response Theory) scores were made available for research (Yang et al., 2003). Achievement in mathematics was measured by a combination of mathematics test items from the subject-specific component and the curriculum-free component. The level of mathematics achievement was indicated by an IRT score which allows for between-grade comparison. I use these IRT mathematics scores for an analysis of the widening-gap phenomenon.

2.2.3. *The second wave of data collection, 2003*

By February 2003, the Grade 11 sample cohort had moved on to the final (Spring) semester of high school and was expected to graduate in June. This was the semester during which the Grade 11 sample follow-up was conducted. In September 2003, students from the Grade 7 sample were entering the Fall semester of Grade 9; their follow-up survey was conducted at this time.

Between the baseline survey of 2001 and the follow-up survey of 2003, most students sampled by TEPS remained in the same school, and even the same classroom. Thus, response rates were extremely high in the 2003 follow-up survey. In the Grade 7 sample, responses were obtained for 95 percent of the original sample in the baseline survey. In the Grade 11 sample, the response rate was 96 percent. In the present study, I use data from the Grade 7 sample in which students were first surveyed in Grade 7 in 2001, and I also use data from the follow-up survey, which was conducted when these students attended Grade 9 in 2003.

3. Methods

3.1. *Identifying the Widening-Gap Phenomenon in Various Birth Cohorts and Levels of Schooling*

To investigate whether the Matthew effect operates between Grade 4 and Grade 9 for students of different birth cohorts, I use data from three birth cohorts. These 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) TEPS students born around 1989, who were first surveyed in 2001 when they were in Grade 7, and were followed up in 2003 when they were in Grade 9. When analyzing data from the TIMSS, I include all participating countries (or jurisdictions), making possible for an international comparison.

3.2. *Identifying the Magnitude and the Characteristics of the Widening-Gap Phenomenon*

TIMSS is designed to compare student performance of a given grade level (Grade 4 or

Grade 8) between survey years. TIMSS is not designed to compare student performance between grade levels. This study, however, is all about making between-grade comparisons for students born around the same year. Fortunately, it is still possible to make a between-grade comparison in TIMSS. Between-grade comparisons can be made within the same set of countries. In this way, while lacking the absolute size of changes between grade levels, we know how much a country changes relative to changes in other countries. In addition, TIMSS provides information regarding the percentages of students reaching advanced, high, intermediate, or low international benchmarks in mathematics performance, which are set at the score of 625, 550, 475, and 400, respectively. These percentages can be compared between grades. To identify the characteristics of the widening-gap phenomenon, I contrast the distribution of mathematics performance in a lower grade level with that in a higher grade level.

3.3. *Identifying the Covariates of the Widening-Gap Phenomenon*

I use data derived solely from the TEPS to determine whether or not those who do better initially progress more rapidly over time, whether the Matthew effect occurs within or between students' differing socioeconomic backgrounds, and whether or not the Matthew effect broadens the urban-rural gap in mathematics performance. I use data from both the TEPS and the TIMSS to address the question regarding whether the Matthew effect occurs within or between genders. To examine whether or not the Matthew effect occurs within classrooms, I use data derived solely from the TIMSS. This is because Taiwanese students usually stay in the same classroom as they progress from Grade 7 to Grade 9, with various courses being delivered to the students in that same classroom. Therefore, in TEPS, all changes in student performance over time necessarily occur within classrooms.

4. Results

4.1. *Widening-Gap Phenomenon in Various Birth Cohorts*

4.1.1. Grade 4 of 2003 and Grade 8 of 2007

A total of 21 jurisdictions participated in TIMSS in both the Grade 4 survey in 2003 and the Grade 8 survey in 2007. The combination of data from these two TIMSS surveys provides an intra-cohort perspective on changes in student performance for those born around the same year (around 1993), as they progressed from Grade 4 to Grade 8. Based on these data, Table 1 displays the intra-cohort changes in average performance in mathematics and variation of mathematics performance for those who attended Grade 4 in 2003 and Grade 8 in 2007. Taiwan stands out by having the most dramatic expansion in the size of performance variation in mathematics from Grade 4 to Grade 8. Compared to their international peers, Taiwanese fourth-graders perform most homogeneously in mathematics. However, four years later, when these Taiwanese students reached Grade 8, their performance in mathematics varied widely. None of the other jurisdictions listed in Table 1 showed a wider variation in performance. As Taiwanese students progressed from Grade 4 to Grade 8, there was a dramatic change in the degree of variation in mathematics performance, as well as impressive progress in average performance. In Grade 4, Taiwanese students ranked fourth in average mathematics performance among countries listed in Table 1. By Grade 8, Taiwanese students were the top-ranked in average performance in mathematics.

As indicated in Table 1, Hong Kong also experienced the widening-gap phenomenon, albeit to a lesser degree. However, unlike Taiwan, the widening-gap phenomenon in Hong Kong did not bring about a substantial increase in average performance. The U.S. exemplifies the situation wherein variation in mathematics performance does not change significantly between Grade 4 and Grade 8. Additionally, students in the U.S. do not become more competitive as they age. Students in Norway performed higher and more homogeneously in mathematics as they progressed from Grade 4 to Grade 8. Tunisia and Morocco have much lower rates of school enrollment in secondary education relative to other TIMSS participating countries.[3] Therefore, in these two countries, eighth-grade students represent a more selective group relative to fourth-grade students, and these eighth-graders not only performed better, but also more homogeneously.

Table 2 reports between-grade changes in the percentage of students reaching the advanced international benchmark, and in the percentage of students failing to reach the low international benchmark. Taiwan made a remarkable improvement in the percentage of students reaching the advanced international benchmark, from 16% to 45%, as students progressed from Grade 4 to Grade 8. Such a significant improvement is exceptional from an international perspective. Despite the remarkable increase in the proportion of students reaching the advanced benchmark, there was an increase in the percentage of students failing to reach the low international benchmark, from 1% to 5%, as Taiwanese students progressed from Grade 4 to Grade 8.

Figure 1 shows a comparison of test score distribution in mathematics between Grade 4 of TIMSS 2003, and Grade 8 of TIMSS 2007, in Taiwan. The test score distribution of Taiwanese eighth-graders shows a wider dispersion as well as higher average performance, compared with that of Taiwanese fourth-graders.

4.1.2. Grade 4 of 2007 and Grade 8 of 2011

Table 3 reports intra-cohort changes in mathematics performance for a four-year younger cohort, born around 1997 -- students who attended Grade 4 in 2007 and Grade 8 in 2011. The widening-gap phenomenon in mathematics performance persisted as this younger cohort of Taiwanese students progressed from Grade 4 to Grade 8. Among 29 jurisdictions listed in Table 3, Taiwan showed the largest expansion in the variation of mathematics performance as students progressed from Grade 4 to Grade 8. As the variation increased,

---

[3] In secondary education, Morocco reportedly has a net school enrollment rate as low as 39% (Sarri 2002), and it is 65% in Tunisia (Dridi 2002). Mullis et al. (2004, 29) report a net secondary school enrollment ratio of 31% in Morocco and 68% in Tunisia.

average performance was significantly heightened. As a result, Taiwanese eighth graders performed as well as those in Singapore and South Korea, which ranked top in eighth-grade mathematics performance in TIMSS 2011. These three top-performing countries did not differ significantly from one another in average performance in Grade 8 mathematics.

Table 3 indicates that fourth-graders in Massachusetts perform equally well with fourth-graders in Taiwan and Japan. As these students progressed to Grade 8, however, students in Massachusetts and Japan perform significantly more poorly, compared to their counterparts in Taiwan. Students in Taiwan and Massachusetts have almost the same distribution in mathematics performance in Grade 4. By Grade 8, however, high-performing Taiwanese students perform much better than high-performing students in Massachusetts.

While the U.S., as a whole, is not among the high-performing countries in mathematics, students in Massachusetts do very well from an international perspective. This suggests that between-state differences in average performance in mathematics can be very significant in the U.S., and calls for future research into whether state-differences in mathematics performance in the U.S. are attributable to demographic differences in student population, or differences in state and local policies.

For fourth-graders in 2007, who had progressed to Grade 8 in 2011, Table 4 reports between-grade changes in the percentage of students reaching the advanced international benchmark and in the percentage of students failing to reach the low international benchmark. Like the previous TIMSS cohort, which was four years older, students of this TIMSS cohort also made remarkable improvement in the percentage reaching the advanced international benchmark, from 24% to 49%, as they progressed from Grade 4 to Grade 8. Such a dramatic improvement is not seen in other countries, as demonstrated in Table 4. While there was a significant increase in the proportion of students reaching the advanced benchmark, there was also an increase in the percentage of students failing to reach the low international benchmark, from 1% to 4%, as these Taiwanese students progressed from

Grade 4 to Grade 8.

4.1.3. Grade 7 of 2001 and Grade 9 of 2003

As indicated in Tables 1 and 3, the widening gap in mathematics performance in Taiwan is exceptionally large compared with that seen in other countries, and it occurred across multiple birth cohorts covered by the TIMSS. To shed additional light upon the phenomenon, I use data from a national educational longitudinal survey in Taiwan, the TEPS, which covered junior high school students born around 1989. Table 5 reports those findings. The standard deviation in student mathematics performance expanded from 1 in Grade 7, to 1.25 in Grade 9, for those who born around 1989. Such an expansion in variation is also apparent in Figure 2, which compares the mathematics test score distribution in Grade 7 with that of Grade 9. As shown in Figure 2, after students progress from Grade 7 to Grade 9, they perform more unequally in math. The expansion in variation exists because low-performing ninth-graders do as poorly as low-performing seventh-graders, but high-performing ninth-graders do much better than high-performing seventh-graders.

This increase in variation in performance is smaller than that reported by Huang (2007, p. 180), who also used data from the TEPS and suggested an expansion in the standard deviation from 1 to 1.4 as students moved from Grade 7 in 2001, to Grade 9 in 2003. The inconsistency may due to the fact that Huang (2007) used a data set released to the public, which contained only 70% of the total TEPS sample. In addition, Huang did not apply sample weights in his analysis. In the Appendix, Table A1 reports the TEPS findings without applying sample weights. When sample weights are not applied, the widening-gap phenomenon is slightly more significant, suggesting an expansion in the standard deviation of mathematics performance from 1 to 1.3 as students progressed over time from Grade 7 to Grade 9.

Until the final year of lower-secondary schooling (Grade 9), the widening-gap phenomenon in mathematics performance in Taiwan is common to those who born around 1989, 1993, and 1997. While these students performed more unequally over time, they also made substantial progress in terms of their average performance and the percentage of reaching advanced international benchmark. Because the improvement Taiwanese students make as they progress from elementary to junior high school, Taiwanese eighth-graders are usually listed in the top rank in terms of average performance in mathematics in international student assessments, such as the TIMSS.

4.2. *Does the Widening-Gap Phenomenon Appear Between or Within Genders?*

The widening-gap phenomenon in Taiwan acted on boys as well as girls. Gender-specific analysis shows that the widening-gap phenomenon in Taiwan occurred within gender, as displayed in Table 5 (for TEPS) and Tables A2-A5 (for TIMSS) in the Appendix. Male students tended to perform more unequally than did female students, but the magnitude of the expansion in variation over time did not differ significantly between male and female students.

4.3. *Does the Widening-Gap Phenomenon Take Place Between or Within Types of Residential Communities?*

Changes in the variation in mathematics performance as students progress through the grades may take place in some residential areas, but not others. Therefore, I distinguish between three kinds of residential communities: city, town, and countryside. Based on longitudinal data from the TEPS, Table 6 displays over-time changes in average mathematics performance and the size of variation in mathematics performance among students of these three residential community types. The two-year widening-gap phenomenon that occurred to students who attended Grade 7 in 2001, and Grade 9 in 2003, took place in all residential community types.

4.4. *Does the Widening-Gap Phenomenon Happen Between or Within Classrooms?*

Various data sets from the TIMSS consistently indicated the existence of a widening-gap phenomenon that begins between Grade 4 and Grade 8 in Taiwan. Such a phenomenon, as indicated in Tables 7 and 8, mainly occurred within classrooms. For example, Taiwanese fourth-graders, in 2003, performed homogeneously within classrooms, despite nearly 84 percent of the total variation in mathematics performance occurring within classrooms. Only Russia, Hong Kong, and Singapore had less within-classroom variation in mathematics performance. Four years later, in 2007, when these Taiwanese students were in Grade 8, their within-classroom performance demonstrated a higher level of inequality than any other country. This dramatic change in the size of within-classroom variation, which occurred over just four years, was exceptional from an international perspective, as indicated in the last column of Table 7.

Table 8 indicates that the same phenomenon was seen in a four-year younger cohort that attended Grade 4 in 2007, and Grade 8 in 2011. Taiwanese fourth-graders started with a low level of within-classroom variation in their performance in mathematics. By Grade 8, these students performed so unequally within classrooms that no country showed a higher level of within-classroom variation. To be sure, the widening-gap phenomenon in Taiwan also led to an increase in between-classroom variation, but this is much less significant than the increase in within-classroom variation.

Table 8 also reports that student performance in mathematics has become more highly stratified by school and classroom, as American students progress from Grade 4 in 2007 to Grade 8 in 2011. The change from a comprehensive system in Grade 4 to a stratified system in Grade 8 is most significant in Massachusetts. In Grade 4, more than three-quarters of the total variation in mathematics performance occurs within classrooms in Massachusetts. By Grade 8, however, only 40 percent of the total variation lies within classrooms. Such a change as occurs in Massachusetts is rather distinctive, from an international perspective. Among 29 jurisdictions displayed in Table 8, only England showed a more significant change than that occurred in Massachusetts.

4.5. *Do Those Who Do Better Initially Progress More Rapidly Over Time?*

A possible explanation for the widening-gap phenomenon in Taiwan is that those who perform better in mathematics initially may progress more rapidly over time. To test this hypothesis, I use data derived solely from the TEPS. This is because the TEPS is a longitudinal study and its mathematical IRT scores are suitable for between-grade comparisons. I divided the sample from an earlier data collection wave into 10 decile groups according to students' mathematics IRT scores, then estimated the average growth in mathematics scores for each of these ten groups in a later follow-up survey. The results of this exercise are found in Table 9. For Taiwanese students who attended Grade 7 in 2001, and Grade 9 in 2003, those who performed better in mathematics in Grade 7 tended to improve more in terms of their test scores two years later. This is particularly evident when the highest-performing and the lowest-performing decile groups are excluded. The growth of the top-performing group may have been suppressed by a ceiling effect, while the lowest-performing group may have had more room to grow than other groups.

4.6. *Does the Widening-Gap Phenomenon Occur Between or Within Socioeconomic Groups?*

I use data from the TEPS to address the question whether the widening-gap phenomenon occur between or within socioeconomic groups. I took the father's level of education to indicate a student's socioeconomic background. The father's level of education described according to one of five categories: less than high school, high school, some college, college, and having a graduate degree.

Table 10 presents the results for the analysis using data from the TEPS. For those who attended Grade 7 in 2001 and Grade 9 in 2003, the widening-gap phenomenon occurred to all students, but the phenomenon was twice as significant among students with less favorable family backgrounds than among students with more favorable backgrounds. Furthermore, students from more favorable family backgrounds progressed more rapidly in mathematics performance as they moved through the grades.

5. Discussions and Conclusions

The widening-gap phenomenon in mathematics performance, which occurred prior to Grade 9 in Taiwan, is unique from an international perspective. It is unique because no other country has as large a widening gap as that found in Taiwan. The phenomenon is unique also because that it accompanies a significant increase in average performance and a remarkable increase in the percentage of students reaching advanced international benchmark. Hong Kong, for example, also had a widening-gap phenomenon in mathematics performance, but mathematics performance did not improve significantly.

Both male and female Taiwanese students in elementary and junior high schools experienced the widening-gap phenomenon in mathematics performance. The widening-gap phenomenon is common to all residential areas, whether they are cities, towns, or the country side, and it occurred mainly within classrooms. No other country showed such a high level of within-classroom variation, moving from a very low level of within-classroom variation in Grade 4, to extremely heterogeneous classrooms in Grade 8. This dramatic change in within-classroom variation between grade levels is distinctive from an international perspective.

The widening-gap phenomenon in Taiwan is associated with findings that those who performed better initially progressed more rapidly over time. Additionally, students with more favorable family backgrounds improved more significantly over time, and they displayed a less significant widening-gap phenomenon amongst themselves.

In Taiwanese elementary and junior high schools, most variation in student performance occurs within classrooms. In Grade 8, most Taiwanese schools and classrooms are extremely heterogeneous internally in terms of mathematics performance. Thus, rather than being concentrated on a few classrooms or schools, poorly-performing students in Taiwan are dispersed widely across classrooms. Therefore, a policy targeting low-performing students within classrooms will be more effective in narrowing the performance gap than a policy targeting at specific schools or residential communities. An effective policy to achieve equity and excellence in Taiwan requires a focus on children who perform poorly at an early age, and on those from socioeconomically disadvantaged families.

Why does the widening-gap phenomenon in Taiwan reappear as each subsequent birth cohort progresses from elementary to junior high school? If so, is this a consequence of the "Basic Competence Test for Junior High School Students" – a national examination administered to nearly all Taiwanese ninth-graders for admission to senior high schools? Taiwanese students compete ferociously for places in the best public academic high schools based on the results of the examination. High schools in Taiwan are subtly ranked according to student performance, and most parents and students know how each high school ranks. Admission to high school depends mainly on student performance on the national examination, which signals multiple and subtle levels of achievement. Each high school has a minimum required test score for admission. The best high schools are for those who do extremely well on the national examination. Those whose test scores fail to reach the minimum requirement for the best high schools strive for entrance to a second- or third-tier high school. Students of all achievement levels, except maybe those at the lowest level, are motivated to work to improve their score to win a place in a better high school.

Starting in 2014, Taiwanese 9th-graders will no longer need to take such a competitive national examination for admission to senior high school because compulsory education will be extended to high school graduation. There will be a new national examination for 9th-graders, but it will signal three levels of achievement only -- advanced, intermediate, and low. Future findings from the TIMSS 2015 will help address the questions (a) whether or not the widening-gap phenomenon in Taiwan will persist and (b) whether the remarkable increase in the percentage of students reaching the advanced benchmark will persist in the absence of the traditional high-school entrance examinations.

Some of the research questions in this study are addressed through international comparisons. These international comparisons yield some findings which have implications for policy and research. For example, in Grade 4, students in Massachusetts and Taiwan perform equally well in mathematics. This finding puts to rest the myth that America's schools cannot be among the world's top-performing school systems. However, four years later, after these students had progressed from Grade 4 to Grade 8, a significant performance gap appears between these two jurisdictions, due to a substantial improvement in performance among Taiwanese students. This performance gap between Taiwan and Massachusetts, which emerges over the course of just four years, may have implications for policy making in the U.S., where mathematics performance has been seen to improve but very slowly as students progress through the grades. In addition, the presence of a remarkable performance gap in mathematics between Massachusetts and other states in the U.S. seems to suggest that local and state policies matter.

American education has been severely criticized for failing to bring its students up to the highest levels of accomplishment in mathematics. It has been found that the percentage of high-achieving students in mathematics in the U.S. is well below that in many of the world's leading industrialized nations (Hanushek, Peterson, and Woessmann, 2010). Some suggests that "America's future depends on how we educate the academically gifted" (Murray, 2008). The relatively low number of high-performing students in the U.S. has raised concerns that "the U.S. cannot afford to neglect high performers in our quest to bring up the bottom. Performance at the top end is no less important, and improvements at both ends reinforce each other, helping to accelerate the growth in productivity of the nation's economy" (Hanushek, Peterson, and Woessmann, 2010, p. 8).

In light of such concerns, Taiwan presents a very special and useful case for future research. In just four years, as Taiwanese students progress from Grade 4 to Grade 8, the percentage of Taiwanese students reaching the advanced international benchmark increases from one-quarter to one-half. Such a dramatic improvement is the envy of the world; it is not seen in other countries, and its causes are yet to be researched.

Reference

Bishop, J. H. (1992). The impact of academic competencies of wages, unemployment, and job performance. *Carnegie-Rochester Conference Series on Public Policy* 37(December): 127-194.

Chang, L. Y. (2002). The introduction of the Taiwan Education Panel Survey. *Survey Research*, 12: 111-118.

Chang, L. Y. (2009). *The handbook of the Taiwan Education Panel Survey*. Taipei: Academia Sinica.

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

Dridi, B. L. (2002). Tunisia. In R. Marlow-Ferguson (Ed.) *World Education Encyclopedia: a Survey of Educational Systems Worldwide*, 2nd ed. Detroit, MI: Gale Group.

Foy, P., & Joncas, M. (2000). TIMSS sample design. In M. O. Martin, K. D. Gregory & S. E. Stemler (Eds.), *TIMSS 1999 technical report* (pp. 29-45). Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.

Garden, R. A., & Smith, T. A. (2000). TIMSS test development. In: M. O. Martin, K. D. Gregory & S. E. Stemler (Eds.), *TIMSS 1999 technical report* (pp. 49-67). Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.

Hanushek, E. A., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, 17(4): 267-321.

Hanushek, E. A., Peterson, P. E., & Woessmann, L. (2010). *U.S. math performance in global perspective: How well does each state do at producing high-achieving students?* PEPG Report No. 10–19. Cambridge, MA: Program on Education Policy and Governance, Harvard University

Huang, M. H. (2007). Variability of math performance within and between classrooms: An international and inter-grade study. *Taiwanese Sociology*, 14: 155-189.

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

Jensen, A. R. (1991). Spearman's *g* and the problem of educational equality. *Oxford Review of Education*, 17, 2: 169-187.

Merton, R. K. (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science*, 159: 56-63.

Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007 international mathematics report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.

Mullis, I.V.S., Martin, M.O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics.* Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report: Findings from IEA'S trends in international mathematics and science study at the fourth and eighth grades.* Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., Chrostowski, S. J., & Smith T. A. (2000). *TIMSS 1999 international mathematics report: Findings from IEA's repeat of the third international mathematics and science study at the eighth grade*. Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.

Murnane, R. J., Willett, J. B., & Levy, F. (1995). The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics*, 77(2): 251-266.

Murray, C. A. (2008). *Real education : four simple truths for bringing America's schools back to reality*. New York: Crown Forum.

OECD (2009). *PISA 2006 technical report*. Paris: OECD Publishing.

Sarri, S. (2002). Morocco. In R. Marlow-Ferguson (Ed.) *World Education Encyclopedia: a Survey of Educational Systems Worldwide*, 2nd ed. Detroit, MI: Gale Group.

Tam, T. (2004). *Methodological issues in the sampling design of the third wave of TEPS: Learning 2000 project technical report*. Taipei: Academia Sinica Center Survey for Research.

Western, B., & Bloome, D. (2009). Variance function regressions for studying inequality. *Sociology Methodology*, 39, 1: 293-326.

Yamamoto, K. (2000). Scaling methodology and procedures for the TIMSS mathematics and science scales. In M. O. Martin, K. D. Gregory & S. E. Stemler (Eds.), *TIMSS 1999 technical report* (pp. 237-263). Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.

Yang, M. L., Tam. T., & Huang, M. H. (2003). *Psychometric report for the Taiwan Panel Education Survey: Learning 2000 project technical report*. Taipei: Academia Sinica Center Survey for Research.

Table 1. Differences between the Grade 4 of 2003 and Grade 8 of 2007 in average mathematics performance and the size of mathematics performance variation (standard deviation and interquartile range), TIMSS

	Grade 4, 2003			Grade 8, 2007			Grade Difference		
	Mean	S. D.	IQR	Mean	S. D.	IQR	Mean	S. D.	IQR
Taiwan	564	63	81	598	106	137	34	43	56
Hong Kong, PRC	575	63	85	572	94	120	-3	31	35
Japan	565	74	97	570	85	113	5	11	16
Singapore	594	84	106	593	93	128	-1	9	22
Hungary	529	77	105	517	85	117	-12	8	12
Lithuania	534	74	98	506	80	108	-28	6	10
Russia	532	78	107	512	83	114	-20	5	7
Cyprus	510	85	115	465	89	119	-45	4	4
Quebec, Canada	506	65	87	528	68	94	22	3	7
Scotland	490	78	103	487	80	112	-3	2	9
United States	518	76	105	508	77	107	-10	1	2
Iran	389	86	120	403	86	115	14	0	-5
Ontario, Canada	511	71	94	517	70	93	6	-1	-1
Australia	499	81	106	496	79	105	-3	-2	-1
Armenia	456	87	116	499	85	106	43	-2	-10
England	531	87	118	513	84	115	-18	-3	-3
Italy	503	82	108	480	76	102	-23	-6	-6
Slovenia	479	78	107	501	72	96	22	-6	-11
Morocco	347	90	129	381	80	115	34	-10	-14
Norway	451	80	109	469	66	92	18	-14	-17
Tunisia	339	100	140	420	67	91	81	-33	-49

Table 2. Differences between Grade 4 of 2003 Grade 8 of 2007 in the percentage of students reaching the advanced international benchmark and in the percentage of students not reaching the low international benchmark, TIMSS

	% reaching advanced benchmark			% not reaching low benchmark		
	Grade 4	Grade 8	Difference	Grade 4	Grade 8	Difference
Taiwan	16	45	29	1	5	4
Hong Kong, PRC	22	31	9	1	6	5
Japan	21	26	5	2	3	1
Quebec, Canada	3	8	5	6	3	-3
Armenia	2	6	4	25	12	-13
Singapore	38	40	2	3	3	0
Slovenia	2	4	2	16	8	-8
Ontario, Canada	5	6	1	6	5	-1
Australia	5	6	1	12	11	-1
Scotland	3	4	1	12	15	3
Iran	0	1	1	55	49	-6
Hungary	10	10	0	6	9	3
Tunisia	0	0	0	72	39	-33
Morocco	0	0	0	71	59	-12
United States	7	6	-1	7	8	1
Norway	1	0	-1	25	15	-10
Russia	11	8	-3	5	9	4
Italy	6	3	-3	11	15	4
Lithuania	10	6	-4	4	10	6
England	14	8	-6	7	10	3
Cyprus	8	2	-6	11	22	11

Table 3. Differences between the Grade 4 sample of 2007 and Grade 8 sample of 2011 in average mathematics performance and the size of mathematics performance variation (standard deviation and interquartile range), TIMSS

	Grade 4, 2007			Grade 8, 2011			Grade Difference		
	Mean	S. D.	IQR	Mean	S. D.	IQR	Mean	S. D.	IQR
Taiwan	576	69	91	609	106	140	33	37	49
Qatar	296	90	127	410	110	158	114	20	31
Georgia	438	88	123	431	106	151	-7	18	28
Hong Kong, PRC	607	67	89	586	84	107	-21	17	18
Iran	402	84	115	415	95	128	13	11	13
Japan	568	76	100	570	85	115	2	9	15
Ukraine	469	84	114	479	90	121	10	6	7
Lithuania	530	76	101	502	79	108	-28	3	7
Dubai, UAE	444	90	122	478	93	129	34	3	7
Massachusetts, US	572	70	92	561	73	98	-11	3	6
Ontario, Canada	512	68	90	512	71	96	0	3	6
Australia	516	83	110	505	85	115	-11	2	5
Sweden	503	66	89	484	68	92	-19	2	3
United States	529	75	102	509	77	105	-20	2	3
Armenia	500	90	120	467	91	126	-33	1	6
Singapore	599	84	111	611	84	113	12	0	2
England	541	86	113	507	85	119	-34	-1	6
Hungary	510	91	122	505	90	119	-5	-1	-3
New Zealand	492	86	117	488	85	120	-4	-1	3
Slovenia	502	71	93	505	70	97	3	-1	4
Russian Federation	544	83	107	539	81	111	-5	-2	4
Alberta, Canada	505	66	87	505	63	86	0	-3	-1
Italy	507	77	101	498	73	99	-9	-4	-2
Kazakhstan	549	84	114	487	80	113	-62	-4	-1
Quebec, Canada	519	67	93	532	62	84	13	-5	-9
Minnesota, US	554	78	105	545	72	98	-9	-6	-7
Morocco	341	95	131	371	86	116	30	-9	-15
Norway	473	76	102	475	65	89	2	-11	-13
Tunisia	327	111	162	425	75	102	98	-36	-60

Table 4. Differences between Grade 4 of 2007 Grade 8 of 2011 in the percentage of students reaching the advanced international benchmark and in the percentage of students not reaching the low international benchmark, TIMSS

	% reaching advanced benchmark			% not reaching low benchmark		
	Grade 4	Grade 8	Difference	Grade 4	Grade 8	Difference
Taiwan	24	49	25	1	4	3
Singapore	41	48	7	2	1	-1
Japan	23	27	4	2	3	1
Ukraine	2	5	3	21	19	-2
Dubai, UAE	2	5	3	31	21	-10
Georgia	1	3	2	33	38	5
Iran	0	2	2	47	45	-2
Qatar	0	2	2	87	46	-41
Quebec, Canada	5	6	1	4	2	-2
Slovenia	3	4	1	8	7	-1
Australia	9	9	0	9	11	2
New Zealand	5	5	0	15	16	1
Ontario, Canada	4	4	0	6	6	0
Alberta, Canada	3	3	0	6	5	-1
Tunisia	0	0	0	72	39	-33
Morocco	0	0	0	74	64	-10
Hungary	9	8	-1	12	12	0
Norway	2	1	-1	17	13	-4
Russia	16	14	-2	5	5	0
Sweden	3	1	-2	7	11	4
Massachusetts., US	22	19	-3	1	2	1
United States	10	7	-3	5	8	3
Italy	6	3	-3	9	10	1
Minnesota., US	18	13	-5	3	3	0
Lithuania	10	5	-5	6	10	4
Armenia	8	3	-5	13	24	11
Hong Kong, PRC	40	34	-6	0	3	3
England	16	8	-8	6	12	6
Kazakhstan	19	3	-16	5	15	10

Table 5. Over-time changes in average mathematics performance and the size of variation in mathematics performance among junior high school students, with sample weights applied, TEPS

	All Students		Female Students		Male Students	
	Mean	S. D.	Mean	S. D.	Mean	S. D.
N=18,655						
Grade 7, 2001	0.02	0.997	0.02	0.95	0.02	1.04
Grade 9, 2003	0.811	1.242	0.793	1.186	0.828	1.295
Growth	0.791	0.245	0.773	0.236	0.808	0.255

Table 6. Over-time changes in average mathematics performance and the size of variation in mathematics performance among students of different residential communities, TEPS

	City		Town		Countryside	
	Mean	S. D.	Mean	S. D.	Mean	S. D.
Junior High, N=18,655						
Grade 7, 2001	0.17	0.97	0.01	0.99	-0.46	0.95
Grade 9, 2003	0.98	1.22	0.79	1.24	0.30	1.18
Growth	0.81	0.25	0.78	0.25	0.76	0.23
Sample Size	10,389		6,465		1,801	

Table 7. Differences between the Grade 4 sample of TIMSS 2003 and Grade 8 sample of TIMSS 2007 in (a) the total variance in mathematics performance, (b) the percentage of performance variance in mathematics lying within classrooms (c) the amount of variance in mathematics performance lying within classrooms

	Grade 4, 2003			Grade 8, 2007			Grade Differences		
	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
Taiwan	3969	83.5	3313	11236	74.7	8394	7267	-8.8	5081
Cyprus	7225	85.8	6199	7921	90.2	7144	696	4.4	945
Russia	6084	51.2	3113	6889	56.3	3879	805	5.1	766
Japan	5476	91.6	5015	7225	77.4	5589	1749	-14.2	574
Lithuania	5476	69.7	3818	6400	68.2	4364	924	-1.5	546
Hong Kong, PRC	3969	72.1	2861	8836	36	3180	4867	-36.1	319
Armenia	7569	66.3	5017	7225	72.9	5269	-344	6.6	252
Hungary	5929	71.4	4233	7225	59.4	4288	1296	-12	55
Morocco	8100	61.7	4996	6400	74.4	4762	-1700	12.7	-234
Italy	6724	61.6	4145	5776	67.2	3883	-948	5.6	-262
Ontario, Canada	5041	73.2	3689	4900	68.5	3357	-141	-4.7	-332
Iran	7396	67.8	5011	7396	58.8	4348	0	-9	-663
Slovenia	6084	83.6	5088	5184	83.1	4309	-900	-0.5	-779
Quebec, Canada	4225	81.4	3440	4624	49.1	2269	399	-32.3	-1171
Singapore	7056	43.5	3066	8649	21.9	1894	1593	-21.6	-1172
United States	5776	64.1	3704	5929	38.3	2271	153	-25.8	-1433
Norway	6400	83.6	5353	4356	83.5	3637	-2044	-0.1	-1716
Australia	6561	68.6	4498	6241	43.5	2714	-320	-25.1	-1784
Scotland	6084	77.1	4688	6400	32.8	2100	316	-44.3	-2588
Tunisia	10000	62	6204	4489	80.1	3594	-5511	18.1	-2610
England	7569	70.8	5356	7056	23.6	1667	-513	-47.2	-3689

Table 8. Differences between the Grade 4 sample of TIMSS 2007 and Grade 8 sample of TIMSS 2011 in (a) the total variance in mathematics performance, (b) the percentage of performance variance in mathematics lying within classrooms (c) the amount of variance in mathematics performance lying within classrooms

	Grade 4, 2007			Grade 8, 2011			Grade Differences		
	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
Taiwan	4761	86.8	4132	11236	75.6	8492	6475	-11.2	4360
Georgia	7744	64.1	4964	11236	65.2	7324	3492	1.1	2360
Morocco	9025	47.5	4283	7396	70.5	5212	-1629	23.0	929
Iran	7056	57.2	4034	9025	54.0	4875	1969	-3.2	840
Japan	5776	88.4	5105	7225	81.8	5911	1449	-6.6	806
Ontario, Canada	4624	68.5	3169	5041	72.2	3638	417	3.6	468
Ukraine	7056	73.6	5192	8100	69.6	5639	1044	-4.0	447
Sweden	4356	79.7	3473	4624	81.3	3760	268	1.6	287
Russian Federation	6889	48.0	3304	6561	54.2	3556	-328	6.2	252
Armenia	8100	69.9	5660	8281	69.0	5717	181	-0.8	58
Italy	5929	64.9	3846	5329	71.8	3826	-600	6.9	-20
Slovenia	5041	83.6	4213	4900	85.0	4167	-141	1.5	-46
Lithuania	5776	73.7	4259	6241	67.3	4201	465	-6.4	-58
Kazakhstan	7056	43.4	3065	6400	46.7	2992	-656	3.3	-73
Qatar	8100	77.4	6270	12100	50.8	6148	4000	-26.6	-122
Hungary	8281	58.9	4881	8100	57.6	4665	-181	-1.4	-217
Alberta, Canada	4356	72.4	3153	3969	72.6	2881	-387	0.2	-272
Dubai, UAE	8100	54.0	4377	8649	47.1	4073	549	-6.9	-304
Hong Kong, PRC	4489	66.3	2976	7056	36.5	2579	2567	-29.7	-397
Norway	5776	77.5	4478	4225	83.0	3508	-1551	5.5	-970
United States	5625	60.3	3394	5929	37.0	2194	304	-23.3	-1200
Quebec, Canada	4489	74.4	3338	3844	54.6	2100	-645	-19.7	-1238
Singapore	7056	45.2	3188	7056	23.0	1624	0	-22.2	-1564
Australia	6889	62.2	4285	7225	37.1	2682	336	-25.1	-1602
Massachusetts, US	4900	77.4	3793	5329	40.8	2172	429	-36.7	-1621
New Zealand	7396	65.7	4858	7225	44.1	3184	-171	-21.6	-1674
Minnesota, US	6084	68.9	4191	5184	40.1	2078	-900	-28.8	-2113
England	7396	78.0	5768	7225	22.6	1635	-171	-55.4	-4134
Tunisia	12321	67.5	8316	5625	69.8	3927	-6696	2.3	-4389

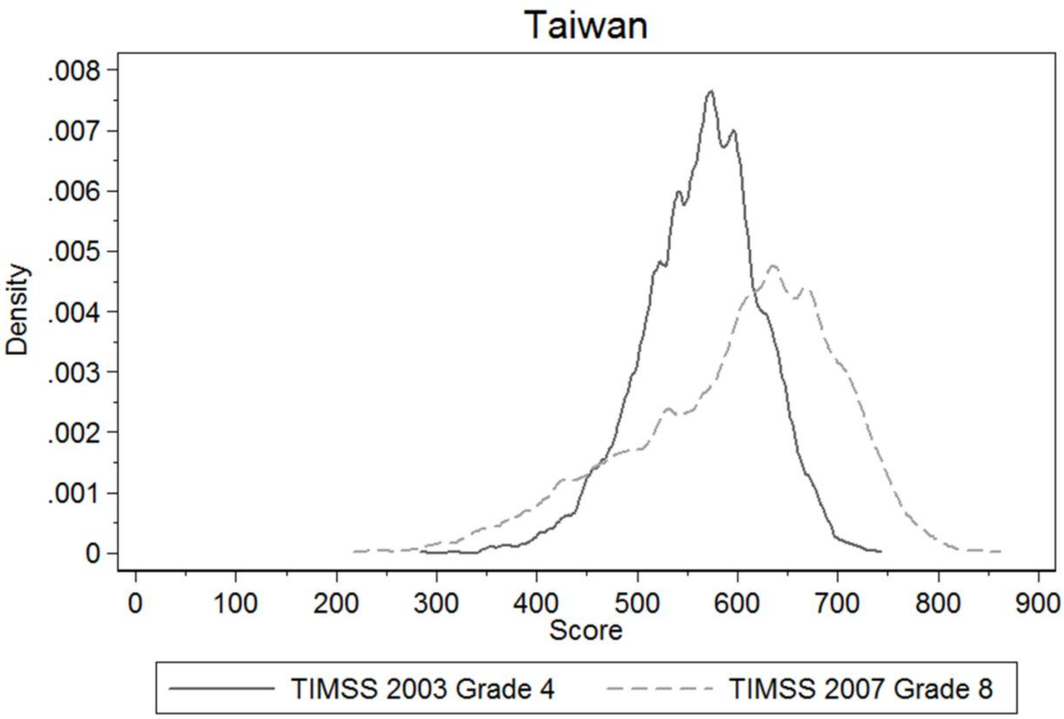
Table 9. Growth in average mathematics IRT test scores between Grade 7 and Grade 9 for each decile group in Grade 7 mathematics performance, TEPS

	Growth
Lowest to 10%	0.92
11-20%	0.63
21-30%	0.64
31-40%	0.68
41-50%	0.77
51-60%	0.83
61-70%	0.85
71-80%	0.85
81-90%	0.87
91% to Top	0.67
Sample Size	18,655

Table 10. Over-time changes in average mathematics performance and the size of variation in mathematics performance among students of different levels of father's education (less than high school, high school, some college, college, and graduate school), TEPS

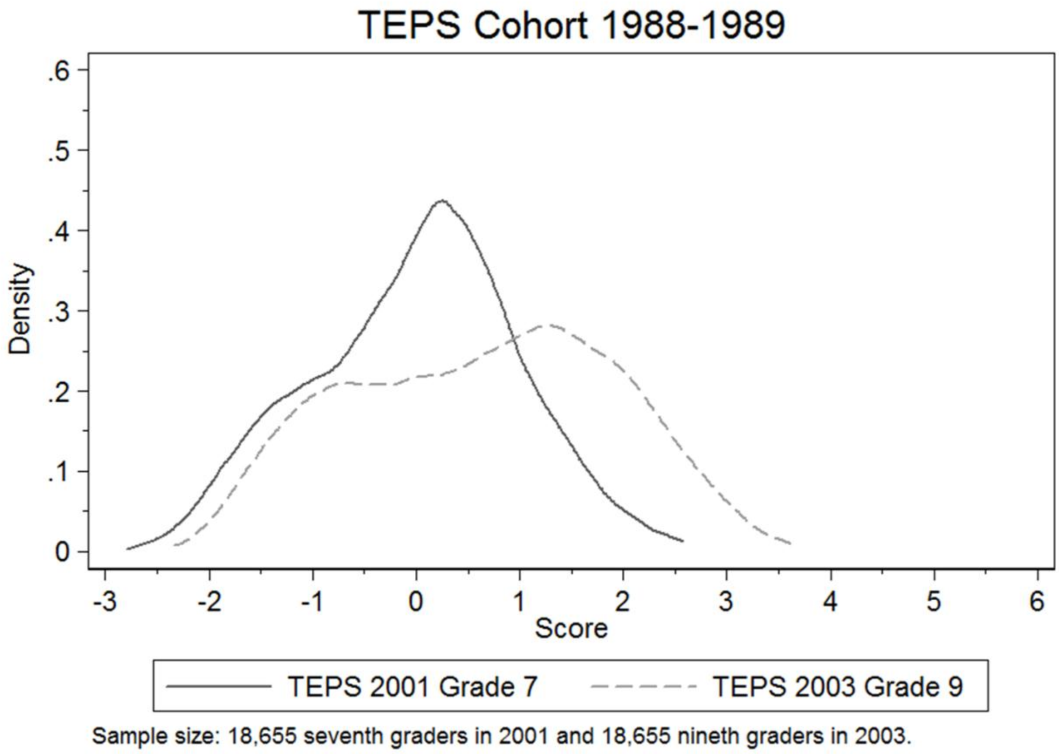
	Less High		High School		Some College		College		Graduate Sch.	
	Mean	S. D.	Mean	S. D.	Mean	S. D.	Mean	S. D.	Mean	S. D.
N=18,655										
Grade 7, 2001	-0.30	0.97	0.03	0.94	0.42	0.90	0.64	0.89	0.83	0.83
Grade 9, 2003	0.40	1.20	0.80	1.17	1.34	1.07	1.64	1.01	1.89	0.97
Growth	0.69	0.23	0.77	0.23	0.92	0.18	1.00	0.13	1.07	0.14
Sample Size	6039		6479		2647		1508		546	

Figure 1



Sample size: 4,661 fourth graders in 2003 and 4,046 eighth graders in 2007.

Figure 2


TEPS Cohort 1988-1989
Density
.6
.5
.4
.3
.2
.1
0
-3
-2
-1
0
1
2
3
4
5
6
Score
TEPS 2001 Grade 7
TEPS 2003 Grade 9
Sample size: 18,655 seventh graders in 2001 and 18,655 nineth graders in 2003.

Appendix

Table A1. Over-time changes in average mathematics performance and the size of variation in mathematics performance among junior and senior high school students, without applying sample weights, TEPS

	All Students		Female Students		Male Students	
	Mean	S. D.	Mean	S. D.	Mean	S. D.
Junior High, N=18,655						
Grade 7, 2001	-0.004	1.005	0.004	0.959	-0.011	1.046
Grade 9, 2003	0.592	1.304	0.612	1.248	0.574	1.355
Growth	0.596	0.299	0.608	0.289	0.585	0.309

Table A2. Differences between the Grade 4 sample of 2003 and Grade 8 sample of 2007 in average mathematics performance and the size of mathematics performance variation (standard deviation and interquartile range), female students only, TIMSS

	Grade 4, 2003			Grade 8, 2007			Grade Difference		
	Mean	S. D.	IQR	Mean	S. D.	IQR	Mean	S. D.	IQR
Taiwan	564	59	75	599	98	126	35	39	51
Hong Kong, PRC	575	60	80	578	87	109	3	27	29
Japan	563	69	90	568	85	107	5	16	17
Singapore	599	80	100	600	88	120	1	8	20
Hungary	527	77	103	517	83	114	-10	6	11
Lithuania	535	73	96	509	78	106	-26	5	10
Scotland	485	74	97	486	78	109	1	4	12
Cyprus	505	82	110	476	84	112	-29	2	2
Russian Federation	530	78	108	514	80	112	-16	2	4
Quebec, Canada	502	64	85	527	66	95	25	2	10
Armenia	462	84	112	501	85	105	39	1	-7
United States	514	74	102	507	75	105	-7	1	3
Iran	394	83	114	407	83	111	13	0	-3
Ontario Canada	505	69	93	513	67	92	8	-2	-1
England	530	85	114	511	82	115	-19	-3	1
Australia	497	79	103	488	75	99	-9	-4	-4
Italy	498	80	103	477	75	100	-21	-5	-3
Slovenia	477	75	101	500	69	94	23	-6	-7
Morocco	344	91	130	377	80	117	33	-11	-13
Norway	449	78	104	471	63	89	22	-15	-15
Tunisia	342	98	139	410	67	91	68	-31	-48

Table A3. Differences between the Grade 4 sample of 2003 and Grade 8 sample of 2007 in average mathematics performance and the size of mathematics performance variation (standard deviation and interquartile range), male students only, TIMSS

	Grade 4, 2003			Grade 8, 2007			Grade Difference		
	Mean	S. D.	IQR	Mean	S. D.	IQR	Mean	S. D.	IQR
Taiwan	564	66	85	598	112	149	34	46	64
Hong Kong, PRC	575	66	87	567	99	133	-8	33	46
Singapore	590	88	112	586	97	137	-4	9	25
Hungary	530	78	107	517	86	119	-13	8	12
Japan	566	78	102	572	86	117	6	8	15
Russian Federation	534	78	106	509	86	117	-25	8	11
Cyprus	514	88	118	455	94	126	-59	6	8
Lithuania	536	76	103	502	81	110	-34	5	7
Quebec, Canada	509	66	88	529	70	93	20	4	5
Iran	386	87	123	400	89	118	14	2	-5
Scotland	496	80	109	489	81	115	-7	1	6
Ontario, Canada	517	72	94	522	73	95	5	1	1
United States	522	78	108	510	78	110	-12	0	2
Australia	500	83	109	504	82	108	4	-1	-1
Armenia	450	89	120	497	85	107	47	-4	-13
Italy	507	83	113	483	78	104	-24	-5	-9
England	532	90	122	516	85	116	-16	-5	-6
Slovenia	481	81	112	503	74	98	22	-7	-14
Morocco	350	89	127	385	80	112	35	-9	-15
Norway	454	82	114	467	68	94	13	-14	-20
Tunisia	337	101	142	431	64	87	94	-37	-55

Table A4. Differences between the Grade 4 sample of 2007 and Grade 8 sample of 2011 in average mathematics performance and the size of mathematics performance variation (standard deviation and interquartile range), female students only, TIMSS

	Grade 4, 2007			Grade 8, 2011			Grade Difference		
	Mean	S. D.	IQR	Mean	S. D.	IQR	Mean	S. D.	IQR
Taiwan	575	65	87	613	100	133	38	35	46
Qatar	307	86	121	415	107	152	108	21	31
Hong Kong, PRC	605	64	85	588	81	100	-17	17	15
Georgia	440	86	117	430	101	144	-10	15	27
Iran	409	79	109	411	92	123	2	13	14
Japan	568	73	97	566	80	110	-2	7	13
Massachusetts, US	567	68	90	558	73	98	-9	5	8
Lithuania	530	73	98	507	76	104	-23	3	6
Australia	513	79	103	500	83	112	-13	4	9
Ukraine	469	81	111	478	85	114	9	4	3
Sweden	499	64	86	486	67	93	-13	3	7
Ontario, Canada	509	66	88	512	69	93	3	3	5
United States	526	74	100	508	76	103	-18	2	3
Dubai, UAE	452	82	110	486	84	116	34	2	6
Slovenia	499	68	91	502	70	98	3	2	7
New Zealand	492	82	110	478	82	114	-14	0	4
Hungary	508	90	122	502	89	118	-6	-1	-4
Alberta, Canada	500	64	86	504	63	85	4	-1	-1
England	541	83	110	508	82	115	-33	-1	5
Singapore	603	80	106	615	78	106	12	-2	0
Russian Federation	548	81	105	539	79	110	-9	-2	5
Armenia	504	90	120	472	87	118	-32	-3	-2
Italy	499	76	101	493	72	97	-6	-4	-4
Kazakhstan	553	82	111	486	78	109	-67	-4	-2
Minnesota, US	551	75	103	545	70	95	-6	-5	-8
Quebec, Canada	515	67	95	531	61	84	16	-6	-11
Morocco	339	94	129	371	87	118	32	-7	-11
Norway	470	76	101	476	64	90	6	-12	-11
Tunisia	337	108	157	417	75	103	80	-33	-54

Table A5. Differences between the Grade 4 sample of 2007 and Grade 8 sample of 2011 in average mathematics performance and the size of mathematics performance variation (standard deviation and interquartile range), male students only, TIMSS

	Grade 4, 2007			Grade 11, 2011			Grade Difference		
	Mean	S. D.	IQR	Mean	S. D.	IQR	Mean	S. D.	IQR
Taiwan	577	73	96	606	111	147	29	38	51
Qatar	285	93	130	404	113	163	119	20	33
Georgia	437	90	127	432	110	156	-5	20	29
Hong Kong, PRC	609	70	91	583	88	114	-26	18	23
Japan	568	79	102	574	89	118	6	10	16
Iran	396	87	123	418	97	132	22	10	9
Ukraine	469	87	118	481	95	128	12	8	10
Dubai, UAE	438	95	137	470	100	144	32	5	7
Armenia	495	89	121	462	94	132	-33	5	11
Lithuania	530	79	104	498	81	111	-32	2	7
Ontario, Canada	514	70	92	512	72	100	-2	2	8
Massachusetts, US	578	71	95	563	73	97	-15	2	2
Singapore	596	88	115	607	90	122	11	2	7
United States	532	77	104	511	78	106	-21	1	2
Australia	519	87	115	509	88	117	-10	1	2
England	542	88	116	505	88	122	-37	0	6
Sweden	506	68	90	482	68	92	-24	0	2
Russian Federation	540	85	111	539	83	113	-1	-2	2
Hungary	511	93	121	508	90	119	-3	-3	-2
Slovenia	504	74	96	507	71	99	3	-3	3
New Zealand	493	90	125	496	87	124	3	-3	-1
Kazakhstan	545	85	117	488	82	117	-57	-3	0
Italy	514	78	100	504	74	100	-10	-4	0
Alberta, Canada	510	68	89	506	64	86	-4	-4	-3
Quebec, Canada	524	67	91	532	62	84	8	-5	-7
Minnesota, US	557	80	107	545	74	102	-12	-6	-5
Norway	477	76	104	473	65	88	-4	-11	-16
Morocco	343	96	132	371	84	115	28	-12	-17
Tunisia	319	113	166	433	75	102	114	-38	-64