

RESEARCH

Open Access



Equity in mathematics education in Hong Kong: evidence from TIMSS 2011 to 2019

Xue-Lan Qiu and Frederick K. S. Leung* 

*Correspondence:
frederickleung@hku.hk

Faculty of Education, The
University of Hong Kong, Hong
Kong, SAR, China

Abstract

This study investigated the status of equity in mathematics education in Hong Kong based on data from TIMSS 2011 to 2019 and the changes in education equity across the three cycles. The effects of various student-, family-, and school-level variables on students' mathematics achievement were examined and compared using multilevel modeling. Results showed that there were significant variances between schools in mathematics achievement in Hong Kong, especially for eighth grade students. Overall, boys performed better in mathematics than girls in these three TIMSS cycles in Hong Kong, except for eighth grade boys in the cycles of 2011 and 2019. The family and school socio-economic status (SES) related to the possession of traditional materials (e.g., desks and books) and parents' highest education level had positive and significant effects on students' mathematics achievement, and the effects of the predictors increased from the 2011 cycle to the 2019 cycle. SES related to the possession of home digital devices (e.g., computers and internet connection) also had positive but much less substantial effects. School location and resource availability did not have significant effects. In general, the mean levels of SES and school resource availability in Hong Kong is improving. However, the effects of some important student- and school-level factors on mathematics achievement increased. Suggestions for improving education equity in Hong Kong are discussed.

Keywords: Education equity, Mathematics education, TIMSS, International large-scale assessment

Introduction

This paper utilizes equity in mathematics education as a lens through which to study the issue of education equity in Hong Kong. According to the Organization for Economic Co-operation and Development (OECD), equity in education refers to “fairness” and “inclusion,” where “fairness” means that personal and social circumstances, such as gender, socio-economic status (SES), ethnic origin, or family background, should not be an obstacle to achieving educational potential, and “inclusion” indicates the assurance of a basic minimum standard of education for all (OECD, 2008). Similarly, Scott (2001) defined equity as “the ways in which systems and individuals habitually operate to ensure that every learner—in whatever learning environment that learner is found—has the greatest opportunity to learn enhanced by the resources and supports necessary

to achieve competence, excellence, independence, responsibility, and self-sufficiency for schools and for life” (p. 6).

Education equity is a key and fundamental component of an equitable society (OECD, 2008). Equity in mathematics education has received increased attention in recent years, partly because of the widely accepted goal of “mathematics for all” (Artiles et al., 2011; Clements et al., 2012). Past studies have suggested that the diversity of students’ backgrounds may affect students’ achievement, particularly in subjects such as mathematics. Such background variables include gender (Fennema et al., 1998; McGraw et al., 2006), SES (Lee & Burkam, 2002; Sirin, 2005; White et al., 1993), ethnicity (Chen & Stevenson, 1995; Cocking & Chipman, 1988; Wang & Goldschmidt, 1999), and the type of schools students attend (Pahlke et al., 2014). For example, Stoet and Geary (2015) found that for overall achievement across reading, mathematics, and science literacy in the Programme for International Student Assessment (PISA), girls outperformed boys in most countries, including many where there are considerable gaps in the political or economic gender equity indices. The results raise doubts about the relationship between national equity policies and mathematics achievement.

Large-scale international comparative studies provide ideal data sets for examining important educational issues, including the issue of equity in education (Leung, 2014; Leung et al., 2005). Based on the results of these studies, policymakers often either examine the recent development of mathematics education in their own countries or initiate educational innovations to improve students’ mathematics achievement (Anderson et al., 2007; Rautalin & Alasuutari, 2009). Among these international comparative studies, Trends in International Mathematics and Science Study (TIMSS) has been providing participating countries with data for the past two decades as evidence for decision-making to improve educational policies. In contrast to studies such as PISA that assess students’ “literacy” drawing on age-based samples, TIMSS draws on grade-based samples and is curriculum-centered, policy-relevant information about the home, school, and classroom contexts for learning. TIMSS results are used by governments and ministries in multiple ways, such as to measure the effectiveness of their educational systems in a global context (Kyriakides, 2006) and to identify gaps in learning resources and opportunities (Kang & Hong, 2008). Because TIMSS adopts a rigorous methodology in data collection, working on TIMSS data relieves researchers of the need to attend to complex sampling and technical issues. It allows them to analyze a rich data set to contribute to our understanding of student learning and yield significant insights into the learning process and situation in which learning happens by identifying the various links between background factors and learning outcomes (Beatty et al., 1999; Chow & Kennedy, 2014).

Studies have examined the relationships between various background variables and students’ achievement using TIMSS data (e.g., Crane, 1996). Student-level variables usually include characteristics of students and their families. For example, it has been found that socio-economic background has a substantial influence on students’ mathematics achievement (Broer et al., 2019; Yang, 2003). Zhu and Leung (2012) studied the mathematics performance of immigrant children in TIMSS compared to native students in Hong Kong.

Home-level factors are an important source of inequity. In TIMSS 2011, it was observed that students’ early home experiences and preschool experiences in learning

mathematics seemed to be crucial in their later development (e.g., Mullis & Martin, 2013). TIMSS 2015 also included a survey of fourth grade students' parents or caregivers to collect information about home resources for learning, language(s) spoken in the home, parental educational expectations, and academic socialization, as well as early literacy, numeracy, and science activities (Mullis & Martin, 2013). Such an evolution of data contributes to the improvement of the methodology of our secondary analyses by allowing us to triangulate the interpretation of student-level variables (e.g., SES of students) with these home-level data, in addition to providing more information to explore family factors. The literature has suggested the significant roles of students' families, family-school relationships, and parental involvement in promoting student achievement (e.g., Dearing et al., 2006; Fan & Chen, 2001; Hill & Taylor, 2004). However, explicit evidence is still lacking on which family factors specifically contribute to student achievement and how they might reflect education equity/inequity.

School-level variables that are believed to be a potential source of education inequity often involve school characteristics, such as resources and school climate. For instance, Hosenfeld et al. (1999) investigated several school-level factors, such as the types of schools, with their results confirming the impact of school-level factors on student achievement. Similar findings were recorded by Leung (2005).

The studies discussed above have suggested potential sources of education inequity, but they have not addressed deeper questions that pertain to education equity in Hong Kong in particular. Hong Kong has performed very well in international studies of mathematics achievement, such as TIMSS and PISA. However, beyond attaining a high level of achievement, it is imperative for Hong Kong to aspire to ensuring education equity. If certain subgroups of Hong Kong students (for example, boys or students of high SES) can excel in these international studies of academic achievement, there is no reason to believe that other subgroups (for example, girls or students of low SES) cannot. Therefore, it is important to examine whether "personal and social circumstances such as gender, socio-economic status, ethnic origin or family background" have become "an obstacle to achieving educational potential" (OECD, 2008, p. 9).

The topic of whether Hong Kong students of different backgrounds perform equally well has rarely been studied. For the few studies that have considered this issue (mainly regarding gender differences), the findings are inconclusive. Lee and Manzon (2014) argued that Hong Kong students' high average scores in PISA illustrated a high quality of education that benefited the whole population, regardless of the socio-economic conditions of the students. This quality of education might be attributed to both the cultural habitus and structural contexts, albeit contested, of policies in educational equity and quality in Hong Kong. Yip et al. (2004) found that Hong Kong boys' and girls' science scores in PISA did not differ overall. However, boys scored higher than girls at the higher percentiles, and they also tended to score higher on tests with more earth and physical science items, understanding of scientific knowledge items, and closed items. In contrast, girls tended to score higher on "recognizing questions" and "identifying evidence" items. Hence, there are hints that gender differences in academic achievement exist in Hong Kong, and the TIMSS 2015 results show that the male advantage of boys at fourth grade may be escalating.

Other than gender differences, background variables in Hong Kong have rarely been analyzed utilizing data from international studies of mathematics achievement. The international reports of these studies do include information on the relationships between these variables and student achievement. For example, in the TIMSS report, the percentages of students in different categories of a certain background variable (e.g., home resources), together with the average achievement scores of students in each category of the variable are reported for each country to describe a relationship between the variable and student achievement. However, because the relationships are not established through comprehensive data analyses performed at the individual student level, the results are not sensitive enough to yield a reliable relationship between background variables and achievement. Therefore, the studies in TIMSS report failed to provide an accurate picture of education equity in Hong Kong, and any recommended changes in educational practices and policies based on such rough relationships remain precarious. Student SES, an important indicator of education equity, offers one example. In TIMSS 2011, it was found that the SES of students was significantly related to students' mathematics achievement (Mullis et al., 2012). Has this relationship changed over the past 10 years? And if it has, has education equity in this respect improved or worsened in Hong Kong, and what are the reasons behind this change? For such questions, studies are still lacking.

To fill the research gaps, this study aims to carry out comprehensive and in-depth analyses of the contextual factors that contribute to students' mathematics achievement in Hong Kong. We focused on the three most recent cycles of TIMSS assessment—the 2011, 2015, and 2019 cycles—because they cover a period of nearly 10 years, which predates Hong Kong's current social and education (in)equality. The following six research questions (RQs) are addressed in this study:

- (1) What is the relationship between gender and mathematics achievement in Hong Kong? How has this relationship changed from 2011 to 2019?
- (2) What is the relationship between SES and mathematics achievement in Hong Kong? How has this relationship changed from 2011 to 2019?
- (3) Does school SES affect students' achievement above and beyond the role of family SES? In particular, are low SES students in double jeopardy if they come from low SES backgrounds and attend low SES schools?
- (4) What is the students' mathematics achievement gap between urban and rural areas? How has that gap changed from 2011 to 2019?
- (5) What is the students' mathematics achievement gap between schools with different resources for mathematics instruction? How has that gap changed from 2011 to 2019?
- (6) How is the students' mathematics achievement gap explained by student and school factors?

This study will make important and unique contributions to research that uses TIMSS for trend analyses. First, constructing a consistent measure of SES across three cycles and two grades is an important contribution of this study. Second, the results, obtained from comprehensive and elaborate data analysis in which student and school factors

Table 1 School distribution of Hong Kong in TIMSS 2011, 2015, and 2019

	2011		2015		2019	
	G4	G8	G4	G8	G4	G8
Schools	136	117	132	133	138	136
Sample size	3464	3250	3315	3234	2669	3342

were considered simultaneously, will help researchers better understand the factors that contribute to student learning and achievement. They will also reveal the status of education equity in the Hong Kong system and the status changes in the years between TIMSS 2011 and 2019. On the basis of such analyses, important information on the types of inequity that existed—and likely still exist—will be identified, and suggestions will be made for policymakers on how to improve education equity in Hong Kong.

Method

Data

This study focuses on the Hong Kong component in the three most recent cycles of TIMSS (TIMSS 2011, TIMSS 2015, and TIMSS 2019). Details of the school distribution are given in Table 1, where invalid cases in which all home background variables were missing were excluded. The number of schools was all greater than 100 for each year and grade, and the sample sizes were approximately or greater than 3000, which is considered sufficient for multilevel analysis.

Variables

The variables in this study were derived from the open-access data sets of TIMSS 2011 to 2019. The dependent variables are five sets of plausible values (PVs) of mathematics achievement, the scores of which were obtained through the methodology of multiple imputation (von Davier et al., 2009) by calibrating responses to mathematics items in TIMSS using the item response theory (IRT) approach. The scores were scaled to have a mean of 500 and a standard deviation of 100 in TIMSS 1995, and they are comparable across TIMSS cycles. Within the TIMSS mathematics framework, the assessment covers three cognitive domains (Knowing, Applying, and Reasoning) and various content domains (for fourth grade, Numbers, Geometric Shapes and Measures, and Data Display; and for eighth grade, Numbers, Algebra, Geometry, and Data and Chance).

The explanatory variables are summarized in Table 2. Due to space constraints, more detailed descriptions of the variables are provided in Additional file 1: Appendix A. At the student/family level, the explanatory variables are those factors from the student and home questionnaires related to education equity. Based on the literature review, the major variables include student gender and family SES factors, which were measured by parents' highest education level (PARED), and home resources to support learning, which comprised five questions on number of books (BOOK), computer for study (COMPUTER), study desk (DESK), own room (ROOM), and internet connection (INTERNET) in the home. The six variables were selected to measure family SES for two reasons: (1) they were found to be important indicators of family SES in previous studies (e.g., Crane, 1996; Sirin, 2005), and (2) they were common variables across cycles

Table 2 List of explanatory variables in the study

Level 1: Student- and family-related variables	Level 2: School-related variables
(1) Gender	(1) School SES
(2) Family SES	(2) School location
Parents' highest education level	(3) Resources for mathematics instruction
Number of books	General school resources
Computer for study at home	Resources for mathematics instruction
Study desk	
Own room	
Internet connection at home	

and grades. Among the variables, both PARED and COMPUTER had five categories, whereas the other four variables were binary. The variables were recoded if necessary such that the higher the scores, the higher the SES level.

At the school/classroom level, three explanatory variables—school SES, school location, and school resources for mathematics instruction—were considered. The school SES was computed based on the family SES, as shown below. Rather than using the original five-category variable, the school location was recoded into a binary variable with 1 for urban and 0 for otherwise, to simplify the variable's meaning.

School resources for mathematics instruction were measured using two sets of questions in the school questionnaires. The first set comprised items about the school's general resources, such as infrastructure and supporting staff, and the second set comprised items concerning resources particularly for mathematics instruction and mathematics teachers. An example item of general school resources concerning instructional material is as follows: "How much the school's capacity to provide instruction is affected by a shortage or inadequacy of instructional material." The four Likert response categories were "not at all," "a little," "some," and "a lot." Because the questions were about the shortage or inadequacy of the resources, the scores were reversely recoded such that higher scores reflect greater availability of school resources for mathematics instruction.

Analysis

The analysis for the study requires a separate file at each grade in each of the three cycles—that is, fourth grade in TIMSS 2011 (T11G4), eighth grade in TIMSS 2011 (T11G8), fourth grade in TIMSS 2015 (T15G4), eighth grade in TIMSS 2015 (T15G8), fourth grade in TIMSS 2019 (T19G4), and eighth grade in TIMSS 2019 (T19G8). To build the data files, students' PVs of mathematics scores, gender, and home background variables were merged into a data file at the student level, and school background variables were merged into a data file at the school level. The student ID and school ID variables were used to link the student-level and school-level data files. As mentioned previously, only cases without missing data in the home background variables were included in the analysis. Sampling weights were not used in the analysis.

To answer the RQs of this study, multilevel modeling was primarily used. However, the data set comprised dozens of variables with different item contents and varying numbers of categories. For example, the six variables for family SES measured

different aspects of a family's possession (e.g., books, desks for learning, own room, and computers), among which some were binary and others had multiple categories. As such, the original responses of the variables could not be used directly in the multilevel models as predictors because there were too many categories and, more importantly, their scales were different. For this reason, before conducting the multilevel modeling, several analyses, as described below, were conducted to transform the explanatory variables into scores that are low-dimensional and in a common scale.

The first step was the descriptive analysis, where the mean and standard deviation of the dependent variables and the percentage distribution of categorical explanatory variables were computed. This step provided a rough, general understanding of the data set. Second, as mentioned earlier, because of the complexities of the variables involved in family SES, principal component analysis (PCA) was used to reduce the dimensions of the data. PCA is commonly used for dimensionality reduction by projecting the original variables (usually many in number) to only a few principal components while keeping as much variation in the data as possible. A common scale for SES across grades and cycles was also established by combining the responses to six variables in the three cycles and two grades when conducting the PCA. Such concurrent analysis gauged the students' SES on the same scale and yielded a consistent measure (component scores) across different cycles and grades, which could be compared directly. The free R package *prcomp* was used to conduct the PCA. Based on the number of components that were chosen, the component scores of family SES for different cycles and grades were added to the multilevel models as predictors, and their effects were examined.

School SES was computed by averaging the family SES of students in the same school (van Ewijk & Sleegers, 2010). This measure was considered because of the concern of so-called double jeopardy for students who come from low SES backgrounds and attend low SES schools. What especially concerned us was two students with equivalent family SES levels, who attended schools of different SES, and whether the one attending the higher SES school would have higher mathematics scores than the one attending the lower SES school.

The third step of the analysis consisted of conducting IRT analysis for school resources by calibrating the six data sets concurrently to obtain estimates that represent school resource availability. In this step, Mokken scale analysis (Mokken, 1971; Sijtsma & Molenaar, 2002) was first used to check the assumption of unidimensionality of the scale, followed by an analysis with a specific unidimensional IRT model to determine if the assumption was met. In Mokken scale analysis, unidimensionality checking can be carried out using Loevinger's coefficient for the scale H (Sijtsma & Molenaar, 2002). The coefficient takes values between 0 and 1, and as a rule of thumb, scales where H exceeds 0.5 are considered unidimensional and strong in a scaling sense. With respect to the IRT analysis, the partial credit model (PCM; Masters, 1982) was used because the questions were polytomous. In PCM, each item is measured by a parameter representing the overall difficulty of the item and by $(J - 1)$ threshold parameters, where J is the number of categories. In this analysis, J equals four because the questions have four categories. The R packages *Mokken* and *TAM* were used for the Mokken scale analysis and IRT analysis, respectively.

The fourth and most important step of the analysis was to examine the effects of student- and school-level factors on mathematics achievement using multilevel modeling (Goldstein, 2010). In large-scale educational assessments, such as TIMSS, two-stage or multistage sampling is often used. Usually, a certain number of schools is sampled first, and a number of students is then selected from each sampled school. This two-stage sampling creates a two-level hierarchical data structure: student level and school level. Because of school characteristics, students in the same school tend to be more homogeneous in the outcome of variables of interest, particularly variables associated with student performance, than students from different schools (Goldstein, 2010). Ignoring multilevel structures will result in undesirable consequences that have been well documented (Goldstein, 2010). For example, the fixed-effect parameters (e.g., the coefficients of predictors), in general, are not biased. However, the variance of the ignored level (e.g., school level) is redistributed to the adjacent levels (e.g., student level), which results in inaccurate estimates of variance–covariance at the student level. More importantly, the standard errors of predictors at the student level will be underestimated if multilevel structures are ignored. The underestimation of standard errors may lead to serious consequences when an important decision or practice is overturned due to the size of a standard error (Goldstein, 2010). Because this study focuses on mathematics achievement, which is a continuous variable, the multilevel models for continuous data (Goldstein, 2010) will be used with the *Mplus* software package (Muthén & Muthén, 1998–2017). Moreover, because the dependent variables were five sets of PVs for mathematics achievement that are randomly drawn from the posterior distribution of students' ability (Mislevy et al., 1992), the option TYPE = IMPUTATION was implemented when running multilevel models in *Mplus*. Therefore, the resultant regression coefficients were pooled as the mean across all five sets of regression coefficients, and their variances were quantified by considering the variances within and between PVs.

Following the practice in multilevel analysis, the null model, which has no explicit predictors (Model 0), was first fitted to the six data sets (T11G4, T11G8, T15G4, T15G8, T19G4, and T19G8) separately. With this model, the variance of the dependent variables (i.e., mathematics achievement) is decomposed into two components—student-level variance and school-level variance. The intraclass correlation coefficient (ICC), defined as the ratio of school-level variance over total variance (i.e., the sum of the student- and school-level variances), and the design effect, defined as $1 + (\bar{N} - 1) \times ICC$, where \bar{N} is the mean sample size across schools (Hox, 2010), were computed. In general, the magnitude of the design effect suggests whether or not there is a necessity of multilevel analysis, with a value greater than 2 indicating the necessity of a multilevel analysis and a value around 2 or less indicating the sufficiency of a single-level analysis.

To answer the RQs, six two-level models with different Level-1 (student-level) and/or Level-2 (school-level) predictor(s) were established and estimated. Specifically, to answer the first RQ, students' gender was included explicitly in Level 1 of the multilevel model as the predictor (Model 1). Similarly, for the second RQ, the PC scores of family SES were included as Level-1 predictors (Model 2). To answer the third RQ, the PC scores of family SES were included in Level 1, and those of school SES were included in Level 2 as predictors (Model 3). As such, this model partitions the effects of SES into the contributions of student- and school-level predictors.

Table 3 Descriptive analysis of mathematics achievement and school resources availability for TIMSS 2011 to 2019

	2011				2015				2019			
	G4		G8		G4		G8		G4		G8	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
PV1	607.1	61.4	592.9	78.0	616.0	64.0	596.1	75.7	602.5	67.9	577.2	86.7
PV2	609.4	60.5	594.9	79.8	616.6	64.2	596.2	76.5	604.7	69.6	578.3	89.0
PV3	608.0	60.3	593.8	81.4	616.5	62.7	597.7	76.7	604.7	67.8	578.5	88.1
PV4	608.4	60.6	594.7	81.1	616.4	63.5	596.7	78.8	604.5	68.1	577.3	88.0
PV5	607.8	60.8	594.0	81.1	616.6	64.6	597.5	77.3	603.7	68.0	579.2	89.4
School resources	1.1	0.8	2.2	1.0	1.9	1.0	2.2	0.9	2.1	1.0	1.9	1.1

PV1 to PV5: the first to the fifth set of plausible values of mathematics achievement; the school resources were computed based on the raw data

To answer the fourth and fifth RQs, school location (Model 4) and school resource availability estimates (Model 5) were included as Level-2 predictors, respectively. Finally, for the sixth RQ, a more complicated model that included family SES as the Level-1 predictor and school SES, school location, and school resource availability estimates as Level-2 predictors was fitted (Model 6). Note that Model 6 focuses on student and school factors that may be partly controllable. Hence, gender was not considered. To illustrate, the *Mplus* code for running Model 6 is provided in Additional file 1: Appendix B.

Results

This section presents the results of the descriptive analysis, PCA for family SES, IRT analysis for the school resources scale, and multilevel modeling. Due to space constraints, we will focus on the results of multilevel modeling. Detailed results of the PCA and IRT analysis are provided in Appendices 1 and 2, respectively.

Descriptive analysis

Table 3 shows the mean and standard deviation of the five sets of PVs for mathematics achievement and school resource availability. In general, Hong Kong students' mathematics scores in the 2015 cycle were the highest, followed by those in the 2011 cycle and the 2019 cycle, and fourth grade students' scores were higher than those of eighth grade students. In terms of the resource availability for mathematics instruction, there was a stable increase from the 2011 cycle to the 2019 cycle for fourth grade. Note that the available resources for fourth grade were less than those for eighth grade for the cycles of 2011 and 2015. However, they were nearly equivalent to those of eighth grade in the 2019 cycle as a result of the aforementioned increase in resource availability for mathematics instruction.

Table 4 shows the percentage of categorical variables for TIMSS 2011 to 2019. The number of girls and boys in the sample was nearly equal. The parents of students in TIMSS 2011 fourth grade seemed to have the lowest education level, with only about 17.1% having received university or higher education. For the home possession of books, students in TIMSS 2019 eighth grade seemed to have the lowest amount, whereas for the

Table 4 Percentage of categorical variables (%) for TIMSS 2011 to 2019

		G4					G8				
		1	2	3	4	5	1	2	3	4	5
2011	Gender	46.8	53.2				49.7	50.3			
	PARED	11.1	23.8	39.8	8.1	17.1	7.9	20.7	45.2	6.1	20.1
	BOOK	13.9	24.7	36.4	14.9	10.2	18.5	24.9	31.8	13.0	11.8
	DESK	26.2	73.8				21.8	78.2			
	ROOM	40.2	59.8				41.3	58.7			
	COMPUTER	1.4	98.6				0.6	99.4			
	INTERNET	7.2	92.8				2.3	97.7			
	Urban [†]	46.9	47.8				45.3	47.1			
2015	Gender	44.9	55.1				47.5	52.5			
	PARED	7.3	11.8	32.2	16.9	31.8	5.4	14.3	43.2	14.0	23.1
	BOOK	12.8	19.9	33.8	18.5	14.9	17.5	26.7	31.5	12.8	11.5
	DESK	29.6	70.4				24.6	75.4			
	ROOM	42.4	57.6				40.8	59.2			
	COMPUTER [#]	4.9	95.1				2.3	97.7			
	INTERNET	26.9	73.1				4.1	95.9			
	Urban [†]	40.1	59.1				48.6	45.8			
2019	Gender	46.6	53.4				45.6	54.4			
	PARED	4.9	9.7	26.7	19.1	39.6	3.5	10.8	40.3	13.8	31.6
	BOOK	15.4	19.7	30.0	18.6	17.3	19.3	26.0	28.3	13.4	13.0
	DESK	16.7	83.3				12.4	87.6			
	ROOM	38.0	62.0				34.5	65.5			
	COMPUTER	4.3	95.7				4.1	95.9			
	INTERNET	18.7	81.3				2.7	97.3			
	Urban [†]	46.0	53.2				41.9	57.4			

PARED: Parents' highest education level; BOOK: Number of books in the home; DESK: Own study desk in the home; ROOM: Has own room in the home; COMPUTER: Own computer for study in the home; INTERNET: Has internet connection in the home. Descriptions of categories of variables can be found in Additional file 1: Appendix A

[#] There were two variables measuring the possession of computers for study in the home in the 2015 cycle, one for own computers and the other for shared computers, whereas there was only one single variable in the cycles of 2011 and 2019. To be consistent across cycles, the two variables in the 2015 cycle were recoded into one variable with 1 = possess own or shared computers, and 0 = possess neither computers

[†] The sum of percentages for the urban variable was not equal to 100% because of missing data

home possession of study desks, own rooms, own computers, and internet connection, students in TIMSS 2015 fourth grade seemed to have the lowest percentages. The results showed that Hong Kong families could usually afford digital devices, such as computers and an internet connection, to support their children's learning. The number of schools from urban and nonurban (suburban and rural) areas was roughly equal.

PCA for family SES

Two PCs were chosen for family SES and were interpreted as traditional SES and home digital devices (HDD) SES, respectively, based on the loadings of variables on the PCs. The multivariate analysis of variance (MANOVA) of component scores for the two PCs showed that overall, the mean level of SES in Hong Kong was increasing, especially for traditional SES. Detailed results from the PCA of family SES variables are provided in Appendix 1.

IRT analysis for school resources scale

The results of the Mokken scale analysis showed that the H coefficient for the school resources scale was 0.524 with a standard error of 0.028. Therefore, the scale was unidimensional. The test reliability was 0.902, which is considered high. The results from the IRT analysis suggested that, generally, sufficient material and resources for mathematics education were available to schools in Hong Kong. Detailed results are provided in Appendix 2.

Multilevel analysis with null model

The ICCs for the null model without any predictors (Model 0) for T11G4, T11G8, T15G4, T15G8, T19G4, and T19G8 were 25.2%, 62.7%, 29.9%, 56.5%, 23.2%, and 49.3%, respectively, and the corresponding design effects were 8.08, 17.79, 8.86, 14.17, 7.10, and 12.62. Because the design effects far exceed the criterion of 2 (Hox, 2010), the results indicate that there were significant between-school variances or school effects, especially for eighth grade students. Therefore, a multilevel analysis was necessary for this study.

Below, we present the results of the study with reference to the RQs.

RQ1: Relationship between gender and mathematics achievement in Hong Kong and changes from 2011 to 2019

As shown in Model 1 in Table 5, the coefficients for gender (male = 1) are positive and significant, suggesting that, in general, boys in Hong Kong performed better in mathematics than girls. Also, the coefficients for the 2019 cycle are relatively smaller than those of the 2015 cycle for both grades, and the coefficients for the 2015 cycle are in turn smaller than those of the 2011 cycle, suggesting that the gaps in mathematics achievement between boys and girls reduced slightly from 2011 to 2019.

RQ2: Relationship between family SES and mathematics achievement in Hong Kong and changes from 2011 to 2019

As presented in Table 5, regarding traditional SES, the coefficients of the predictor on mathematics achievement for fourth grade students are 1.49 ($p > 0.05$), 4.95 ($p < 0.00$), and 7.22 ($p < 0.00$) in the cycles of 2011, 2015, and 2019, respectively, whereas for eighth grade students, the corresponding coefficients are 0.25 ($p > 0.05$), 0.78 ($p > 0.05$), and 1.30 ($p > 0.05$), respectively. Therefore, traditional SES seems to have more positive and significant effects on the mathematics achievements of fourth grade students than on eighth grade students. In addition, the effects for fourth grade students increased from 2011 to 2019. For example, for a one-unit increase in traditional SES, students' mathematics scores were predicted to be higher by 1.49 points, 4.95 points, and 7.22 points in the three cycles, respectively.

In terms of HDD SES, the coefficients of the predictor on mathematics achievement were not significant, except for fourth grade students in the 2015 cycle, where, though significant, the magnitude of the coefficient was relatively small (1.54, $p < 0.05$). The

Table 5 Multilevel analysis of student-, family-, and school-level variables with mathematics achievement

Model	Predictor	2011			2015			2019					
		G4			G4			G4					
		Est	SE	G8	Est	SE	G8	Est	SE	G8			
1	Gender	12.34**	2.01	20.67**	1.93	11.19**	2.01	16.43**	2.28	9.20**	2.47	12.02**	2.64
2	Traditional SES (Family)	1.49	1.20	0.25	1.06	4.95**	1.03	0.78	1.04	7.22**	1.35	1.30	1.27
	HDD SES (Family)	0.72	1.45	0.44	2.09	1.54*	0.69	2.54	1.32	1.62	1.07	0.90	1.68
3	Traditional SES (Family)	0.82	1.21	0.72	1.05	3.66**	1.02	0.24	1.00	5.35**	1.34	0.24	1.26
	HDD SES (Family)	0.41	1.47	0.05	2.08	1.81*	0.70	2.48	1.31	2.01	1.08	0.98	1.68
	Traditional SES (School)	20.16**	4.68	62.22**	17.98	29.88**	3.09	54.92**	12.30	29.15**	3.55	60.10**	6.29
	HDD SES (School)	0.88	8.87	26.18	43.38	12.97*	5.22	45.06	31.83	2.27	5.70	6.92	13.40
4	School Location	2.70	7.14	18.26	13.43	16.13**	6.29	9.45	11.06	21.51**	6.12	25.37*	11.99
5	School Resources	4.02	5.37	6.08	3.49	7.30**	1.97	4.18	3.06	2.20	2.09	4.80	3.28
6	Traditional SES (Family)	0.85	1.21	0.18	1.07	3.66**	1.03	0.36	1.04	5.45**	1.36	0.52	1.27
	HDD SES (Family)	0.67	1.46	0.03	2.12	1.86**	0.71	2.31	1.34	2.05	1.08	1.10	1.69
	Traditional SES (School)	19.33**	4.80	60.82**	18.27	30.23**	3.65	54.61**	13.31	27.74**	3.64	58.77**	6.26
	HDD SES (School)	1.26	9.93	26.79	45.05	13.40*	5.25	44.37	33.47	2.45	5.59	7.42	13.67
	School Location	2.58	5.96	10.22	9.91	3.12	5.06	3.73	10.27	8.98	4.89	12.04	8.49
	School Resources	4.43	3.95	1.94	2.67	1.38	1.42	1.74	2.61	2.21	1.21	0.98	1.91

* *p*-values are smaller than 0.05; ** *p*-values are smaller than 0.01. The option TYPE = IMPUTATION was implemented when running the six models in the Mplus

results suggest that traditional SES is a more important predictor of mathematics achievement than HDD SES, especially for fourth grade students.

RQ3: Effects of school SES on mathematics achievement above and beyond the role of family SES

This RQ aimed to investigate the effects of school SES on mathematics achievement after considering the effects of family SES. The most substantial results for Model 3 in Table 5 concern traditional school SES; the coefficients of the predictor of mathematics achievement were consistently positive and significant in the three cycles and two grades. These results suggest that students from high SES families who were also attending high SES schools were predicted to have higher mathematics scores. The same relationship was found for low SES, with low school and family SES predicting lower mathematics scores, putting low SES students in double jeopardy in Hong Kong. In general, the magnitude of the predictor's coefficients increased from the 2011 cycle to the 2019 cycle, especially for the lower grade students. For example, for fourth grade students, the coefficients were 20.16, 29.88, and 29.15 for the three cycles, respectively, and they were 62.22, 54.92, and 60.10, respectively, for eighth grade students. These findings deserve further attention.

Regarding HDD school SES, except for the coefficient (12.97, $p < 0.05$) for fourth grade students in the cycle 2015, all the coefficients were not statistically significant.

RQ4/RQ5: Relationships between school location/school resources for mathematics instruction and mathematics achievement in Hong Kong and changes from 2011 to 2019

The results of Model 4 in Table 5 show that school location (urban = 1) had significant effects on fourth grade students' mathematics scores in the 2015 and 2019 cycles, suggesting that school location was likely to affect fourth grade students more than eighth grade students. The results of Model 5 show that school resources did not seem to have significant effects on students' mathematics achievement in Hong Kong because only the coefficient in T15G4 was significant.

RQ6: Relationships between student- and school-level predictors and mathematics achievement in Hong Kong and changes from 2011 to 2019

The results of Model 6 in Table 5 show the effects of student- and school-level predictors on mathematics achievement when they were simultaneously considered. Several findings deserve attention here.

First, consistent with the findings in RQ3, traditional SES predictors, particularly school SES, had positive and significant effects on mathematics achievement for students across cycles and grades; in contrast, HDD SES predictors had relatively minor effects on students' mathematics achievement.

Second, as mentioned earlier, the effects of traditional SES on students' mathematics achievement seemed to have increased from cycle 2011 to cycle 2019, especially for the effects of traditional school SES on fourth grade students' mathematics achievement.

Third, the coefficients of school location and school resources for mathematics instruction were not statistically significant in all six data sets, suggesting that after taking the SES predictors into account, school location and school resources are no longer important for predicting students' mathematics achievement.

Conclusions and discussion

How well has Hong Kong been addressing the issue of equity?

Hong Kong has been performing very well in international studies of academic achievement, such as TIMSS and PISA, especially in the subject of mathematics. From the background information collected in these studies, it is evident that Hong Kong is a relatively affluent society, but there seems to be polarization in terms of the family SES of students. This wealth disparity has a significant impact on student achievement. The results of this study show that family SES makes a significant difference in students' mathematics achievement in TIMSS (see Table 5), and the effect of family SES on mathematics achievement has increased from 2011 to 2019 for both grades, where the effects on fourth grade students are particularly substantial. Since fourth grade is the younger of the two cohorts, the findings may indicate that Hong Kong is moving toward more inequity in the influence of family SES on students' mathematics achievements. More research needs to be conducted on subsequent cycles of TIMSS to confirm or refute this trend.

There are also substantial disparities in terms of mathematics achievement among students in other areas of student background. Overall, boys in Hong Kong performed better than girls in mathematics, but it is interesting to find that the gender gap has seemed to narrow over the years. Further studies on students' performance in subsequent cycles of these international studies should be conducted to track this trend in the gender gap.

Regarding school-related variables, our study shows that there were also strong effects of school SES on students' mathematics achievement (see Table 5). The other two school-related variables, school location (urban versus suburban and rural) and school resources for mathematics instruction, were found not to have significant effects on students' mathematics achievement in Hong Kong. This finding seems to suggest equity among different kinds of schools in Hong Kong. Moreover, our in-depth analysis (see Model 6 in Table 5) shows that after taking the SES predictors into account, school location and school resources were no longer important for predicting students' mathematics achievement. In other words, SES had such a strong influence on student achievement that school-related variables, such as school location and school resources for mathematics instruction, mattered much less.

Suggestions for improving education equity in Hong Kong

Although the influence of SES on student achievement is a more or less universal phenomenon, our study suggests that Hong Kong is no better off than many other systems in the world. What can the various stakeholders in education do to alleviate the impact of SES on student achievement?

Hong Kong is a capitalistic society, and inequity, including inequity in education provisions, seems to be tolerated by many people. To cite just one example, the Direct Subsidy Scheme (DSS) introduced in Hong Kong 30 years ago is intended to tap into private resources in funding education and provide more alternatives in educational provisions for parents and students to choose from. Under the DSS scheme, schools are allowed to collect fees on top of the government subsidy so that they will be better resourced to provide higher-quality education. At the same time, DSS schools are given slightly more autonomy in determining their school curricula. The government requires that a certain

proportion of the fees collected should be used as scholarships for needy students, but the net result of the scheme is that DSS schools admit students from wealthier families, and their students perform better academically. Given the inequity in terms of mathematics achievement in TIMSS, as shown by this study, further research is needed to investigate the effects of initiatives such as DSS on both student achievement and possible inequity of achievement. The Hong Kong government should adopt an evidence-based approach to review the effects of the DSS and other similar schemes to strike a balance between choice and equity in educational provision. In particular, while tapping into resources from the private sector to fund education is a right move toward equity, more measures, in terms of government funding toward education, should be put in place to ensure that students from low SES families and low SES schools receive more resources and support in learning.

One specific way to improve education equity in Hong Kong is to reduce the double jeopardy for students who come from low SES families and attend low SES schools. As mentioned above, our study shows that traditional school SES has a significant impact on students' mathematics achievement in TIMSS beyond the family SES. The need to narrow the disparities between schools is imperative in bringing about a more equitable society in relation to children's learning. The government, with the support of certain nongovernmental organizations, has already begun some work on this, but the results of our study show that this is still inadequate in addressing the disparity in learning opportunities among Hong Kong children.

Conclusion

Hong Kong is an affluent society, and Hong Kong students have been performing very well in international studies of educational achievement, such as TIMSS and PISA. However, Hong Kong society seems to benefit the "haves" much more than the "have-nots." While this phenomenon exists in many places around the world, it is not something that Hong Kong people should assume is normal or acceptable. We should strive to alleviate the influence of SES on student achievement and make Hong Kong an even more equitable society.

Appendix 1: Principal component analysis of family SES factors

As mentioned above, because family SES factor is measured by six variables and these variables have different number of categories, the principal component analysis (PCA) was used to reduce the dimensions of the data. Meanwhile, because the focus of this study was to investigate the equity in mathematics education in Hong Kong across the years, to avoid confusion that may be caused by amounts of results of PCA, the detailed PAC results are shown in this Appendix.

Figure 1 shows the scree plot (left) which presents the decreasing rate at which variance is explained by additional principal components (PCs) and the cumulative variance plot (right) which presents the proportion of variance that the components explain. These two plots help determine the number of PCs. As a good rule of thumb, PCs with eigenvalues that are greater than 1 are retained. In these results, the first two PCs have eigenvalues greater than 1 and they explain about 50% of the variation

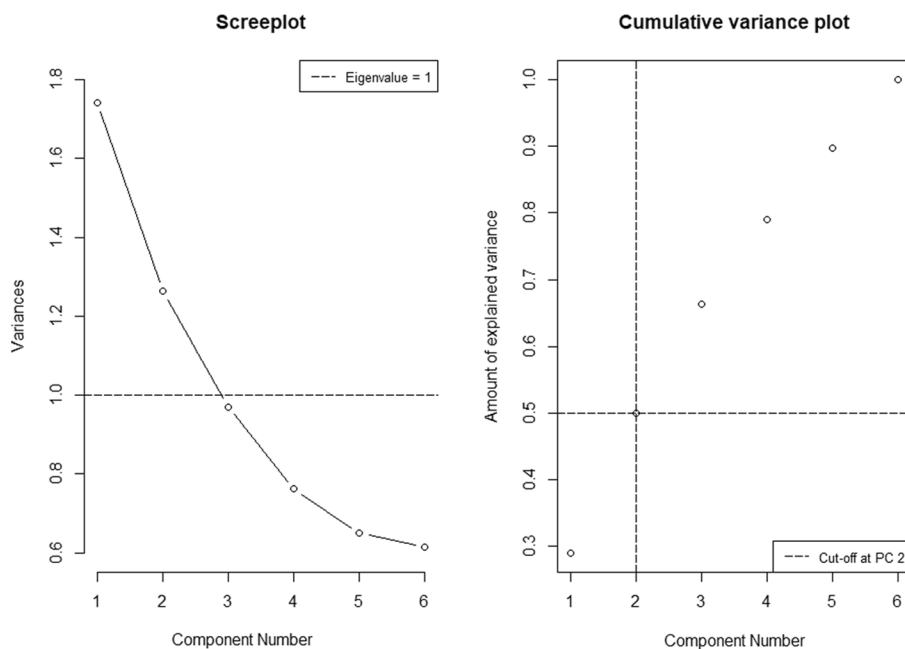


Fig. 1 Screenplot and cumulative variance plot from principal component analysis of family socio-economic status

Table 6 Factor loadings from principal components analysis of family socio-economic status

Variable	PC1	PC2	PC3	PC4	PC5	PC6
PARED	0.485 [†]	-0.335	0.369	-0.041	-0.183	-0.693
DESK	0.485 [†]	0.056	-0.492	-0.017	0.708	-0.136
BOOK	0.476 [†]	-0.287	0.445	-0.110	0.134	0.680
ROOM	0.465 [†]	0.027	-0.546	0.109	-0.661	0.190
INTERNET	0.186	0.668 [†]	0.156	-0.696	-0.097	-0.043
COMPUTER	0.227	0.596 [†]	0.319	0.700	0.034	-0.010

PARED: Parents' highest education level; DESK: Study desk in the home; ROOM: Has own room in the home; BOOK: Number of books in the home; INTERNET: Has internet connection in the home; COMPUTER: Computer for study in the home

[†] Factor loadings are greater than .400 in the first two principal components which have eigenvalue larger than 1

in the data. The amount of variation is adequate enough to explain the total variation in the data. Hence, the first two PCs are chosen.

The factor loadings of each variable on a particular PC are given in Table 6, which is useful in interpreting the meaning of the components. In principle, larger loadings (in absolute value) indicate that a particular variable has a stronger relationship to a particular PC, and the sign of a loading indicates a positive or negative relationship. Thus, the interpretation of PCs relies on the magnitude and direction of the loadings for the original variables. Inspecting Table 6, it was found that the first PC has strong and positive associations with four variables, as in, DESK (own desk for learning), ROOM (own room), PARED (parents' highest education level), and BOOK (amount of books in home). Therefore, this component primarily measures the level

Table 7 Results of multivariate analysis of variance for comparing mean of six dataset

Source	Sum of square	df	Mean square	F	p
Traditional SES	543.38	5	108.68	63.60	0.000
HDD SES	2120.81	5	424.16	394.78	0.000
Error (Traditional SES)	32,922.81	19,268	1.71		
Error (HDD SES)	20,702.06	19,268	1.07		

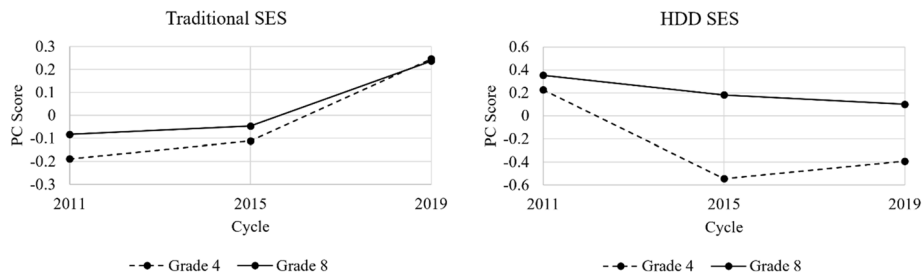


Fig. 2 Mean level of components scores of family socio-economic status across cycles and grades

of SES that are related to traditional family possession or backgrounds. The second PC has large positive associations with COMPUTER (own computer for learning) and INTERNET (internet connection in home). Hence, this component primarily measures the level of SES that are related to the possessions of home digital devices. For this reason, the two PCs were referred to as traditional SES and home digital device (HDD) SES, respectively.

Multivariate analysis of variance (MANOVA) was conducted to examine the mean differences of SES component scores across cycles and grades. Overall, there was a statistically significant difference in SES component scores, $F(10, 38,534) = 226.00$, $p < 0.000$; Wilk's $\Lambda = 0.892$, partial $\eta^2 = 0.055$. The detailed results from MANOVA are shown in Table 7. The results of post-hoc comparison also showed that mean levels for traditional SES (PC1) were statistically significantly different between all groups except between T11G4 and T15G4, between T11G8 and T15G4, between T11G8 and T15G8, and T15G4 and T15G8; mean levels for HDD SES (PC2) were statistically significantly different between all groups except between T11G4 and T15G8.

The changes can be easily visualised in Fig. 2. For the traditional SES (left panel), the mean levels of both grades increased from the 2011 cycle to the 2019 cycle. Additionally, the differences between fourth grade and eighth grade reduced. Specifically, in the 2011 cycle, the traditional SES of fourth grade students was about 0.1 unit lower than that of eighth grade students, and the difference nearly disappeared in the 2019 cycle. This information may indicate the education equity in Hong Kong increased in this regard.

In terms of HDD SES (right panel), the mean levels decreased from the 2011 cycle to the 2019 cycle, especially for fourth grade students. The levels of fourth grade students were consistently lower than that of eighth grade students across three cycles. Moreover, the difference between two grades was the smallest in the 2011 cycle and

Table 8 Item parameter estimates for school resources availability scale

Question	δ	SE	τ_1	SE	τ_2	SE	τ_3	SE
1. Instructional materials	-2.027	0.061	-8.463	0.147	2.334	0.102	2.736	0.126
2. Supplies	-1.702	0.065	-9.259	0.178	2.608	0.107	2.957	0.140
3. School building	-2.352	0.063	-7.909	0.137	1.512	0.099	2.798	0.115
4. Heating system	-1.926	0.060	-8.666	0.150	3.015	0.107	2.210	0.126
5. Instructional space	-2.448	0.063	-7.724	0.132	1.630	0.099	2.547	0.112
6. Technical staff	-2.125	0.068	-8.580	0.171	1.205	0.105	2.830	0.113
7. Computer technology	-2.306	0.067	-8.092	0.153	1.023	0.104	2.682	0.108
8. Resources for disabilities [†]	-1.947	0.104	-9.480	0.353	1.279	0.145	3.589	0.202
9. Mathematics teachers	-2.353	0.065	-7.967	0.141	1.511	0.101	2.414	0.108
10. Computer software for mathematics instruction	-2.173	0.076	-8.756	0.221	0.388	0.121	2.814	0.110
11. Library material for mathematics instruction	-1.957	0.075	-9.135	0.235	0.651	0.115	3.188	0.118
12. Audio/visual resources for mathematics instruction	-2.110	0.074	-8.824	0.210	0.517	0.115	3.096	0.113
13. Calculators	-1.668	0.069	-9.565	0.205	2.081	0.105	3.409	0.144
14. Concrete objects [†]	-2.434	0.091	-8.556	0.238	1.456	0.141	2.923	0.164

Questions 1 to 8 measure general school resources and questions 9 to 14 measure school resources for mathematics. δ : difficulty parameter; τ : Step parameter; [†]: For TIMSS 2015 and TIMSS 2019 only

the largest in the 2019 cycle, indicating that the equity worsened from the 2011 cycle to the 2015 cycle and then increased in the 2019 cycle.

Appendix 2: Item response theory model analysis of school resources availability scale

The responses to school resources availability scale were analyzed using the PCM (Masters, 1982). As shown in Table 8, the item difficulty parameters are all negative, suggesting that these items are “easy” to Hong Kong schools. In this context, the results suggest that generally, sufficient material and resources for mathematics education were available to the schools in Hong Kong.

Abbreviations

- BOOK Books at home
- COMPUTER Computers for learning
- DESK Desk for learning
- EST Estimate
- HDD Home digital device
- ICC Intraclass correlation coefficient
- INTERNET Internet connection
- IRT Item response theory
- MANOVA Multivariate analysis of variance
- OECD The Organization for Economic Co-operation and Development
- PARED Parents’ highest education level
- PC Principal components
- PCA Principal component analysis
- PCM Partial credit model
- PISA Programme for International Student Assessment
- PV Plausible values
- ROOM Own room
- RQ Research questions
- SD Standard deviation
- SE Standard errors
- SES Socio-economic status
- TIMSS Trends in International Mathematics and Science Study

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40536-022-00121-z>.

Additional file 1. Appendix A. List and descriptions of explanatory variables. Appendix B. Mplus code for multilevel analysis of student-, family-, and school-level variables with mathematics achievement.

Acknowledgements

The authors would like to thank Dr. Chen Gaowei, Dr. Mok Ida Ah Chee, and Miss Yang Yawei for their helpful comments on the draft of the manuscript. We also thank the managing editor and two anonymous reviewers for their constructive comments on earlier drafts of this article.

Author contributions

XQ developed the research methods, conducted data compilation, the statistical analysis and interpretation of results and drafted significant parts of the manuscript. FKSL developed the research questions, conducted the literature research and drafted significant parts of the manuscript. Both authors read and approved the final manuscript.

Authors' information

FKSL is the national research coordinator of Hong Kong component of TIMSS 2011, 2015 and 2019.

Funding

The Hong Kong component of TIMSS 2011, 2015 and 2019 are funded by the Education Bureau of Hong Kong (EDB), and this study is funded under the General Research Fund (GRF) of the Research Grant Council (RGC) of Hong Kong (RGC ref. no.: 17602918). But the views expressed in this paper are those of the authors and do not necessarily represent the views of EDB or RGC.

Availability of data and materials

The TIMSS data that support the findings can be found in the repository of IEA (<https://www.iea.nl/data-tools/repository>).

Declarations

Ethics approval and consent to participate

The dataset used in this study (i.e., TIMSS data) is publicly available. Since this study did not involve data collection, no ethics approval was needed.

Consent for publication

I, as the corresponding author of the submission titled 'Equity in Mathematics Education in Hong Kong: Evidence from TIMSS 2011 to 2019', hereby affirm that: (a) neither the article nor portions of it have been previously published elsewhere, (b) the manuscript is not under consideration for publication in another journal, and will not be submitted elsewhere until the LSAE editorial process is completed, and (c) all authors consent to the publication of the manuscript in LSAE, should the article be accepted by the Editor-in-chief upon completion of the refereeing process.

Competing interests

The authors declare that they have no competing interests.

Received: 15 April 2021 Accepted: 19 May 2022

Published online: 08 June 2022

References

- Anderson, J. O., Lin, H.-S., Treagust, D. F., Ross, S. P., & Yore, L. D. (2007). Using large-scale assessment datasets for research in science and mathematics education: Programme for International Student Assessment (PISA). *International Journal of Science and Mathematics Education*, 5(4), 591–614. <https://doi.org/10.1007/s10763-007-9090-y>
- Artiles, A. J., Kozleski, E. B., & Waitoller, F. R. (Eds.). (2011). *Inclusive Education: Examining equity on five continents*. Harvard Education Press.
- Beatty, A., Paine, L. W., & Ramirez, F. O. (1999). *Next steps for TIMSS: Directions for secondary analysis*. National Academies Press.
- Broer M., Bai Y., & Fonseca F. (2019). Methodology: Constructing a socio-economic index for TIMSS trend analyses. In: Socioeconomic inequality and educational outcomes. IEA Research for Education (A Series of In-depth Analyses Based on Data of the International Association for the Evaluation of Educational Achievement (IEA)), vol 5. Springer, Cham. https://doi.org/10.1007/978-3-030-11991-1_3
- Chen, C., & Stevenson, H. W. (1995). Motivation and mathematics achievement: A comparative study of Asian-American, Caucasian-American, and East Asian high school students. *Child Development*, 66(4), 1215–1234. <https://doi.org/10.2307/1131808>
- Chow, K. F., & Kennedy, K. J. (2014). Secondary analysis of large-scale assessment data: An alternative to variable-centred analysis. *Educational Research and Evaluation*, 20(6), 469–493. <https://doi.org/10.1080/13803611.2014.976831>
- Clements, M. A., Keitel, C., Bishop, A. J., Kilpatrick, J., & Leung, F. K. S. (2012). From the few to the many: Historical perspectives on who should learn mathematics. In M. Clements, A. Bishop, C. Keitel, J. Kilpatrick, & F. Leung (Eds.), *Third international handbook of mathematics education*. Springer.

- Cocking, R., & Chipman, S. (1988). Conceptual issues related to mathematics achievement of language minority children. In R. R. Cocking & J. P. Mestre (Eds.), *Linguistic and cultural influences on learning mathematics* (pp. 17–46). Hillsdale.
- Crane, J. (1996). Effects of home environment, SES, and maternal test scores on mathematics achievement. *The Journal of Educational Research*, 89(5), 305–314.
- Dearing, E., Kreider, H., Simpkins, S., & Weiss, H. B. (2006). Family involvement in school and low-income children's literacy: Longitudinal association between and within families. *Journal of Educational Psychology*, 98(4), 653–664. <https://doi.org/10.1037/0022-0663.98.4.653>
- Fan, X., & Chen, M. (2001). Parental involvement and students' academic achievement: A meta-analysis. *Educational Psychology Review*, 13(1), 1–22. <https://doi.org/10.1023/A:1009048817385>
- Fennema, E., Carpenter, T. P., Jacobs, V. R., Franke, M. L., & Levi, L. W. (1998). A longitudinal study of gender differences in young children's mathematical thinking. *Educational Researcher*, 27(5), 6–11. <https://doi.org/10.2307/1176733>
- Goldstein, H. (2010). *Multilevel statistical models* (4th ed.). Wiley.
- Hill, N. E., & Taylor, L. C. (2004). Parental school involvement and children's academic achievement: Pragmatics and issues. *Current Directions in Psychological Science*, 13(4), 161–164. <https://doi.org/10.1111/j.0963-7214.2004.00298.x>
- Hosenfeld, I., Köller, O., & Baumert, J. (1999). Why sex differences in mathematics achievement disappear in German secondary schools: A reanalysis of the German TIMSS-data. *Studies in Educational Evaluation*, 25(2), 143–161. [https://doi.org/10.1016/S0191-491X\(99\)00018-8](https://doi.org/10.1016/S0191-491X(99)00018-8)
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). Routledge.
- Kang, N.-H., & Hong, M. (2008). Achieving excellence in teacher workforce and equity in learning opportunities in South Korea. *Educational Researcher*, 37(4), 200–207. <https://doi.org/10.3102/0013189X08319571>
- Kyriakides, L. (2006). Using international comparative studies to develop the theoretical framework of educational effectiveness research: A secondary analysis of TIMSS 1999 data. *Educational Research and Evaluation*, 12(6), 513–534. <https://doi.org/10.1080/13803610600873986>
- Lee, V. E., & Burkam, D. T. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school*. Economic Policy Inst.
- Lee, W., & Manzoni, M. (2014). The issue of equity and quality of education in Hong Kong. *Asia-Pacific Education Researcher*, 23(4), 823–833. <https://doi.org/10.1007/s40299-014-0214-1>
- Leung, F. K. S. (2005). Some characteristics of East Asian mathematics classrooms based on data from the TIMSS 1999 video study. *Educational Studies in Mathematics*, 60(2), 199–215. <https://doi.org/10.1007/s10649-005-3835-8>
- Leung, F. K. S. (2014). What can and should we learn from international studies of mathematics achievement? *Mathematics Education Research Journal*, 26(3), 579–605. <https://doi.org/10.1007/s13394-013-0109-0>
- Leung, F. K. S., Yung, B. H. W., & Tso, A. S. F. (2005). Digging below the surface: Secondary analysis of TIMSS data for Hong Kong and its implications for educational reform. *Canadian Journal of Science, Mathematics and Technology Education*, 5(3), 329–360. <https://doi.org/10.1080/14926150509556665>
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174. <https://doi.org/10.1007/BF02296272>
- McGraw, R., Lubienski, S., & Strutchens, M. (2006). A closer look at gender in NAEP mathematics achievement and affect data: Intersections with achievement, race/ethnicity, and socioeconomic status. *Journal for Research in Mathematics Education*, 37(2), 129–150. <https://doi.org/10.2307/30034845>
- Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29(2), 133–161. <https://doi.org/10.1111/j.1745-3984.1992.tb00371.x>
- Mokken, R. J. (1971). *A theory and procedure of scale analysis*. De Gruyter.
- Mullis, I. V. S., & Martin, M. O. (Eds.). (2013). *TIMSS 2015 assessment frameworks*. Chestnut Hill.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill.
- Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (Version 8.5) [Computer software]. <https://www.statmodel.com/>.
- OECD. (2008). Equity and quality in education: Supporting disadvantaged students and schools. OECD. In Olson, J. F., Martin, M. O., & Mullis, I. V. S. (Eds.). *TIMSS 2007 technical report*. Chestnut Hill.
- Pahlke, E., Hyde, J. S., & Allison, C. M. (2014). The effects of single-sex compared with coeducational schooling on students' performance and attitudes: A meta-analysis. *Psychological Bulletin*, 140(4), 1042–1072. <https://doi.org/10.1037/a0035740>
- Rautalin, M., & Alasuutari, P. (2009). The uses of the national PISA results by Finnish officials in central government. *Journal of Education Policy*, 24(5), 539–556. <https://doi.org/10.1080/02680930903131267>
- Scott, B. (2001, March). Coming of age. IDRA Newsletter. Retrieved from <http://www.idra.org/Newsltr/2001/Mar/Bradley.htm>
- Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory*. Sage.
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417–453. <https://doi.org/10.3102/00346543075003417>
- Stoet, G., & Geary, D. C. (2015). Sex differences in academic achievement are not related to political, economic, or social equality. *Intelligence*, 48, 137–151. <https://doi.org/10.1016/j.intell.2014.11.006>
- van Ewijk, R., & Sleegers, P. J. C. (2010). Peer ethnicity and achievement: A meta-analysis into the compositional effect. *School Effectiveness and School Improvement*, 21(3), 237–265. <https://doi.org/10.1080/09243451003612671>
- von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful. *IERI Monograph Series*, 2, 9–36.
- Wang, P., & Goldschmidt, P. (1999). Opportunity to learn, language proficiency, and immigrant status effects on mathematics achievement. *Journal of Educational Research*, 93(2), 101–111.
- White, S. B., Reynolds, P. D., Thomas, M. M., & Gitzlaff, N. J. (1993). Socioeconomic status and achievement revisited. *Urban Education*, 28(3), 328–343. <https://doi.org/10.1177/0042085993028003007>

- Yang, Y. (2003). Dimensions of socio-economic status and their relationship to mathematics and science achievement at individual and collective levels. *Scandinavian Journal of Edu Research*, 47(1), 21–41. <https://doi.org/10.1080/00313830308609>
- Yip, D. Y., Chiu, M. M., & Ho, E. S. C. (2004). Hong Kong student achievement in OECD-PISA study: Gender differences in science content, literacy skills, and test item formats. *International Journal of Science and Mathematics Education*, 2, 91–106. <https://doi.org/10.1023/B:IJMA.0000026537.85199.36>
- Zhu, Y., & Leung, F. K. S. (2012). Homework and mathematics achievement in Hong Kong: Evidence from the TIMSS 2003. *International Journal of Science and Mathematics Education*, 10(4), 907–925. <https://doi.org/10.1007/s10763-011-9302-3>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
