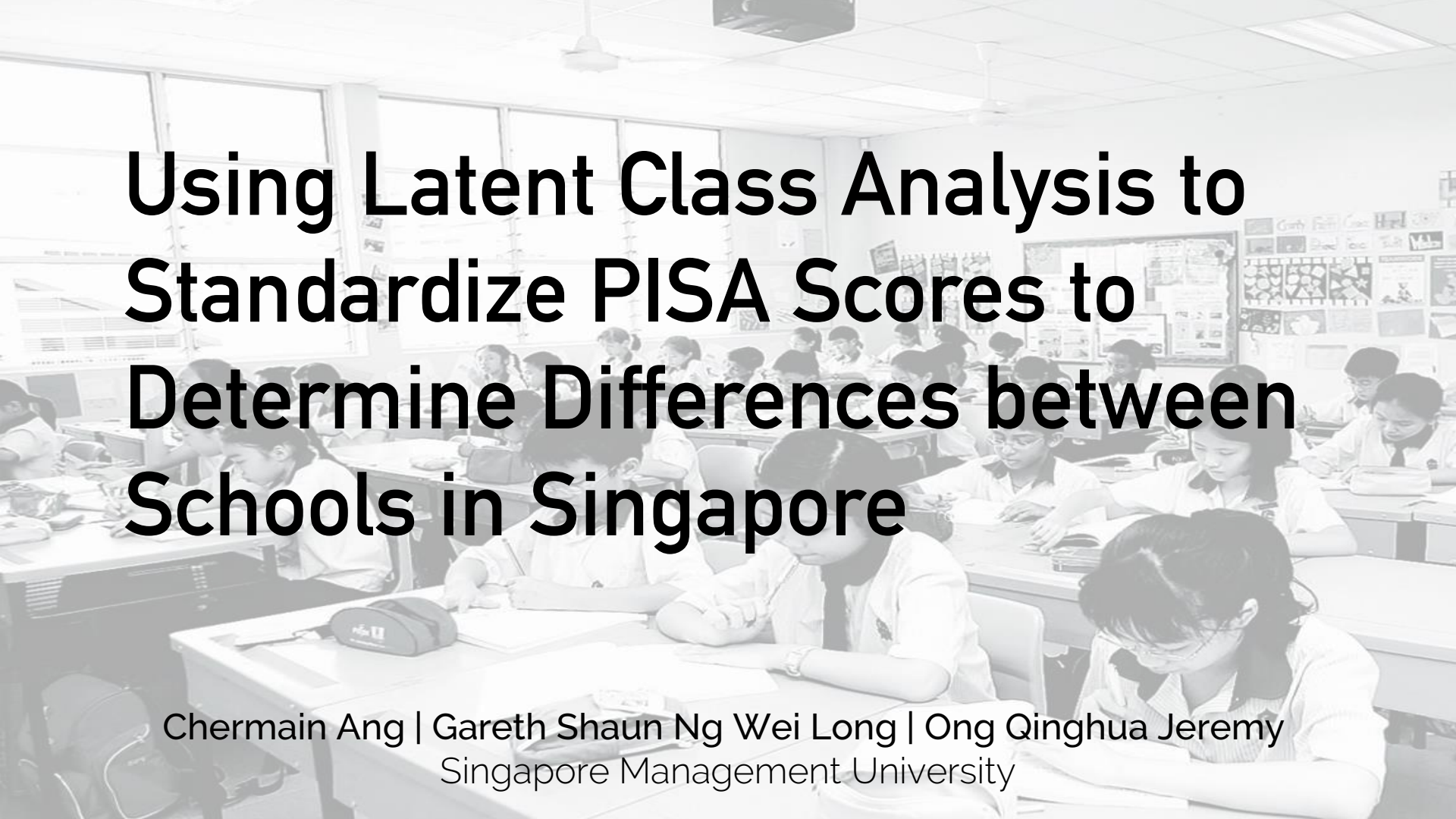


Using Latent Class Analysis to Standardize PISA Scores to Determine Differences between Schools in Singapore

Chermain Ang | Gareth Shaun Ng Wei Long | Ong Qinghua Jeremy
Singapore Management University

Agenda

Project Overview

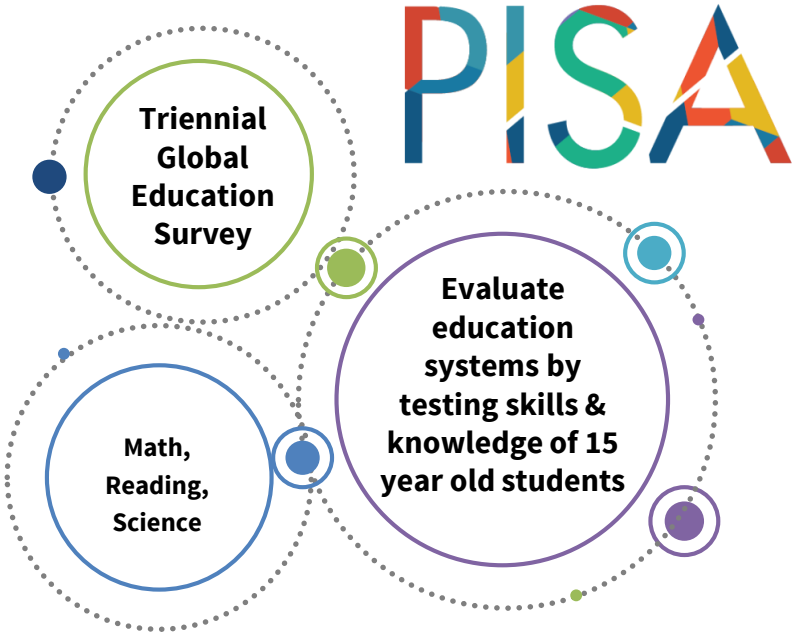
Data Preparation

Data Analysis

Insights

Conclusion

Background



HOW DID SINGAPORE STUDENTS FARE?

#1

Reading
Math
Science

Background & Motivation

Ministry of Education
SINGAPORE

"*Every school a good school*"

"*Singapore managed to achieve excellence without wide differences between children from wealthy and disadvantaged families.*"

- OECD education director Andreas Schleicher

Source: BBC

Background & Motivation

1. Are there differences across schools in Singapore?

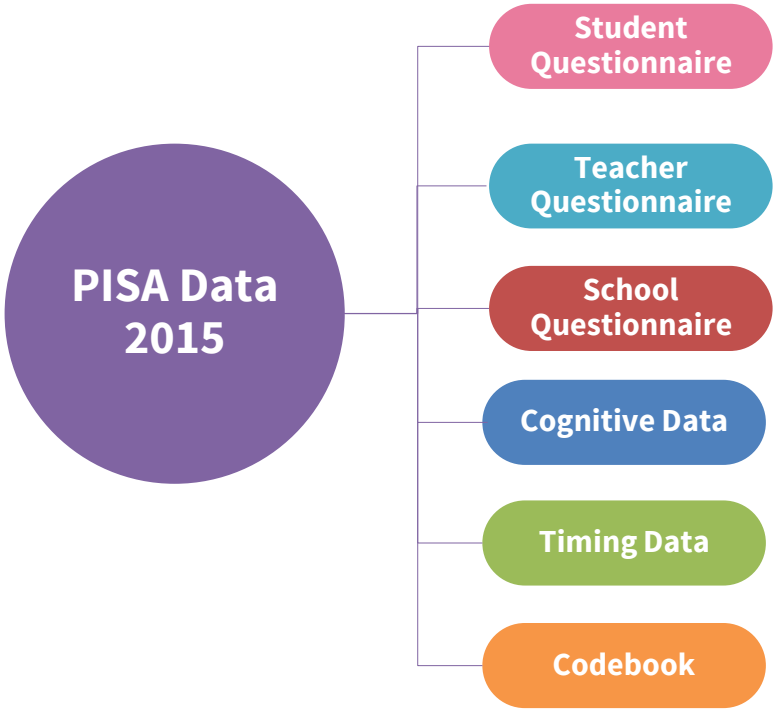
2. Find out the contributing factors for schools who did well in the test

3. Identify **factors** that affect how well students do in Reading, Mathematics, Science and overall

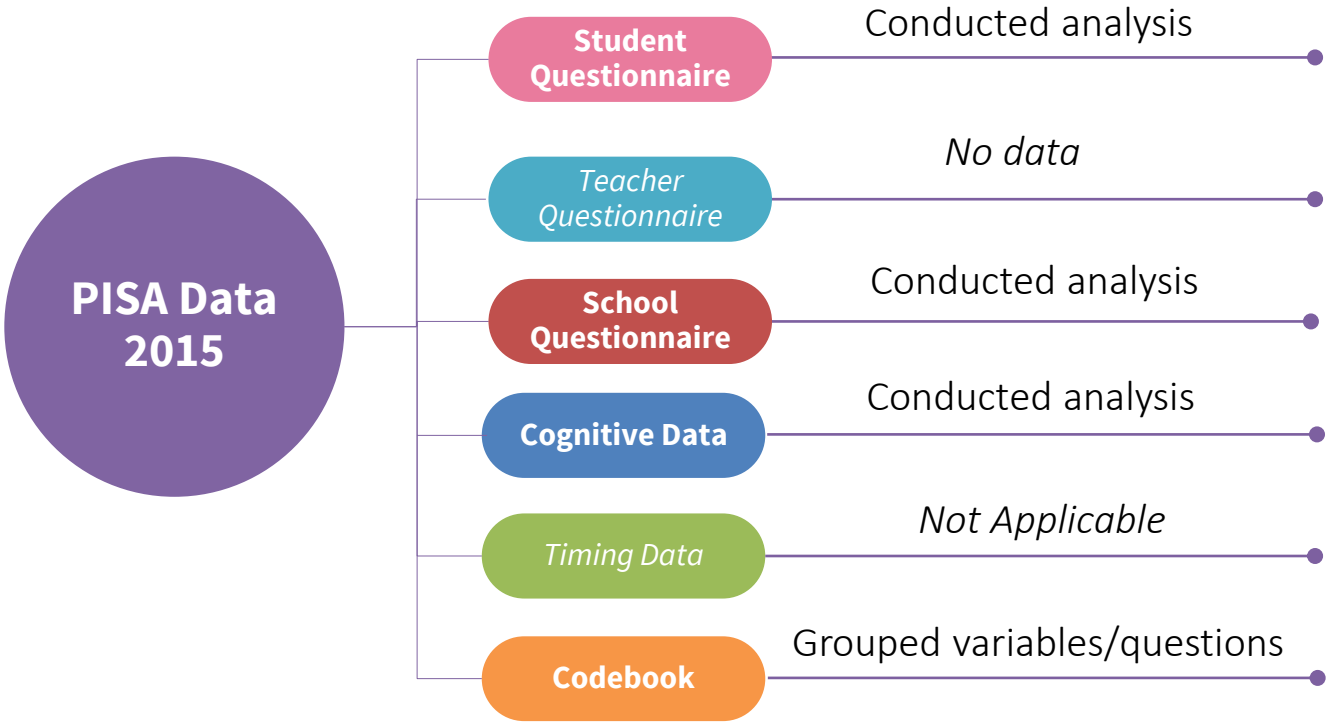
Objective

We seek to determine if there are differences between schools in Singapore based on their PISA performance

2015 PISA Data



2015 PISA Data

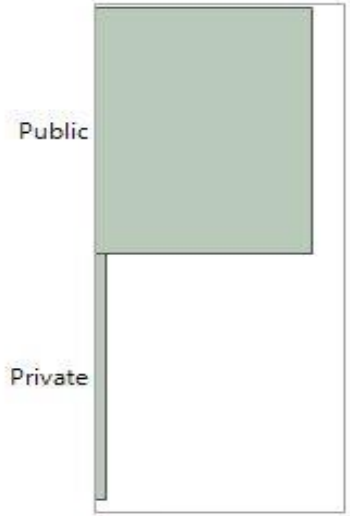


2015 PISA Data

- 6115 Students
- 168 Public Schools
  - 5,825 Students (95%)
- 9 Private Schools
  - 290 Students (5%)

Distributions

School Type



Frequencies

Level	Count	Prob
Private	290	0.04742
Public	5825	0.95258
Total	6115	1.00000

N Missing 0

2 Levels

2015 PISA Data

- 66 Booklets

Booklet ID	Reading	Math	Science
31 - 42	✓		✓
43 - 54		✓	✓
55 - 66	✓	✓	✓
67 - 78		✓	✓
79 - 90	✓		✓
91 - 96			✓

Different total score for each booklet

Methodology

Data Preparation

1. Filtering to Singapore Data
2. Removing Columns with no Response
3. Keeping Scored and Coded Responses
4. Adjusting Scores
5. Transposing questions
6. Binning Scoring Classifications

Data Analysis

1. Latent Class Analysis
2. Standardized Scores

Insights & Recommendations

Latent Class Analysis (LCA)

- Statistical method for finding subtypes of related cases (latent classes)
- Common areas for the use of LCA are in health research, sociology, psychology, and education.

Data Preparation

1. Filtering to Singapore data
2. Removing columns with no response or same value in all entries
3. Kept Scored and Coded responses from cognitive data

DS601Q02R Sustainable Fish Farming - Q02 (Raw Response)
CS601Q02S Sustainable Fish Farming - Q02 (Scored Response)
CS601Q02T Sustainable Fish Farming - Q02 (Timing)
CS601Q02A Sustainable Fish Farming - Q02 (Number of Actions)

Data Preparation

4. Adjust Scores

- 0 for no credit
- 0.5 for partial credit
- 1 for full credit
- 9999 for missing values

0	0 - No credit
1	1 - Partial credit
2	2 - Full credit
6 / .R	Not Reached
7 / .N	Not Applicable
8 / .I	Invalid
9 / .M	No Response
SYSTEM MISSING	Missing

0	00 - No credit
11	11 - Partial credit
12	12 - Partial credit
13	13 - Partial credit
21	21 - Full credit
22	22 - Full credit
23	23 - Full credit
96 / .R	Not Reached
97 / .N	Not Applicable
99 / .M	No Response
SYSTEM MISSING	Missing

Data Preparation

5. Transposed Questions

	CNTSCHID	STRATUM	CNTSTUID	CBASCI	BOOKID	Question	Subject	N(0)	N(0.5)	N(1)	N(9999)
1	70200001	SGP0203	70200112	4	31	DR219Q01AC	R	0	0	1	0
2	70200001	SGP0203	70200112	4	31	DR219Q01BC	R	0	0	1	0
3	70200001	SGP0203	70200112	4	31	DR219Q01CC	R	0	0	1	0
4	70200001	SGP0203	70200112	4	31	DR219Q01DC	R	0	0	1	0
5	70200001	SGP0203	70200112	4	31	DR219Q01EC	R	0	0	1	0
6	70200001	SGP0203	70200112	4	31	DR219Q01C	R	0	0	1	0
7	70200001	SGP0203	70200112	4	31	DR219Q02C	R	0	0	1	0
8	70200001	SGP0203	70200112	4	31	CR067Q01S	R	0	0	1	0
9	70200001	SGP0203	70200112	4	31	DR067Q04C	R	0	0	1	0

- Removal of missing responses

Data Preparation

6. Bin Scoring Classifications

	Questions	% of (0)	% of (0.5)	% of (1)	% of (9999)	% of (0) Binned	% of (0.5) Binned	% of (1) Binned	% of (9999) Binned
1	DR219Q01AC	5.12%	0.00%	87.65%	7.23%	0.00%—10.00%	0.00%—10.00%	80.00%—90.00%	0.00%—10.00%
2	DR219Q01BC	2.23%	0.00%	90.67%	7.10%	0.00%—10.00%	0.00%—10.00%	90.00%—100.00%	0.00%—10.00%
3	DR219Q01CC	1.97%	0.00%	91.72%	6.31%	0.00%—10.00%	0.00%—10.00%	90.00%—100.00%	0.00%—10.00%
4	DR219Q01DC	6.70%	0.00%	86.47%	6.83%	0.00%—10.00%	0.00%—10.00%	80.00%—90.00%	0.00%—10.00%
5	DR219Q01EC	14.19%	0.00%	78.19%	7.[illegible]	[illegible]	[illegible]—10.00%	70.00%—80.00%	0.00%—10.00%
6	DR219Q01C	14.59%	0.00%	79.63%	5.[illegible]	[illegible]	[illegible]—10.00%	70.00%—80.00%	0.00%—10.00%
7	DR219Q02C	14.06%	0.00%	81.60%	4.34%	10.00%—20.00%	0.00%—10.00%	80.00%—90.00%	0.00%—10.00%
8	CR067Q01S	10.51%	0.00%	89.09%	0.39%	10.00%—20.00%	0.00%—10.00%	80.00%—90.00%	0.00%—10.00%
9	DR067Q04C	19.45%	33.51%	44.15%	2.89%	10.00%—20.00%	30.00%—40.00%	40.00%—50.00%	0.00%—10.00%
10	DR067Q05C	20.76%	8.94%	66.75%	3.55%	20.00%—30.00%	0.00%—10.00%	60.00%—70.00%	0.00%—10.00%
11	DR102Q04C	70.70%	0.00%	22.21%	7.10%	70.00%—80.00%	0.00%—10.00%	20.00%—30.00%	0.00%—10.00%
12	DR102Q05C	43.89%	0.00%	53.48%	2.63%	40.00%—50.00%	0.00%—10.00%	50.00%—60.00%	0.00%—10.00%

- Assigning questions to bin based on performance

Data Analysis

LATENT CLASS ANALYSIS

1 2 to 5 clusters to determine the best model fit

Latent Class Model	BIC statistic
2 Clusters	2990.5
3 Clusters	**2927.15**
4 Clusters	**2921.4**
5 Clusters	2954.2

Easy, Medium, Hard Difficulty

Data Analysis

LATENT CLASS ANALYSIS

2 Profiling of Clusters

Cluster 1:

- 60% to 80% of (1)
- 20% to 40% of (0)

Medium Difficulty

Transposed Parameter Estimates

Column	Category	Cluster 1	Cluster 2	Cluster 3
% of (0) Binned	0.00% — 10.00%	0.0003	0.1493	0.0004
% of (0) Binned	10.00% — 20.00%	0.0687	0.4064	0.0008
% of (0) Binned	20.00% — 30.00%	0.4151	0.0003	0.0320
% of (0) Binned	30.00% — 40.00%	0.5145	0.0008	0.0320
% of (0) Binned	40.00% — 50.00%	0.0004	0.4419	0.0913
% of (0) Binned	50.00% — 60.00%	0.0003	0.0004	0.3687
% of (0) Binned	60.00% — 70.00%	0.0003	0.0003	0.3478
% of (0) Binned	70.00% — 80.00%	0.0003	0.0003	0.0951
% of (0) Binned	80.00% — 90.00%	0.0003	0.0003	0.0320
% of (0.5) Binned	0.00% — 10.00%	0.9740	0.9637	0.8628
% of (0.5) Binned	10.00% — 20.00%	0.0244	0.0101	0.0401
% of (0.5) Binned	20.00% — 30.00%	0.0005	0.0088	0.0533
% of (0.5) Binned	30.00% — 40.00%	0.0005	0.0086	0.0431
% of (0.5) Binned	40.00% — 50.00%	0.0005	0.0088	0.0007
% of (1) Binned	0.00% — 10.00%	0.0003	0.0003	0.0319
% of (1) Binned	10.00% — 20.00%	0.0003	0.0003	0.0950
% of (1) Binned	20.00% — 30.00%	0.0003	0.0003	0.1793
% of (1) Binned	30.00% — 40.00%	0.0003	0.0101	0.3247
% of (1) Binned	40.00% — 50.00%	0.0003	0.0761	0.3671
% of (1) Binned	50.00% — 60.00%	0.1160	0.3886	0.0005
% of (1) Binned	60.00% — 70.00%	0.5225	0.0003	0.0004
% of (1) Binned	70.00% — 80.00%	0.3596	0.0021	0.0004
% of (1) Binned	80.00% — 90.00%	0.0003	0.4058	0.0004
% of (1) Binned	90.00% — 100.00%	0.0003	0.1161	0.0004
% of (9999) Binned	0.00% — 10.00%	0.9757	0.9986	0.9141
% of (9999) Binned	10.00% — 20.00%	0.0243	0.0014	0.0859

Data Analysis

LATENT CLASS ANALYSIS

2 Profiling of Clusters

Cluster 2:

- 80% to 100% of (1)
- 0% to 20% of (0)

Easy Difficulty

Transposed Parameter Estimates

Column	Category	Cluster 1	Cluster 2	Cluster 3
% of (0) Binned	0.00% — 10.00%	0.0003	0.1493	0.0004
% of (0) Binned	10.00% — 20.00%	0.0687	0.4064	0.0008
% of (0) Binned	20.00% — 30.00%	0.4151	0.0003	0.0320
% of (0) Binned	30.00% — 40.00%	0.5145	0.0008	0.0320
% of (0) Binned	40.00% — 50.00%	0.0004	0.4419	0.0913
% of (0) Binned	50.00% — 60.00%	0.0003	0.0004	0.3687
% of (0) Binned	60.00% — 70.00%	0.0003	0.0003	0.3478
% of (0) Binned	70.00% — 80.00%	0.0003	0.0003	0.0951
% of (0) Binned	80.00% — 90.00%	0.0003	0.0003	0.0320
% of (0.5) Binned	0.00% — 10.00%	0.9740	0.9637	0.8628
% of (0.5) Binned	10.00% — 20.00%	0.0244	0.0101	0.0401
% of (0.5) Binned	20.00% — 30.00%	0.0005	0.0088	0.0533
% of (0.5) Binned	30.00% — 40.00%	0.0005	0.0086	0.0431
% of (0.5) Binned	40.00% — 50.00%	0.0005	0.0088	0.0007
% of (1) Binned	0.00% — 10.00%	0.0003	0.0003	0.0319
% of (1) Binned	10.00% — 20.00%	0.0003	0.0003	0.0950
% of (1) Binned	20.00% — 30.00%	0.0003	0.0003	0.1793
% of (1) Binned	30.00% — 40.00%	0.0003	0.0101	0.3247
% of (1) Binned	40.00% — 50.00%	0.0003	0.0761	0.3671
% of (1) Binned	50.00% — 60.00%	0.1160	0.3886	0.0005
% of (1) Binned	60.00% — 70.00%	0.5225	0.0003	0.0004
% of (1) Binned	70.00% — 80.00%	0.3596	0.0021	0.0004
% of (1) Binned	80.00% — 90.00%	0.0003	0.4058	0.0004
% of (1) Binned	90.00% — 100.00%	0.0003	0.1161	0.0004
% of (9999) Binned	0.00% — 10.00%	0.9757	0.9986	0.9141
% of (9999) Binned	10.00% — 20.00%	0.0243	0.0014	0.0859

Data Analysis

LATENT CLASS ANALYSIS

2 Profiling of Clusters

Cluster 3:

- 0% to 50% of (1)
- 40% to 90% of (0)
- 10% to 50% of (0.5)

Hard Difficulty

Transposed Parameter Estimates

Column	Category	Cluster 1	Cluster 2	Cluster 3
% of (0) Binned	0.00% — 10.00%	0.0003	0.1493	0.0004
% of (0) Binned	10.00% — 20.00%	0.0687	0.4064	0.0008
% of (0) Binned	20.00% — 30.00%	0.4151	0.0003	0.0320
% of (0) Binned	30.00% — 40.00%	0.5145	0.0008	0.0320
% of (0) Binned	40.00% — 50.00%	0.0004	0.4419	0.0913
% of (0) Binned	50.00% — 60.00%	0.0003	0.0004	0.3687
% of (0) Binned	60.00% — 70.00%	0.0003	0.0003	0.3478
% of (0) Binned	70.00% — 80.00%	0.0003	0.0003	0.0951
% of (0) Binned	80.00% — 90.00%	0.0003	0.0003	0.0320
% of (0.5) Binned	0.00% — 10.00%	0.9740	0.9637	0.8628
% of (0.5) Binned	10.00% — 20.00%	0.0244	0.0101	0.0401
% of (0.5) Binned	20.00% — 30.00%	0.0005	0.0088	0.0533
% of (0.5) Binned	30.00% — 40.00%	0.0005	0.0086	0.0431
% of (0.5) Binned	40.00% — 50.00%	0.0005	0.0088	0.0007
% of (1) Binned	0.00% — 10.00%	0.0003	0.0003	0.0319
% of (1) Binned	10.00% — 20.00%	0.0003	0.0003	0.0950
% of (1) Binned	20.00% — 30.00%	0.0003	0.0003	0.1793
% of (1) Binned	30.00% — 40.00%	0.0003	0.0101	0.3247
% of (1) Binned	40.00% — 50.00%	0.0003	0.0761	0.3671
% of (1) Binned	50.00% — 60.00%	0.1160	0.3886	0.0005
% of (1) Binned	60.00% — 70.00%	0.5225	0.0003	0.0004
% of (1) Binned	70.00% — 80.00%	0.3596	0.0021	0.0004
% of (1) Binned	80.00% — 90.00%	0.0003	0.4058	0.0004
% of (1) Binned	90.00% — 100.00%	0.0003	0.1161	0.0004
% of (9999) Binned	0.00% — 10.00%	0.9757	0.9986	0.9141
% of (9999) Binned	10.00% — 20.00%	0.0243	0.0014	0.0859

Data Analysis

STANDARDIZED SCORING

1 Adjust weight of questions

- Most likely cluster is derived based on the column with the highest probability.
- Adjustment of question's weight and calculation of total score for every student

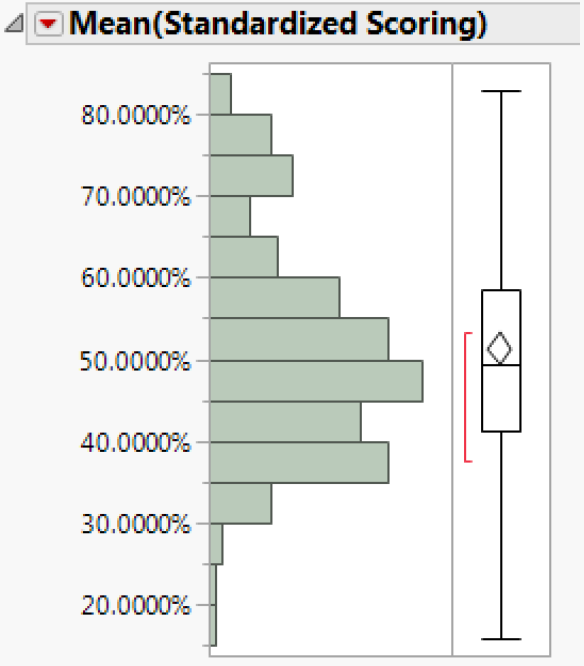
Prob in Medium	Prob in Easy	Prob in Hard	Most Likely Cluster
8.2362003e-7	0.9999973249	1.8515159e-6	Easy
2.7115109e-6	0.9999889343	8.3541687e-6	Easy
2.7115109e-6	0.9999889343	8.3541687e-6	Easy
8.2362003e-7	0.9999973249	1.8515159e-6	Easy
0.0000500252	0.9999309148	0.0000190599	Easy
0.0000500252	0.9999309148	0.0000190599	Easy
0.0000394369	0.999941914	0.0000186491	Easy
0.0000394369	0.999941914	0.0000186491	Easy
0.9866573206	0.0013065138	0.0120361656	Medium
0.9942583343	0.0057390453	2.6203851e-6	Medium
4.864051e-6	5.9006039e-6	0.9999892353	Hard
0.9999262397	5.7220175e-7	0.0000731881	Medium
0.0000394369	0.999941914	0.0000186491	Easy
1.3304652e-6	1.6132772e-6	0.9999970563	Hard
0.9999428389	1.3412866e-6	0.0000558198	Medium
0.9999262397	5.7220175e-7	0.0000731881	Medium

Data Analysis

STANDARDIZED SCORING

2 Data Exploration

- Mean score for all schools – 51.28%
- Middle 50% had a range score of 41.23% to 58.39%

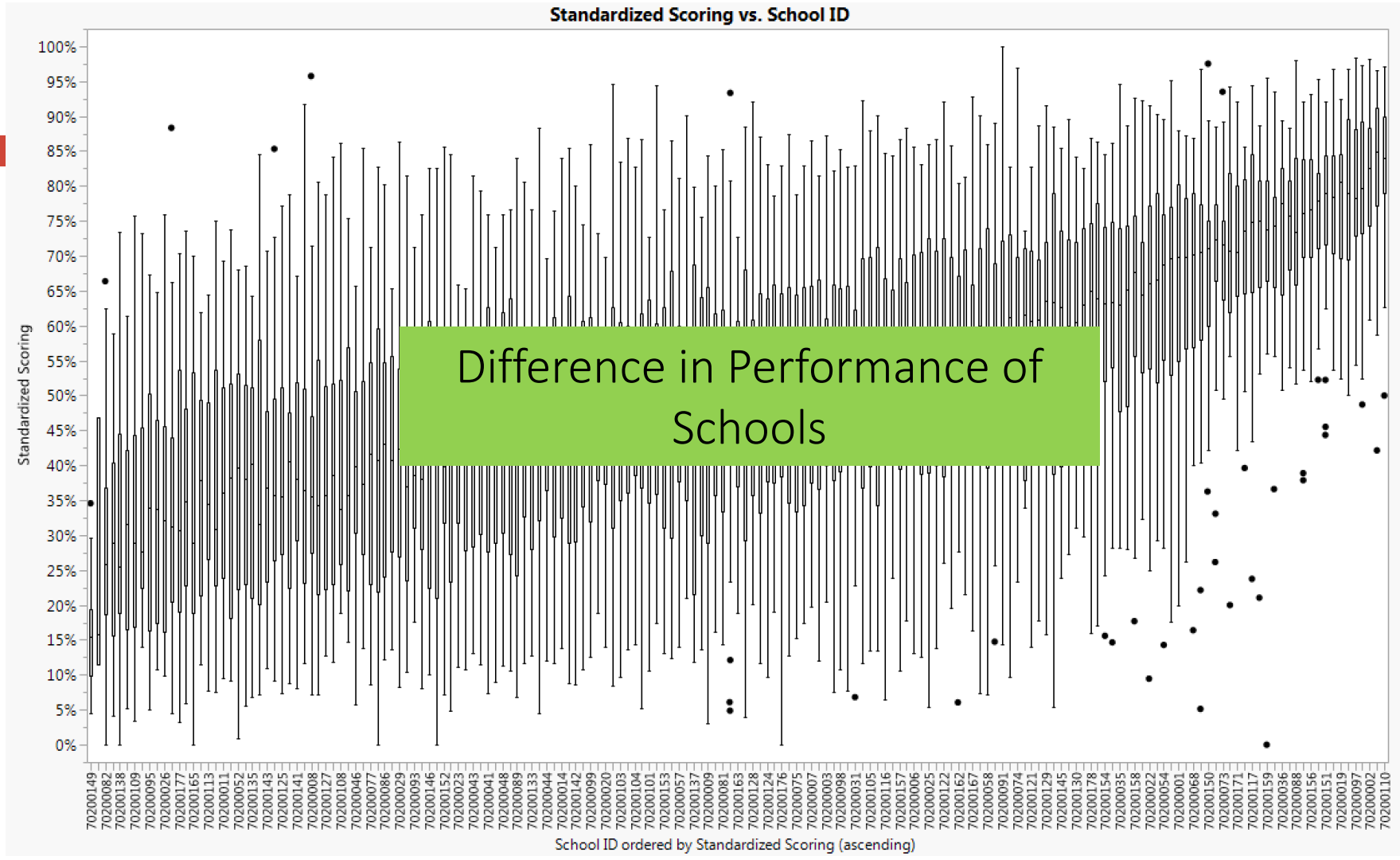


Quantiles

100.0%	maximum	82.8971%
99.5%		82.8971%
97.5%		78.8448%
90.0%		72.2012%
75.0%	quartile	58.3913%
50.0%	median	49.2170%
25.0%	quartile	41.2267%
10.0%		36.5593%
2.5%		29.7756%
0.5%		15.7393%
0.0%	minimum	15.7393%

Summary Statistics

Mean	0.5128268
Std Dev	0.1313274
Std Err Mean	0.0098712
Upper 95% Mean	0.5323079
Lower 95% Mean	0.4933457
N	177

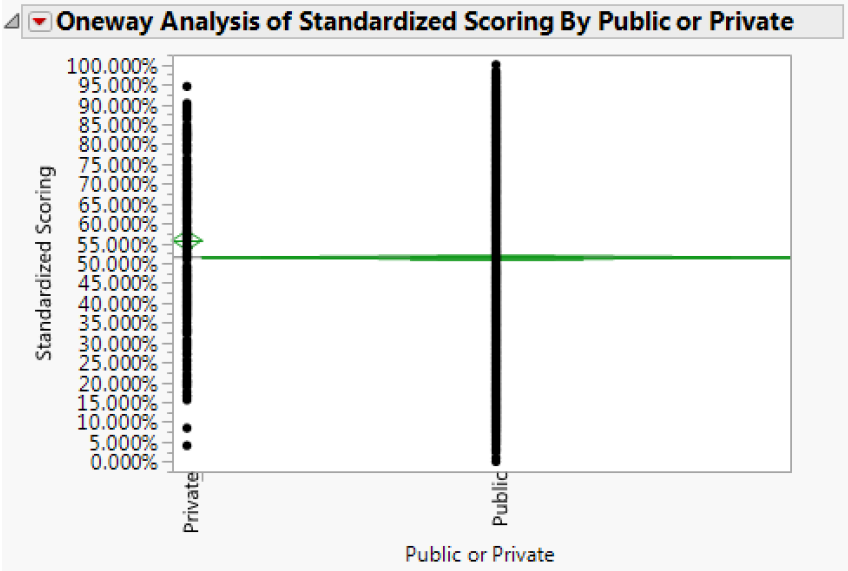

Standardized Scoring vs. School ID
Standardized Scoring
100%
95%
90%
85%
80%
75%
70%
65%
60%
55%
50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
Difference in Performance of Schools
School ID ordered by Standardized Scoring (ascending)


Standardized Scoring vs. School ID
Standardized Scoring
100%
95%
90%
85%
80%
75%
70%
65%
60%
55%
50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
There are good performing schools but there are more outliers
School ID ordered by Standardized Scoring (ascending)

Data Analysis

STANDARDIZED SCORING

3 Analysis of Variance



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Public or Private	1	0.51273	0.512735	11.1130	0.0009*
Error	6113	282.04463	0.046138		
C. Total	6114	282.55736			

Significant Difference between the performance of public and private schools

Conclusion

Through our analysis of the 2015 PISA global education survey, there is indeed a difference between Schools in Singapore based on their performance.

The factors affecting the performances of schools and students will be discussed in the next 2 sessions.

1. **An Analysis of Singapore's School Performance in the PISA Global Education Survey**
2. **Using Partition Models to Identify Key Differences between Top Performing and Poor Performing Students**


Thank You