**An analysis of Singapore student performance in the OECD Programme for International Student Assessment**

Group 15

Jin Hui Poh | Nguyen Le Hong Ngoc | Zhao Yazhi

**Final report**

Table of Contents

Executive summary 3

1. Introduction 4

1.1. Business Problem 4

1.2. Introduction of PISA 4

2. Project Overview 4

2.1 Project motivation 4

2.2 Scope of work 5

2.3 Business objectives & Analytical problems 5

2.3.1 Business Objectives 5

2.3.2 Analytical Problems 5

3. Dataset 6

3.1 Data retrieval 6

3.2 Data extraction 6

3.3 Data preparation 7

4. Methodology 9

4.1 Framework of analysis 9

4.2 Techniques of analysis and variable selection 10

4.2.1 K-means clustering 10

4.2.2 Partition analysis for school profiling 11

4.2.3 Constructing regression model 13

5. School-level Findings 16

5.1 Exploratory Data Analysis 16

5.2 Profiling high performance and low performance schools 18

5.2.1 Decision tree analysis 18

5.3 Multiple regression model 20

5.3.1 Numeric factors that affect test subject performance 20

5.3.2 Categorical factors that affect test subject performance 22

6. Student-level Findings 23

6.1 Exploratory Data Analysis 23

6.2 Multiple regression model 24

6.2.1 Variable selection 24

6.2.2 Multiple regression 25

6.2.3 Factors that affect low performing schools 26

7. Discussion 28

7.1 General recommendation 28

7.2 Gaps identified in data for future research efforts 28

7.3 Conclusion 29

References 30

Appendix 31

Appendix A: Decision tree analysis results 31

Appendix B: Multiple regression results for numeric attributes (School level analysis) 33

Appendix C: Attributes selected for student level analysis 34

Appendix D: Multiple regression results for student-related factors for overall & different subjects 35

# Executive summary

This project aims to investigate the effects of various personal and environmental factors on the cognitive abilities of secondary school students in Singapore, in particular their family background and school environment. We make use of publicly available data collected by the Organization for Economic Co-operation and Development (OECD) through their “Programme for International Student Assessment” (PISA) survey. The triennial international survey aims to evaluate education systems worldwide by testing the skills and knowledge of 15-year-old students. Using Singapore’s data from the most recent 2012 survey results, key findings would serve as insights for policy makers and educators to improve existing practices.

Our analysis revealed significant variance in performance across schools, with low performing schools showing greater inconsistency. While clustering and multiple regression analysis indicate that quality of teaching staff and student attendance is key to good school performance, more than a quarter of schools reported a shortage of teaching staff as the top resource shortages faced. Closer examination of students’ background revealed that good access and appropriate usage of educational resources like books and computers enhances test performance. Those whose parents have at least an O level qualification or are working full time also tend to do better. However, with the right learning attitude, unfavourable external conditions can be overcome. It is important for students to be confident in their abilities and believe that schools are capable of preparing them for life. For those in low performing schools, access to adequate educational resources and having a strong internal locus of control becomes important, exceeding the impact of classroom management efforts. Personal factors seem to play a larger role in performance of students from low performing schools. Overall, school, family and personal factors all have a combined effect on the student and so schools need to adopt a holistic view to education. We recommend that schools take a 2-pronged approach of detection and prevention to improve performance.

Although we are able to gain some general insights into the education landscape in Singapore, we noted several data gaps that need to be addressed in order to reap greater insights in future research efforts:

1. Data for shortage and learning hindrance are highly aggregated categorical data, and are difficult to be used in analysis if not quantified. There is also little detail of school spending, which is important given how closely related it is to resource allocation in schools and the environment created for students. More detailed numeric data will need to be collected in these areas.
2. We are also unable to identify factors to explain student truancy, the most distinguishing factor of school performance. Surveys are recommended to address the knowledge gap in this area.
3. The economic-socio index used for the international scale is not suitable for the local context and locally available data will need to be provided to derive a more relevant attribute for economic-socio status of families in Singapore.

While Singapore schools do relatively well internationally, greater access to local data will aid research efforts required for more relevant insights to improving performance, allowing every school to achieve high and consistent performance.

# 1. Introduction

## 1.1. Business Problem

The Ministry of Education (MOE) of Singapore collects and analyses data from schools island wide to continually improve policies and practices in education. However, most of this data are not publicly available for research and analysis by those outside the Ministry. Hence, our sponsor seeks to gain insights about education in Singapore from the publicly available data collected by the OECD through their “Programme for International Student Assessment” (PISA) survey. The PISA is a triennial international survey, which aims to evaluate education systems worldwide by testing the skills and knowledge of 15-year-old students. The most recently published results are from the assessment in 2012.

## 1.2. Introduction of PISA

The Programme for International Student Assessment (PISA) is an international survey that aims to evaluate education systems worldwide by testing the skills and knowledge of 15-year-old students. To date, students representing more than 70 economies have participated in the assessment. The most recently published results are from the assessment in 2012.

Around 510,000 students in 65 economies took part in the PISA 2012 assessment of reading, mathematics and science representing about 28 million 15-year-olds globally. Given that PISA is an ongoing triennial survey, countries and economies participating in successive surveys can compare their students' performance over time and assess the impact of education policy decisions.

Since the year 2000, every three years, fifteen-year-old students from randomly selected schools worldwide take tests in the key subjects: reading, mathematics and science, with a focus on one subject in each year of assessment. Students take a test that lasts 2 hours. The tests are a mixture of open-ended and multiple-choice questions that are organized in groups based on a passage setting out a real-life situation. A total of about 390 minutes of test items are covered. Students take different combinations of different tests. The students and their school principals also answer questionnaires to provide information about the students' backgrounds, schools and learning experiences and about the broader school system and learning environment.

# 2. Project Overview

## 2.1 Project motivation

There is a need for data-driven approach to better understand education system in Singapore, and potentially to find ways to improve student performance. However, Ministry of Education (MOE) in Singapore, although possessing a rich collection of data, does not make them publicly available for extensive studies. In an effort to encourage more data sharing by MOE, a proof-of-concept is needed in order to convey the capability of analytics to education policy-making. This can be done by using publicly available alternative data sources such as survey data collected by PISA. With the insights derived from PISA data, we hope to encourage MOE to share more education data with the public, as this would deepen the analysis and give a more extensive knowledge discovery.

## 2.2 Scope of work

The scope of this project is largely determined by the available data from the PISA2012 survey, specifically for Singapore. The survey takes sample from 172 secondary schools in Singapore; each with randomly selected 35-40 students. Data collected during the survey includes family background, parents’ education, student possession at home, school funding, staff headcounts and profile, facilities and prominent issues (truancy and shortage of resources). A 2-hour test was conducted to measure students’ competency in Mathematics, Science and Reading. Overall, the PISA2012 survey results serve as a rich source of data to conduct our analysis for the education landscape in Singapore.

## 2.3 Business objectives & Analytical problems

### 2.3.1 Business Objectives

The project aims to investigate the effects of various personal and environmental factors on the cognitive abilities of secondary school students in Singapore, especially their family background and overall academic environment. We would like to find out, in particular, which factors matter and relatively how important are each of them towards student performance. Key findings in this project can serve as insights for policy makers and educators to further improve current practices for the creation of better learning environments and processes.

### 2.3.2 Analytical Problems

To achieve the above business object, the following questions need to be addressed

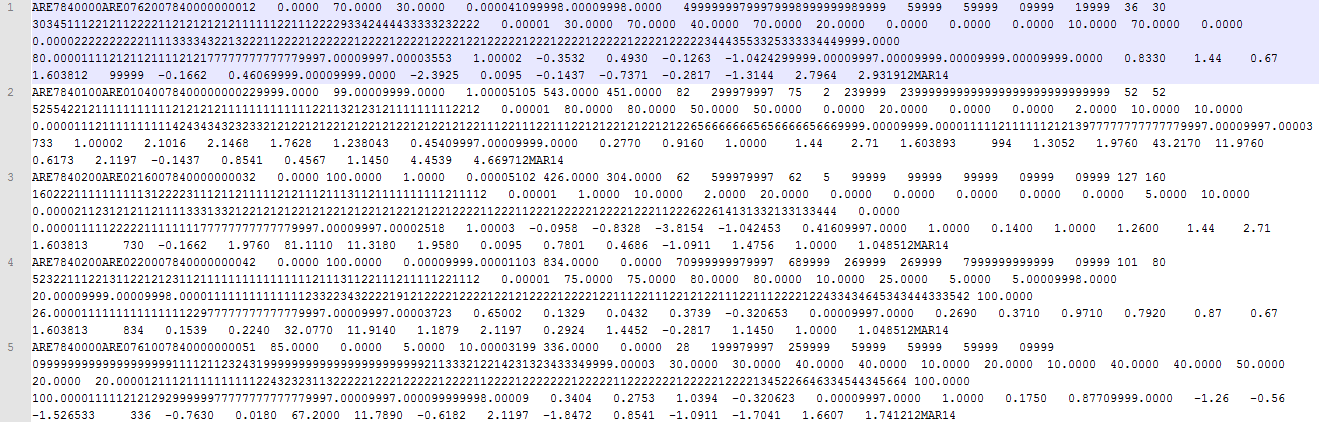
* What factors pertaining to school environment, family background and students might affect academic performance?
* How important are each of those factors?

The above are over-arching questions to be addressed. Within the methodology, our analysis will break down the analytical problems above into questions more specific to aspects of the data.

# 3. Dataset

## 3.1 Data retrieval

The data used in this project is questionnaire result from the latest PISA survey in 2012. All raw data files are publicly available on PISA website[[1]](#footnote-1). However, the raw data is in flat file text format, where a fix number of characters represent a value (e.g. first 3 letters indicate country code), as seen in Figure 3:



*Figure 3: Raw text data of PISA 2012 questionnaire survey result*

Raw data in this form is not ready for cleaning and analysis. PISA database has scripts to convert the raw text data into table forms. The steps to convert the data are as follows:

1. Download raw questionnaire results (zipped text files) from PISA 2012 website and extract
2. Retrieve SAS programs for appropriate data files
3. Open the SAS scripts in SAS Enterprise Guide
4. Ensure that the path to raw text files are correct



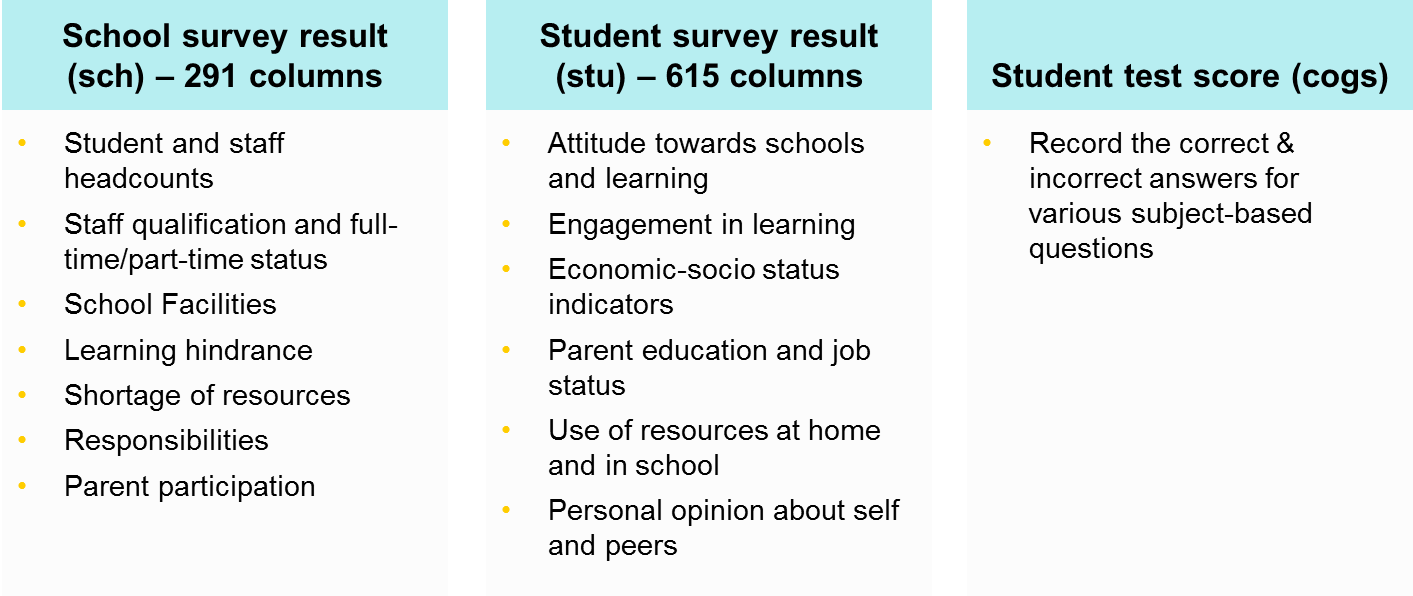
1. Run the programs in SAS Enterprise Guide to get output SAS data table
2. Export the output SAS data table in desired formats (.sas7bdat, .csv and so on); display labels as column names for easy interpretation later on.

## 3.2 Data extraction

Only Singapore data is of interest for our scope of project, therefore only the records with Country code ‘SGP’ are extracted. This process gives us the following for analysis:

* School survey results - 172 secondary schools
* Student survey results - 5,546 records, approx. 35 students per school
* Student test score in Math, Science, Reading, Computer-based assessment

Each of the 3 mentioned data tables have a rich set of attributes. The summary below shows an overview of aspects that the data covers.



*Figure 4: General data table attributes*

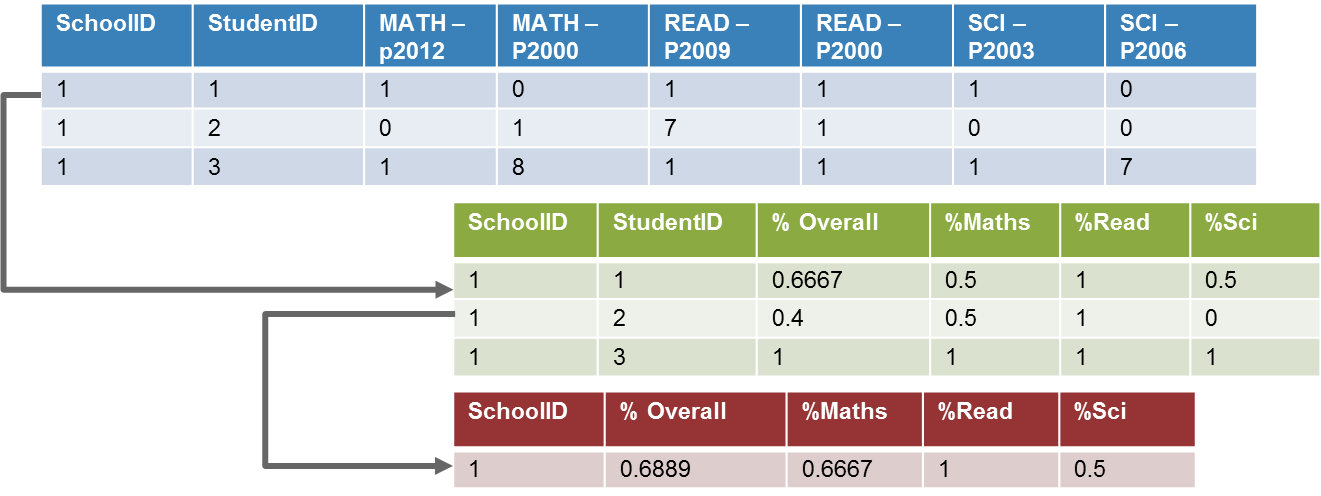
## 3.3 Data preparation

The data preparation follows 3 basic steps.

**Step 1: Aggregate student test score**

The original student test score table records the correct and incorrect answers by each student. We convert these into percentage score of each student by dividing the total number of correct answers by the total number of questions student attempted. Questions that are labeled 7 and 8 are not applicable to that student, and thus not taken into account when calculating percentage score.

Finally, we calculate the average test score for each school, in each subject based on student percentage score.

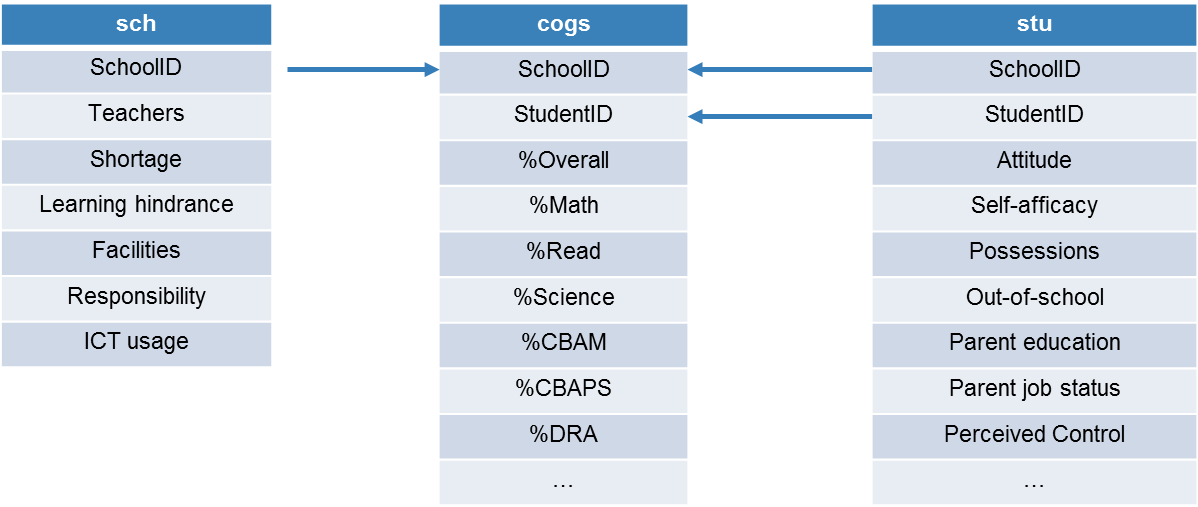


*Figure 5: Joining of data table for test scores*

**Step 2: Joining data tables**

Score data and school data tables are joined by matching school ID. Since the granularity of these 2 tables is not the same – for score table, each record represents one student; for school table, each record represents one school – hence the need to aggregate data in score table in Step 1. The aggregated score will then be joined with school data table.

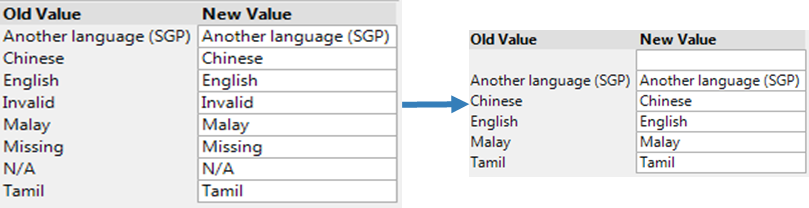
Joining score table and student table is more straightforward, simply by matching school ID and student ID. As the result of this step, student score will be used as measure for their academic performance, and other factors from school and student data tables will be used to help explain the difference in their performance.



*Figure 6: Joining of data tables for school, student and test scores*

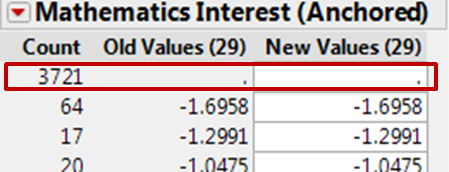
**Step 3: Data cleaning and standardization**

Missing records in PISA data follow a consistent set of labels: ‘I’/’Invalid’, ‘M’/ ‘Missing’ and ‘N’/’No response’. In JMP Pro, we used Standardize Attribute to convert these values into a common null, so that JMP will recognize them all as missing data and process accordingly when building models.



*Figure 7: Recode missing records into a common null value*

Upon checking for missing data pattern, some attributes are revealed to have more than 50% null values. These will be excluded from analysis. An example is given below: out of 5,546 records in student table, 3,721 records in Mathematics Interest (Anchored) are missing, so this attribute will not be chosen for analysis.



*Figure 8: Columns with >50% null data excluded*

Lastly, attributes that do not help differentiating the entities in data are also excluded. For example, all attributes regarding “Acculturation” aspects only have ‘N’ values (No response) and these too will be excluded.

# 4. Methodology

## 4.1 Framework of analysis

Our analysis assumes that school, family and personal factors can all affect student performance. Figure 9 describes the steps taken for our school level analysis.

*Figure 9: Framework for school level analysis*

With this framework, we want to first get an overview of school performance in Singapore, before expanding our analysis to consider more factors as we gain greater understanding of the data and context, and thereby contemplate our business questions in greater depth. This is an iterative process as we continuously refine our models.

First, to get an overview of academic performance of Singapore schools, we perform a distribution analysis on school test scores. Thereafter, having identified the spread of performance, we separate schools into different segments, based on their academic results, and study each of the segments to observe how school-related factors might cause the difference in student performance. Features that separate high-performing from low-performing schools are to be identified, so that we can profile schools using those features. Finally, a regression model is constructed to explain in greater detail the extent to which selected factors might cause the difference in performance.

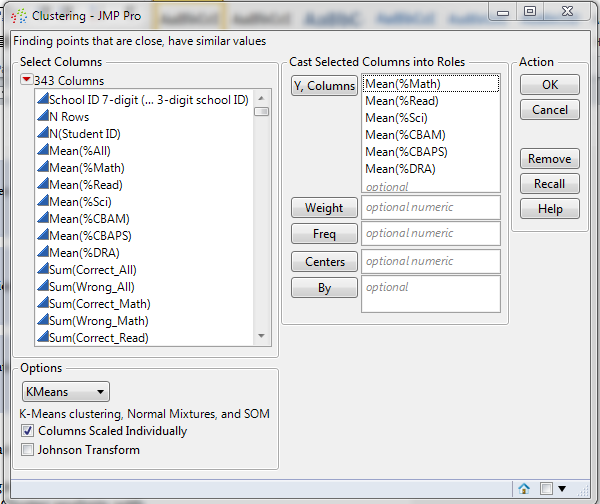
At the student level, the framework and techniques applied are largely the same as with schools. However, given that student data is more atomic and has a richer set of attributes, there is no need to segment them. Instead, we use their scores as the direct measure to build models. Figure 10 shows the framework for our student level analysis.

*Figure 10: Framework for student level analysis*

## 4.2 Techniques of analysis and variable selection

### 4.2.1 K-means clustering

K-means clustering is used to segment the schools based on student performance. To this end, in JMP Pro, we perform Cluster analysis, with the average school scores in each subject as response variables. In terms of k-value, numbers 3-6 are chosen. This is due to the consideration that upon segmenting the schools, we will study the profile of each cluster and ultimately devise recommendation accordingly; hence clusters should come with a reasonably large number of data points, and that too many clusters are not favourable. The configuration in JMP Pro is shown in Figure 11:



*Figure 11: Clustering configuration in JMP*

Running the analysis gives us the following clustering results in Table 1:

**Table 1: Results from k-means clustering**

|  |  |  |  |
| --- | --- | --- | --- |
| *No.of clusters* | *CCC* | *Best* | *Remarks* |
| 3 | -2.370384428 |  | k=3 is chosen because it is more logical to segment the schools as low-performing, high-performing and average. K=4 does not give additional insights; the cluster breakdown gets more detailed but does not reflect any discrepancies in the score ranking among the clusters. We have a clear-cut order of which cluster is better with k=3. |
| 4 | -2.229977536 | Optimal CCC |
| 5 | -2.902106982 |  |
| 6 | -3.180532388 |  |

Statistically speaking, with a Cubic Clustering Criterion (CCC) of -2.23, k = 4 gives the best clustering result. Nevertheless, k = 3 is chosen for further analysis as closer examination of cluster details revealed little value-add in having 4 clusters.

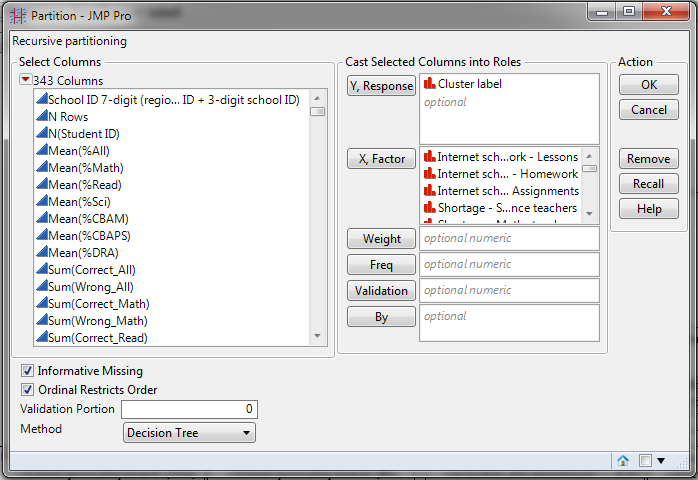
As seen from Table 2, when using 4 clusters, the lowest performing schools simply get further segmented but this does not give much additional insights. Hence, dividing the schools into 3 clusters of low, average and high performance would yield more meaningful results with the clearer distinction and order in school performance.

**Table 2: Results from k-means clustering (k=3 & k=4)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *k-means no.of cluster = 3* | *Mean %Math* | *Mean %Read* | *Mean %Science* | *Mean %CBAM* | *Mean %CBAPS* | *Mean %DRA* | *Count* |
| 1 (Low-performing) | 0.48252 | 0.54326 | 0.49250 | 0.39294 | 0.53031 | 0.56579 | 50 |
| 3 (Average) | 0.60683 | 0.66375 | 0.61202 | 0.52589 | 0.65167 | 0.68950 | 85 |
| 2 (High-performing) | 0.80116 | 0.81298 | 0.78686 | 0.71849 | 0.78564 | 0.82540 | 37 |
|  |  |  |  |  |  |  |  |
| *k-means no.of cluster = 4* | *Mean %Math* | *Mean %Read* | *Mean %Science* | *Mean %CBAM* | *Mean %CBAPS* | *Mean %DRA* | *Count* |
| 1 | 0.57477 | 0.63036 | 0.58107 | 0.49136 | 0.62745 | 0.65810 | 66 |
| 2 | 0.84181 | 0.84644 | 0.82819 | 0.76052 | 0.80726 | 0.85248 | 23 |
| 3 | 0.47261 | 0.53453 | 0.48121 | 0.38095 | 0.51544 | 0.55603 | 41 |
| 4 | 0.68275 | 0.73034 | 0.68172 | 0.60456 | 0.71108 | 0.75234 | 42 |

### 4.2.2 Partition analysis for school profiling

Using the clusters in the previous step as response variable, we use the decision tree analysis to profile the schools. Figure 12 shows the configuration of the analysis on JMP. The data is split according to the school clusters, selecting attributes with the highest LogWorth, since it would have the greatest effect in differentiating the schools.



*Figure 12: Partitioning configuration in JMP*

In order to observe the combined effect of all school factors on the student performance, we use 225 attributes, all of which have been checked for missing data (attributes with more than 50% missing data are excluded). Running the analysis results in a decision tree with RSquare value of 0.703 and a total of 20 splits.

**Table 3: Decision tree analysis results**

|  |  |  |  |
| --- | --- | --- | --- |
| Term | No. of splits | G2 | Portion |
| Learning Hindrance - Students truancy | 1 | 47.04693 | 0.1758 |
| Student-Teacher ratio | 1 | 26.98794 | 0.1009 |
| Index of school responsibility for curriculum and assessment | 2 | 25.84174 | 0.0966 |
| Proportion of girls at school | 1 | 19.14672 | 0.0716 |
| Parental achievement pressure | 1 | 17.19589 | 0.0643 |
| Parent Participation - Progress discussion teacher initiative | 1 | 16.30177 | 0.0609 |
| Parent Participation - School government | 2 | 15.89513 | 0.0594 |
| Internet schoolwork - Assignments | 1 | 14.75293 | 0.0551 |
| Learning Hindrance - Heterogeneous classes | 1 | 14.69938 | 0.0549 |
| Index of school responsibility for resource allocation | 1 | 13.52389 | 0.0505 |
| Learning Hindrance - Students being late | 1 | 11.48591 | 0.0429 |
| Responsibility - Course content - Teachers | 1 | 11.36385 | 0.0425 |
| Teacher participation in leadership | 1 | 11.10997 | 0.0415 |
| Ratio of computers for education and number of students in the <national modal grade for 15-year-olds> | 1 | 7.226757 | 0.027 |
| Teacher intentions - Social development | 1 | 4.75314 | 0.0178 |
| Responsibility - Salary increase - <School governing board> | 1 | 3.845087 | 0.0144 |
| Extracurricular creative activities at school | 1 | 3.277932 | 0.0123 |
| Parent Participation - Library volunteering | 1 | 3.107483 | 0.0116 |

Table 3 shows the attributes selected based on the decision tree analysis. It is possible to identify the features that differentiate the schools into high-performing and low-performing clusters. Decision tree results will be discussed in detail in our findings.

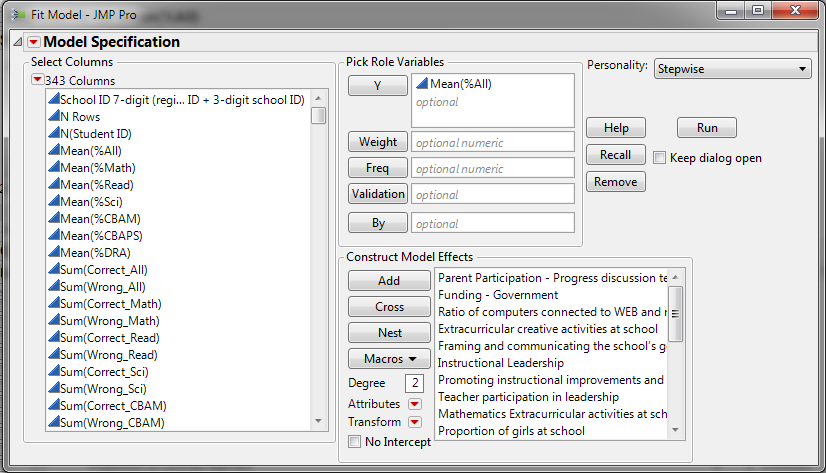
### 4.2.3 Constructing regression model

#### 4.2.3.1 Creating dummy variables for categorical attributes

For regression analysis to account for categorical attributes, they need to be converted into numeric form. Partition analysis can also be used to generate dummy variables for this purpose. Using school cluster (for school-level analysis) and student score (for student-level analysis) as response variables, Partition analysis will generate dummy variables for selected categorical attributes, such that the difference between nodes will be greatest.

#### 4.2.3.2 School-level

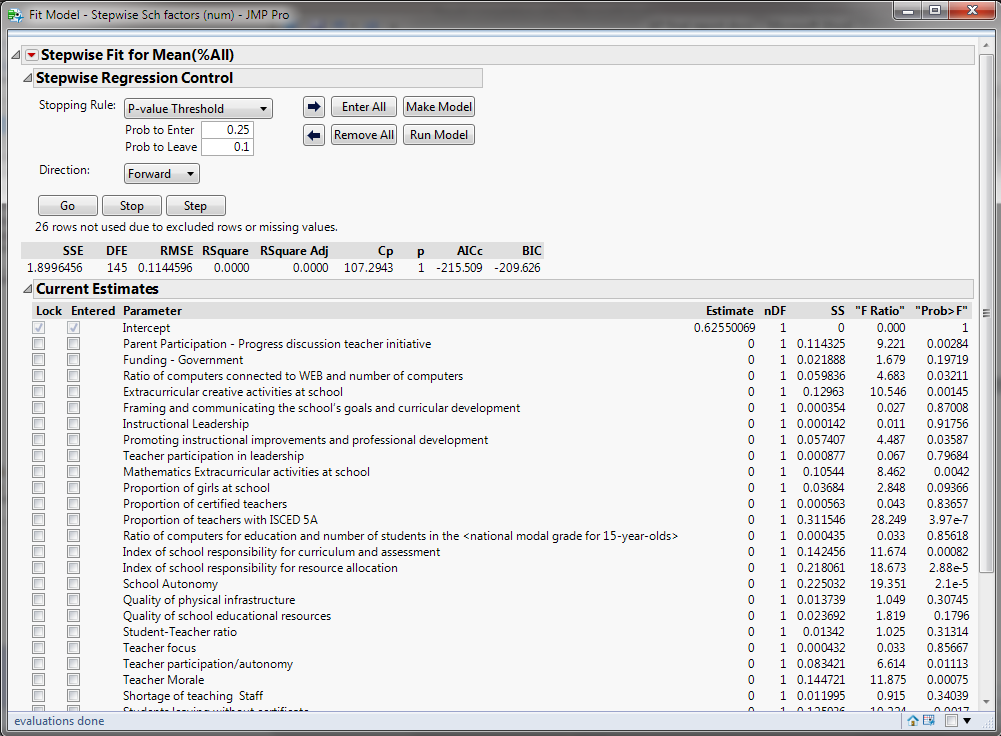
There are several regression models to be built. To analyze performance at the school level, we carried out the configuration on JMP as shown in Figure 13 below.



*Figure 13: Fit model configuration in JMP for school level analysis*

Response variable in this case is the overall mean score of each school, but we extend the analysis to cover all subject scores, which are Mathematics, Reading, Science and Computer-based assessment. The purpose is to compare the impact, if any, that various factors have on different subject performance by students. We will also be able to compare the extent of impact and how relatively important the factors. We used Stepwise regression, with p-value threshold as the stopping rule. This allows us to interactively refine the regression model by adding/ removing attributes as necessary. Attributes selected to build regression model should have p-value of 0.01 or less.

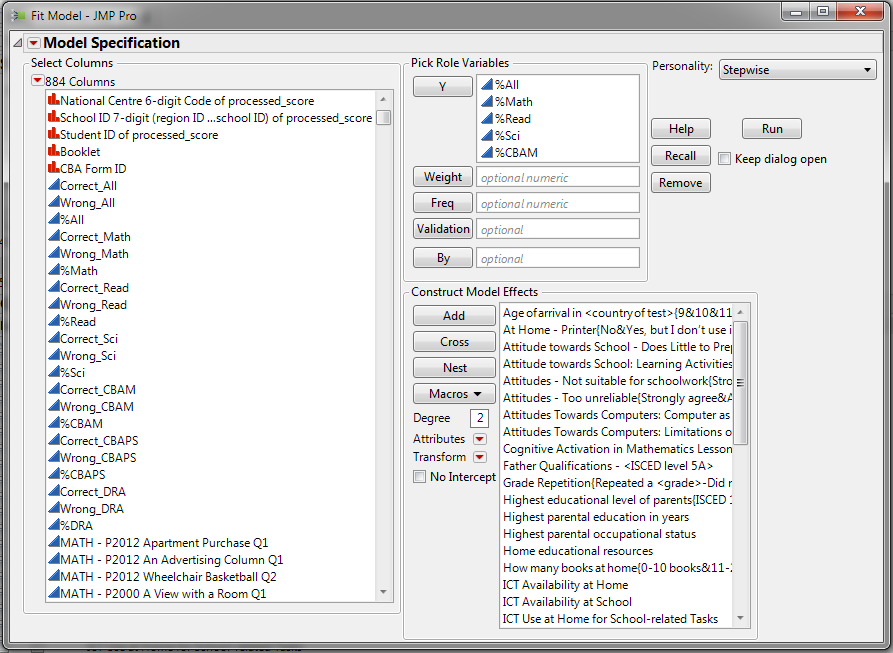
Figure 14 shows a snapshot of the configuration process for attribute selection on JMP.



*Figure 14: Fit model results in JMP for school level analysis*

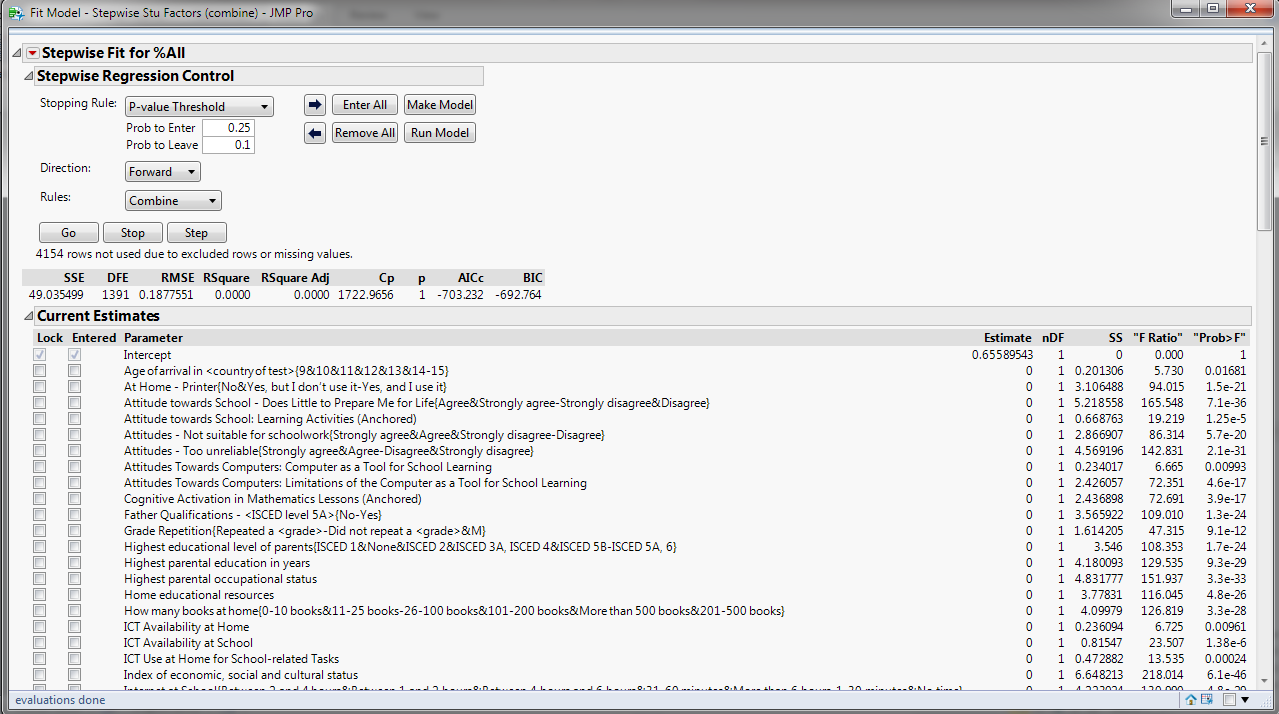
#### 4.2.3.3 Student –level

At the student level of analysis, the subject scores are used as response variables, and the factors used to build regression models are numeric attributes and dummy variables of categorical attributes generated in the previous section. The stepwise configuration for student-level regression analysis is shown in Figure 15 below.



*Figure 15: Fit model configuration in JMP for student level analysis*

Similar to the school-level analysis, we use p-value threshold as the stopping rule. Variables selected to run models must have p-value (Prob>F) less than 0.01.



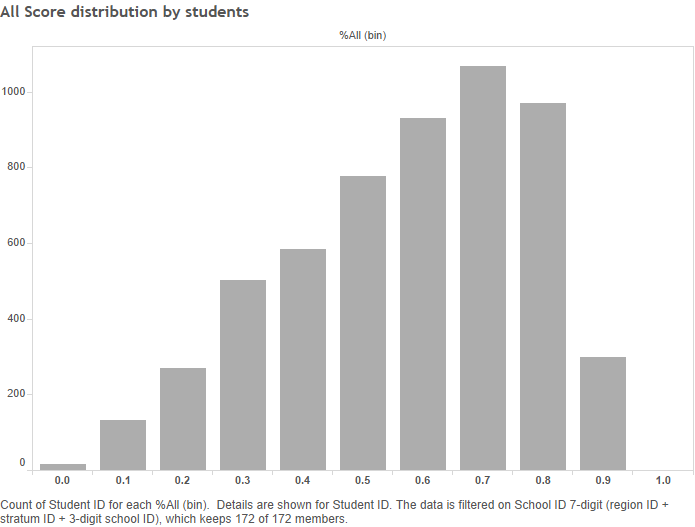
*Figure 16: Fit model results in JMP for student level analysis*

Using the techniques discussed above, we expanded the student-level analysis by segregating students into those from low performing and high performing schools, and built separate regression models for each group. This is done to help us gain greater insight into the different factors that cause the differences in student performance, beyond the school factors. We are also interested in studying the outliers in low-performing schools – students from low performing schools that can still perform well. Understanding how other factors come into play students are placed in less than ideal environments could be the key to improving performances of students in Singapore.

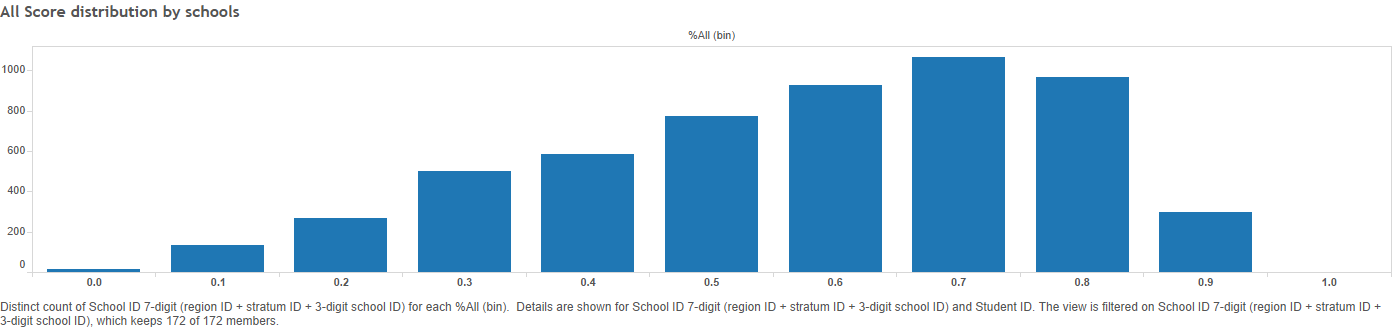
# 5. School-level Findings

## 5.1 Exploratory Data Analysis

We conducted a distribution analysis and based on our findings, we created a Tableau software visualization to enable users to analyze data from the school level and eventually do a deep dive into the performance of individual schools. In this section, we will discuss our findings.

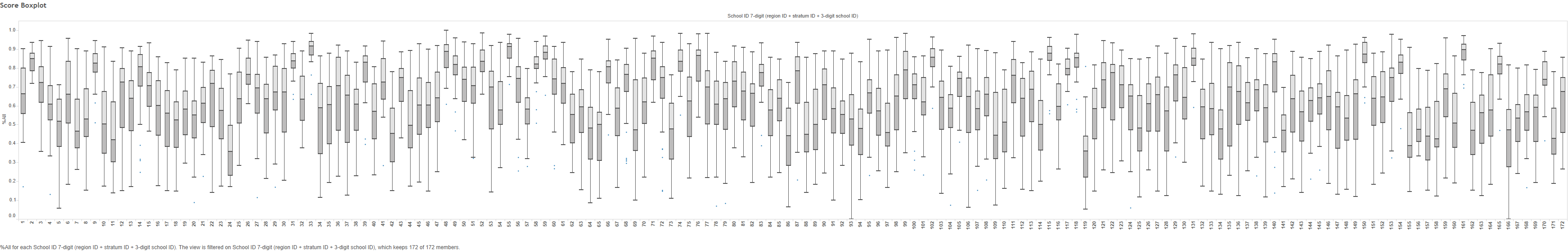


*Figure 17a: Student score distribution & School score distribution*



*Figure 17b: Student score distribution & School score distribution*

Majority of schools in Singapore are performing well academically, as seen in Figures 17a and 17b above, where the distribution of scores on both student and school levels are right-skewed.

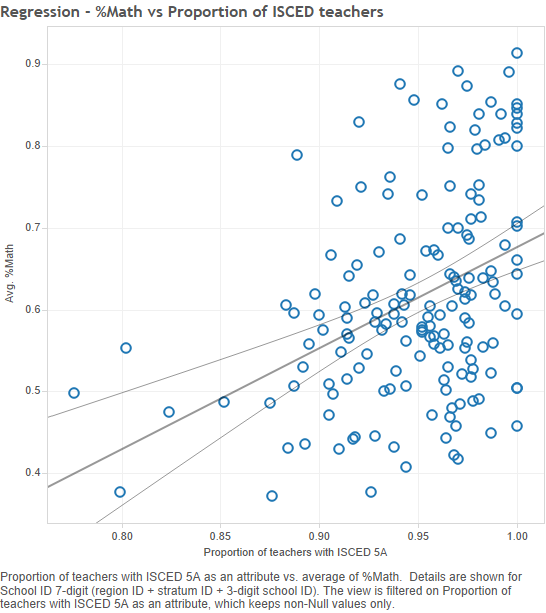


*Figure 18: Boxplot of student performance in schools*

However, we also found significant variation of performance amongst the schools. Figure 18 above shows the boxplot of the individual school performances.

|  |  |
| --- | --- |
|  |  |
| *Figure 19a: Boxplot of scores in top 10 schools* | *Figure 19b: Boxplot of scores in bottom 10 schools* |

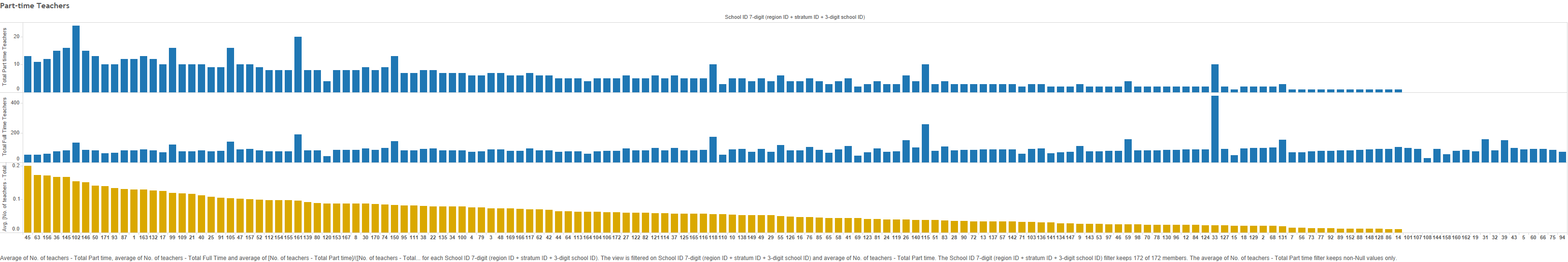
Students in high-performing schools are seen to perform well consistently, with their scores concentrated around the range of about 0.3, as seen in Figure 19a above. However, those of the low performing schools show a much greater disparity (Figure 19b), with a score range of 0.9. Some students in low performing schools score as well as those in high performing schools at about 0.8 to 0.9, while others have scores nearing zero. This suggests that there is a significant disparity in school performance that requires further investigation.



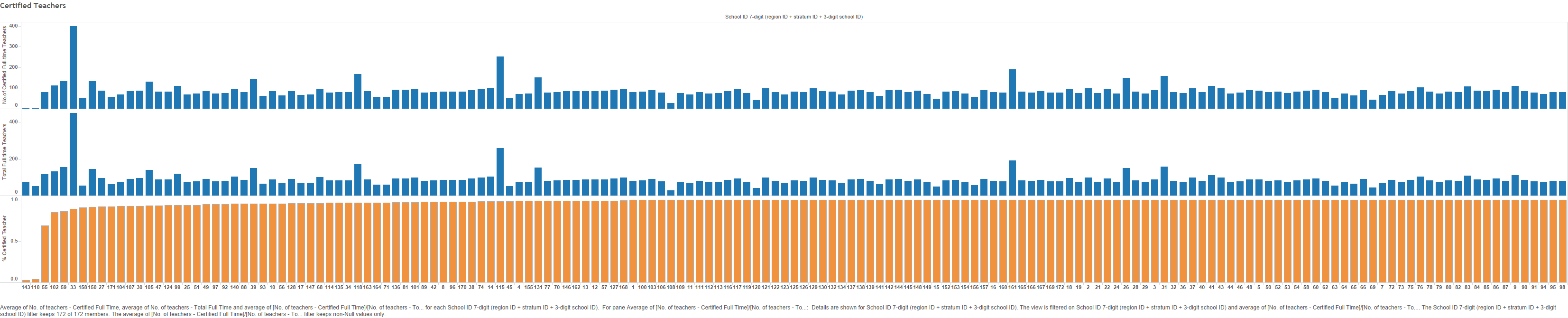
*Figure 20: Scatterplot of Math scores & Proportion of qualified teachers*

We can see that proportion of teachers is essential to the performance schools, especially in subjects like Mathematics. Figure 20 shows a relatively strong positive correlation between student performance in Mathematics and proportion of qualified teachers in schools.

*Figure 21a: Distribution of proportion of part-time teachers to total teacher population*



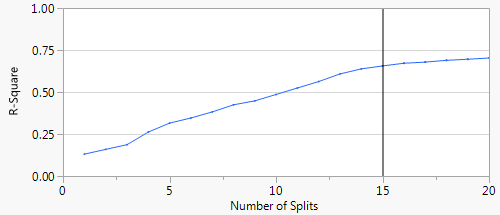
*Figure 21b: Distribution of proportion of certified teachers to total teacher population*



Figures 21a and 21b illustrate the proportionate staffing levels of part-time and full-teachers in schools. Not all schools have desirable staffing levels, with some having as much as 20% of their teachers working part-time.

## 5.2 Profiling high performance and low performance schools

### 5.2.1 Decision tree analysis



*Figure 22: Split history of decision tree and resulting RSquare*

As addressed in previous section, based on the clusters, we conducted a decision tree analysis to identify the different profiles of the schools in each cluster. Figure 22 shows the split history with the resulting R-square of each split. Splitting can go up to 20 levels, but we decided on 15 splits, since there is marginal improvement in R-square beyond that point. We then used the tree to profile the characteristics of high performing and low performing schools.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 4: Column Contributions by partition analysis (Outcome = Cluster)** | | | |
| Term | No. of splits | G2 | Portion |
| Learning Hindrance - Students truancy | 1 | 47.0469306 | 0.1912 |
| Student-Teacher ratio | 1 | 26.98793766 | 0.1097 |
| Index of school responsibility for curriculum and assessment | 2 | 25.8417386 | 0.105 |
| Proportion of girls at school | 1 | 19.1467187 | 0.0778 |
| Parental achievement pressure | 1 | 17.19588522 | 0.0699 |
| Parent Participation - Progress discussion teacher initiative | 1 | 16.30176539 | 0.0662 |
| Internet schoolwork - Assignments | 1 | 14.75292863 | 0.0599 |
| Learning Hindrance - Heterogeneous classes | 1 | 14.6993767 | 0.0597 |
| Index of school responsibility for resource allocation | 1 | 13.52388809 | 0.055 |
| Learning Hindrance - Students being late | 1 | 11.48590648 | 0.0467 |
| Responsibility - Course content - Teachers | 1 | 11.36385015 | 0.0462 |
| Teacher participation in leadership | 1 | 11.10997139 | 0.0451 |
| Parent Participation - School government | 1 | 9.41657479 | 0.0383 |
| Ratio of computers for education and number of students in the <national modal grade for 15-year-olds> | 1 | 7.226757448 | 0.0294 |

Table 4 above shows the top factors that differentiate between the low, average and high performing schools*.* Based on the results, quality of student attendance is the biggest differentiator of school performance as the distinction between the 2 clusters is biggest when it comes to truancy. Majority of good schools have zero cases of truancy and better ones have no cases of late attendance (Visual decision tree in Appendix A). This is consistent with studies that show better attendance is related to higher academic achievement for students of all backgrounds (Epstein & Sheldon, 2002).

|  |  |
| --- | --- |
| High performing schools | Low performing schools |
| * No incidence of truancy * Highly responsible towards curriculum & assessment | * Higher rate of truancy * Insufficient attention to course curriculum * Lack of parent initiative to supervise child’s progress |
| *Figure 23: Profiles of low and high performing schools* | |

Figure 23 highlights the key characteristics that define and differentiate high performing schools from low performing ones. High performing schools have no incidence of truancy and take charge of their curriculum to cater to their unique circumstances, whereas low performing ones see a higher rate of truancy and pay less attention to their course curriculum. Parents of students in those schools are also less involved in their child’s learning process.

## 5.3 Multiple regression model

### 5.3.1 Numeric factors that affect test subject performance

**Table 5: Multiple regression model using numeric factors (for overall score)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| R-Square = 0.404553  R-Square Adj = 0.381503  Observations = 162 | Y = Score (%All) | | | | |
| Parameters | Estimate | Std Error | T Ratio | P-value | VIF |
| Proportion of teachers with ISCED 5A | 0.852302 | 0.18538 | 4.6 | <.0001 | 1.087449 |
| Index of school responsibility for resource allocation | 0.036807 | 0.010833 | 3.4 | 0.0009 | 1.084648 |
| Extracurricular creative activities at school | 0.03174 | 0.010892 | 2.91 | 0.0041 | 1.03161 |
| Teacher Morale | 0.031614 | 0.007767 | 4.07 | <.0001 | 1.063242 |
| Parent Participation - Progress discussion teacher initiative | -0.0007 | 0.000233 | -2.99 | 0.0033 | 1.026621 |
| Students leaving without certificate | -0.02007 | 0.006969 | -2.88 | 0.0045 | 1.041224 |

Using multiple regression model of the numeric factors, we identified the factors that affect the overall school test scores (See Table 5). The model shows a good fit with a R-square of about 0.404, covering 40.4% of the data given. The p-value is the probability of obtaining the estimated value of the parameter if the actual parameter is zero. The smaller the p value, the more significant the parameter is and the less likely it is that the actual parameter value is zero. As seen in Table 5, the p-value of all factors are below 0.05 and VIF is relatively small, indicating that all above factors are statistically significant with little collinearity between them. (Full results in Appendix B)

Looking at the estimates, we see a strong positive correlation between proportion of teachers with tertiary education and school test scores, indicating that quality teaching staff is key to good school performance. It is also important to ensure teacher morale is high, with its positive estimate of 0.031614.

**Table 6: Significance of selected numeric attributes to school performance scores**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameters | Y = %All | | Y = %Math | | Y = %Read | | Y = %Science | | Y = %CBAM | |
|  | *p* | Est. | *p* | Est. | *p* | Est. | *p* | Est. | *p* | Est. |
| Proportion of teachers with ISCED 5A | 0.0000 | 0.8523 | 0.0000 | 0.8805 | 0.0001 | 0.7136 | 0.0001 | 0.7640 | 0.0014 | 0.7720 |
| Index of school responsibility for resource allocation | 0.0009 | 0.0368 | 0.0283 | 0.0289 | 0.0003 | 0.0388 | 0.0041 | 0.0338 |  |  |
| Extracurricular creative activities at school | 0.0041 | 0.0317 | 0.0067 | 0.0326 | 0.0205 | 0.0245 | 0.0083 | 0.0303 |  |  |
| Teacher Morale | 0.0001 | 0.0316 | 0.0000 | 0.0352 | 0.0009 | 0.0249 | 0.0001 | 0.0316 | 0.0001 | 0.0397 |
| Parent Participation - Progress discussion teacher initiative | 0.0033 | -0.0007 | 0.0040 | -0.0007 | 0.0054 | -0.0006 | 0.0053 | -0.0007 | 0.0013 | -0.0010 |
| Students leaving without certificate | 0.0045 | -0.0201 | 0.0111 | -0.0193 | 0.0002 | -0.0263 | 0.0049 | -0.0207 |  |  |
| Index of school responsibility for curriculum and assessment |  |  |  |  |  |  | 0.0895 | 0.0163 |  |  |
| Proportion of girls at school |  |  |  |  | 0.0865 | 0.0585 |  |  |  |  |
| School Autonomy |  |  | 0.1134 | 0.0200 |  |  |  |  | 0.0001 | 0.0547 |
| Student-Teacher ratio |  |  |  |  | 0.0113 | 0.0036 |  |  |  |  |

This remains consistent when we extended the analysis to different subjects – Math, Science, Reading and Computer-based assessments – as seen in Table 6 above. For comparison, we looked to the log worth, p-value and estimates of each of the parameters to determine their importance across the subjects. While all factors affecting the overall score are significant for the individual subjects, some factors are more pertinent to specific subjects. For instance, having a good student-teacher ratio is important for reading scores and it is important for the schools to have control over their operations for students to do well in science and computer based assessments. Nevertheless, proportion of tertiary-educated teachers remains the top factor across the board and thus, having well-qualified teachers with high staff morale is really crucial for school performance.

Surprisingly, 27% of Singapore schools reported a shortage of teaching staff, especially for science teachers. And it is top amongst the all resource shortages reported by schools. However, as the data is highly aggregated and categorical– they only report the shortages in categories (e.g. “a lot”, “to some extent”, “very little” and “not at all”) – it is very difficult for us to examine the extent of the shortage if we are not able to quantify them. It is thus subjected to the school’s opinion and our analysis results would not be comparable.

### 5.3.2 Categorical factors that affect test subject performance

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 7: Significance of selected categorical attributes to school performance scores** | | | | | | | | | | | | | | | |
| Parameters | Overall score | | | Math score | | | Reading score | | | Science score | | | Computer-based score | | |
| Rsquare | 0.441771 | | | 0.392877 | | | N/A | | | 0.39406 | | | 0.281219 | | |
|  | p-value | Estimate | VIF | p-value | Estimate | VIF | p-value | Estimate | VIF | p-value | Estimate | VIF | p-value | Estimate | VIF |
| Learning Hindrance - Students truancy{A lot&To some extent&Very little-Not at all} | 0.0000 | -0.0610 | 1.1502 | 0.0000 | -0.0677 | 1.1480 |  |  |  | 0.0000 | -0.0618 | 1.1480 | 0.0000 | -0.0830 | 1.0397 |
| Parental achievement pressure{Largely absent&Minority of Parents-Many Parents} | 0.0002 | -0.0288 | 1.0854 | 0.0002 | -0.0310 | 1.0479 |  |  |  | 0.0001 | -0.0311 | 1.0479 | 0.0036 | -0.0286 | 1.0397 |
| Activities - School play {Yes-No} | 0.0024 | -0.0241 | 1.0573 |  |  |  |  |  |  |  |  |  |  |  |  |
| Learning Hindrance - Students being late{A lot&To some extent-Very little&Not at all} | 0.1210 | -0.0173 | 1.0619 |  |  |  |  |  |  |  |  |  |  |  |  |
| Responsibility - Budget allocation - <School governing board> | 0.1436 | -0.0105 | 1.0169 |  |  |  |  |  |  |  |  |  |  |  |  |
| Activities - Chess club {Yes-No} |  |  |  | 0.0008 | 0.0667 | 1.1333 |  |  |  | 0.0004 | 0.0674 | 1.1333 |  |  |  |

We repeated the multiple regression analysis for categorical attributes and Table 7 shows the most significant categorical factors that affect school performance. Student attendance remains the most significant factor, as student truancy has the largest estimate and smallest p value indicating that it is the most significant factor across the board. Student lateness is also significant to the overall test scores. From Table 7, we can see that students also perform better in the presence of parental achievement pressure, when parents show concern for their child’s progress and give them the pressure to do well academically. Some school activities actually show to have a positive effect on results, related to Mathematics, chess and theatre, possibly because these activities allow children to engage in subject-related activities or train their logic and strategic thinking (Smith & Cage, 2000).

# 6. Student-level Findings

Given that the multiple regression models for school level analysis only cover less than half (40.4%) of the dataset examined, looking at school environmental factors will not provide us with the complete picture. We will need to look into factors related to the family background and home environment of students to give more insights in explaining the differences in student performance.

## 6.1 Exploratory Data Analysis

|  |  |
| --- | --- |
|  |  |
| *Figure 24a: Boxplot of scores (parents’ education)* | *Figure 24b: Boxplot of scores (parents’ job status)* |

Results in Figures 24a and 24b show that parents’ education levels and employment status do impact student performance. From Figure 24a, we can see that students with parents that have at least an O level qualification tend to do better. Interestingly, those with parents working full time also tend to do better as seen from Figure 24b.

## 6.2 Multiple regression model

### 6.2.1 Variable selection

For our analysis, we considered a combination of family and personal factors in the following areas as seen in Figure 11, such as the student’s attitude towards school and learning, parents’ education and occupation, availability of educational resources, computer habits and subject-related activities engaged in and out of school. (Full list of attributes in Appendix C)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Student attitude | Family & academic background | Access & usage of resources | Subject learning behavior | Classroom behavior |
| * Attitude towards School - Does Little to Prepare Me for Life * Attitude towards School: Learning Activities * Attitudes - Not suitable for schoolwork * Attitudes - Too unreliable * Attitudes Towards Computers: Computer as a Tool for School Learning * Attitudes Towards Computers: Limitations of the Computer as a Tool for School Learning * Mathematics Behaviour * Mathematics Intentions * Mathematics Self-Concept * Mathematics Self-Efficacy * Mathematics Work Ethic * Perceived Control - Perform Poorly Regardless * Subjective Norms in Mathematics | * Age of arrival in <country of test> * Father Qualifications - <ISCED level 5A> * Grade compared to modal grade in country * Grade Repetition * Highest educational level of parents * Highest parental education in years * Highest parental occupational status * Index of economic, social and cultural status * Language at home * Miss 2 months of <ISCED 2> * Mother<Highest Schooling> | * At Home - Printer * Home educational resources * How many books at home * ICT Availability at Home * ICT Availability at School * ICT Use at Home for School-related Tasks * Internet at School * Out-of-school - Use email * Possessions – literature * Time of computer use (min) * Use of ICT at School | * Maths Behaviour - Computer programming * Use of ICT in Mathematic Lessons | * Cognitive Activation in Mathematics Lessons * Mathematics Teacher's Classroom Management * Student Orientation - Plans Classroom Activities * Teacher Behaviour: Formative Assessment * Teacher Behaviour: Student Orientation * Teacher Behaviour: Teacher-directed Instruction * Vignette Classroom Management - Students Frequently Interrupt/Teacher Arrives Early |

*Figure 25: Attributes selected for student level analysis*

### 6.2.2 Multiple regression

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 8: Regression analysis results (selected)** | | | |
| **Parameters** | **Y=%All** | | |
|  | **Estimate** | **p-value** | **VIF** |
| Index of economic, social and cultural status |  |  |  |
| How many books at home | -0.018842067 | <.0001 | 1.20935752 |
| Internet at School | -0.028495637 | <.0001 | 1.12370357 |

Research shows that financial conditions will have a substantial impact on their academic achievement as those from poor families are likely to have adverse home environments which would affect their development (Dahl & Lochner, 2005). However, we observe from Table 8 that economic socio status of schools is actually insignificant to student’s test performance. This might be due to the nature of the survey. The PISA survey is designed for use on an international scale and since extreme poverty is largely absent in Singapore unlike other countries, the economic-socio index used might not be distinguishing enough for the purpose of our analysis. Nevertheless, results do show that having better access and appropriate usage of educational resources like books and computers will improve test performance.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 9: Regression analysis results (selected)** | | | | | | |
| Parameters | Overall | | | | | |
| Rsquare | 0.555544 | | | | | |
|  | Estimate | Std Error | t Ratio | p-value | VIF |
| Mathematics Self-Efficacy | 0.0532 | 0.0039 | 13.7700 | <.0001 | 1.3053 |
| Miss 2 months of <ISCED 2> | -0.0354 | 0.0077 | -4.5900 | <.0001 | 1.0600 |
| Grade Repetition | -0.0344 | 0.0077 | -4.4400 | <.0001 | 1.0537 |
| Internet at School | -0.0285 | 0.0040 | -7.2100 | <.0001 | 1.1237 |
| Attitude towards School - Does Little to Prepare Me for Life | -0.0243 | 0.0038 | -6.4100 | <.0001 | 1.1419 |
| Attitudes - Too unreliable | -0.0224 | 0.0041 | -5.4600 | <.0001 | 1.3228 |
| How many books at home | -0.0188 | 0.0040 | -4.6900 | <.0001 | 1.2094 |
| Perceived Control - Perform Poorly Regardless | -0.0183 | 0.0045 | -4.0800 | <.0001 | 1.2159 |
| Student Orientation - Plans Classroom Activities | -0.0174 | 0.0047 | -3.6800 | 0.0002 | 1.0902 |
| Mathematics Intentions | -0.0166 | 0.0037 | -4.5400 | <.0001 | 1.1282 |
| At Home - Printer | -0.0161 | 0.0045 | -3.6200 | 0.0003 | 1.3634 |
| Possessions - literature | -0.0149 | 0.0038 | -3.9000 | 0.0001 | 1.2663 |
| ICT Availability at Home | -0.0146 | 0.0049 | -2.9700 | 0.0030 | 1.4248 |
| Attitudes - Not suitable for schoolwork | -0.0127 | 0.0037 | -3.4100 | 0.0007 | 1.2559 |
| ICT Availability at School | -0.0127 | 0.0040 | -3.1800 | 0.0015 | 1.1270 |
| Attitudes Towards Computers: Computer as a Tool for School Learning | -0.0101 | 0.0038 | -2.6600 | 0.0080 | 1.1594 |

While material factors do have an influence on performance, students’ attitude towards school and learning is more crucial to their performance, as seen from Table 9 above. Personal attitude factors such as self-efficacy have a larger estimate as compared to home possession factors such as availability of books, printers and computers at home and in school. It is important for students to be confident in their abilities and believe in the usefulness of schools in preparing them for real life. (Full results in Appendix D)

### 6.2.3 Factors that affect low performing schools

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 10: Multiple regression results for student performance in high and low performing schools** | | | | | | | | | | |
| Parameters | **Students from high-performing schools** | | | | | **Students from low-performing schools** | | | | |
| Rsquare | 0.1324 | | | | | 0.4399 | | | | |
|  | **Estimate** | **Std Error** | **t Ratio** | **p-value** | **VIF** | **Estimate** | **Std Error** | **t Ratio** | **p-value** | **VIF** |
| ICT Use at Home for School-related Tasks |  |  |  |  |  | 0.0274 | 0.0076 | 3.6100 | 0.0003 | 1.2568 |
| Attitudes Towards Computers: Computer as a Tool for School Learning |  |  |  |  |  | -0.0187 | 0.0074 | -2.5200 | 0.0121 | 1.1352 |
| Use of ICT in Mathematic Lessons |  |  |  |  |  | -0.0191 | 0.0066 | -2.9200 | 0.0037 | 1.0959 |
| Attitude towards School - Does Little to Prepare Me for Life{Agree&Strongly agree-Strongly disagree&Disagree} |  |  |  |  |  | -0.0248 | 0.0068 | -3.6400 | 0.0003 | 1.0928 |
| Teacher Behaviour: Student Orientation |  |  |  |  |  | -0.0253 | 0.0070 | -3.6000 | 0.0003 | 1.1514 |
| Vignette Classroom Management - Students Frequently Interrupt/Teacher Arrives Early{Strongly agree&Agree-Disagree&Strongly disagree} |  |  |  |  |  | -0.0255 | 0.0068 | -3.7600 | 0.0002 | 1.0897 |
| Language at home (3-digit code){Malay-Tamil&Another language (SGP)&Chinese&English} |  |  |  |  |  | -0.0257 | 0.0080 | -3.2100 | 0.0014 | 1.0827 |
| Perceived Control - Perform Poorly Regardless{Strongly agree&Agree-Disagree&Strongly disagree} |  |  |  |  |  | -0.0302 | 0.0075 | -4.0200 | <.0001 | 1.1410 |
| Attitudes Towards Computers: Limitations of the Computer as a Tool for School Learning |  |  |  |  |  | -0.0314 | 0.0075 | -4.1800 | <.0001 | 1.1047 |
| Internet at School{Between 2 and 4 hours&Between 1 and 2 hours&Between 4 hours and 6 hours&31-60 minutes&More than 6 hours-1-30 minutes&No time} | -0.0180 | 0.0051 | -3.5200 | 0.0005 | 1.0103 | -0.0329 | 0.0072 | -4.5800 | <.0001 | 1.1234 |
| Use of ICT at School |  |  |  |  |  | -0.0337 | 0.0079 | -4.2700 | <.0001 | 1.3001 |
| How many books at home{0-10 books&11-25 books-26-100 books&101-200 books&More than 500 books&201-500 books} |  |  |  |  |  | -0.0349 | 0.0067 | -5.2400 | <.0001 | 1.0422 |
| Age of arrival in <country of test>{9&10&11&12&13&14-15} | -0.0651 | 0.0147 | -4.4200 | <.0001 | 1.0024 |  |  |  |  |  |
| Attitudes - Not suitable for schoolwork{Strongly agree&Agree-Strongly disagree&Disagree} | -0.0241 | 0.0046 | -5.2000 | <.0001 | 1.0170 |  |  |  |  |  |
| Highest parental education in years | 0.0057 | 0.0017 | 3.3000 | 0.0010 | 1.0253 |  |  |  |  |  |
| Miss 2 months of <ISCED 2>{Yes, once&Yes, twice or more-No, never} | -0.0456 | 0.0123 | -3.7200 | 0.0002 | 1.0119 |  |  |  |  |  |

As the difference in student performance between low performing and high performing schools is significant (difference of 0.4 in their mean scores), factors that affect their scores may be different. Hence, we built separate multiple regression models for students in low and high performing schools to better identify factors that will truly affect their performance.

Table 10 shows the most significant factors identified through the regression analysis. All factors featured are statistically significant and there is little collinearity. Once again, we see that control over student usage of resources is important as an overuse of Internet at school will negatively affect the performance of all students, particularly those in low performing schools. Personal attitude and family background are more significant to student performance in low performing schools as compared to high performing ones.

**Table 11: Most significant factors to student performance in low performing schools (descending order of significance)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Parameters | **Students from low-performing schools** | | | | |
| Rsquare | 0.4399 | | | | |
|  | **Estimate** | **Std Error** | **t Ratio** | **p-value** | **VIF** |
| How many books at home{0-10 books&11-25 books-26-100 books&101-200 books&More than 500 books&201-500 books} | -0.0349 | 0.0067 | -5.2400 | <.0001 | 1.0422 |
| Use of ICT at School | -0.0337 | 0.0079 | -4.2700 | <.0001 | 1.3001 |
| Internet at School{Between 2 and 4 hours&Between 1 and 2 hours&Between 4 hours and 6 hours&31-60 minutes&More than 6 hours-1-30 minutes&No time} | -0.0329 | 0.0072 | -4.5800 | <.0001 | 1.1234 |
| Perceived Control - Perform Poorly Regardless{Strongly agree&Agree-Disagree&Strongly disagree} | -0.0302 | 0.0075 | -4.0200 | <.0001 | 1.1410 |
| Attitudes Towards Computers: Limitations of the Computer as a Tool for School Learning | -0.0314 | 0.0075 | -4.1800 | <.0001 | 1.1047 |
| ICT Use at Home for School-related Tasks | 0.0274 | 0.0076 | 3.6100 | 0.0003 | 1.2568 |
| Language at home (3-digit code){Malay-Tamil&Another language (SGP)&Chinese&English} | -0.0257 | 0.0080 | -3.2100 | 0.0014 | 1.0827 |
| Vignette Classroom Management - Students Frequently Interrupt/Teacher Arrives Early{Strongly agree&Agree-Disagree&Strongly disagree} | -0.0255 | 0.0068 | -3.7600 | 0.0002 | 1.0897 |
| Teacher Behaviour: Student Orientation | -0.0253 | 0.0070 | -3.6000 | 0.0003 | 1.1514 |
| Attitude towards School - Does Little to Prepare Me for Life{Agree&Strongly agree-Strongly disagree&Disagree} | -0.0248 | 0.0068 | -3.6400 | 0.0003 | 1.0928 |
| Use of ICT in Mathematic Lessons | -0.0191 | 0.0066 | -2.9200 | 0.0037 | 1.0959 |
| Attitudes Towards Computers: Computer as a Tool for School Learning | -0.0187 | 0.0074 | -2.5200 | 0.0121 | 1.1352 |

For low performing schools, it is important for students to have good access to resources, as seen from Table 11 where the availability of books and computers are parameters with the largest estimates with significant p-values. Use of computers in school and home to aid academic related tasks will help students score better.

Next would be the student’s personal attitude and mindset. Those that do well tend to have a strong internal locus of control, meaning they believe that their performance is primarily determined by through their own efforts. Those who believe that they will perform poorly regardless, tend to do worse. The effect of these personal factors exceeds the impact of classroom management, such as the amount of disruption in class and teachers’ efforts to engage students.

# 7. Discussion

## 7.1 General recommendation

Overall, all 3 aspects - school, family & personal factors - have a combined effect on the student and so a holistic view is required to improve student performance.



*Figure 26: Recommended general approach to improving school performance*

In general, based on our research, we recommend schools take a 2-pronged approach to improving student performance as shown in Figure 26. The first step to “detect” is for schools to identify high-risk students based on various factors such as their attitudes and self-esteem, access to educational resources, their computer habits and family background (Figure 26). Findings from our research will be able to help schools better understand such factors. Early detection of students at risk of poor performance to provide assistance as early as possible will strengthen the safety net in schools.

The next step is to “prevent”. While detecting and helping high-risk students helps to improve performance, it is important for schools to continuously improve their learning environment and increase the overall standard of their school to improve the performance of the entire student population and not just those in need. In particular, addressing the problem of shortage of quality staff is crucial for all schools since having good teaching staff is the top factor that impacts school test scores. Improving access to resources and increasing parent participation will also help improve academic performance of students in an overall encouraging and engaging learning environment.

## 7.2 Gaps identified in data for future research efforts

Through our project we also noted the following gaps in the data that can be addressed for future research efforts.

**Lack of numeric data for attributes**

Data for resource shortages in schools and learning hindrances (e.g. student lateness, teacher absenteeism, skipping events) are highly aggregated categorical data. Obtaining quantities for such attributes will be useful for analysis purposes. In addition, there is little detail about the usage of school funding. Such information on school spending is closely related to their resource allocation which will have an impact on school environment and hence, student performance. Hence, for these gaps, more detailed numeric data will need to be collected in order to achieve a more in-depth analysis of factors affecting school performance.

**Lack of data for investigation into student truancy**

From our decision tree results at the school level analysis, student truancy is biggest differentiator of school performance, however, we are unable to identify factors that explain student truancy, as the data does not provide information in this aspect. Hence we recommend that surveys can be directed to address the gap in this aspect.

**Lack of local data**

Data for the economic socio index used in an international survey is not differentiating enough for our in-depth analysis of Singapore’s local education landscape. Despite availability of data related to access to educational resources, it may not be an accurate indicator of wealth in Singapore’s context due to the ease of access to books, computers and other educational resources through public facilities and financial assistance schemes available. In addition, due to Singapore’s housing policy, even the poorest typically have a roof over their heads (Chan & Basu., 2013). Thus, more accurate data such as family income and family’s participation in any financial assistance schemes will be able to give more accurate results for our model.

## 7.3 Conclusion

Overall, Singapore schools have performed well in the international arena, but more can be done to ensure that all schools are able to achieve high and consistent academic performance, achieving MOE’s goal of making “every school a good school”.

While data from the PISA survey is rich for analysis, the nature of the survey catered to the international scale means that data collected tends to be highly aggregated and generalized for comparability. Hence, there still exists a substantial data gap that needs to be filled in order for academics to conduct an in-depth analysis of Singapore’s education landscape and provide more meaningful and relevant insights to improving the education practice in Singapore.

# References

1. Epstein, J. L., & Sheldon, S. B. (2002). *Present and accounted for: Improving student attendance through family and community involvement.*Journal of Educational Research, 95(5), 308-318. Retrieved from http://www.childtrends.org/?indicators=student-absenteeism#\_edn1
2. Gordon Dahl & Lance Lochner (August 2005). *The Impact of Family Income on Child Achievement.* Retrieved from http://www.irp.wisc.edu/publications/dps/pdfs/dp130505.pdf
3. R. Chan & R. Basu. (2013, October 26). *The Invisible Poor*? Retrieved from http://lkyspp.nus.edu.sg/wp-content/uploads/2013/06/straitstimes.com-The\_invisible\_poor.pdf
4. Smith, J.P. and Cage, B. N. (2000). The effects of chess instruction on the mathematics achievements of southern, rural, black secondary students. Research in the Schools, 7, 19-26.

# Appendix

## Appendix A: Decision tree analysis results

## 

## Appendix B: Multiple regression results for numeric attributes (School level analysis)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameters | Overall score | | | Math score | | | Reading score | | | Science score | | | Computer-based score | | |
| Rsquare | 0.404553 | | | 0.40834 | | | 0.415055 | | | 0.38989 | | | 0.294477 | | |
|  | Log  Worth | p-value | Est | Log  Worth | p-value | Est | Log  Worth | p-value | Est | Log  Worth | p-value | Est | Log  Worth | p-value | Est |
| Proportion of teachers with ISCED 5A | 5.0550 | 0.0000 | 0.8523 | 4.6620 | 0.0000 | 0.8805 | 4.0640 | 0.0001 | 0.7136 | 3.9090 | 0.0001 | 0.7640 | 2.8650 | 0.0014 | 0.7720 |
| Index of school responsibility for resource allocation | 3.0640 | 0.0009 | 0.0368 | 1.5480 | 0.0283 | 0.0289 | 3.6040 | 0.0003 | 0.0388 | 2.3890 | 0.0041 | 0.0338 |  |  |  |
| Extracurricular creative activities at school | 2.3880 | 0.0041 | 0.0317 | 2.1750 | 0.0067 | 0.0326 | 1.6880 | 0.0205 | 0.0245 | 2.0820 | 0.0083 | 0.0303 |  |  |  |
| Teacher Morale | 4.1280 | 0.0001 | 0.0316 | 4.3620 | 0.0000 | 0.0352 | 3.0320 | 0.0009 | 0.0249 | 3.8620 | 0.0001 | 0.0316 | 4.1170 | 0.0001 | 0.0397 |
| Parent Participation - Progress discussion teacher initiative | 2.4850 | 0.0033 | -0.0007 | 2.3940 | 0.0040 | -0.0007 | 2.2700 | 0.0054 | -0.0006 | 2.2770 | 0.0053 | -0.0007 | 2.8940 | 0.0013 | -0.0010 |
| Students leaving without certificate | 2.3430 | 0.0045 | -0.0201 | 1.9530 | 0.0111 | -0.0193 | 3.7990 | 0.0002 | -0.0263 | 2.3120 | 0.0049 | -0.0207 |  |  |  |
| Index of school responsibility for curriculum and assessment |  |  |  |  |  |  |  |  |  | 1.0480 | 0.0895 | 0.0163 |  |  |  |
| Proportion of girls at school |  |  |  |  |  |  | 1.0630 | 0.0865 | 0.0585 |  |  |  |  |  |  |
| School Autonomy |  |  |  | 0.9460 | 0.1134 | 0.0200 |  |  |  |  |  |  | 4.2080 | 0.0001 | 0.0547 |
| Student-Teacher ratio |  |  |  |  |  |  | 1.9460 | 0.0113 | 0.0036 |  |  |  |  |  |  |

## Appendix C: Attributes selected for student level analysis

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student attitude** | **Family & academic background** | **Access & usage of resources** | **Subject learning behavior** | **Classroom behavior** |
| * Attitude towards School - Does Little to Prepare Me for Life{Strongly agree&Agree-Disagree&Strongly disagree} * Attitude towards School: Learning Activities (Anchored) * Attitudes - Not suitable for schoolwork{Strongly agree&Agree&Strongly disagree-Disagree} * Attitudes - Not suitable for schoolwork{Strongly agree&Agree-Strongly disagree&Disagree} * Attitudes - Too unreliable{Strongly agree&Agree-Disagree&Strongly disagree} * Attitudes Towards Computers: Computer as a Tool for School Learning * Attitudes Towards Computers: Limitations of the Computer as a Tool for School Learning * Mathematics Behaviour * Mathematics Intentions * Mathematics Self-Concept (Anchored) * Mathematics Self-Efficacy * Mathematics Work Ethic (Anchored) * Perceived Control - Perform Poorly Regardless{Strongly agree&Agree-Strongly disagree&Disagree} * Subjective Norms in Mathematics | * Age of arrival in <country of test>{9&10&11&12&13&14-15} * Father Qualifications - <ISCED level 5A> * Grade compared to modal grade in country * Grade Repetition{Repeated a <grade>-Did not repeat a <grade>&M} * Highest educational level of parents{ISCED 1&None&ISCED 2&ISCED 3A, ISCED 4&ISCED 5B-ISCED 5A, 6} * Highest parental education in years * Highest parental occupational status * Index of economic, social and cultural status * Language at home (3-digit code){Malay&Tamil-Chinese&Another language (SGP)&English} * Language at home (3-digit code){Malay-Tamil&Another language (SGP)&Chinese&English} * Miss 2 months of <ISCED 2>{Yes, once&Yes, twice or more-No, never} * Mother<Highest Schooling>{She did not complete <ISCED level 1>&<ISCED level 1>&<ISCED level 2>-<ISCED level 3A>} | * At Home - Printer{No-Yes, but I don’t use it&Yes, and I use it} * Home educational resources * How many books at home{0-10 books&11-25 books-26-100 books&101-200 books&More than 500 books&201-500 books} * ICT Availability at Home * ICT Availability at School * ICT Use at Home for School-related Tasks * Internet at School{Between 2 and 4 hours&Between 1 and 2 hours&Between 4 hours and 6 hours&31-60 minutes&More than 6 hours-1-30 minutes&No time} * Internet at School{Between 4 hours and 6 hours&31-60 minutes&Between 2 and 4 hours&Between 1 and 2 hours-1-30 minutes&No time&More than 6 hours} * Out-of-school 8 - Use email{Never or hardly ever&Once or twice a month-Once or twice a week&Almost every day&Every day} * Possessions – literature * Time of computer use (mins) * Use of ICT at School * Use of ICT in Mathematic Lessons | * Maths Behaviour - Computer programming{Always or almost always&Often&Sometimes-Never or rarely} | * Cognitive Activation in Mathematics Lessons (Anchored) * Mathematics Teacher's Classroom Management (Anchored) Student Orientation - Plans Classroom Activities{Every Lesson&Most Lessons-Never or Hardly Ever&Some Lessons} * Teacher Behaviour: Formative Assessment * Teacher Behaviour: Student Orientation * Teacher Behaviour: Teacher-directed Instruction * Vignette Classroom Management - Students Frequently Interrupt/Teacher Arrives Early{Strongly agree&Agree-Strongly disagree&Disagree} |

## Appendix D: Multiple regression results for student-related factors for overall & different subjects

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Parameters | Overall | | | | |
| Rsquare | 0.555544 | | | | |
|  | Estimate | Std Error | t Ratio | p-value | VIF |
| Mathematics Self-Efficacy | 0.0532 | 0.0039 | 13.7700 | <.0001 | 1.3053 |
| Cognitive Activation in Mathematics Lessons (Anchored) | 0.0200 | 0.0062 | 3.2100 | 0.0014 | 2.0599 |
| Mathematics Self-Concept (Anchored) | 0.0170 | 0.0066 | 2.5900 | 0.0096 | 2.5094 |
| Attitudes Towards Computers: Computer as a Tool for School Learning | -0.0101 | 0.0038 | -2.6600 | 0.0080 | 1.1594 |
| Out-of-school 8 - Use email{Never or hardly ever&Once or twice a month-Once or twice a week&Almost every day&Every day} | -0.0120 | 0.0035 | -3.3800 | 0.0007 | 1.1176 |
| ICT Availability at School | -0.0127 | 0.0040 | -3.1800 | 0.0015 | 1.1270 |
| Attitudes - Not suitable for schoolwork{Strongly agree&Agree&Strongly disagree-Disagree} | -0.0127 | 0.0037 | -3.4100 | 0.0007 | 1.2559 |
| ICT Availability at Home | -0.0146 | 0.0049 | -2.9700 | 0.0030 | 1.4248 |
| Possessions - literature{No-Yes} | -0.0149 | 0.0038 | -3.9000 | 0.0001 | 1.2663 |
| Attitude towards School: Learning Activities (Anchored) | -0.0156 | 0.0056 | -2.8000 | 0.0051 | 1.9001 |
| At Home - Printer{No&Yes, but I don’t use it-Yes, and I use it} | -0.0161 | 0.0045 | -3.6200 | 0.0003 | 1.3634 |
| Mathematics Intentions | -0.0166 | 0.0037 | -4.5400 | <.0001 | 1.1282 |
| Student Orientation - Plans Classroom Activities{Every Lesson&Most Lessons-Never or Hardly Ever&Some Lessons} | -0.0174 | 0.0047 | -3.6800 | 0.0002 | 1.0902 |
| Perceived Control - Perform Poorly Regardless{Strongly agree&Agree-Disagree&Strongly disagree} | -0.0183 | 0.0045 | -4.0800 | <.0001 | 1.2159 |
| Mathematics Work Ethic (Anchored) | -0.0183 | 0.0074 | -2.4800 | 0.0132 | 3.0954 |
| Father Qualifications - <ISCED level 5A>{No-Yes} | -0.0184 | 0.0041 | -4.4500 | <.0001 | 1.2088 |
| How many books at home{0-10 books&11-25 books-26-100 books&101-200 books&More than 500 books&201-500 books} | -0.0188 | 0.0040 | -4.6900 | <.0001 | 1.2094 |
| Maths Behaviour - Computer programming{Always or almost always&Often-Sometimes&Never or rarely} | -0.0191 | 0.0051 | -3.7300 | 0.0002 | 1.1221 |
| Attitudes - Too unreliable{Strongly agree&Agree-Disagree&Strongly disagree} | -0.0224 | 0.0041 | -5.4600 | <.0001 | 1.3228 |
| Attitude towards School - Does Little to Prepare Me for Life{Agree&Strongly agree-Strongly disagree&Disagree} | -0.0243 | 0.0038 | -6.4100 | <.0001 | 1.1419 |
| Vignette Classroom Management - Students Frequently Interrupt/Teacher Arrives Early{Strongly agree&Agree-Disagree&Strongly disagree} | -0.0253 | 0.0039 | -6.4300 | <.0001 | 1.3280 |
| Internet at School{Between 2 and 4 hours&Between 1 and 2 hours&Between 4 hours and 6 hours&31-60 minutes&More than 6 hours-1-30 minutes&No time} | -0.0285 | 0.0040 | -7.2100 | <.0001 | 1.1237 |
| Language at home (3-digit code){Malay-Chinese&Another language (SGP)&English&Tamil} | -0.0309 | 0.0056 | -5.5300 | <.0001 | 1.1896 |
| Grade Repetition{Repeated a <grade>-Did not repeat a <grade>&M} | -0.0344 | 0.0077 | -4.4400 | <.0001 | 1.0537 |
| Miss 2 months of <ISCED 2>{Yes, once&Yes, twice or more-No, never} | -0.0354 | 0.0077 | -4.5900 | <.0001 | 1.0600 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Parameters | Mathematics | | | | |
| Rsquare | 0.495877 | | | | |
|  | Estimate | Std Error | t Ratio | p-value | VIF |
| Mathematics Self-Efficacy | 0.06355565 | 0.004497923 | 14.13 | <.0001 | 1.291439896 |
| Mathematics Self-Concept (Anchored) | 0.038025632 | 0.007382508 | 5.15 | <.0001 | 2.379960281 |
| Attitudes Towards Computers: Computer as a Tool for School Learning | -0.017849285 | 0.004484621 | -3.98 | <.0001 | 1.16689831 |
| ICT Availability at School | -0.016545168 | 0.00458705 | -3.61 | 0.0003 | 1.126186144 |
| Attitudes - Not suitable for schoolwork{Strongly agree&Agree&Strongly disagree-Disagree} | -0.015133412 | 0.004391044 | -3.45 | 0.0006 | 1.268902135 |
| Student Orientation - Plans Classroom Activities{Every Lesson&Most Lessons-Never or Hardly Ever&Some Lessons} | -0.023847936 | 0.005713031 | -4.17 | <.0001 | 1.20768003 |
| Perceived Control - Perform Poorly Regardless{Strongly agree&Agree-Disagree&Strongly disagree} | -0.01572456 | 0.005147967 | -3.05 | 0.0023 | 1.227161779 |
| Mathematics Work Ethic (Anchored) | -0.018715822 | 0.007168867 | -2.61 | 0.0091 | 2.255805858 |
| How many books at home{0-10 books&11-25 books-26-100 books&101-200 books&More than 500 books&201-500 books} | -0.027517748 | 0.004509552 | -6.1 | <.0001 | 1.152313393 |
| Maths Behaviour - Computer programming{Always or almost always&Often-Sometimes&Never or rarely} | -0.018906937 | 0.005818697 | -3.25 | 0.0012 | 1.116478914 |
| Attitudes - Too unreliable{Strongly agree&Agree-Disagree&Strongly disagree} | -0.024011777 | 0.004802672 | -5 | <.0001 | 1.347251871 |
| Attitude towards School - Does Little to Prepare Me for Life{Agree&Strongly agree-Strongly disagree&Disagree} | -0.02524204 | 0.004411727 | -5.72 | <.0001 | 1.156857574 |
| Vignette Classroom Management - Students Frequently Interrupt/Teacher Arrives Early{Strongly agree&Agree-Disagree&Strongly disagree} | -0.017321172 | 0.004482415 | -3.86 | 0.0001 | 1.266129185 |
| Internet at School{Between 2 and 4 hours&Between 1 and 2 hours&Between 4 hours and 6 hours&31-60 minutes&More than 6 hours-1-30 minutes&No time} | -0.034184526 | 0.004614457 | -7.41 | <.0001 | 1.130912362 |
| Language at home (3-digit code){Malay-Chinese&Another language (SGP)&English&Tamil} | -0.02874052 | 0.006193277 | -4.64 | <.0001 | 1.136780895 |
| Grade Repetition{Repeated a <grade>-Did not repeat a <grade>&M} | -0.040918296 | 0.008778054 | -4.66 | <.0001 | 1.043065169 |
| Miss 2 months of <ISCED 2>{Yes, once&Yes, twice or more-No, never} | -0.025882111 | 0.008461063 | -3.06 | 0.0023 | 1.059870167 |
| Highest parental education in years | 0.007374244 | 0.001490002 | 4.95 | <.0001 | 1.190098759 |
| ICT Use at Home for School-related Tasks | 0.018686617 | 0.004897707 | 3.82 | 0.0001 | 1.275452204 |
| Teacher Behaviour: Formative Assessment | -0.011390615 | 0.004565397 | -2.49 | 0.0127 | 1.239033833 |
| Use of ICT in Mathematic Lessons | -0.012023988 | 0.004016706 | -2.99 | 0.0028 | 1.098798812 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Parameters | Reading | | | | |
| Rsquare | 0.395314 | | | | |
|  | Estimate | Std Error | t Ratio | p-value | VIF |
| Mathematics Self-Efficacy |  |  |  |  |  |
| Cognitive Activation in Mathematics Lessons (Anchored) | 0.019373028 | 0.007646379 | 2.53 | 0.0114 | 1.494616457 |
| ICT Availability at School | -0.015051658 | 0.005951838 | -2.53 | 0.0116 | 1.191500716 |
| ICT Availability at Home |  |  |  |  |  |
| Possessions - literature{No-Yes} | -0.026029539 | 0.005602129 | -4.65 | <.0001 | 1.258786354 |
| Attitude towards School: Learning Activities (Anchored) | -0.018390979 | 0.007376519 | -2.49 | 0.0128 | 1.536922418 |
| Perceived Control - Perform Poorly Regardless{Strongly agree&Agree-Disagree&Strongly disagree} | -0.03241512 | 0.006165869 | -5.26 | <.0001 | 1.109559901 |
| Father Qualifications - <ISCED level 5A>{No-Yes} | -0.020741294 | 0.006009038 | -3.45 | 0.0006 | 1.201715973 |
| How many books at home{0-10 books&11-25 books-26-100 books&101-200 books&More than 500 books&201-500 books} | -0.021531211 | 0.005853551 | -3.68 | 0.0002 | 1.169782398 |
| Attitudes - Too unreliable{Strongly agree&Agree-Disagree&Strongly disagree} | -0.025380267 | 0.00551444 | -4.6 | <.0001 | 1.100774594 |
| Attitude towards School - Does Little to Prepare Me for Life{Agree&Strongly agree-Strongly disagree&Disagree} | -0.032767605 | 0.005487391 | -5.97 | <.0001 | 1.112264791 |
| Vignette Classroom Management - Students Frequently Interrupt/Teacher Arrives Early{Strongly agree&Agree-Disagree&Strongly disagree} | -0.035188694 | 0.005648482 | -6.23 | <.0001 | 1.250136482 |
| Internet at School{Between 2 and 4 hours&Between 1 and 2 hours&Between 4 hours and 6 hours&31-60 minutes&More than 6 hours-1-30 minutes&No time} | -0.020896157 | 0.006114225 | -3.42 | 0.0007 | 1.224586947 |
| Language at home (3-digit code){Malay-Chinese&Another language (SGP)&English&Tamil} | -0.032695645 | 0.00827183 | -3.95 | <.0001 | 1.145913263 |
| Grade Repetition{Repeated a <grade>-Did not repeat a <grade>&M} | -0.036848931 | 0.010810334 | -3.41 | 0.0007 | 1.058685924 |
| Miss 2 months of <ISCED 2>{Yes, once&Yes, twice or more-No, never} | -0.067019228 | 0.01155094 | -5.8 | <.0001 | 1.030911072 |
| Age of arrival in <country of test>{9&10&11&12&13&14-15} | -0.048552489 | 0.018876082 | -2.57 | 0.0103 | 1.034459874 |
| ICT Use at Home for School-related Tasks | 0.020126329 | 0.006721353 | 2.99 | 0.0028 | 1.435830455 |
| Use of ICT at School | -0.0214773 | 0.006441332 | -3.33 | 0.0009 | 1.573545609 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Parameters | Science | | | | |
| Rsquare | 0.412564 | | | | |
|  | Estimate | Std Error | t Ratio | p-value | VIF |
| Mathematics Self-Efficacy |  |  |  |  |  |
| Cognitive Activation in Mathematics Lessons (Anchored) | 0.02558363 | 0.007781851 | 3.29 | 0.001 | 1.61878652 |
| Out-of-school 8 - Use email{Never or hardly ever&Once or twice a month-Once or twice a week&Almost every day&Every day} | -0.016013974 | 0.005062914 | -3.16 | 0.0016 | 1.076184979 |
| Attitude towards School: Learning Activities (Anchored) | -0.017899676 | 0.007214698 | -2.48 | 0.0132 | 1.504221686 |
| Perceived Control - Perform Poorly Regardless{Strongly agree&Agree-Disagree&Strongly disagree} | -0.016008021 | 0.006317403 | -2.53 | 0.0114 | 1.134461784 |
| How many books at home{0-10 books&11-25 books-26-100 books&101-200 books&More than 500 books&201-500 books} | -0.028068874 | 0.005679444 | -4.94 | <.0001 | 1.197476161 |
| Maths Behaviour - Computer programming{Always or almost always&Often-Sometimes&Never or rarely} | -0.030889486 | 0.007179033 | -4.3 | <.0001 | 1.067040777 |
| Attitudes - Too unreliable{Strongly agree&Agree-Disagree&Strongly disagree} | -0.028116919 | 0.005461661 | -5.15 | <.0001 | 1.126569149 |
| Attitude towards School - Does Little to Prepare Me for Life{Agree&Strongly agree-Strongly disagree&Disagree} | -0.042017046 | 0.005467601 | -7.68 | <.0001 | 1.15090582 |
| Vignette Classroom Management - Students Frequently Interrupt/Teacher Arrives Early{Strongly agree&Agree-Disagree&Strongly disagree} | -0.031813901 | 0.005697979 | -5.58 | <.0001 | 1.319373275 |
| Internet at School{Between 2 and 4 hours&Between 1 and 2 hours&Between 4 hours and 6 hours&31-60 minutes&More than 6 hours-1-30 minutes&No time} | -0.031069492 | 0.005516935 | -5.63 | <.0001 | 1.046417635 |
| Language at home (3-digit code){Malay-Chinese&Another language (SGP)&English&Tamil} | -0.046831548 | 0.007547166 | -6.21 | <.0001 | 1.118998269 |
| Grade Repetition{Repeated a <grade>-Did not repeat a <grade>&M} | -0.037236915 | 0.011278087 | -3.3 | 0.001 | 1.042469532 |
| Miss 2 months of <ISCED 2>{Yes, once&Yes, twice or more-No, never} | -0.039115788 | 0.010716467 | -3.65 | 0.0003 | 1.075899285 |
| Highest parental education in years | 0.010311158 | 0.001861837 | 5.54 | <.0001 | 1.217447195 |
| Teacher Behaviour: Student Orientation | -0.014116893 | 0.005345494 | -2.64 | 0.0084 | 1.125206709 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Parameters | Computer-based assessment | | | | |
| Rsquare | 0.334303 | | | | |
|  | Estimate | Std Error | t Ratio | p-value | VIF |
| Perceived Control - Perform Poorly Regardless{Strongly agree&Agree-Disagree&Strongly disagree} | -0.053218273 | 0.012436628 | -4.28 | <.0001 | 1.132992249 |
| Internet at School{Between 2 and 4 hours&Between 1 and 2 hours&Between 4 hours and 6 hours&31-60 minutes&More than 6 hours-1-30 minutes&No time} | -0.037080655 | 0.011787117 | -3.15 | 0.0018 | 1.153254193 |
| Language at home (3-digit code){Malay-Chinese&Another language (SGP)&English&Tamil} | -0.046061831 | 0.015795572 | -2.92 | 0.0038 | 1.117065232 |
| Grade Repetition{Repeated a <grade>-Did not repeat a <grade>&M} | -0.065071278 | 0.023033607 | -2.83 | 0.005 | 1.063393547 |
| Attitudes Towards Computers: Limitations of the Computer as a Tool for School Learning | -0.046413001 | 0.012158315 | -3.82 | 0.0002 | 1.058900795 |
| Highest parental occupational status | 0.002207032 | 0.000574915 | 3.84 | 0.0001 | 1.137953602 |
| Home educational resources | 0.047689135 | 0.011477989 | 4.15 | <.0001 | 1.267909738 |
| Teacher Behaviour: Teacher-directed Instruction | -0.032345232 | 0.011639054 | -2.78 | 0.0057 | 1.097474407 |
| Use of ICT at School | -0.038048504 | 0.011472505 | -3.32 | 0.001 | 1.20783465 |

1. Data can be retrieved from <https://pisa2012.acer.edu.au/downloads.php> [↑](#footnote-ref-1)