Planning Investigations Mark Scheme

Q1.

Question	Answer	Additional guidance	Mark
	B1 for each of five correct comments e.g. • the hypothesis should be a statement not a question • time intervals are not exhaustive (e.g. there is no option for 0 or 7+) • time intervals need to be longer as many students are likely to spend more than 6 hours a week • a scatter graph is not appropriate (as horizontal axis is not numerical data) • it might be difficult for students to know the times as they might go on social media multiple times a day for short periods • it is not necessary to know the name of the student • he has not planned to collect the relevant information for a time series graph • calculating estimates of the means is sensible • finding the gender of the student is relevant	B1 for each correct comment on the appropriateness of the hypothesis or plans for collecting, processing and presenting the data	(5)

Q2.

Question number	Answer	Additional guidance	Mark
	Collecting data B1 for identifying one appropriate thing that should be included in the plan for collecting data **and** B1 for explaining why this aspect is appropriate OR B1 for deciding what data to collect and/or how to collect and record it **and** B1 for an appropriate reason OR B1 for a strategy to process data **and** B1 for an appropriate reason OR B1 for designing a collection method for primary/secondary data **and** B1 for an appropriate reason OR B1 for appreciating the importance of acknowledging sources **and** B1 for an appropriate reason OR B1 for recognising where issues of sensitivity may influence data availability **and** B1 for an appropriate reason **Processing and presenting** B1 for planning to organise and/or process data **and** B1 for an appropriate reason OR	B1B1B1 for each of three planned elements and B1B1B1 for each of three appropriate reasons from their three things in the statistical enquiry cycle. Maximum 4 marks if only one aspect (from Collecting data, Processing and presenting, Interpreting and Evaluating) is referenced. B1 for e.g. use amount of time measured to the nearest minute **and** B1 for e.g. this is sufficient as there will be a large range of times B1 for e.g. collect data for 23 boys and 23 girls **and** B1 for e.g. this will make the calculation of quartiles easier B1 for e.g. use random sampling **and** B1 for e.g. this reduces bias as Gary's friends/class/peers may generally watch the same programs B1 for e.g. use primary data **and** B1 for e.g. this increases reliability as Gary will know how the data was collected B1 for e.g. A student (Gary) should collect the data **and** B1 for e.g. students are more likely to give an honest answer to a fellow student (less threatening) B1 for e.g. use a grouped frequencies table for the data **and** B1 for e.g. as this will enable a quick way of estimating the mean or this can be used to draw a histogram	(6)

Question number	Answer	Additional guidance	Mark
continued	B1 for planning to generate diagrams and/or visualisations to represent the data **and** B1 for an appropriate reason OR B1 for planning to generate statistical measures to compare data **and** B1 for an appropriate reason **Interpreting** B1 for planning to interpret diagrams and/or calculations/measures **and** B1 for an appropriate reason OR B1 for planning to make an inference and/or prediction **and** B1 for an appropriate reason **Evaluating** B1 for planning to identify weaknesses in approach or representation **and** B1 for an appropriate reason OR B1 for planning to refine the processes to elicit further clarification of the hypothesis **and** B1 for an appropriate reason	B1 for e.g. use box plots **and** B1 for e.g. these will enable the comparison of both the medians and the IQRs (i.e. the distributions) of the data B1 for e.g. interpret results for each individual school year **and** B1 for e.g. as different years could have different watching habits B1 for e.g. by comparing means **and** B1 for e.g. you can see whether the amount of time that boys spend watching TV is greater, in general, than the amount of time that girls spend watching TV B1 for e.g. use the results from the school to predict the results nationally **and** B1 for e.g. as students in different parts of the country are likely to have the same watching habits B1 for e.g. choose not to display the information in histograms **and** B1 for e.g. as "the target audience" may not know how to interpret them B1 for e.g. consider using more than one type of visual representation (for the same information) **and** B1 for e.g. as different representations focus on different aspects of the data	

Q3.

Question	Answer	Additional guidance	Mark
	Data collection B1 for identifying one appropriate thing that should be included in the plan for data collection **and** B1 for explaining why this aspect is appropriate		(6)
	B1 for deciding what data to collect and/or how to collect and record it **and** B1 for an appropriate reason OR B1 for designing a collection method for primary/secondary data **and** B1 for an appropriate reason	B1 for e.g. use distance to holiday to the nearest mile **and** B1 for e.g. this is sufficient as there will be a large range of distances B1 for e.g. use quota sampling **and** B1 for e.g. this ensures that you have data for each price range OR B1 for e.g. collecting data for prices of only one type of holiday, for example all inclusive **and** B1 for e.g. this means that the prices aren't affected by factors other than distance	
	OR B1 for recognising where issues of sensitivity may influence data availability B1 for an appropriate reason	B1 for e.g. find prices from a holiday company website rather than asking people how much they paid for their holiday **and** B1 for e.g. as people may not want to say how much they spent on their holiday	
	Presenting and interpreting diagrams B1 for identifying an appropriate method for presenting data in a diagram **and** B1 for explaining how this will be interpreted in the context of the investigation	B1 for e.g. use a scatter diagram for the data **and** B1 for e.g. as this will allow you to see if there is a correlation (between the distance travelled and the price of the holiday)	

Question	Answer	Additional guidance	Mark
	Presenting and interpreting calculations B1 for identifying an appropriate method to generate statistical measures to compare data **and** B1 for explaining how this will be interpreted in the context of the investigation	B1 for e.g. calculate the mean point to plot an accurate line of best fit **and** B1 for e.g. as you could use the line of best fit to make predictions about the cost of holidays of a certain distance OR B1 for e.g. calculate the equation of the line of best fit **and** B1 for e.g. to be able to see the relationship between distance travelled and price OR B1 for e.g. use a statistical package to calculate Spearman's rank correlation coefficient **and** B1 for e.g. to measure the strength of any correlation between distance travelled and price	

Q4.

Question	Answer	Additional guidance	Mark
(a)	B1 9.8		(1)
(b)	B1B1 for considering the data eg • 'small data set' • 'data set may not be representative (of all adult cats)' • 'not a random sample' B2 for naming an appropriate average with a supporting reason eg • 'she should use the mean as it uses all the data' • 'she should use the median as it is less affected by extreme values' B1 for considering the effect of the outlier eg • 'outliers affect the mean' • 'outlier will have less of an effect the median' • 'including the outlier may lead to an overestimate' • 'it would be appropriate to include the outlier as it may be a real data value'	B1 for each correct consideration of the data (up to B2) B2 for correct of average and supporting reason (B1 for naming an appropriate average and an attempt at a reason) B1 for an appropriate comment about the effect of the outlier	(5)

Q5.

Question	Answer	Additional guidance	Mark
	B1B1B1 for planning for each of 3 uses of technology Rebecca can use technology to **Plan for Processing and Presenting** • e.g. sort/order data columns by age/number of appointments • e.g. generate random numbers • e.g. remove extraneous data (phone number/address) columns • e.g. identify missing data • e.g. remove extraneous symbols • e.g. automate the calculation of summary statistics • e.g. automate the production of visual representations • e.g. remove outliers B1B1B1 for justifying why it is appropriate/advantageous **Advantages of using technology** • e.g. saves time • e.g. can reduce errors (in lengthy calculations) • e.g. facilitates use of all data (unbiased) • e.g. easier to correct a mistake • e.g. more visually appealing	B1B1B1 for each use of technology B1B1B1 for each advantage over processing data by hand Ignore extraneous non-contradictory comments	(6)

Q6.

Question	Answer	Additional guidance	Mark
(a)	B1 Scatter (diagram) is most suitable graph	B1 for correct choice of graph	(5)
	B1 ... as data is **bivariate**/paired	B1 for reason explaining appropriateness for graph choice. Allow equivalent reference to bivariate data eg she was expecting to see correlation	
	B1 line of best fit drawn (through the (double) **mean point**)	B1 for recognising that the double mean point should be used in drawing a best fit line	
	B1 **gradient** (of line of best fit) is 0.8 / will tell her the relative rates of change in gross pay	B1 for correctly recognising that statistical reasoning for the comparison is based on gradient. Accept equivalent, eg steepness	
	B1 points will be close to the line of best fit, OR there is strong **correlation** (so evidence is strong)	B1 for recognising how the high value of pmcc supports the validity of the conclusion OR that points will be close to a straight line	
(b)	B1 Correct reason from • Old data / out of date • Small sample • Secondary data so don't know how collected • May be affected by (varying) exchange rates	B1 for an equivalent reason for conclusion to be inappropriate. Ignore excess statements if not contradictory.	(1)