Chinese/English Journal of Educational Measurement and Evaluation | 教育测量与评估双语季刊

Volume 2 | Issue 4 Article 1

2021

Using Data Mining to Explore Factors That Distinguish Between Students With High and Low Mathematical Literacy Performance — An Example With Socio-Economically Disadvantaged and Advantaged Students in Macao

Wa Kit Sou
University of Macau

Kwok Cheung Cheung
University of Macau

Man Kai Ieong
University of Macau

Soi Kei Mak
University of Macau

Using Data Mining to Explore Factors That Distinguish Between Students With High and Low Mathematical Literacy Performance—An Example With Socio-Economically Disadvantaged and Advantaged Students in Macao

Wa Kit Sou [a], Kwok Cheung Cheung [a], Man Kai Ieong [a], and Soi Kei Mak [a]

[a]University of Macau

Abstract
Using Macao-PISA 2012 data collected from socio-economically disadvantaged and advantaged students, this study identified two sets of important learning factors that distinguished between low- and high-performing disadvantaged students, and between low- and high-performing advantaged students, respectively. The findings of this research contribute to a better understanding of the reasons for Macao's high-quality and equitable education as compared to other regions with high mathematical literacy performance while also revealing the crux of small inequities in its education system. The analysis method used in this paper provides a paradigm for data mining research using large-scale assessment data and helps researchers better grasp the state of education at the local level.

Keywords
PISA;
Classification and regression trees;
Socioeconomic status;
Mathematical literacy;
Data mining

1 Problem Statement

In the era of big data, mathematics is fundamental to various fields and disciplines, and students should be well-equipped with mathematical literacy skills to undertake challenges in future work and studies (Li, 2012). However, not every student has the opportunity to learn mathematics well; socioeconomic status (SES) can impact students' academic performance (Coleman, 1968). Parents with higher SES can help their children study, provide sufficient financial support and home resources, and positively influence the family and school environments, thereby effectively promoting children's cognitive development. By contrast, students coming from socio-economically disadvantaged families tend to have lower performance and difficulties standing out academically, which in turn undermine their future development (Roberts et al., 2007; Sirin, 2005).

However, studies have also indicated that some socio-economically disadvantaged students are able to overcome socioeconomic adversity and achieve the same outstanding performance in international assessments or academic studies as those who are socio-economically advantaged. This creates conditions conducive to attaining higher educational and career goals and increases the odds of breaking down barriers to social mobility (Cheung, 2017; Ieong, 2015; Organization for Economic Co-operation and Development [OECD], 2013b). At the same time, some socio-economically advantaged students are far behind their peers in academic studies, showing lower performance despite having an advantaged background (Cheung, 2017). Thus, both advantaged and disadvantaged students are likely to struggle or attain high academically. It is important to identify learning factors that play a decisive role in shaping student performance.

The Programme for International Student Assessment (PISA) is one of the most influential international educational comparative studies. It assesses 15-year-old students from participating countries and economies in mathematics, science, and reading, and collects information (e.g., student background information and learning history) indicative of educational equity and quality for analysis. It further evaluates the effectiveness of each country's education system, explores reasons for the system's success

CONTACT: Soi Kei Mak. Faculty of Education, University of Macau. skmak@um.edu.mo.
Translator: Tianying Feng. tfeng@cresst.org. 300 Charles E. Young Drive North, Los Angeles, CA 90095.

and failure, and provides an important reference for governments to formulate education policies. Having participated in PISA since 2003, Macao-China has been regarded as an economy with a high-quality and equitable education system by the OECD. In PISA 2018, it was named as the only economy in the world with continuous and rapid advances in educational quality. The quality of its basic education is recognized internationally (OECD, 2020). Macao has a high-quality and fair basic education system because SES has relatively little impact on student academic achievement, and there are more high-performing socio-economically disadvantaged students (also known as Academic Resilient Student), who are able to overcome socioeconomic difficulties and achieve the same level of performance as socio-economically advantaged students (Ieong, 2015).

Yet, small inequities still exist in Macao's basic education system due to the remaining impact of SES on academic performance and disadvantaged students having greater difficulty succeeding in school. In Macao's basic education system, a substantial number of socio-economically disadvantaged students have significantly better mathematical literacy performance than their peers. However, there are also students who have similar socioeconomic backgrounds but are falling behind, which raises the question of what factors underly such differences in mathematical literacy performance. Meanwhile, although many socio-economically advantaged students are benefitted from their more affluent backgrounds and outperform their peers, there are also advantaged students slackening and performing at a level incommensurate with their SES. This raises the question of what factors underly the phenomenon of high achievement or academic slackening of the advantaged students, and ultimately, to what extent these factors differ between advantaged and disadvantaged students.

Past research on the mathematical literacy performance of disadvantaged and advantaged students has been limited to a set of specific variables, while many large-scale international assessment databases (such as PISA, TIMSS) contain rich information. Using a small set of variables based on specific theories may run the risk of ignoring other influential facets and failing to utilize the rich resources in the databases. However, processing big data is a complicated process, involving problems of multicollinearity, outliers, and missing values, which often pose challenges to many researchers.

The classification and regression tree (CART) algorithm is a data mining technique. It can identify factors in order of decreasing relative importance, categorize massive data in a database, and facilitate the feature extraction and identification of the rules of the classification process, thereby yielding the most suitable decision tree model. It has been widely used for commercial and clinical purposes (Zeng et al., 2005) and has been gradually applied in educational studies in recent years. Because of the large number of variables affecting mathematical literacy performance and the complex relationships among them, this study used CART to analyze and identify factors that would distinguish between low and high performers in mathematical literacy in socio-economically disadvantaged and advantaged student groups.

2 Method

2.1 Participants

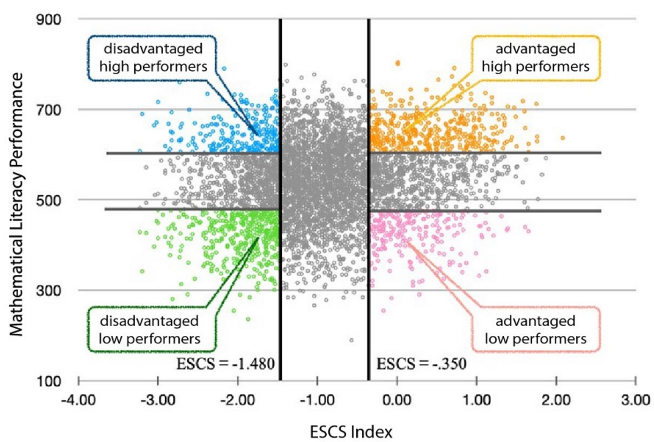
This study focused on students from Macao-China, who participated in PISA 2012. The total number of 15-year-old students was 5,335. According to OECD standards, students within the lowest and highest 25% in the Economic, Social and Cultural Status (ESCS) index were identified as disadvantaged students and advantaged students, respectively; students with the lowest and highest 25% mathematical literacy performance in Macao were identified as low- and high-performing students, respectively. Specifically, there were 242 disadvantaged high-performing students, 419 disadvantaged low-performing students, 434 advantaged high-performing students, and 248 advantaged low-performing students (Figure 1).

2.2 Variable Selection

2.2.1 Dependent Variable

The study analyzed data from the ESCS disadvantaged student group and ESCS advantaged student group, separately. The dependent variable for each group was a binary variable derived from five sets of plausible values of Macao-PISA 2012 mathematical literacy. For each set of plausible values, student performance was categorized as high, medium, or low level. The data for students categorized into the middle level were not used in analyses and hence dropped. Then, the dichotomous variable was derived using a median split of the five sets of values. For the disadvantaged group, the dependent variable indicated whether a disadvantaged student was high- or low-performing; for the advantaged group, the dependent

Figure 1
Identification of Macao's High- and Low-Performing Socio-Economically Disadvantaged and Advantaged Students



variable indicated whether an advantaged student was high- or low-performing.

2.2.2 Independent Variables

There are different theories and interpretations about what affect students' mathematical achievement. One theory alone cannot fully explain all the variations of mathematical literacy, but these theories can be roughly categorized into two frameworks: individual learning and school effectiveness. Among them, the self-regulated learning theory is one of the most important learning-motivation theories and has been widely verified and supported (Boekaerts, 1999; Zimmerman, 2008). Discussions and research on school effectiveness have undergone rapid development after Coleman published the Equality of Educational Opportunity report. The current mainstream view tends to support the "input-process-output" evaluation model (Scheerens, 1990). Therefore, based on theories focusing on individual learning and school effectiveness, along with OECD's (2013a) two-dimensional classification framework used to explain educational outcomes and predict students' mathematical literacy performance, we classified factors affecting mathematical literacy performance into the following: self-regulated learning factors, background and input factors, and teaching process factors. With the classification criteria, we then selected 101 variables from the PISA 2012 database as the independent variables (Table 1).

2.3 Analysis

This study used SPSS 25.0 and IDB analyzer 4.0 as research tools, and used CART, χ^2 tests, and independent sample t-tests for analysis. The analysis process involved resolving issues such as weighting, standard error, and plausible values, in multilevel modeling (Cheung & Keeves, 1990; OECD, 2009).

Compared with other data mining techniques, CART has the following advantages: (a) it can analyze hundreds of independent variables, which can be continuous or categorical; (b) the results are easy to understand and explain; (c) it is a nonparametric statistical technique and does not require making specific assumptions about the sample distribution; (d) it is not affected by the issue of multicollinearity between variables; (e) it is not influenced by outliers; (f) it can effectively handle missing values (Allore et al., 2005; Jin et al., 2017).

CART is essentially a binary recursive partitioning technique, which divides the current sample set (parent node) into two child sets (child nodes), and the child node can become a new parent node to be further divided. The recursion terminates according to the Gini index or the number of observations in a child node (this study set Gini $< .001$ or the number of observations in a child node less than 20 as the stopping criteria) (Lewis, 2000; Strobl et al., 2007, 2009). The Gini index measures the degree of impurity contained in a node. The lower the impurity, the more homogenous the node is. CART uses the Gini index to find nodes with greater impurities while also automatically selecting independent variables that would maximize the homogeneity of the nodes for further partitioning. The process is repeated until the termination conditions are satisfied, and a decision tree model is obtained.

The study also used k-fold cross-validation to estimate the error in the decision tree and improve its accuracy (Brieman et al., 1984). K-fold cross-validation divides the whole sample into k sub-samples ($k = 10$ in this study), selects k-1 sub-samples as the training data, and uses the remaining sub-sample as the testing data. Repeating the process k times would yield k decision tree models, and the one with the highest accuracy is selected as the basis for classification.

Lastly, when performing secondary analysis on data from large-scale international assessments, such as PISA and TIMSS, it is necessary to resolve issues including two-stage stratified cluster sampling and multilevel data. For example, when analyzing PISA data, weighting and error estimation

Table 1
Variables Selected from PISA 2012 Data

Influential factors	Variable category	Variable label	Variable name
Self-regulated learning factors	Motivation and belief	Mathematics self-efficacy	MATHEFF
		Mathematics self-concept	ANCSCMAT*
		Mathematics anxiety	ANXMAT
		Attributions to failure in mathematics	FAILMAT
		Subjective norms in mathematics	ANCSUBNORM*
		Mathematics intentions	MATINTFC
		Mathematics interest	ANCINTMAT*
		Instrumental motivation for mathematics	ANCINSTMOT*
		Openness for problem solving	OPENPS
	Prior knowledge	Familiarity with mathematical concepts	FAMCON
		Experience with applied mathematics tasks at school	EXAPPLM
		Experience with pure mathematics tasks at school	EXPUREM
	Effort and work ethic	Perseverance	PERSEV
		Mathematics work ethic	ANCMATWKETH*
		Mathematics behavior	MATBEH
Background and input factors	Individual	Gender	ST04Q01
		Years attending kindergarten	ST05Q01
		Age when started primary school	ST06Q01
		Country of birth - self	ST20Q01
		Grade repetition	REPEAT
		Immigration status	IMMIG
	Family	Family structure	FAMSTRUC
		Family wealth possessions	WEALTH
		Cultural possessions	CULTPOS
		Home educational resources	HEDRES
		Educational level of father	FISCED
		Educational level of mother	MISCED
		Highest educational level of parents	HISCED
		Expected completed levels of education	PQEDASP
		Occupational aspirations for the child	PQOCCASP
		Mathematics career	PQMCAR
		Parent attitudes toward mathematics	PQMIMP
		Country of birth - father	PQCOB_F
		Country of birth - mother	PQCOB_M
		Country of birth - paternal grand-father	PQCOB_PGF

Influential factors	Variable category	Variable label	Variable name
Background and input factors	Family	Country of birth - paternal grand-mother	PQCOB_PGM
		Country of birth - maternal grand-father	PQCOB_MGF
		Country of birth - maternal grand-mother	PQCOB_MGM
		Parental citizenship - father	PQCITIZF
		Parental citizenship - mother	PQCITIZM
	School	Public or private	SC01Q01
		Funding - government	SC02Q01
		Funding - student fees	SC02Q02
		Funding - benefactors	SC02Q03
		Funding - other	SC02Q04
		Class size	CLSIZE
		School size - total school enrollment	SCHSIZE
		Math teacher-student ratio	SMRATIO
		Proportion of girls at school	PCGIRLS
		Proportion of certified teachers	PROPCERT
		Proportion of mathematics teachers with an ISCED 5A qualification	PROPMA5A
		Proportion of math teachers	PROPMATH
		Proportion of teachers holding a bachelor's degree	PROPQUAL
		Shortage of teaching staff	TCSHORT
		Quality of physical infrastructure	SCMATBUI
		Quality of school educational resources	SCMATEDU
	ICT availability	Availability at home	ICTHOME
		Availability at school	ICTSCH
		Ratio of computers available	RATCMP15
		Ratio of computers connected to the internet	COMPWEB
Teaching process factors	Study time	Average time per week on mathematics (minutes)	MMINS
		Out of school study time on mathematics (hours)	ST55Q02
		Out-of-school study time per week (hours)	OUTHOURS
		Number of times arriving late for school in past two weeks	ST08Q01
		Number of times skipping the whole school day in past two weeks	ST09Q01
	Mathematics teacher and instruction	Mathematics teacher's support	ANCMTSUP*
		Teacher-directed instruction	TCHBEHTD
		Student oriented instruction	TCHBEHSO
		Formative assessment	TCHBEHFA
		Cognitive activation in mathematics lessons	ANCCOGACT*

Influential factors	Variable category	Variable label	Variable name
Teaching process factors		Disciplinary climate	DISCLIMA
		Mathematics teacher's classroom management	ANCCLSMAN*
	ICT usage and attitudes	Use of ICT at school	USESCH
		Use of ICT in mathematic lessons	USEMATH
		ICT use at home for entertainment	ENTUSE
		ICT use at home for school-related tasks	HOMSCH
		Attitudes towards computers: limitations of the computer as a tool for school learning	ICTATTNEG
		Attitudes towards computers: computer as a tool for school learning	ICTATTPOS
	School policy and leadership	Academic selectivity	SCHSEL
		Ability grouping in mathematics classes	ABGMATH
		School responsibility for curriculum and assessment	RESPCUR
		School responsibility for resource allocation	RESPRES
		School autonomy	SCHAUTON
		Teacher participation/autonomy	TCHPARTI
		Framing and communicating the school's goals and curricular development	LEADCOM
		Instructional leadership	LEADINST
		Promoting instructional improvements and professional development	LEADPD
		Teacher participation in leadership	LEADTCH
		Creative extra-curricular activities at school	CREACTIV
		Mathematics-related extra-curricular activities at school	MACTIV
		Mathematics extension courses offered at school	MATHEXC
	School climate	Student-related factors affecting school climate	STUDREL
		Teacher-related factors affecting school climate	TEACCLIM
		Teacher morale	TCMORALE
		Teacher focus	TCFOCST
	Attitudes towards school	Teacher-student relation	ANCSTUDRL*
		Sense of belonging to school	ANCBELONG*
		Attitude towards school: learning outcomes	ANCATSCHL*
		Attitude towards school: learning activities	ATTLNACT
	Parent involvement	Parental involvement in the child's school	PARINVOL
		Parents current support of the child	PARSUPP

*Scale scores were adjusted based on anchoring vignettes to account for students' different response styles caused by social and cultural differences.

should be carried out correctly according to OECD's (2009) recommendations. When plausible values are used (PISA 2012 has five sets, PISA 2015 and 2018 have ten sets), the statistical analysis should be conducted on each set of plausible values to obtain estimates per set. The finalized estimate and total error (including sampling and measurement errors) should be calculated according to the algorithm provided by OECD (2009, p.100). This research performed data processing in accordance with the aforementioned requirements. Specifically, the research used two categorical variables (disadvantaged/advantaged high-performing students and disadvantaged/advantaged low-performing students) as the dependent variables for the CART analysis as opposed to directly using the plausible values.

3 Results

3.1 Important Factors That Distinguish Between Social-Economically Disadvantaged Students With Low and High Mathematical Literacy Performance

Figure 2 shows that 63.4% of the sample were disadvantaged and low-performing students, and 36.6% were disadvantaged and high-performing students (see node 0). With more than 100 variables inputted into the CART model, variables that distinguished low performers from high performers in order of decreasing importance were: mathematics self-efficacy, grade repetition, familiarity with mathematical concepts, and mathematics anxiety.

Mathematics self-efficacy in PISA 2012 was based on Ajzen's (1991) theoretical framework of planned behavior, which explored the motivations and beliefs affecting students' literacy performance. According to Ajzen's theory of planned behavior, behavioral intentions are shaped by an individual's attitudes and subjective norms (i.e., values), as well as by perceived behavioral control (i.e., expectations). Mathematics self-efficacy in PISA 2012 referred to students' evaluation of their ability to solve given math problems, while mathematics anxiety reflected students' self-confidence in solving math problems, such as whether they would be worried about their scores or felt nervous when doing math homework. Carroll (1963) and Abedi et al. (2006) also proposed the concept of opportunity to learn to identify whether students had enough guidance and time for learning in cross-country comparisons, which has been proved to be closely related to student performance. Therefore, the index of familiarity with mathematical concepts, constructed according to the PISA questionnaire, aimed to understand students' familiarity with relevant concepts in the learning process, such as familiarity with functions and equations. PISA also asked 15-year-old students to indicate whether they had ever repeated a grade to facilitate further discussion on the impact of grade repetition on equity.

As shown in Figure 2, when the mathematics self-efficacy was less than or equal to $-.100$, the proportion of disadvantaged low-performing students increased from 63.4% to 83.4% (see node 1), meaning up to 83.4% of these students did not have the index of mathematics self-efficacy exceeding $-.100$. However, if mathematics self-efficacy was greater than $-.100$, the proportion of disadvantaged high-performing students would increase drastically from 36.6% to 75.2% (see node 2). Next, we interpret results from the decision tree, starting with the left-hand side.

The left-hand side of the decision tree showed that node 1 was split into nodes 3 and 4 by grade repetition. If students had ever repeated a grade, the proportion of disadvantaged high performers would decrease from 16.6% to 2.7% (see node 3). For students who did not repeat a grade, the proportion of disadvantaged high performers increased from 16.6% to 47.1% (see node 4).

Node 4 was further divided into nodes 7 and 8 by students' familiarity with mathematical concepts. If this index was less than or equal to .303, the proportion of disadvantaged high-performing students would decrease from 47.1% to 24.4% (see node 7); if the index was greater than .303, then the proportion would increase from 47.1% to 58.2% (see node 8).

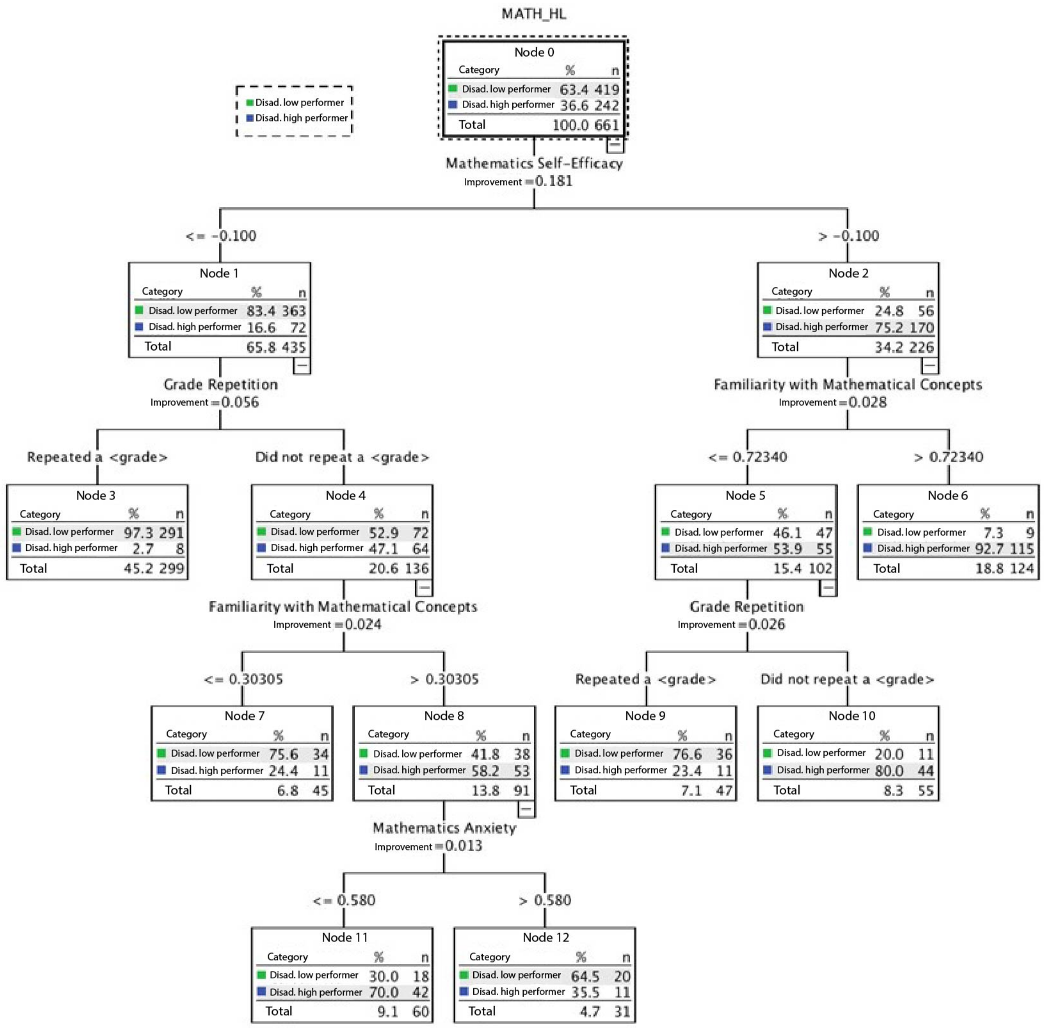
Node 8 was divided into nodes 11 and 12 by mathematics anxiety. If students' mathematics anxiety indices were less than or equal to .580, the proportion of disadvantaged high-performing students would increase from 58.2% to 70.0% (see node 11); if the same indices were greater than .580, the proportion would then decrease from 58.2% to 35.5% (see node 12).

Node 8 was divided into nodes 11 and 12 by mathematics anxiety. If students' mathematics anxiety indices were less than or equal to .580, the proportion of disadvantaged high-performing students would increase from 58.2% to 70.0% (see node 11); if the same indices were greater than .580, the proportion would then decrease from 58.2% to 35.5% (see node 12).

Lastly, node 5 was divided into node 9 and node 10 by grade repetition. For students who repeated a grade, the proportion of disadvantaged high performers decreased

Figure 2

Important Factors Identified from CART Results Distinguishing Between Social-Economically Disadvantaged Students With Low and High Mathematical Literacy Performance



from 53.9% to 23.4% (see node 9); for students who did not repeat any grade, the proportion of high performers increased from 53.9% to 80.0% (see node 10).

The results above showed that for socio-economically disadvantaged students, factors, such as having confidence in their ability to solve math tasks, learning math without being overly anxious, and having no grade repetition, were all important for predicting or explaining whether a student was more likely to become a high performer in mathematics.

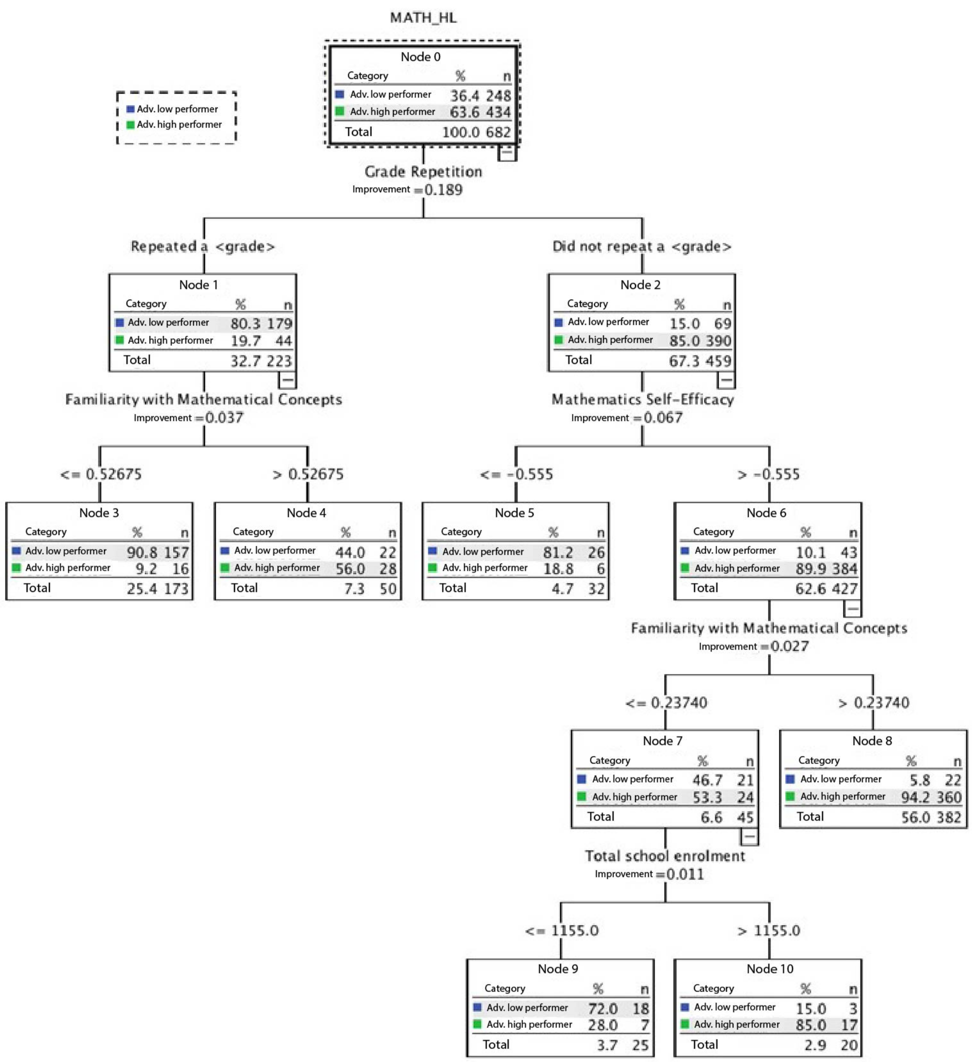
3.2 Important Factors That Distinguish Between Social-Economically Advantaged Students With Low and High Mathematical Literacy Performance

Figure 3 shows that 36.4% of the sample were advantaged low-performing students, and 63.6% were advantaged high-performing students (see node 0). With more than 100 variables inputted into the CART model, variables that distinguished low performers from high performers in order of decreasing importance were: grade repetition, mathematics self-efficacy, familiarity with mathematical concepts, and total school enrollment.

For students who repeated a grade, the proportion of low

Figure 3

Important Factors Identified from CART Results Distinguishing Between Social-Economically Advantaged Students With Low and High Mathematical Literacy Performance



performers increased from 36.4% to 80.3% (see node 1). For students who did not repeat any grade, the proportion of low performers decreased from 36.4% to 15.0% (see node 2).

On the left-hand side of the decision tree, node 1 was divided into node 3 and node 4 by familiarity of mathematical concepts. If this index was less than or equal to .527, the proportion of advantaged low performers would increase from 80.3% to 90.8% (see node 3). However, if the index was greater than .527, the proportion would decrease from 80.3% to 44.0% (see node 4).

On the right-hand side of the decision tree, node 2 was divided into nodes 5 and 6 by mathematics self-efficacy. If the index was less than or equal to −.555, the proportion of low-performing students would increase from 15.0% to 81.2% (see node 5). If the same index was greater than −.555, the proportion of low-performing students would decrease from 15.0% to 10.1% (see node 6).

Node 6 was further divided into nodes 7 and 8 by familiarity with mathematical concepts. If the index was less than or equal to .237, the proportion of low performers would increase from 10.1% to 46.7% (see node 7). If the index was greater than .237, and the proportion of low performers would decrease from 10.1% to 5.8% (see node

8).

Lastly, node 7 was divided into node 9 and node 10 by total school enrollment. If the total enrollment value was less than or equal to 1,155, the proportion of low performers would increase from 46.7% to 72.0% (see node 9). If it was greater than 1,155, then the proportion of low performers would decrease from 46.7% to 15.0% (see node 12).

These results showed that, for socio-economically advantaged students, factors, including grade repetition, mathematics self-efficacy, familiarity with mathematical concepts, were important for predicting or explaining whether a student was more likely to be a high performer in mathematics. Moreover, students attending schools with higher enrollment rates tended to receive more educational resources and hence were more likely to attain high in their studies.

3.3 Comparing Important Learning Factors Between Disadvantaged and Advantaged Students in Macao-China

The analysis results from CART showed that the overall prediction accuracy for the disadvantaged student model was 88.0%, and for the advantaged student model was 88.9%. Both models achieved satisfactory correct prediction rates with satisfactory risk values. The most important learning factors for disadvantaged students were mathematics self-efficacy, grade repetition, familiarity with mathematical concepts, and mathematics anxiety. For advantaged students, the most important factors were grade repetition, mathematics self-efficacy, familiarity with mathematical concepts, and total school enrollment. Both groups had the following factors in common: grade repetition, mathematics self-efficacy, and familiarity with mathematical concepts. The findings indicate that regardless of their socioeconomic statuses, students who did not repeat a grade, had a higher sense of self-efficacy in mathematics, and were more familiar with mathematical concepts, were more likely to become high performers. The presence of these common important learning factors suggests that mathematical literacy performance depends on personal beliefs about learning and the accumulation of past learning outcomes, rather than directly depending on other resources and circumstances.

The aforementioned research findings are overall consistent with the findings of Zimmerman and Martinez-Pons (1990), reflecting the close relationship between self-efficacy and learning effectiveness. In addition, Schunk (1989) also found that self-efficacy could boost perseverance and determination, which directly or indirectly helped students master the learning content. Regarding affective attitudes, a stronger sense of self-efficacy in tackling tasks is linked to greater resilience to stress and lower degree of anxiety and depression (Bandura, 1997). Findings from the analysis indicate that regardless of their socioeconomic statuses, students can succeed through personal beliefs and efforts. Self-efficacy is beneficial for student development. It fosters intrinsic interest, allows students to exert control over their own motivation and behavior, which helps drive students to set their own goals, as well as to self-monitor and evaluate their learning progress (Ho, 2004).

On the other hand, although the mathematical literacy performance of Macao students is not directly determined by resource availability, Table 2 shows that, when compared to socio-economically advantaged students, disadvantaged students had a significantly lower sense of mathematics self-efficacy and familiarity with mathematical concepts, and a much higher percentage of students repeating a grade. The findings reflect the presence of disparities in the common learning factors between the advantaged and disadvantaged students. Such disparities, in turn, result in an overall lower mathematical literacy performance of the disadvantaged students and small inequities in a fair education system. Thus, to help disadvantaged students develop confidence in their ability to learn mathematics and promote positive self-evaluations of capability, schools should consider two aspects—individual learning and school effectiveness—as a starting point for reducing achievement gaps between students. Schools should also provide early support and more effective remedial opportunities for disadvantaged students to reduce grade repetition, thereby further eliminating inequities in the education system.

4 Implications and Suggestions

In today's era of big data, mathematics and human life are closely intertwined, with mathematics becoming an increasing necessity for a number of fields. Talents with better mathematical literacy have more career development choices, while those with poor mathematical literacy will likely be confronted with limited life and career opportunities. Hence, it is crucial to foster students' mathematical literacy (Li, 2012). From the viewpoints of individual learning and school effectiveness, we provide the following recommendations based on our study results.

Table 2
Differences in Teaching Process Factors Between ESCS Disadvantaged and Advantaged Students in Macao

Variable	ESCS disadvantaged	ESCS advantaged	Statistics
	Percentage		χ^2
Grade repetition			
Did not repeat a grade	43.6	56.4	69.828***
Repeated a grade	60.1	39.9	
	M (*SD*)		*t*
Mathematics self-efficacy	−0.025 (0.947)	0.374 (0.991)	−8.236***
Familiarity with mathematical concepts	0.497 (0.946)	0.871 (1.192)	−7.295***

Note. Data were processed according to OECD's (2009) suggestions in the PISA Data Analysis Manual.
***$p < .001$.

4.1 Individual Learning

4.1.1 Enhancing Students' Self-Efficacy

Self-efficacy refers to an individual's particular set of beliefs in their ability to accomplish certain tasks and achieve goals. If individuals believe that they have the ability to tackle tasks and overcome obstacles, they will be more willing to accept challenges and persist in their efforts until they accomplish their goals (Bandura, 1977). For socio-economically disadvantaged and advantaged students, mathematics self-efficacy plays a major role in influencing their mathematical literacy performance. Therefore, teachers can refer to Artino's suggestions to promote the development of mathematics self-efficacy. The main practices include: (a) help students set clear and specific goals, (b) encourage students to reach challenging yet proximal goals, (c) provide realistic and explicit feedback to cultivate students' efficacy beliefs, (d) assist students in accurately estimating their self-efficacy, and (e) use peer modeling to establish self-efficacy (Artino, 2012).

4.1.2 Alleviating Mathematics Anxiety in Socio-Economically Disadvantaged Students

Mathematics anxiety refers to the feeling of tension and anxiety that hinders the manipulation of numbers and the solving of mathematical problems in daily and academic settings. It is one of the important factors undermining the development of mathematical literacy in disadvantaged students. It is recommended that schools and teachers adopt the following strategies in daily teaching and tutoring sessions to prevent and alleviate students' mathematics anxiety: (a) connect mathematics to daily life; (b) avoid emphasizing competition between students and encourage cooperative learning in teaching; (c) focus on the problem-solving process rather than the correct answer or calculation speed; (d) avoid using mathematics as a means of punishment; (e) adapt to different students' ways of learning; (f) offer students support and encouragement; (g) avoid putting students in awkward situations; and (h) use alternative forms of assessments, such as oral questions, observations, presentations, learning logs, reports, and so forth.

4.1.3 Developing Students' Familiarity With Mathematical Concepts

Mathematical concepts are an important element of "basic knowledge" in the "four basics" of Chinese mathematics education. Without a solid understanding of mathematical concepts, it is impossible to accurately, swiftly, and flexibly apply mathematical thinking strategies in solving mathematical problems. According to findings of this study, regardless of their socioeconomic standing, students who were well-acquainted with mathematical concepts had a greater chance of becoming high performers. A thorough understanding of the concepts will help students develop an in-depth and comprehensive understanding of the nature of mathematical problems, enabling students to solve unfamiliar problems. It is suggested that teachers can use the pedagogy of variation to allow students to discern the invariants from the variants, highlight the intension of a concept through non-standard variation, and clarify the extension of a concept through unessential variation, thereby promoting students' understanding of mathematical concepts (Bao et al., 2003).

4.2 School Effectiveness

4.2.1 Implementing Effective Remedial Instruction

Often used by many countries as a policy to ensure education quality, grade repetition refers to a type of remediation that holds students in the same grade for an extra year due to students failing on the previous year (Jimerson, 2001). However, more recent empirical studies have found numerous problems associated with grade repetition, which negatively impact academic achievement (Martin, 2009).

Results from PISA also showed that (a) grade repetition had no obvious benefit to the overall performance of an education system, (b) the chance of repeating was greater for disadvantaged than for advantaged students, and (c) grade repetition may exacerbate inequities in the education system (OECD, 2011). According to our findings, grade repetition is an important factor distinguishing between students with high and low mathematical literacy performance (regardless of SES). Repeating a grade is not an appropriate arrangement for students when teachers cannot provide proper guidance and instructional support following the repetition. The right approach is to provide effective remedial instruction to the underachievers as early as possible before resorting to grade repetition (Sit et al., 2015).

Thus, it is recommended that teachers refer to Rathvon's (2008) Three-Tier Response to Intervention (RTI) model for the implementation of remediation. Tier 1 consists of universal screening of all students in a class and evidence-based instruction. Tier 2 employs "targeted or strategic interventions" in small-group settings in addition to the regular school curriculum, focusing on students whose performance falls below expectations. Tier 3 provides individualized interventions to students who are falling significantly behind when compared to their peers.

4.2.2 Supporting the Development of Smaller Schools

Results of this study showed that total school enrollment was an important factor influencing the mathematical literacy performance of socio-economically advantaged students. The larger the school, the more resource-rich and well-equipped it is. This also entails better curriculum planning and better overall academic performance of the students. It is recommended that the government improves its subsidy policies and supports smaller schools' development to reduce cross-school disparities. The government can also introduce measures to attract and encourage cross-school cooperation, optimize the use of educational resources, and promote school improvement.

4.3 Conclusion

The implementation of education policy cannot have the same effect on all students in the same school, nor can it have the same effect on all education systems, local circumstances, or schools. Researchers can also draw different conclusions based on what type of results is emphasized (Kyriakides & Tsangaridou, 2004). This study shows that even though Macao is regarded by OECD as a region with relatively high-quality and equitable education, there are still areas for continued improvement. For example, compared with advantaged students, Macao's disadvantaged students have a lower sense of self-efficacy and familiarity with mathematical concepts while also having a higher percentage of them repeating a grade. Therefore, the research on educational effectiveness should consider the different effects and adaptations brought forth by the interaction with input factors (Creemers & Kyriakides, 2008; Scheerens, 2000; Teddlie & Reynolds, 2000). A better way to support students with learning difficulties is that schools provide additional instruction time and adjust instructional methods to meet the needs of low-achieving students, thereby allowing these students to catch up to their peers and spending the efforts on where they are needed most. Future research can have a more in-depth and focused investigation regarding school effectiveness.

References

Abedi, J., Courtney, M., Leon, S., Kao, J., & Azzam, T. (2006). *English language learners and math achievement: A study of opportunity to learn and language accommodation* (Technical Report 702). National Center for Research on Evaluation, Standards, and Student Testing. `https://files.eric.ed.gov/fulltext/ED495848.pdf`.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179–211.

Allore, H., Tinetti, M. E., Araujo, K. L., Hardy, S., & Peduzzi, P. (2005). A case study found that a regression tree outperformed multiple linear regression in predicting the relationship between impairments and social and productive activities scores. *Journal of Clinical Epidemiology*, *58*(2), 154–161.

Artino, A. R. (2012). Academic self-efficacy: From educational theory to instructional practice. *Perspectives on Medical Education*, *1*, 76–85.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*, 191–215.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.

Bao, J., Huang, R., Yi, L., & Gu, L. (2003). Research on the pedagogy of variation (in Chinese). *Mathematics Education*, *1*, 6–12.

Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research*, *31*, 445–457.

Brieman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth.

Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, *64*(8), 723–733.

Cheung, K. C. (2017). The effects of resilience in learning variables on mathematical literacy performance: A study of learning characteristics of the academic resilient and advantaged low achievers in Shanghai, Singapore, Hong Kong, Taiwan and Korea. *Educational Psychology*, *37*(8), 965–982.

Cheung, K. C., & Keeves, J. P. (1990). Hierarchical linear modeling. *International Journal of Educational Research*, *14*(3), 289–297.

Coleman, J. S. (1968). Equality of educational opportunity. *Equity & Excellence in Education*, *6*(5), 19–28.

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. Routledge.

Ho, S. (2004). *An analysis of the relationship between learning strategies and learning effectiveness of Hong Kong middle school students using PISA data* (in Chinese). The Chinese University of Hong Kong: Hong Kong Institute of Educational Research.

Ieong, M. K. (2015). *In search of resilience in learning variables mediating the effects of ESCS on mathematical literacy performance of adolescents in Macao: Through the lens of PISA 2012* [Unpublished doctoral dissertation]. University of Macau, Macao.

Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, *30*(3), 420–437.

Jin, S. L., Cheung, K. C., & Sit, P. S. (2017). Task- and non-task-specific factors classifying problem-solving experts and novices: Comparing students of the top ten high-performing eastern and western economies in PISA 2012. *Contemporary Educational Research Quarterly*, *25*(3), 71–103.

Kyriakides, L., & Tsangaridou, N. (2004). *School effectiveness and teacher effectiveness in physical education [Paper presentation]*. The 85th Annual Meeting of the American Educational Research Association, San Diego, CA, USA.

Lewis, R. J. (2000). *An introduction to classification and regression tree (cart) analysis [Paper presentation]*. The Annual Meeting of the Society for Academic Emergency Medicine, San Francisco, CA, USA.

Li, Z. (2012). The significance of mathematics and the value of mathematics education (in Chinese). *Curriculum, Teaching Material, & Method*, *32*(1), 58–62.

Martin, A. J. (2009). Age appropriateness and motivation, engagement, and performance in high school: Effects of age within cohort, grade retention, and delayed school entry. *Journal of Educational Psychology*, *101*(1), 101–114.

OECD. (2009). *PISA data analysis manual: SPSS* (2nd ed.). OECD Publishing.

Organization for Economic Co-operation and Development [OECD]. (2011). When students repeat grades or are transferred out of school: What does it mean for education systems? *PISA in Focus*, *6*. OECD Publishing.

Organization for Economic Co-operation and Development [OECD]. (2013a). *PISA 2012 assessment and analytical framework: mathematics, reading, science, problem solving and financial literacy*. OECD Publishing.

Organization for Economic Co-operation and Development [OECD]. (2013b). *PISA 2012 results: Excellence through equity: Giving every student the chance to succeed* (Vol. II). OECD Publishing.

Organization for Economic Co-operation and Development [OECD]. (2020). *PISA 2018 results: What students know and can do* (Vol. I). OECD Publishing.

Rathvon, N. (2008). *Effective school interventions: Evidence-based strategies for improving student outcomes* (2nd ed.). Guilford.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, *2*(4), 313–345.

Scheerens, J. (1990). School effectiveness research and the development of process indicators of school functioning. *School Effectiveness and School Improvement*, *1*(1), 61–80.

Scheerens, J. (2000). *Improving school effectiveness (Fundamentals of Education Planning, No. 68)*. IIEP-UNESCO.

Schunk, D. H. (1989). Social cognitive theory and self-regulated learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theory, research, and practice* (pp. 83–110). Springer.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*(3), 417–453.

Sit, P. S., Cheung, K. C., Cheong, W. C., Mak, S. K., Soh, K. C., & Ieong, M. K. (2015). What studying problems are faced by the adolescent grade repeaters in Macao: Uncovering underlying mechanisms based on evidence from the PISA 2012 study. *Asia Pacific Education Review*, *16*(3), 367–377.

Strobl, C., Boulesteix, A. L., & Augustin, T. (2007). Unbiased split selection for classification trees based on the Gini index. *Computational Statistics & Data Analysis*, *52*, 483–501.

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, *14*(4), 323–348.

Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research*. Falmer Press.

Zeng, X., Cai, X., Su, D., Zeng, Q., & Wang, Q. (2005). *Data mining* (in Chinese). Flag Publishing.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, *45*(1), 166–183.

Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology*, *82*(1), 51–59.