

2020 VCE Applied Computing: Data Analytics examination report

General comments

In 2020 the Victorian Curriculum and Assessment Authority produced an examination based on the *VCE Applied Computing: Data Analytics Adjusted Study Design for 2020 only*. The 2020 VCE Applied Computing: Data Analytics exam saw a change in the allocation of marks for Sections B and C from previous years. Section B now has a total of 20 marks and Section C has a total of 60 marks. The examination's format remained the same: Section A (Multiple-choice questions), Section B (Short-answer questions) and Section C (Case study).

The multiple-choice questions were answered well. Areas in which students performed well were data visualisation/infographics and data integrity. Students had more difficulty with queries and database design tools.

It was evident from student responses to the short-answer questions that students understood data integrity, project management, data collection and validation techniques.

Areas for improvement include students' understanding of basic definitions from the study design, such as verification, formats and conventions, and correct referencing standards. Students would benefit from having clear, separate lists of what a format and convention is for each software tool, to avoid confusing the two.

Another area for improvement is the writing of evaluation criteria (Section C, Question 7b.). While an evaluation criterion does not have to be in the form of a question, it may assist students who struggle with the concept to formulate it as a question. The criterion must be measurable; however, a large number of students made criterion statements that were not measurable or consisted of one or two words, and therefore could not be awarded marks.

Other areas that should be addressed include the following:

- When a question requires students to justify a selection or action identified, the answer must include both reasons for and reasons against the selection to receive full marks. It was clear that most students did not understand this requirement and could not be awarded full marks as a result of not comparing.
- When quoting from legislation, student must include the full name of the Act and the correct year.
- When two or more items need to be addressed, indicated by 'and' in the question (Section C, Questions 4b. and 9), students must address all the items to receive full marks.

Specific information

Student responses reproduced in this report have not been corrected for grammar, spelling or factual information.

This report provides sample answers or an indication of what answers may have included. Unless otherwise stated, these are not intended to be exemplary or complete responses.

The statistics in this report may be subject to rounding resulting in a total more or less than 100 per cent.

Section A – Multiple-choice questions

The table below indicates the percentage of students who chose each option. The correct answer is indicated by shading.

Question	% A	% B	% C	% D	Comments
1	16	75	6	3	
2	11	22	34	32	A is incorrect as NAS does not need extra space for individuals. B is incorrect as NAS is not affected by local hardware failures. C is incorrect as NAS is not part of their local computer.
3	15	79	4	2	
4	1	0	89	10	
5	91	6	1	1	
6	9	1	20	69	
7	2	88	7	3	
8	1	93	5	1	
9	3	27	67	2	B is incorrect as the data needs to be easily manipulated.
10	8	7	78	7	
11	27	68	3	2	A is incorrect because you don't protect data from authorised access, you protect it from unauthorised access.
12	96	2	2	0	
13	6	6	10	78	
14	84	10	1	4	
15	3	46	14	38	Answer needs to include members that have purchased 6 or more tickets, but there is a limit of 8. B does not cover the upper limit.
16	77	12	9	1	
17	10	8	32	49	C is incorrect as it is irrelevant to what the data formats and conventions are, as they can be changed when manipulating the data.
18	79	11	8	2	
19	17	3	38	42	D is incorrect as an annotated data structure is not a design tool in the study design.
20	0	2	7	90	

Section B

Question 1a.

Marks	0	1	Average
%	20	80	0.8

One feature missing from the Gantt chart is milestone(s).

Question 1b.

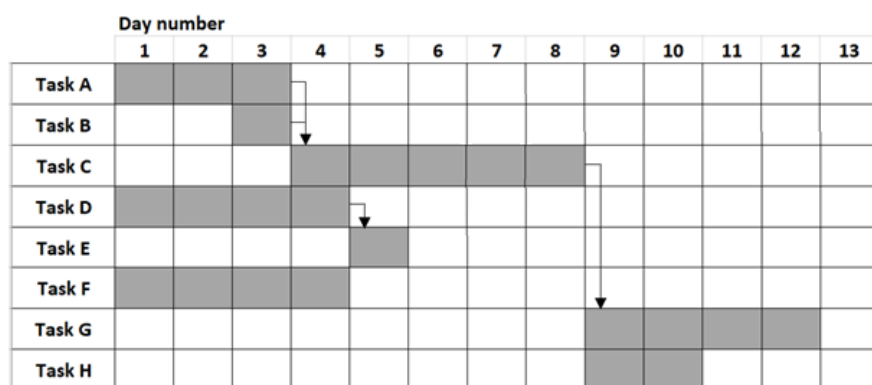
Marks	0	1	Average
%	78	22	0.2

The tasks on the critical path are A, C and G.

Question 1c.

Marks	0	1	Average
%	28	72	0.7

The majority of students who gained a full mark also moved Task H an extra two days, which was not necessary to obtain the mark.



Question 1d.

Marks	0	1	2	Average
%	14	37	49	1.4

To obtain full marks, students needed to explain what a critical path is (the longest path that has no slack time) and then state the project would run two days longer. A suggested response could be as follows:

A critical path is the longest possible path through a project, without any gaps. If task C is extended, then the project will run two days late from that point on.

Question 2a.

Marks	0	1	2	Average
%	29	29	42	1.1

Students who gained full marks included a method (survey or interview) and then explained why that method would be suitable: for example, 'Survey – because they can get data from many people quickly'.

A number of students were able to identify the method but could not explain why it was suitable.

Question 2b.

Marks	0	1	Average
%	89	11	0.1

Students needed to identify a constraint for the data visualisation and relate it to the question. A large number of students misinterpreted the question and identified a constant for collecting the data. No marks were awarded for these answers, nor for responses that identified a constraint but did not include an explanation.

Possible answers:

- Economic: the due date for the infographic to be completed
- Technical: must display on multiple devices
- Social: technical ability of the residents
- Usability: the ability of residents to use the infographic.

Students are encouraged to familiarise themselves with the Problem-solving methodology specifications in the study design.

Question 2c.

Marks	0	1	2	Average
%	79	16	5	0.3

Students who gained full marks suggested that they could look at the data collected from the residents and Meals by Delivery and then brainstorm ideas to create a list.

Other responses that gained marks included coding the data into lists that would allow for a list to be created.

Question 3a.

Marks	0	1	Average
%	47	53	0.5

Students were asked to consider the presentation and give a reason as to why the data needed to be checked and verified. The most common answers were 'NSW is not a city' or 'NSW is way out of proportion to the other two cities'.

Question 3b.

Marks	0	1	2	Average
%	69	15	16	0.5

Students who obtained marks recognised that verification related to proofreading and that the data entered needed to be matched against the source data.

Students needed to distinguish between verification and validation, both of which have clear definitions in the study design. It was evident that many students did not know the difference between the two. Responses that gained no marks discussed validation techniques, such as range or data type checks, instead of verification.

Question 4a.

Marks	0	1	2	Average
%	15	41	44	1.3

Students needed to outline two reasons why Arnis would want to search existing data sets from government organisations. Appropriate reasons included the following:

- saving time – no need to waste time collecting the data and analysing it himself
- saving effort – easier to use data that has already been collected
- greater number of responses – likely to give a more realistic representation of the situation / potential sample size would be greater using an existing data set
- wider access to more data
- more likely to be a better cross-section / less bias
- reputable organisations that should have authentic data
- reliable and authentic sources.

Marks were not awarded if a student mentioned saving money.

Question 4b.

Marks	0	1	2	Average
%	37	46	17	0.8

Many students could only explain one reason why acquiring this government data would be important to Arnis's organisation.

Acceptable answers included:

- As Arnis manages a water conservation organisation, gaining data on droughts can aid his organisation in determining their course of action.
- Acquiring this government data would be important to Arnis' organisation as it would provide them with accurate data that will make his organisation more reputable as it is an authentic source.

Question 5

Marks	0	1	2	3	Average
%	51	33	12	4	0.7

Students needed to describe three different types of information other than text. A list of acceptable answers could have been drawn from the 2020 VCE Applied Computing: Data Analytics Software tools and functions document.

Acceptable types of information were:

- maps (i.e. flow, density, heat, cartogram)
- charts/graphs (i.e. ring, word cloud, matrix, pie, stacked bar, bar, histogram, scatter plots, line, timelines, time series, sparkline, waterfall, funnel)
- diagrams (i.e. node-link, alluvial, Venn, sunburst, tree, polar area)
- tables
- images/pictures/photos
- sound/audio/video.

Students needed to describe how each type of information related to water safety. If the students just listed three types, they were awarded one mark in total for the question.

Section C

Question 1a.

Marks	0	1	2	Average
%	56	23	21	0.7

Students were asked to write a research question that would help Evan present his findings about young travellers.

A good research question must be clear, focused, concise, complex and arguable, and should not be answered with a yes or no.

Possible responses include:

- Why has the number of young travellers staying in homestay accommodation increased in recent years?
- Why has the preference for young travellers' accommodation increased in recent years?
- Why do young travellers prefer homestay accommodation to hotels?
- What extra activities at the hotel would encourage more young travellers to book holidays there?

Question 2

Marks	0	1	2	3	Average
%	7	6	84	3	1.8

Element	Category
Only data regarding young travellers is used	Scope
Information displayed should be easy to read	Non-functional requirement
Only data from 2015–2019 is to be used	Scope (does not fit any constraint definition)

Students should note that just because there are three options, they do not always have to be used.

Question 3a.

Marks	0	1	Average
%	23	77	0.8

The most common acceptable responses were:

- low costs / cheaper to run
- convenient data gathering
- good statistical significance
- little or no observer subjectivity
- precise results
- little effort from survey setters
- faster than other ways
- more responses
- can reach a large number of people in a shorter time.

Marks were not awarded if the students just mentioned efficiency or effectiveness.

Question 3b.

Marks	0	1	Average
%	31	69	0.7

Students needed to suggest an appropriate technique for collecting secondary data and also explain why it would be useful.

The most common correct responses were:

- download from an existing repository – useful because the data has already been collected and organised
- look at websites – a quick way to get the data you need
- use the internet – an easy way to find the data you need.

Question 3c.

Marks	0	1	Average
%	23	77	0.8

The only acceptable answers were American Psychological Association or APA.

A significant number of students wrote Harvard as an answer, which suggests they were not aware of the correct referencing standard. Teachers should refer to the key knowledge in the current study design.

Question 4a.

Marks	0	1	2	Average
%	16	38	46	1.3

Students were asked to identify two reasons why there may be concerns about the integrity of the data the study found. The most common acceptable responses were:

- The study is 10 years old, so it is not timely.
- The study refers to visitors to Melbourne on business trips, so it is not relevant.
- Landmarks and places of interest may not be attractive to young travellers (relevant).
- It references a different category of visitor (i.e. business), so it is not relevant.
- It is focused on the city of Melbourne, so it is not relevant.
- The study analyses tourist sites, not hotel facilities, so it is not relevant.
- Data from 2015 might not be relevant now.

To gain full marks, students needed to include a data integrity characteristic in their statement, and the only acceptable characteristics were 'reasonableness', 'relevance' and 'timeliness'.

Question 4b.

Marks	0	1	2	Average
%	58	34	8	0.5

In their answer to this question, students were expected to recommend a technique for improving data integrity for both searching and analysing.

Many students recommended two techniques for either searching or analysing and therefore could not be awarded full marks.

The most common acceptable answers were:

- searching – check the source of the data from a website to see if the data is authentic
- analysing – using validation techniques such as existence and range checking to make sure the data is accurate.

Students needed to link the responses to either authenticity, accuracy or reasonableness.

The following is an example of a high-scoring response.

Searching: When analysing data, check the source of where it came from to verify that it is a genuine and trustworthy source. i.e. government or education website.

Analysing: Data can be checked using an existence check so that no data is missing, this will assist in making the data accurate.

Question 5

Marks	0	1	2	3	Average
%	18	45	33	5	1.2

Students were required to justify why an infographic was more suitable than a dynamic data visualisation. Full marks were awarded to responses that indicated a reason why the infographic was more suitable and why dynamic data visualisation was less suitable, and that also related this information to the case study.

The majority of students did not state why the dynamic data visualisation was not suitable and could not gain full marks.

The following is an example of a high-scoring response.

An infographic contains multiple visualisations and thus can present a much larger breadth of information. As Evan wants to look at both travel and interests of young travellers at least, an infographic is more suitable. An infographic also utilises more information types such as text, images and graphs, which can be used to present information clearly creating a more effective solution.

Question 6a.

Marks	0	1	Average
%	50	50	0.5

Acceptable answers were breach of *Copyright Act 1968* or breach of copyright laws.

When students refer to a specific Copyright Act, it is important that they also include the year. General references to copyright laws do not necessarily require specific reference to an Act or year.

Too many students confused copyright laws with plagiarism in relation to images and charts. Some students incorrectly referred to plagiarism laws due to the use of an image without permission.

Question 6b.

Marks	0	1	Average
%	38	62	0.6

Students were required to state what needed to be done to ensure that the legal requirements were met. Acceptable answers focused on seeking written permission from the owner.

Some students incorrectly assumed that the use of images and charts only required a source citation. This is only true if the charts and images are free; at all other times, one needs to seek permission to use the images and charts.

Question 7a.

Marks	0	1	2	Average
%	11	61	28	1.2

The most common acceptable answer was 'To give Huang and Lucinda a choice so that they can pick what solution best suits them'.

Students who gained full marks explained why Evan would draw two design ideas and related it to the case study. Many students gave just an explanation.

The following is an example of a high-scoring response.

Evan would draw two design ideas as it gives them more options as to what they want. If they don't like some features of one they can use the other. They can also see a range of ideas produced as there is 2 options.

Question 7b.

Marks	0	1	Average
%	70	30	0.3

Acceptable responses were:

- Which design is more visually appealing?
- Which design is easier to understand?
- Which is more readable on mobile phone screen?

The criterion statement should have included words such as visual impact, ease of understanding, clarity, balance or readability. A one-word answer did not receive a mark.

Question 7c.

Marks	0	1	2	3	Average
%	5	26	53	17	1.8

As with all 'justify' question responses, there needed to be a comparison between the two choices. Students needed to select Design A or Design B, state why that design was selected and explain why the other design was not selected.

The following is an example of a high-scoring response.

Design A should be chosen as the variety of shapes of the text boxes and the chart is visually appealing and attractive, to the users. Design A only uses rectangle text boxes and therefore visually boring and not attractive, to the users. The use of the guitar on design B is relevant to the topic, design A lacks images.

Question 7d.

Marks	0	1	2	Average
%	39	36	26	0.9

Design principles needed to relate to appearance; the only acceptable answers were alignment, balance, image use, space, and text and table formatting.

Question 8a.

Marks	0	1	2	3	4	Average
%	39	13	29	8	11	1.4

Acceptable software functions were:

- import/export (e.g. Excel, CSV, Word Table)
- set up tables and manually enter data
- use of sort, filters or queries to remove unwanted data
- perform calculations
- create relationships.

To gain marks, students needed to provide an appropriate description of how each function would work in relation to the case study.

Here too the responses could have come directly from the 2020 VCE Applied Computing: Data Analytics Software tools and functions document.

Question 8b.

Marks	0	1	2	Average
%	66	26	8	0.4

Students were required to identify and outline one format and one convention that could be applied to the spreadsheet.

Acceptable answers were:

- format – creating a table, chart or columns in the spreadsheet
- convention – heading at top, currency formatting to two decimal places for dollars and cents, numbers to right, text to left, bold text for headings.

Student responses indicated that many were unclear on the difference between formats and conventions, with a fair number of students writing conventions in the format section.

Question 8c.

Marks	0	1	2	Average
%	35	13	52	1.2

Only one type of validation technique was possible with this question, and that was an existence check. Students then needed to describe how the validation technique avoided the blank cells issue.

The following is an example of an acceptable response:

An existence check avoids blank cells so as to ensure that fields are filled in as necessary, improving and validating the completeness of the records.

Question 8d.

Marks	0	1	2	Average
%	86	9	5	0.2

The question asked students to describe a test that Evan could perform to ensure that another validation rule will restrict the ages in the data to between 8 and 25. To obtain full marks, students needed to state five items of test data (outside, on, inside, on and outside the range) and then indicate that the actual results needed to be compared to the predicted outcome. For example, 'Evan should come up with a set of test data, such as 17, 18, 19, 25, 26. He should enter these one at a time, expecting 17 and 26 to be rejected as they are out of range'. In this example, 17 is below the lowest range and tests that they have the correct rules.

Question 9

Marks	0	1	2	3	4	Average
%	47	14	24	7	8	1.2

Students who gained full marks outlined one technique that could be used for effectiveness and one technique for efficiency, as stated in the question; however, many students outlined two techniques for efficiency or two for effectiveness.

High-scoring responses must have included one of each of the following:

- for efficiency: time or effort
- for effectiveness: clarity, communication of message, readability or relevance.

Acceptable techniques were as follows: count if, conditional formatting, sorting, filtering and queries.

The following is an example of a high-scoring response.

Conditional formatting could be used to highlight and identify the top five responses, which could then effectively be extracted and turned into an infographic.

Data could be sorted in each column to be organised for highest to lowest value. The top five responses could then easily be identified and selected to be used in the infographic

Question 10

Marks	0	1	2	3	Average
%	65	22	9	4	0.5

Students were required to describe how the infographic was to be tested. Acceptable answers needed to include the element to be tested, mention the expected and actual results and compare the two.

Many students were confused between testing and evaluating the infographic and described ways that the infographic could be evaluated.

Accepted responses were:

- Evan should compare the infographic with what he expected it to look like at the design stage.
- Open the infographic in different formats or devices to see if it opens as intended. Compare this to what was expected.
- Print the infographic and compare the printed version to see if what was expected actually was printed.

The following is an example of a high-scoring response.

Evan should create a testing tables outlining the test data, expected output and actual output. To test the ages are between 18 and 25, they should test below the range (17), on the range (18), in the range (22), on the upper range (25) above the upper range (26). Evan should get error messages for 17 and 26 and no errors for 18, 22 and 25.

Question 11

Marks	0	1	2	Average
%	57	38	5	0.5

Students who scored highly were able to outline an evaluation strategy that includes a suitable timeframe for the evaluation to have been conducted (3–6 months). For example, 'After the infographic has been viewed by users for 3–6 months, they should seek feedback from users via a survey'.

Other acceptable data collection methods were:

- website traffic – monitor the number of visitors to the infographic webpage
- subscriber growth – plot subscriber numbers to determine subscriber growth
- search rankings – monitor rankings to determine popularity ratings
- time on site – measure time each visitor spends on the infographic web page
- social media followers – record the number of social media followers regularly to determine popularity status
- social media shares – determine the number of times the link to the infographic webpage is shared with other users
- clickthrough rate (CTR)
- feedback
- focus group
- interviews
- observations.

Question 12a.

Marks	0	1	Average
%	72	28	0.3

The correct legislation was the *Privacy Act 1988*. Whenever they refer to a specific Act, students should take care to include the year.

Question 12b.

Marks	0	1	2	Average
%	19	56	25	1.1

For this question, students needed to outline a possible legal consequence and what would happen to the hotel.

The most common correct responses were:

- fines – the hotel could be fined and could lose business as a result
- legal action – guests could take legal action against the hotel, which could cost it a lot of money.

'Loss of reputation' was not an acceptable answer.

Question 13

Marks	0	1	2	3	4	Average
%	7	12	37	33	11	2.3

Students were required to name a possible threat and suggest a strategy to improve the current practice. The strategy should include a number of controls that will assist in protecting the hotel from that threat.

The most common possible deliberate threats were:

- phishing
- malware (spyware, ransomware, virus, worm)
- hacking
- unauthorised access
- disgruntled employee.

Possible strategies may have included:

- installing a firewall
- installing antivirus software
- educating hotel users on suspicious sites and emails.

To gain full marks, the strategy must have related to the threat. Many students could identify a threat but suggested a strategy that was not related to that threat.

The following is an example of a high-scoring response.

Deliberate threat: Hackers

Strategy: To improve the current practices, the hotel should have to implement the use of username and password for both the Wi-Fi and the computer system, to limit the access to authorised users. This would reduce the chance of hackers obtaining information. They should also implement a firewall to reduce access unauthorised access to the network.

Question 14

Marks	0	1	2	3	4	Average
%	20	8	49	8	15	1.9

Students were required to recommend two software security controls and describe how each will improve the hotel's data and information security practices.

Appropriate examples of software security controls were:

- firewall – to control incoming/outgoing data
- usernames and passwords – to only allow authorised users
- hierarchical access – to protect sensitive data
- use of security protocols like WPA2 – to authenticate users
- filter MAC addresses – to allow only authorised access
- adding an extra antivirus software program.

Responses that suggested updating the current antivirus did not gain any marks because it does not improve the hotel's security, since the antivirus is already in place.

Students needed to demonstrate an understanding of what the control does or how it works, rather than just identifying the software security control.

The following is an example of a high-scoring response.

Software security control	Description of improved security practice
Implement usernames and passwords	If Lucinda and Huan implement username and password, it will ensure that only those with the login can access the system. This prevents unauthorised people accessing the data and information.
Firewall	A firewall should be installed, a firewall examines incoming and packets and blocks any threatening or unrequested data packets. This prevents threats from getting into the network by blocking them, thus improving the security practices.

Question 15a.

Marks	0	1	2	3	Average
%	36	21	20	23	1.3

Some acceptable components of the disaster recovery plan were:

- evacuation/emergency
- restoration/recovery (of data)
- test plan.

Question 15b.

Marks	0	1	2	Average
%	20	42	39	1.2

High-scoring responses referred to all of these appropriate back-up strategies:

- frequency (daily, weekly)
- location (Cloud, offsite)
- media (tape, SSD, cloud)
- type (full, any type of partial – incremental, differential).

A strategy is a plan of action designed to achieve the required aim. In this case, students were asked to recommend an appropriate back-up strategy. Including only one of the above four options is not a strategy.

The following is an example of a high-scoring response.

Notrow Hotel should implement a weekly full backup, and a daily incremental back up. The backed-up data should be stored offsite on a portable SSD. This SSD can be rotated with 6 other hard drives.