HONG KONG ASSOCIATION FOR SCIENCE AND MATHEMATICS EDUCATION
REPORT ON SCIENCE ASSESSMENT TEST 2018

INTRODUCTION

The Science Assessment Test (SAT) was developed by the Hong Kong Association for Science and Mathematics Education (HKASME) for evaluating the ability of students in Hong Kong, as well as those in nearby regions, in learning science at Junior Secondary level. It was first implemented in 2014, and thereafter improvements have been made to the design of the test and to the analysis of the results. The test consists of a 1-hour test paper comprising 24 multiple-choice questions in Section A and 2 short-response questions in Section B. With the SAT, the HKASME hopes to provide feedback to schools and to the education administration on the strengths and weaknesses of students in learning science at Junior Secondary level.

In order to keep the SAT to be in line with the most up-to-date approach of science learning, the questions are so set that the participants are not required to recall a lot of scientific facts. Instead, the SAT questions aim at testing students' scientific understanding as well as science process skills, namely observing, classifying, planning and designing, experimenting, interpreting and communicating. In fact, many of the SAT questions were set to help students appreciate the relevancy of science to daily life.

THE PARTICIPANTS

In 2018 SAT, the total number of participants was 2592, with 2373 from the Hong Kong SAR and 219 from the Macau SAR. The table below lists the breakdown of the 2592 participants according the class attended and sex:

Participants	Secondary 2 (S2)	Secondary 3 (S3)	Total
Boys	793	623	1416
Girls	673	503	1176
Total	1466	1126	2592

PARTICIPANTS' PERFORMANCE

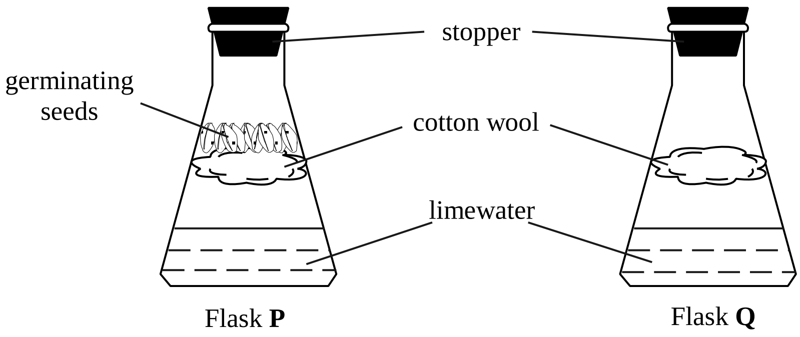
Section A

The mean score and the standard deviation of the 2592 participants in the 24 multiple-choice questions were 14.2 and 3.44 respectively. Readers can refer to Appendix I for an analysis of the participants' responses to the multiple-choice questions.

As revealed from the item analysis, participants showed weaknesses in the areas outlined below. For each of these areas, a few items were selected to help illustrate the participants' weaknesses. In these items, the key is marked with an asterisk (*) and the popularities of the options are shown in parentheses.

(1) Appreciation of the Techniques Involved in Carrying out a Scientific Investigation

Q.4 The set-up shown below was used in a certain experiment:



After a few days, the limewater in **P** turned milky while that in **Q** remained clear. The purpose of this experiment is to show that

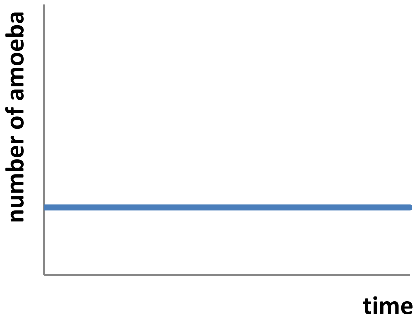
(1) germination of seeds requires oxygen.
(2) germination of seeds produces carbon dioxide.
(3) carbon dioxide can turn limewater milky.

A.	(1) only	(4%)
*B.	(2) only	(28%)
C.	(1) and (3) only	(9%)
D.	(2) and (3) only	(59%)

The purpose of this experiment is to show that germination of seeds produces carbon dioxide. The anticipated experimental result is that the limewater in Flask **P** will turn milky after some time, while that in Flask **Q** will remain clear. That is, the use of limewater is to show that carbon dioxide is formed during germination. A majority of participants wrongly perceived (3) to be a purpose of this experiment. If one wants to show that carbon dioxide can turn limewater milky, one can simply pass carbon dioxide into limewater and it is not necessary to use such a complicated set-up. Participants' performance showed that they were weak in designing experiment.

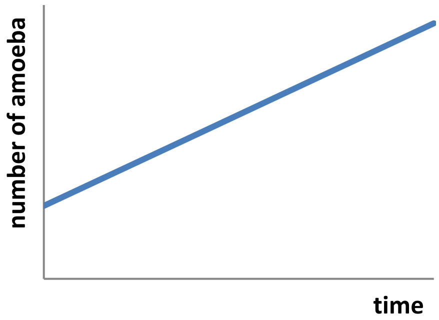
Q.9 Amoeba multiplies by cell division. A sample of amoeba is allowed to multiply in a petri dish under controlled experimental conditions. Which of the following graphs best represents the variation of the number of amoeba in the petri dish with time?

A.



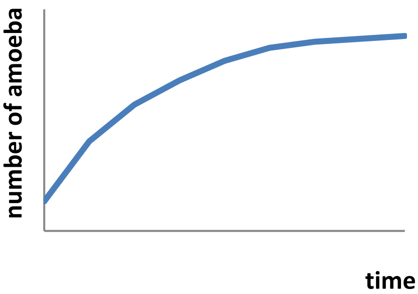
(1%)

B.



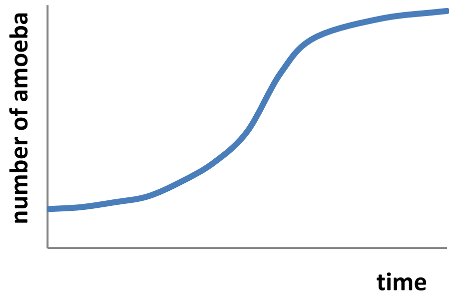
(56%)

C.



(19%)

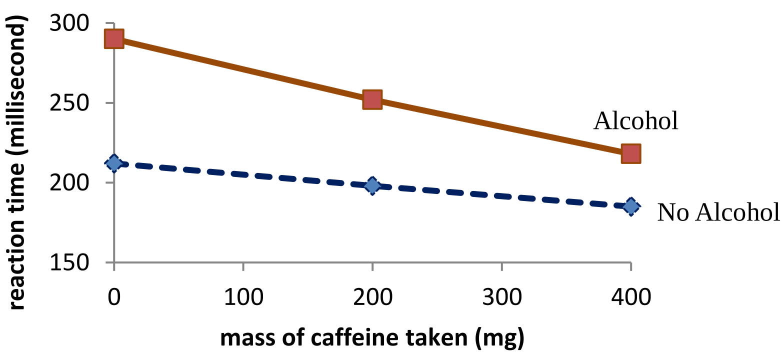
*D.



(24%)

This question was set on the interpretation of experimental results presented in the form of a graph. Most participants correctly realized that the number of amoeba in the petri dish would increase with time. However, many wrongly perceived that the number of amoeba in the petri dish would increase linearly, i.e. option B. Only a few of them were able to recognize that the multiplication of amoeba would initially follow an exponential curve, and after that the rate of multiplication of amoeba would diminish when nutrients present in the petri dish were consumed, i.e. option D.

Q.24 A scientist investigated the effect of consumption of caffeine on the reaction time of people who have drunk and who have not drunk alcohol. The graph below shows the results of the investigation:



Which of the following statements can be deduced from the results of the investigation?

(1) The reaction time of people increases after drinking alcohol.
(2) The reaction time of people increases after taking caffeine.
(3) People's judgment will be greatly affected after taking caffeine together with alcohol.

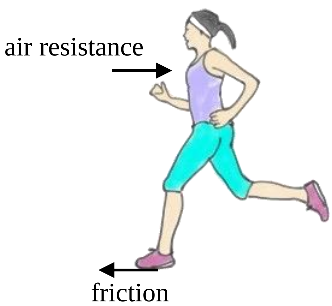
*A. (1) only (37%)
B. (2) only (14%)
C. (1) and (3) only (29%)
D. (2) and (3) only (20%)

Like Q.8, this question was also set to test participants' ability in the interpretation of graphical data. From the graph, we can only deduce that the reaction time of people increases after drinking alcohol. However, quite a number of participants approached this question by making use of their general knowledge and wrongly thought that statement (3) could also be deduced from the results of the investigation. In fact, the results of the investigation only shows that the reaction time of people would be lengthened after drinking alcohol and would be reduced after taking caffeine.

(2) <u>Understanding of Abstract Scientific Concepts</u>

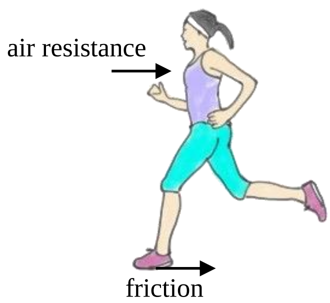
Q.21 Sandy is running on a horizontal road. Which of the following diagrams best illustrates the directions of the air resistance and the friction acting on her foot?

*A.



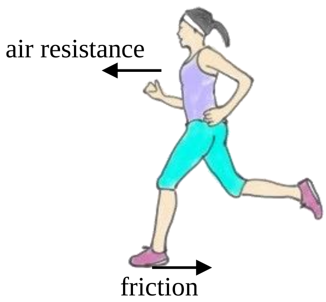
(28%)

B.



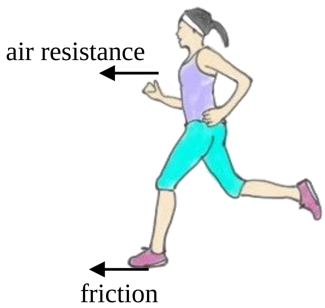
(67%)

C.



(4%)

D.

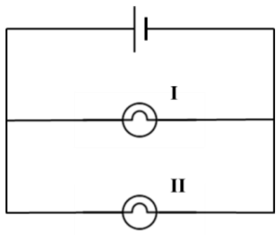


(1%)

Most of the participants knew that the air resistance was always acting against the motion of the runner. However, quite a number of them did not realize that when a runner runs, his/her feet, while stepping on the ground, would exert a backward force on it. According to

Newton's Second Law of Motion, the frictional force exerted by the ground would act against this motion, i.e. it would be pointing forward, i.e. option A.

Q.23 Consider the following circuit, in which the two light bulbs, **I** and **II**, are of different resistance:



Which of the following descriptions about the circuit is correct?

(1) The current passing through **I** and that through **II** are the same.
(2) The current passing through **I** and that through **II** are different.
(3) The voltage across **I** and that across **II** are the same.
(4) The voltage across **I** and that across **II** are different.

A.	(1) and (3) only	(34%)
B.	(1) and (4) only	(23%)
*C.	(2) and (3) only	(25%)
D.	(2) and (4) only	(18%)

Both Qs. 21 and 23 were testing abstract concepts in science, namely force and electricity, and participants' performance was unsatisfactory. In Q.23, the popularities of the four options are quite similar. This might be due to the fact that participants did not have a good grasp of the concepts of current and voltage, which are quite abstract to students in junior secondary. Participants should realize that the voltages across **I** and **II** should be the same as it is actually a measure of the voltage of the battery, whereas the currents passing **I** and **II** are different as the amount of charges passing through an electrical conductor depends on its resistance.

(3) <u>Understanding of Science Knowledge</u>

Q.10 When we inhale, what happen to our diaphragm muscle and the pressure inside our lungs?

	Diaphragm muscle	Pressure inside our lungs	
A.	relaxes	increases	(17%)
B.	relaxes	decreases	(11%)
C.	contracts	increases	(54%)
*D.	contracts	decreases	(18%)

Most of the participants probably did not realize that our diaphragm has a dome shape. With this in mind, they should realize that when we inhale, the diaphragm will contract and the pressure inside our lungs will decrease, i.e. option D.

Section B

In Section B, Q.1 was set to test participants' understanding of experimental procedures, communicative and graph-plotting skills, and interpretation of experimental data. Q.2 was a comprehension question set on a local environmental problem. Participants were required to demonstrate their understanding of written information, and to propose solutions to solve the solid wastes problems in Hong Kong based on the given geographical and meteorological information. The maximum marks for Q.1 and Q.2 were 10 and 8 respectively. The table below lists the mean score and standard deviation for the whole group.

Question No.	Mean	S.D.
1	5.6 (56%)	2.30 (23.0%)
2	4.0 (50%)	1.80 (22.5%)
Overall	9.6 (53%)	3.45 (19.2%)

Participants' performance in the two short-response questions in Section B is outlined below:

Q.1 (a) Excellent performance. Most participants (85%) were able to correctly arrange the listed steps in their proper order as in the experiment.

(b) Fair performance. About 46% of the participants were able to point out that stirring speed, size/type of the coarse salt or type of water is the variable needed to keep constant (controlled variable) in the experiment. Though the question asks for a controlled variable other than the volume of water and the mass of the coarse salt sample, yet about 6% of the participants still quoted these two items as the answer. Participants should learn to read questions more carefully. Some common mistakes included:

- wrongly mentioned room temperature or atmospheric pressure as the controlled variable *(In reality, these two conditions are external factors that cannot changed by us.)*,
- considered using the same apparatus such as the same thermometer, the same beaker, etc. in the experiment as a controlled variable *(In fact, the apparatus used have no direct effect on the dissolving speed of the coarse salt sample.)*, and
- incorrectly stated that water temperature is a controlled variable *(The water temperature is actually the variable to be changed in the experiment, i.e. the independent variable.)*.

(c) Fair performance. About 34% of the participants were able to obtain full score in constructing data table. The mean score in this part was 1 out of 2. Quite a high proportion (31%) of the participants either gave irrelevant table/graph/chart/answers or showed nil response. It is apparent that these participants had no idea of what a data table is. A data table is an organized arrangement of data in labeled rows and columns. It contains column headings with units of measurements. Shown below is a presentation of the data table for the experimental results obtained:

Temperature of water (°C)	Dissolving time (s)
20	112
30	80
40	60
50	48
60	40

The dissolving time needs to be processed into the same units, i.e. "s" before entering into the table.
Some common mistakes included:

- units of measurements were missing in the column heading,
- units of measurements were repeated in the table (both in the column heading and after the numerical data in the individual cells), and
- data of dissolving time were not processed into the same units.

(d)(i) Satisfactory performance. The mean score of the participants was 1.7 out of 3. About 26% of the participants were able to obtain full score in plotting of the graph. The following are some common mistakes made:

- plotting 'temperature of water' against 'dissolving time' (*Some participants did not realize that "dissolving time" is the dependent variable and to be placed on y-axis.)*,
- plotting graphs with the scale in one of the axes not equally spaced,
- plotting graphs with a poor scale, which is highly compressed or falls out of the graph paper,
- plotting graphs by direct transferring the numbers from the data table onto the two axes,
- having all or some of the points incorrectly plotted,
- drawing a best straight line instead of a curve passing through all points, and
- connecting the first data point of the curve to origin by a straight line.

(d)(ii) Satisfactory performance. About 63% of the participants were able to estimate the dissolving time correctly.

(d)(iii) Well answered. About 84% of the participants were able to give concise and precise conclusion statement.

(e) Poorly answered. Only a small proportion (7%) of the participants were able to point out either one of the following reasons for not repeating the experiment at 80°C:

- very short dissolving time leading to a large error in time measurement,
- difficult to keep the water temperature at 80°C steadily as the heat loss to the surrounding is great.

Some participants mentioned that the dissolving time of the coarse salt sample was too short to be measured. They did not realize that at 80°C, the dissolving time was approximately 30 seconds (as estimated from the curve) that could still be recorded by a stop-watch. Other common mistakes included:

- evaporation of water at 80°C leads to a reduction in volume of water in the beaker, thus it affects the results,
- the salt solution becomes saturated at 80°C and cannot dissolve additional amount of salt,
- the dissolving speeds at 80°C and 60°C are very close, thus it is not necessary to repeat the experiment at 80°C, and
- using the terms "melting" and "dissolving" interchangeably when describing the mixing of salt with water to form a salt solution.

It is apparent that participants were weak in answering questions related to planning and design of experiment.

Q.2 The test statistics of this question, with the mean and standard deviation equal to 50% and 22.5% respectively, indicates that the question has a good discrimination power. Most of the sub-questions have a mean mark close to 50% further shows that the discrimination power is good up to the sub-question level.

(a) Most of the participants recognized that due to land shortage, it is difficult for Hong Kong to build a new landfill site for dumping its solid wastes. For this two-mark question, participants were expected to give, in addition to land shortage, a reason to support their answers, such as the decomposition of solid wastes takes a long time, or the rate of waste production is extremely fast in a metropolitan like Hong Kong. As a majority of the participants were unable to give a reason to support their answers, most of them could get only 1 mark in this question.

(b)(i) The participants were too young and might not know the use of incineration in treating solid wastes in Hong Kong in the 1990s. However, with their science/general knowledge, most of them were able to correlate why incineration was abandoned to air pollution issues. Some participants mistakenly considered carbon dioxide to be an air pollutant. Although carbon dioxide is a main contributor to global warming, it is also important in maintaining the Earth's temperature and as a food for green plants. Due to its benefits, carbon dioxide is not regarded as an air pollutant.

(b)(ii) This is the least well-answered sub-question with a mean score of 0.3.

Many participants were unable to suggest an advantage of new generation incinerator in treating solid wastes, namely it emits only a very small amount of air pollutants that meets the very strict international standards, and the heat generated from incineration can be recovered for other uses. Quite a number of participants gave irrelevant answers like the new generation incinerator would be built far away from the densely populated areas. Such answers were considered irrelevant as the question asked for an advantage of the new generation incinerator *over the previously used ones*, rather than a reason for choosing a suitable location for the incinerator. Participants should learn to read questions more carefully.

(c) There should be two points in this question: (1) the proposed sites for the new incinerator were far away from residential areas, and (2) for most of the time, the wind directions in Hong Kong would disperse the flue gases emitted from the incinerator to the sea. About 70% of the participants were able to give the first point and 30% the second. The test results might indicate that participants were not strong in reading maps and in making inference from given information.

(d) There are quite a number of means that the government can adopt to reduce solid wastes. However, less than 30% of the participants were able to score 2 points. Some participants vaguely stated the principles involved in reducing solid wastes such as the 3R's, which is not regarded to be a *means* that can be adopted. Some participants proposed educating the public and promotion through broadcasting on waste reduction as two different means. These two suggestions were considered as belonging to the same category – educating the public. Participants giving such answers would receive 1 mark rather than 2 in this sub-question.

FURTHER ANALYSIS OF THE TEST RESULTS

The 24 multiple-choice questions in Section A and the sub-questions of 2 short-response questions in Section B all had high marking reliability with Cronbach's $\alpha = 0.77$. The test results were further analyzed by comparing (a) the performance of S2 and S3 students, and (b) the performance of boys and girls in the test.

(a) Comparison of the performance of S2 and S3 students

The table below lists the mean and standard deviation of mark awards of the S2 and S3 students in Sections A and B.

	Section A		Section B	
	Mean	S.D.	Mean	S.D.
S2	13.6 (57%)	3.35 (14.0%)	9.1 (50%)	3.25 (18.0%)
S3	14.9 (62%)	3.43 (14.3%)	10.2 (57%)	3.60 (20.0%)

As compared with the results of the 2017 SAT, participants' performance showed a very slight decline in Section A, but a significant improvement in Section B. The test statistics shows that S3 students performed better than S2 students in both Sections A and Section B. Moreover, the spread of marks for the S3 students is greater than that for the S2 students.

(b) Comparison of the performance of the boys and the girls

For Section A, the median score of the participants was 14 (correct to the nearest unit digit) out of the 24 multiple-choice questions. The table below lists the boys and girls (in percentage) having 14 or more multiple-choice questions correct in the test:

	S2	S3	Whole group
Boys	52%	68%	59%
Girls	53%	70%	60%

As compared with the results of the 2017 SAT, the girls made a significant improvement over the boys. This may be related to the different test samples in the two tests and/or other reasons and may worth further studies into it.

Appendix IIIa shows the marks distributions of Section A for the boys and the girls in S2, S3 and the whole group. As revealed from the test statistics, the girls performed slightly better than the boys in Section A.

For Section B, the median score of the participants was 10 (correct to the nearest unit digit) out of 18 marks. The table below lists the boys and girls (in percentage) having 10 marks or more in this part.

	S2	S3	Whole group
Boys	45%	60%	52%
Girls	51%	74%	61%

As revealed from the test statistics, the girls performed better than the boys in Section B. This difference was more significant in S3. Appendix IIIb shows the marks distributions of Section B for the boys and the girls in S2, S3 and the whole group.

THE AWARD SCHEME

Participants who demonstrate competency in science learning will be given an award. There are four levels of awards in the 2018 SAT, namely Diamond (highest), Gold, Silver and Bronze awards. The HKASME has set up an expert group to decide on the cut-off criteria, based on the performance of the participants, for each of these awards. In order to receive an award, a participant needs to get a minimum overall mark as well as to demonstrate a balanced performance in Sections A and B.

For 2018 SAT, about 5% of the best-performed participants were given the Diamond award. The table below lists the criteria for participants to receive the Diamond, Gold, Silver and Bronze awards in 2018 SAT:

	Diamond	Gold	Silver	Bronze
Paper score [1] / marks	$\geq$ 42.0	36.5 – 41.5	32.0 – 36.0	24.5 – 31.5
Score in Section A / MCQs	$\geq$ 10	$\geq$ 10	$\geq$ 10	$\geq$ 10
Score in Section B / marks	$\geq$ 5	$\geq$ 5	$\geq$ 5	$\geq$ 5

(1) In SAT, the paper score = 1.5 $\times$ score in Section A + score in Section B

Readers can refer to Appendix IV for the awards given out in the 2018 SAT as well as the overall performance of the participants in the test. It is worth mentioning that the purpose of SAT is not for the discrimination of the participants and/or their schools according to their achievement in the SAT. It aims at helping teachers/schools to identify the strengths and weaknesses of students in learning science, and as such appropriate means can be implemented to help students make improvements in their learning.

Appendix I

Analysis of Participants’ Responses to the Multiple-choice Questions (MCQs)

No. of schools: 77
No. of participants: 2592

	Whole Group (WG)
Mean score: *(out of 24)*	14.2 (59%)
Standard deviation:	3.44 (14.3%)

Performance of the whole group (WG) in Section A of 2018 SAT

Q. No	**Skills assessed**	**Key**	**Correct %**	**Strength / weakness**
1	Interpreting data; Inferring	A	72.0	-
2	Planning & Design; Predicting	C	56.4	-
3	Understanding; Inferring	B	84.5	-
4	Planning & Design; Understanding	B	27.6	***D***
5	Predicting	A	85.3	-
6	Identifying variables	D	76.4	-
7	Understanding;	C	55.2	-
8	Making hypothesis	B	55.2	-
9	Interpreting graph; Predicting	D	24.4	***D***
10	Understanding; Predicting	D	17.6	***D***
11	Understanding	D	50.2	-
12	Interpreting graph; Inferring	C	69.0	-
13	Choosing apparatus	D	74.8	-
14	Understanding	B	69.3	-
15	Understanding; Inferring	C	77.5	-
16	Interpreting graph; Inferring & Predicting	A	80.7	-
17	Understanding	B	56.6	-
18	Understanding; Predicting	D	73.4	-
19	Interpreting data; Understanding	B	86.6	-
20	Understanding	A	61.9	-
21	Understanding	A	27.5	***D***
22	Interpreting graph; Inferring	C	72.0	-
23	Understanding	C	25.2	***D***
24	Interpreting graph; Inferring	A	37.4	-

Note:

- Questions that are poorly answered by the whole group (≤1/3 correct) are represented by “***D***”.

Appendix II

1. Analysis of Participants' Performance in Short-response Questions

No. of schools: 77
No. of participants: 2592

The table below gives the mean and standard deviation of the whole group (WG) in the two questions in Section B of the 2018 SAT:

	Whole Group (WG)	
	Mean	**SD**
Question 1 *(out of 10)*	5.6 (56%)	2.30 (23.0%)
Question 2 *(out of 8)*	4.0 (50%)	1.80 (22.5%)
Section B *(out of 18)*	9.6 (53%)	3.45 (19.2%)

2. Double-digit Coding Marking System in 2018 SAT

In Section B, a double-digit coding system was adopted in marking the short-response questions. During marking, two digits were used to represent the performance of a participant in each part/sub-part of a question. The first digit indicates the correctness level of the answer while the second identifies the approaches used in answering the questions or the types of errors made. Shown below are the coding scheme adopted in marking Q1(b), (c), (d)(i) and (e) of the 2018 SAT and some common mistakes found.

<u>Double-digit coding system for Q1(b)</u>

Q1(b) *"Other than the volume of water and the mass of the coarse salt sample, state another variable that James needed to keep constant in the experiment."*

The question tests for the ability in identifying variables.

CODE	**ITEM**	**Number of Response**
Correct Responses		***1190 (45.9%)***
10	stirring speed	923 (35.6%)
11	size/ type/ brand of the coarse salt	247 (9.5%)
12	type of water e.g. tap water	20 (0.8%)

CODE	ITEM	Number of Response
Incorrect Responses/Nil Response		***1402 (54.1%)***
90	volume of water/ mass of coarse salt	166 (6.4%)
91	using same apparatus e.g. thermometer, glass rod, beaker, balance, stop-watch	390 (15.0%)
92	room temperature/ atmospheric pressure	233 (9.0%)
93	water temperature	239 (9.2%)
94	reaction time to start/ stop the stop-watch	66 (2.5%)
98	Other irrelevant answers	227 (8.8%)
99	*Unattempt*	81 (3.1%)

Double-digit coding system for Q1(c)

Q1(c) *"Shown below are the notes taken by James regarding the experiment:*
……
Present James' results in the form of a data table."

The question tests for the ability in constructing data table.

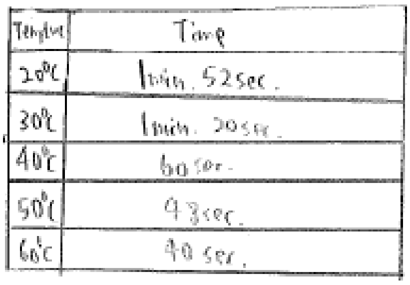
CODE	ITEM	Number of Response
Correct Responses		***889 (34.3%)***
20	Table with appropriate headings + units; Correct data arranged in pairs	624 (24.1%)
21	Table with appropriate headings (without units); Correct data arranged in pairs; units go after data	135 (5.2%)
22	Table with appropriate headings + units; Correct data arranged in pairs; units repeated in the table	130 (5.0%)
Partially Correct Responses		***889 (34.3%)***
10	Table with appropriate headings; Units after headings/ units after data, or both; Data without processing into seconds/ minutes	820 (31.6%)
11	Table with appropriate headings; Units after headings/ units after data, or both; Some wrong data in the table	69 (2.7%)
Incorrect Responses/Nil Response		***814 (31.4%)***
90	Showing a list of data without headings (or incomplete headings)	119 (4.6%)
98	Irrelevant table/ graph/ chart/answers, or incomplete data	277 (10.7%)
99	*Unattempt*	418 (16.1%)

Mean Score and Marks Distribution for Q1(c) on constructing data table:

		WG
Mean Score *(out of 2)*:		1.0
Marks Distribution:	**2**	889 (34.3%)
	1	889 (34.3%)
	0	814 (31.4%)

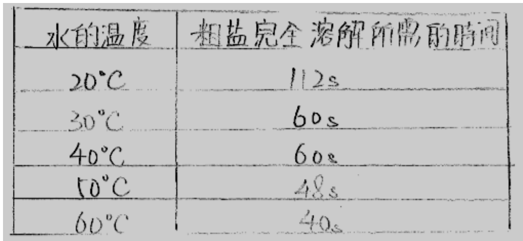
Common Mistakes of Participants in Constructing Data Table

- ***Examples of some partially correct responses***



Temp(°C)	Time
20°C	1min. 52sec.
30°C	1min. 20sec.
40°C	60sec.
50°C	48sec.
60°C	40 sec.

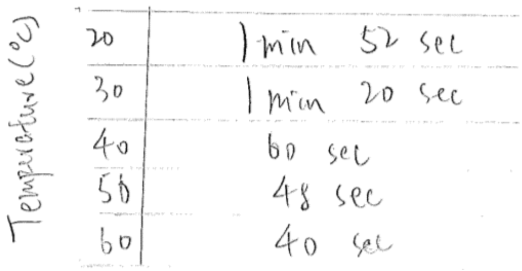
(Code: 10)



水的温度	粗盐完全溶解所需的时间
20°C	112s
30°C	60s
40°C	60s
50°C	48s
60°C	40s

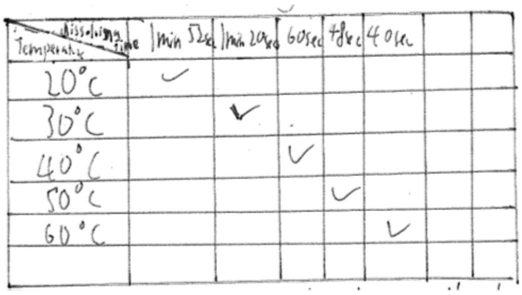
(Code: 11)

- ***Examples of some incorrect responses***



Temperature (°C)	
20	1 min 52 sec
30	1 min 20 sec
40	60 sec
50	48 sec
60	40 sec

(Code: 90)



Temperature \ dissolving time	1 min 52sec	1 min 20sec	60sec	48sec	40sec
20°C	✓				
30°C		✓			
40°C			✓		
50°C				✓	
60°C					✓

(Code: 98)

<u>Double-digit coding system for Q1(d)(i)</u>

Q1(d)(i) *"Plot a graph to show the relationship of the data in (c)."*

The question tests for the ability in plotting graph from given data.

Three aspects were looked for in marking the graph:

- correct labeling of **A**xes (code: **A**1x, **A**9x);
- appropriate choice of **S**cale (code: **S**1x, **S**9x); and
- **G**raph showing a curve passing through the points (code: **G**1x, **G**9x).

Correct labeling of axes:

CODE	ITEM	Number of Response
Correct Responses		***1996 (77.0%)***
A10	Plotting 'dissolving time' against 'temperature of water', and both axes are correctly labeled with correct units	1333 (51.4%)
A11	Plotting 'dissolving time' against 'temperature of water', and both axes are correctly labeled with some units (s/ºC) missing	128 (4.9%)
A12	Plotting 'temperature of water' against 'dissolving time', and both axes are correctly labeled with correct units	473 (18.2%)
A13	Plotting 'temperature of water' against 'dissolving time', and both axes are correctly labeled with some units (ºC /s) missing	62 (2.4%)
Incorrect Responses/Nil Response		***596 (23.0%)***
A91	Plotting 'dissolving time' against 'temperature of water', and both axes are correctly labeled with no/incorrect units	56 (2.2%)
A92	Plotting 'temperature of water' against 'dissolving time', and both axes are correctly labeled with no/incorrect units	21 (0.8%)
A93	Graph with one of the axes unlabeled/ incorrectly labeled	140 (5.4%)
A98	Other mistakes related to axes (e.g. double axis)	4 (0.2%)
A99	*Graph without showing labelled axes/ Unattempt*	375 (14.5%)

Appropriate choice of scale:

CODE	ITEM	Number of Response
Correct Responses		***1141 (44.0%)***
S10	Graph with appropriate scale in both x- and y-axes	1141 (44.0%)
Incorrect Responses/Nil Response		***1451 (56.0%)***
S91	Graph with the scale in one of the axes not equally spaced	901 (34.8%)
S92	Graph with a poor scale (e.g. the graph is highly compressed or falls out of the graph paper) in either axis	146 (5.6%)
S93	Scale using direct entry in either axis	80 (3.1%)
S94	The scale of 'time' axis (or 'temperature axis') in reverse order	66 (2.5%)
S98	Other mistakes related to scale	41 (1.6%)
S99	*Graph without scale in both axes/ Unattempt*	217 (8.4%)

Graph showing a curve passing through the points:

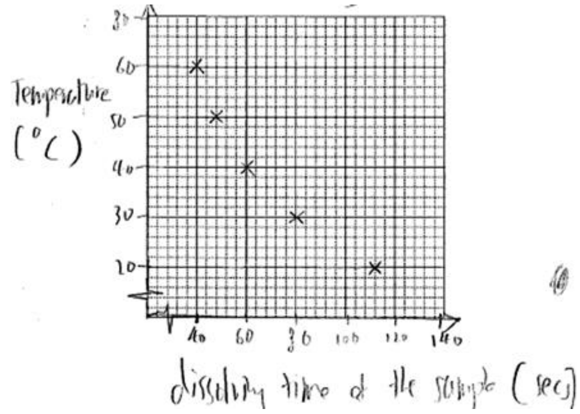
CODE	ITEM	Number of Response
Correct Responses		***1285 (49.6%)***
G10	A curve passing through all points	416 (16.0%)
G11	A graph with line segments joining all points	869 (33.5%)
Incorrect Responses/Nil Response		***1307 (50.4%)***
G90	No line/curve shown on the data points	65 (2.5%)
G91	All or some of the points are incorrectly plotted (and failure to give a curve)	517 (19.9%)
G92	A straight line instead of a curve passing through all points	115 (4.4%)
G93	The 1st data point of the curve connected to the origin by a straight line	142 (5.5%)
G94	The 1st data point of the curve connected to the y-axis / the last data point connected to the x-axis by a straight line	79 (3.0%)
G95	A graph showing a histogram/bar chart/pie chart instead of a curve	91 (3.5%)
G98	Other mistakes related to graph (e.g. multiple lines)	64 (2.5%)
G99	*Unattempt*	234 (9.0%)

Mean Score and Marks Distribution for Q1(d)(i) on plotting graph:

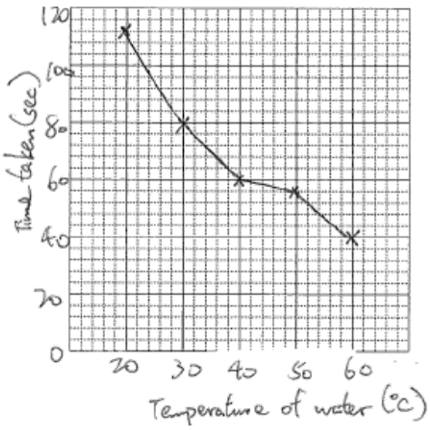
		WG
Mean Score (out of 3):		1.7
Marks Distribution:	**3**	666 (25.7%)
	2	901 (34.8%)
	1	622 (24.0%)
	0	403 (15.5%)

Common Mistakes of Participants in Plotting Graphs

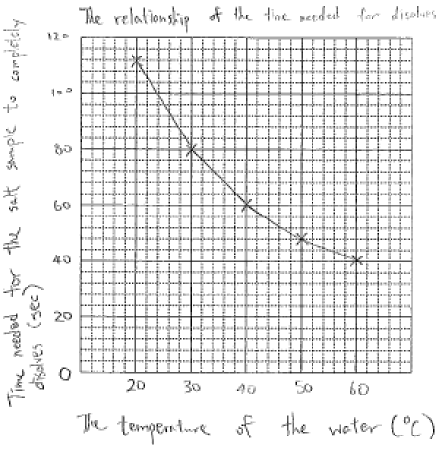
- ***Examples of some partially correct responses***



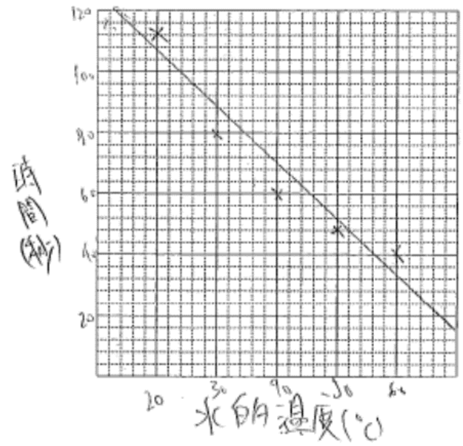
(Code: A12, S10, G90)



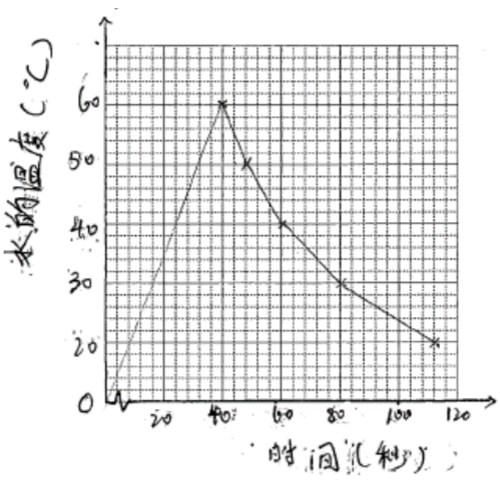
(Code: A10, S91, G91)



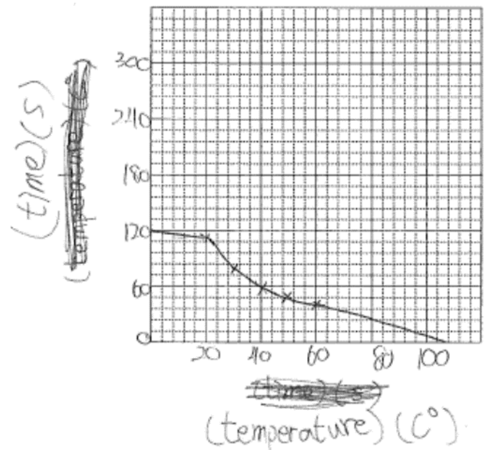
(Code: A10, S91, G11)



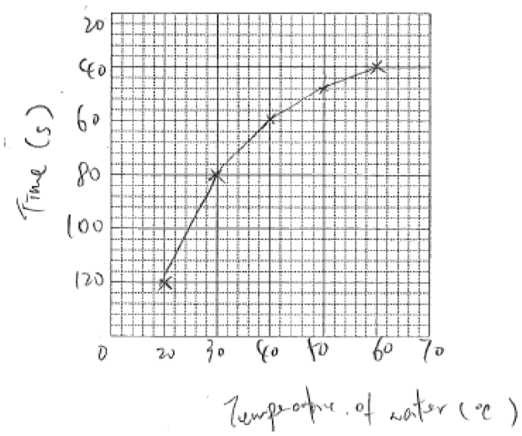
(Code: A10, S91, G92)



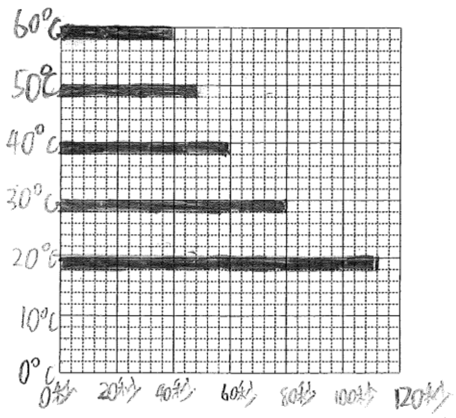
(Code: A12, S91, G93)



(Code: A10, S92, G94)

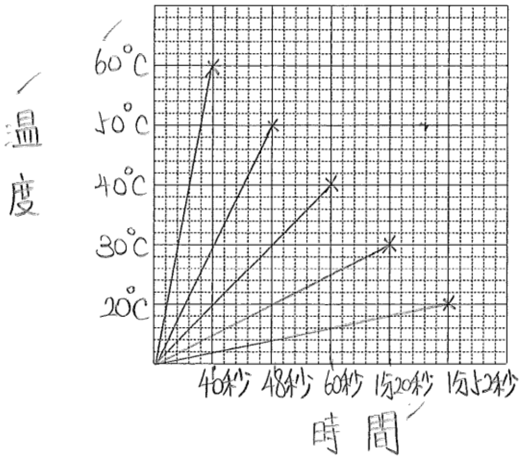


(Code: A10, S94, G91)

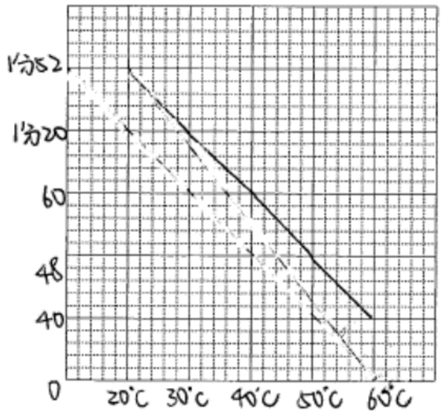


(Code: A99, S10, G95)

- ***Examples of some incorrect responses***



(Code: A98, S93, G98)



(Code: A99, S93, G91)

Double-digit coding system for Q1(e)

Q1(e) *"Suggest why the experiment would NOT give satisfactory results if it is repeated at 80°C."*

The question tests for the ability in Planning and Design of Experiment.

CODE	ITEM	Number of Response
Correct Responses		***176 (6.8%)***
10	Very short dissolving time leads to a greater error in time measurement	143 (5.5%)
11	It is difficult to keep the temperature of the water at 80°C steadily as the heat loss to the surrounding is great.	33 (1.3%)
Incorrect Responses/Nil Response		***2416 (93.2%)***
90	Evaporation/ boiling off/ loss of water/ volume of water becomes smaller	332 (12.8%)

CODE	ITEM	Number of Response
91	The time measured is close to that of 60ºC (small differences)	231 (8.9%)
92	Incomplete answer - the coarse salt dissolve immediately/too fast without mentioning the error in time measurement	693 (26.7%)
93	Misconceptions – - the high temperature may destroy coarse salt/ - at high temperature, it is difficult to dissolve more coarse salt/ close to a saturated solution/ - the temperature is too hot, it breaks the beaker or glass rod, etc.	163 (6.3%)
94	already getting the trend at 20-60ºC, no need to repeat the experiment	30 (1.2%)
98	Other irrelevant answers	690 (26.6%)
99	*Unattempt*	277 (10.7%)

Examples of some incorrect responses

因為80°C接近水的沸點，水的體積會減少，不是公平測試

(Code: 90)

因為80°C攪拌的時候溶解的時間和60°C的相差很小。

(Code: 91)

水溫太高，鹽可能放進去後立刻就溶解了，無法測量。

(Code: 92)

因為水接近變成飽和溶液，難以再溶解更多鹽。

(Code: 93)

因為已經沒有必要，先前已做過5次實驗

(Code: 94)

因為線條圖顯示了溫度越高，效果越不明顯，所以在80°C重複該實驗可能效果不大。

(Code: 98)

因為做了多次，難以確保數據的真實。

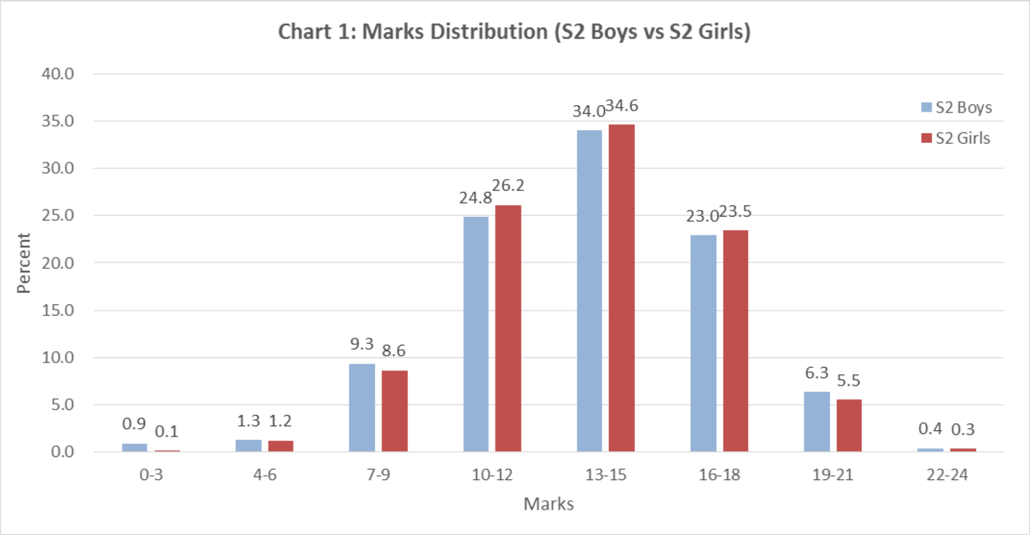
(Code: 98)

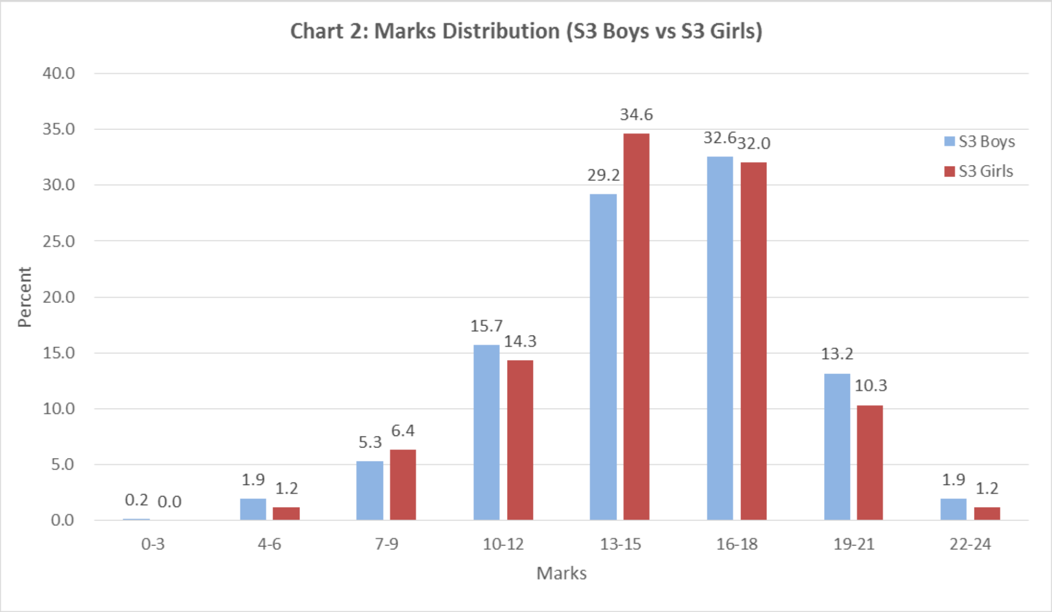
It was because the temperture was too high, so that the result become not satisfactory.

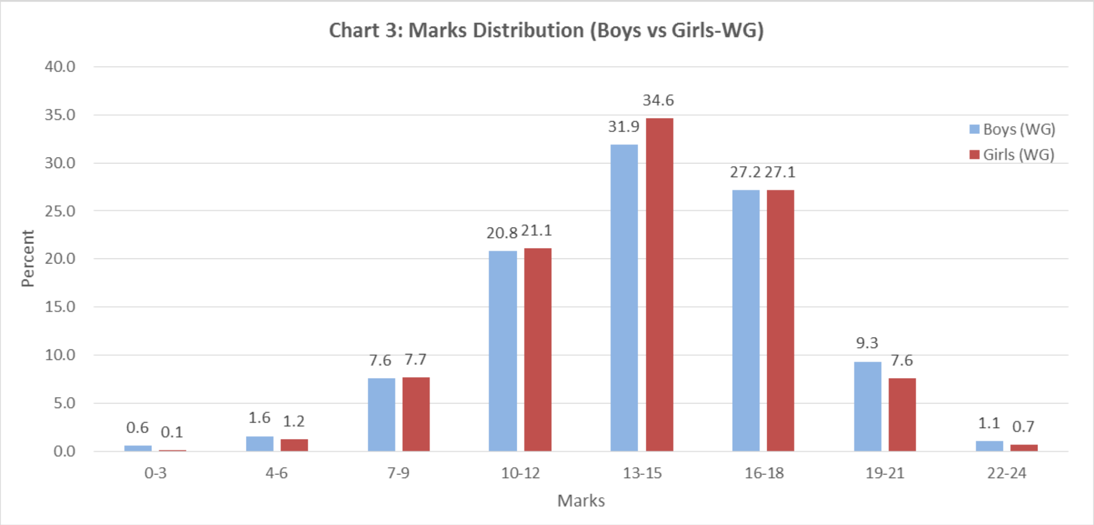
(Code: 98)

Appendix IIIa

Performance of Boys and Girls in Section A of 2018 SAT

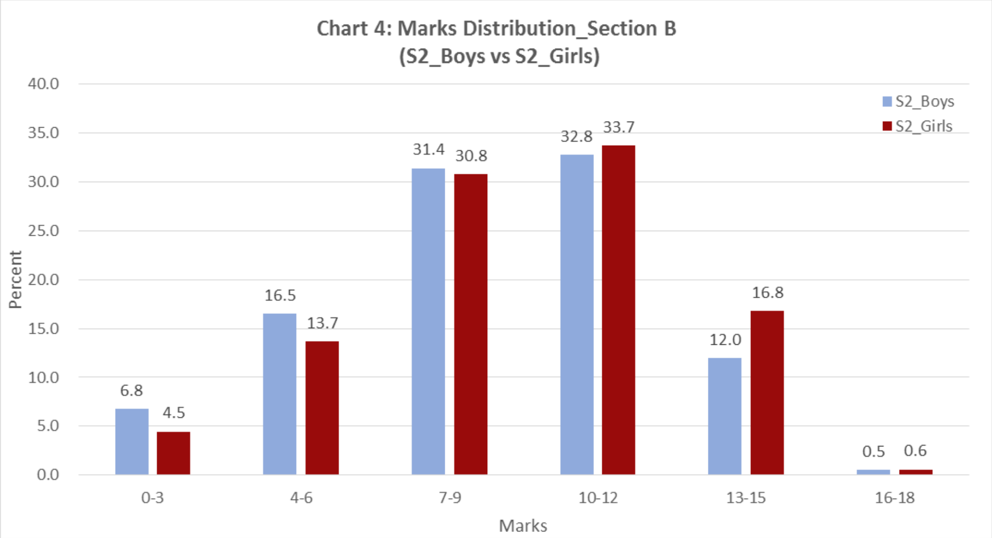


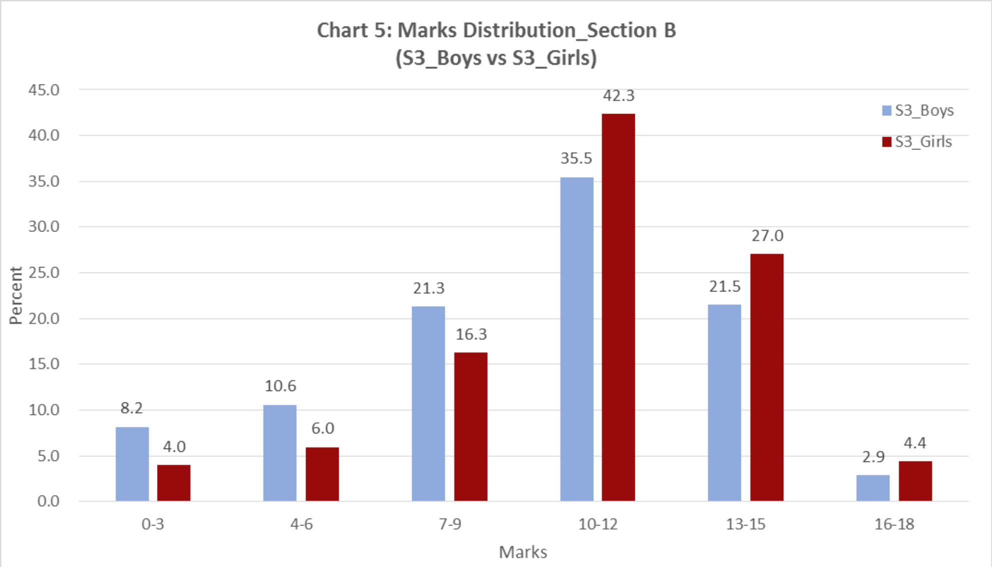


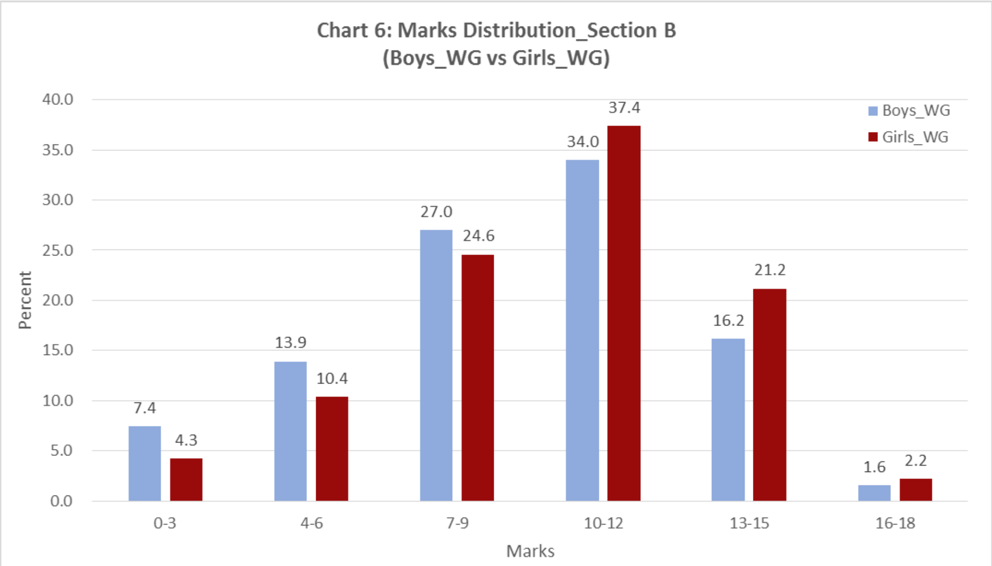


Appendix IIIb

Performance of Boys and Girls in Section B of 2018 SAT







1. Diamond, Gold, Silver and Bronze Awards given out in 2018 SAT

Award type	Number
Diamond	147 (5.7%)
Gold	497 (19.2%)
Silver	610 (23.5%)
Bronze	832 (32.1%)

2. Overall Performance of the Participants

	S2	S3	Whole Group (WG)
Mean score: *(out of 54)*	29.5 (55%)	32.6 (60%)	30.8 (57%)
Standard deviation:	7.21 (13.3%)	7.70 (14.3%)	7.58 (14.0%)