Example Candidate Responses

Cambridge International AS & A Level

Cambridge International AS and A Level Biology

9700

Paper 5 – Planning, Analysis and Evaluation

Cambridge Advanced

CAMBRIDGE International Examinations

In order to help us develop the highest quality Curriculum Support resources, we are undertaking a continuous programme of review; not only to measure the success of our resources but also to highlight areas for improvement and to identify new development needs.

We invite you to complete our survey by visiting the website below. Your comments on the quality and relevance of Cambridge Curriculum Support resources are very important to us.

https://www.surveymonkey.co.uk/r/GL6ZNJB

Do you want to become a Cambridge consultant and help us develop support materials?

Please follow the link below to register your interest.

http://www.cie.org.uk/cambridge-for/teachers/teacherconsultants/

Contents

Introduction

The main aim of this booklet is to exemplify standards for those teaching Cambridge International AS and A Level Biology (9700), and to show how different levels of candidates' performance (high, middle and low) relate to the subject's curriculum and assessment objectives.

In this booklet candidate responses have been chosen to exemplify a range of answers. Each response is accompanied by a brief commentary explaining the strengths and weaknesses of the answers.

For each question, each response is annotated with a clear explanation of where and why marks were awarded or omitted. This, in turn, is followed by examiner comments on how the answer could have been improved. In this way it is possible for you to understand what candidates have done to gain their marks and what they will have to do to improve their answers. At the end there is a list of common mistakes candidates made in their answers for each question.

This document provides illustrative examples of candidate work. These help teachers to assess the standard required to achieve marks, beyond the guidance of the mark scheme. Some question types where the answer is clear from the mark scheme, such as short answers and multiple choice, have therefore been omitted.

The questions, mark schemes and pre-release material used here are available to download as a zip file from Teacher Support as the Example Candidate Responses Files. These files are:

Question Paper 22, June 2016	
Question paper	9700_s16_qp_22.pdf
Mark scheme	9700_s16_ms_22.pdf
Question Paper 33, June 2016	
Question paper	9700_s16_qp_33.pdf
Mark scheme	9700_s16_ms_33.pdf
Question Paper 41, June 2016	
Question paper	9700_s16_qp_41.pdf
Mark scheme	9700_s16_ms_41.pdf
Question Paper 52, June 2016	
Question paper	9700_s16_qp_52.pdf
Mark scheme	9700_s16_ms_52.pdf

Past papers, Examiner Reports and other teacher support materials are available on Teacher Support at **https://teachers.cie.org.uk**

How to use this booklet

Example candidate response – high	Examiner comments
Answer **all** the questions. 1 Statements **A** to **E** are about the structure and functioning of enzymes. State the correct term to match each of the statements **A** to **E**. 1 …that needs to be overcome by reactants in order … …es on the active site being partially flexible and …e. …n… …n enzyme, with a tertiary or quaternary structure that results in an approximately spherical shape. Globular **D** The term for enzymes that function outside cells. Extracellular **E** The concentration of substrate that enables an enzyme to achieve half the maximum rate of reaction. Km value [5] [Total: 5]	1 This candidate has responded as requested and given answers that are concise and are **Total mark awarded = 5 out of 5**

Answers by real candidates in exam conditions. These show you the types of answers for each level.

Discuss and analyse the answers with your learners in the classroom to improve their skills.

Examiner comments are alongside the answers, linked to specific part of the answer. These explain where and why marks were awarded. This helps you to interpret the standard of Cambridge exams and helps your learners to refine their exam technique.

How the candidate could have improved their answer

Stating for **E** the 'Michaelis-Menten constant' wou…
However, knowledge that this is also referred to a…
was able to gain full marks.

This explains how the candidate could have improved their answer and helps you to interpret the standard of Cambridge exams and helps your learners to refine exam technique.

Common mistakes candidates made in this question

A. Some candidates only gave the term 'activation' … strictly correct it was allowed.

B. Some candidates gave a mixture of terms, such … 'induced substrate', 'lock and key fit'. The examiner …

C. Named globular proteins were incorrectly given as a response. Of these, 'haemoglobin' was most commonly seen. The spellings of 'globular' were not always correct.

This lists the common mistakes candidates made in answering each question. This will help your learners to avoid these mistakes at the exam and give them the best chance of achieving a high mark.

Assessment at a glance

Candidates for Advanced Subsidiary (AS) certification take Papers 1, 2 and 3 (either Advanced Practical Skills 1 or Advanced Practical Skills 2) in a single examination series.

Candidates who, having received AS certification, wish to continue their studies to the full Advanced Level qualification may carry their AS marks forward and take Papers 4 and 5 in the examination series in which they require certification.

Candidates taking the full Advanced Level qualification at the end of the course take all five papers in a single examination series.

Candidates may only enter for the papers in the combinations indicated above.

Candidates may not enter for single papers either on the first occasion or for resit purposes.

All components will be externally assessed.

Component	Weighting	
	AS Level	A Level
Paper 1 Multiple Choice **1 hour** This paper consists of 40 multiple choice questions, all with four options. All questions will be based on the AS Level syllabus content. Candidates will answer all questions. Candidates will answer on an answer sheet. [40 marks]	31%	15.5%
Paper 2 AS Level Structured Questions **1 hour 15 minutes** This paper consists of a variable number of questions, of variable mark value. All questions will be based on the AS Level syllabus content. Candidates will answer all questions. Candidates will answer on the question paper. [60 marks]	46%	23%
Paper 3 Advanced Practical Skills **2 hours** This paper requires candidates to carry out practical work in timed conditions. This paper will consist of two or three experiments drawn from different areas of the AS Level syllabus. Candidates will answer all questions. Candidates will answer on the question paper. [40 marks]	23%	11.5%
Paper 4 A Level Structured Questions **2 hours** This paper consists of a variable number of structured questions each with a variable mark value (Section A) and a choice of one free response style question worth 15 marks (Section B). All questions will be based on the A Level syllabus but may require knowledge of material first encountered in the AS Level syllabus. Candidates will answer on the question paper. [100 marks]	–	38.5%
Paper 5 Planning, Analysis and Evaluation **1 hour 15 minutes** This paper consists of a variable number of questions of variable mark value based on the practical skills of planning, analysis and evaluation. Candidates will answer on the question paper. [30 marks]	–	11.5%

Teachers are reminded that the latest syllabus is available on our public website at **www.cie.org.uk** and Teacher Support at **https://teachers.cie.org.uk**

Paper 5 – Planning, analysis and evaluation

Question 1

Example candidate response – high

1 Grassland is an important breeding habitat for some birds. These birds feed on plant material and invertebrates. Biodiversity of the habitat is maintained by domestic herbivores, such as sheep, cows and goats, grazing on growing plant material.

A group of students investigated the effect of grazing by domestic herbivores on the plant biodiversity of a grassland as measured by Simpson's Index of Diversity. They investigated two areas. One area was grazed by herbivores and the other area was not grazed for many years because it was surrounded by a fence to keep out the herbivores.

(a) State the data that the students would have collected from the grazed and ungrazed areas to calculate Simpson's Index of Diversity.

The number of individuals in each plant species in the grazed area and the ungrazed area separately
The total number of individuals (from all species) in the grazed area and total number of individuals (from all species combined) in the ungrazed area. [2]

(b) Describe a random (unbiased) method which the students could have used to collect the data needed to calculate the biodiversity of the plant species in the two areas.

The description of your method should be detailed enough for another person to follow.

With a measuring tape, measure the dimensions of the fence surrounding the ungrazed area – Using the same dimension, (length and width), mark out the ungrazed area with a tape – This is to ensure the perimeters of both the ~~area~~ grazed and ungrazed area are kept ~~the~~ same – Now place quadrats of the same size each time (e.g 1m x 1m) randomly scattered within the determined boundaries (marked) of the grazed land – ~~Keep~~ Use a random number generator app to determine the coordinates of where to place the quadrats to avoid bias – In each quadrat, identify the different species of plants carefully and tabulate the number of ~~each~~ plants in each species from all the quadrats. We do not need to know the name of the specie

Examiner comments

1 The candidate gains credit for descriptions of both sets of data needed to calculate Simpson's Index of Diversity.

Mark for (a) = 2/2

2 The candidate gains credit for describing a suitable method of marking a study area, stating that the area should be the same and specifying both grazed and ungrazed grasslands should be sampled.

3 Further credit is gained for a correct choice of apparatus to use to collect data and for specifying this should be a standard size.

4 The candidate describes a suitable method of obtaining random numbers and how these may be used to place quadrats.

5 The candidate makes an acceptable statement about the collection of data from the quadrats. By this point they have achieved maximum credit for this part of the question.

Example candidate response – high, continued

of a certain plant, just be able to identify that they are two different species of plant – Using the same total number of quadrats, repeat this procedure inside the fence that is, the ungrazed land – The table should look as follows:– 6

Species	No. of individuals grazed land	No. of individuals ungrazed land
A		
B		
C		

We might have to use a magnifying glass to identify some plant species – We will now use the formula for Simpson's Index of Diversity to calculate the species (plant) diversity in the grazed and ungrazed land separately – Formula = $1-\left(\frac{\Sigma n}{N}\right)^{2}$ where 'n' is the number of individuals of a specie and 'N' is the total number of plants for all species in grazed/ungrazed land – The answer obtained will be a numerical value from 0 to 1 – A value close to zero shows low species diversity – A value closer to 1 shows high plant biodiversity – We will obtain two values for the Simpson's Index of Diversity, one for grazed land, one for ungrazed land – 7

→ for example if table was like this:–

species	grazed land (number of individuals
A	20
B	30

we'd calculate Simpson's Index by $1-\left[\left(\frac{20}{50}\right)^{2}+\left(\frac{30}{50}\right)^{2}\right]$ for this grazed land [8]

6 Credit could also have been awarded for standardising the number of quadrats used for each type of grassland.

7 The remainder of the answer describing how to use the data to calculate Simpson's Index of Diversity is not relevant.

Mark for (b) = 8/8

Example candidate response – high, continued

The students also investigated the effect grazing had on the height of one particular species of plant. Their hypothesis was:

The mean height of the plant is greater in the ungrazed grassland than the grazed grassland.

(c) State the independent and the dependent variables in this investigation.

independent variable ungrazed or grazed (grassland)

dependent variable mean height of the plant 8 [1]

(d) Table 1.1 shows the results of their investigation.

Table 1.1

sample number	height of plant/mm	
	grazed area	ungrazed area
1	586	~~858~~
2	549	~~873~~
3	526	~~864~~
4	589	901
5	545	~~847~~
6	538	~~862~~
7	573	864
8	549	~~879~~
9	604	~~864~~
10	611	888
mean	567	870
mode	549	864 9
median	561	864

(i) Complete Table 1.1 by writing the values of the mode and median for the ungrazed area. [1]

862
847, 858, 864, 864, 864, 873, 879, 888, 901

Examiner comments

8 Credit is awarded here, although the dependent variable is actually the height of the plant. The mean height is a calculated variable.

Mark for (c) = 1/1

9 The candidate calculates both values correctly.

Mark for (d) (i) = 1/1

Example candidate response – high, continued	Examiner comments

(ii) Use the information and formula below to calculate the standard error for these results.

Give your answers to 3 significant figures.

$S_M = \frac{s}{\sqrt{n}}$

S_M = standard error

s = standard deviation

n = sample size (number of observations)

grazed area: $s = 29.5$

ungrazed area: $s = 15.7$

standard error, grazed area = 9.33

standard error, ungrazed area = 4.96 [2]

10

Standard error is used to calculate 95% Confidence Intervals (CI).

The values for the grazed area are 548.3 mm to 585.7 mm.

(iii) Use the formula below to calculate the confidence intervals for the **ungrazed** area.

95% CI = mean ± 2 S_M

Show your working.

870 + 2(4.96) and 870 − 2(4.96)

= 879.9 and 860.1

ungrazed area 860.1 mm to 879.9 mm [2]

11

(iv) State what information is gained by calculating the confidence intervals.

A 95% confidence interval means that we can be 95% certain that the true value for mean (within the range of) lies ^ above or below two times the standard error – for example, for grazed area, if another sample is collected, we'll be 95% certain the mean ^ would be between 548.3 mm and 585.7 mm [2]

12

10 The candidate calculates the correct values and presents them to an appropriate number of significant figures.

Mark for (d) (ii) = 2/2

11 The candidate calculates the correct values and presents them to an appropriate number of significant figures.

Mark for (d) (iii) = 2/2

12 The candidate shows an understanding of the role of confidence intervals in setting the level of certainty that can be used in assessing data, in this case the mean value. The example that follows makes the same point. Other roles of confidence intervals, such as assessing statistical significance are not considered.

Mark for (d) (iv) = 1/2

Example candidate response – high, continued

(e) The students used the mark-release-recapture method to estimate the population of an invertebrate animal found living on the grassland. They used the formula:

$$\frac{\text{number of animals marked in the first sample} \times \text{total number of animals in the second sample}}{\text{number of marked animals in the second sample}}$$

State two precautions the students should have taken to ensure that the results they obtained were valid.

13

1. The animals that they marked were given sufficient time to mix with the other grassland animals randomly (when they were first released).
2. The markers that they used did not affect the future survival of the animals when they were released. [2]

(f) The population of an invertebrate that feeds on seeds was estimated in both the grazed and ungrazed areas. Predict which area would have the greatest population and give a reason for your choice.

14

choice the grazed one

reason Because animals remove plants (graze on them) (continued below) [1]

Answer continued

→ sometimes by uprooting the whole plants or grasses so that their seeds are no longer covered with soil. The seeds and embryos are exposed like this, also when soil erosion occurs so the invertebrates are ~~also~~ able to feed on many of these that are scattered on bare or almost bare (grazed land).

[Total: 21]

Examiner comments

13 The candidate is awarded marks for both of their answers.

Mark for (e) = 2/2

14 This answer is not awarded any marks. The description of the effect of grazing on plants is not valid. Grazing constantly removes growing shoots, so the production of flowers and seeds is reduced.

Mark for (f) = 0/1

Total marks awarded = 19 out of 21

How the candidate could have improved their answer

(a) The answer was clear, although the two phrases in brackets were a critical part of the answer and would have been better given outside the brackets.

(b) The candidate went into the detail of calculating Simpson's Index of Diversity from the results; it would have been better to omit this as the question only related to a method for collecting the data.

(c) It would have been better to omit the word 'mean' from the answer for dependent variable, as what is being measured is the height of the plants and a mean is a calculated value.

(d) All the calculations were correct. However, in **(iv)** the answer gave the same information twice. The candidate could have improved their answer by commenting on the reliability of the confidence intervals for the grazed and ungrazed grassland.

(f) The candidate's explanation was incorrect, but in general the answer needed to be much shorter. Only one answer line was provided here, indicating that only a minimum number of words were needed.

Mark awarded = **(a) 2/2**
Mark awarded = **(b) 8/8**
Mark awarded = **(c) 1/1**
Mark awarded = **(d) (i) 1/1, (ii) 2/2, (iii) 2/2, (iv) 1/2**
Mark awarded = **(e) 2/2**
Mark awarded = **(f) 0/1**

Total marks awarded = 19 out of 21

Example candidate response – middle

1 Grassland is an important breeding habitat for some birds. These birds feed on plant material and invertebrates. Biodiversity of the habitat is maintained by domestic herbivores, such as sheep, cows and goats, grazing on growing plant material.

A group of students investigated the effect of grazing by domestic herbivores on the plant biodiversity of a grassland as measured by Simpson's Index of Diversity. They investigated two areas. One area was grazed by herbivores and the other area was not grazed for many years because it was surrounded by a fence to keep out the herbivores.

(a) State the data that the students would have collected from the grazed and ungrazed areas to calculate Simpson's Index of Diversity.

n = Number of individuals of a particular species (Herbivores) // (Plant species) 1

N = Total number of all organisms in the area of investigation. 2

[2]

(b) Describe a random (unbiased) method which the students could have used to collect the data needed to calculate the biodiversity of the plant species in the two areas.

The description of your method should be detailed enough for another person to follow:

① Two different areas are sampled. One area that 3 was grazed by herbivores and another area not grazed by herbivores for many years. Ensure that sampling occurs in these 2 distinct areas. Areas of these descriptions are chosen

② Diversity is calculated using Simpson's Index of Diversity. Formula = $1 - \Sigma\left(\frac{n}{N}\right)^2$

③ The same student should carry out random sampling 4 in each of the 2 areas. The shape and size of quadrat should be the same. A square of $1m^2$ is used. Samples are taken at the same time of day, for example, in the morning.

④ Use quadrat sampling technique. A student, with 5 eyes closed, randomly throws a quadrat in one of the 2 areas. The area in which the quadrat lands is observed. The number of different and

Examiner comments

1 This answer does not earn credit as the candidate refers to only one species, rather than the number in each of the species present.

2 This description of 'N' is not specific enough. The question is related to the effect of grazing on plants; 'organisms' could mean species other than plants.

Mark for (a) = 0/2

3 The candidate gains credit by making a clear reference to sampling in the two types of grassland.

4 The candidate does not give a method for randomising, but does gain credit for using the correct apparatus of a standard size.

5 The method of placing quadrats is not credited. Examiners expected candidates to know how to use a method of randomising within the study area to act as co-ordinates for placing quadrats.

Example candidate response – middle, continued	Examiner comments
6 distinct plant species that is in the quadrat is noted and written down as numerals. Plant species that are not part of the quadrat are omitted. ⑤ step 4 is repeated for a further 4 times at different 7 positions in the area grazed by herbivores and the area not grazed by herbivores. formula is used to calculate Diversity of area. ⑥ few assumptions are made. Number of organisms present in quadrat in the experiments are representative of total population in a particular area. Throwing of quadrat should be completely random. ⑦ Low risk experiment: Ensure that only 1 person throws quadrat and all other students are a 8 considerable distance away to avoid being hit by quadrat. ⑧ 5 times throw of quadrat is repeated 3 times 9 and the average values from the experiment and of Simpson's Biodiversity Index is calculated. ⑨ same person should calculate the number of 10 plant species in each quadrat. This is to avoid biasness. Sampling is done at same time of day to give the same temperature. Ensure that sampling in grazed area is done when there are no herbivores grazing so as to not affect hurt herbivores and for them not to interfere with experiment. ⑩ A control experiment is set up on an area other than a grassland. Ensure for ungrazed area that quadrat is not thrown out of fence. Carry out experiment during the day for easy visualisation of number of organisms. [8]	6 Credit is not awarded here as it is not clear that it is the number in each species that is recorded, not the number of different species. 7 The candidate gains credit for stating that the same number of quadrats would be used in each area. 8 Credit is given for a safety consideration, even though the method of using the quadrat is not appropriate. 9 Credit is not given here as the candidate does not make it clear that the replicates would be in different areas of the grazed and ungrazed grasslands. 10 The rest of the candidate's answer is not relevant and shows some misconceptions about controlling variables in a field investigation. Mark for (b) = 5/8

Example candidate response – middle, continued	Examiner comments

The students also investigated the effect grazing had on the height of one particular species of plant. Their hypothesis was:

The mean height of the plant is greater in the ungrazed grassland than the grazed grassland.

(c) State the independent and the dependent variables in this investigation.

independent variable The type of grassland (Grazed or ungrazed). presence or absence of herbivores.

dependent variable Mean height of a particular species of plant. [1]

11

11 This is an acceptable answer.

Mark for (c) = 1/1

(d) Table 1.1 shows the results of their investigation.

Table 1.1

sample number	height of plant/mm	
	grazed area	ungrazed area
1	586	858
2	549	873
3	526	864
4	589	901
5	545	847
6	538	862
7	573	864
8	549	879
9	604	864
10	611	888
mean	567	870
mode	549	864
median	561	864

12

12 The candidate calculates both values correctly.

(i) Complete Table 1.1 by writing the values of the mode and median for the ungrazed area. [1]

847, 858, 862, 864, 864, 864, 873, 879, 888, 901

Mark for (d) (i) = 1/1

Example candidate response – middle, continued

(ii) Use the information and formula below to calculate the standard error for these results.

Give your answers to 3 significant figures.

$S_M = \frac{s}{\sqrt{n}}$

S_M = standard error

s = standard deviation

n = sample size (number of observations)

grazed area: $s = 29.5$

ungrazed area: $s = 15.7$

$S_{m\,grazed} = \frac{29.5}{\sqrt{10}}$

$S_{m\,ungrazed} = \frac{15.7}{\sqrt{10}}$

standard error, grazed area = 9.33

standard error, ungrazed area = 4.96 [2]

13

Standard error is used to calculate 95% Confidence Intervals (CI).

The values for the grazed area are 548.3 mm to 585.7 mm.

(iii) Use the formula below to calculate the confidence intervals for the **ungrazed** area.

95% CI = mean ± 2 S_M

Show your working.

$= 870 \pm (4.96)2$

$= 870 + 9.92$ or $= 870 - 9.92$

$= 879.92$ $= 860.08$

ungrazed area 860.1 mm to 879.9 mm [2]

14

(iv) State what information is gained by calculating the confidence intervals.

Whether the difference between 2 means is significantly different if difference between means is significantly different, then those differences have occured not by chance. If differences are not significant, they have occured by chance. [2]

15

To ascertain the probabilities or values at which the means are considered to be significantly different.

Examiner comments

13 The candidate correctly calculates both values and uses an appropriate number of significant figures.

Mark for (d) (ii) = 2/2

14 The candidate = calculates both confidence intervals correctly.

Mark for (d) (iii) = 2/2

15 Although the candidate's answer shows an understanding of one role of confidence intervals, no marks are awarded. The answer is too general and not related to the actual values in the question. Examiners expected candidates to realise that the confidence intervals for the mean height of the plants in the two areas do not overlap so the difference in the mean is significant.

Mark for (d (iv) = 0/2

Example candidate response – middle, continued	Examiner comments

(e) The students used the mark-release-recapture method to estimate the population of an invertebrate animal found living on the grassland. They used the formula:

$$\frac{\text{number of animals marked in the first sample} \times \text{total number of animals in the second sample}}{\text{number of marked animals in the second sample}}$$

State two precautions the students should have taken to ensure that the results they obtained were valid.

1. Animals don't lose their marks. Enough time is given for marked and unmarked animals to intermingle. Marks don't hurt animals.

2. Nothing has happened to upset the balance of the number of animals. Examples are predation, migration, mortality. [2]

(f) The population of an invertebrate that feeds on seeds was estimated in both the grazed and ungrazed areas. Predict which area would have the greatest population and give a reason for your choice.

choice Ungrazed areas.

reason Height of plants increases and they can reach a greater reproductive age and undergo pollination. This produces seeds. Seeds consumed by invertebrate. Taller plants have more leaves, a longer stem and more fruits. Fruits produces seeds. [1]

[Total: 21]

16 Marks are awarded for each of the first two statements in the first two lines. The question specifies two, and only the first two are marked, irrespective of the numbered lines. In this case the candidate benefits, as their third sentence and all the statements in number 2 are not creditworthy.

Mark for (e) = 2/2

17 The candidate is awarded a mark for this answer, although the reason is rather long.

Mark for (f) = 1/1

Total marks awarded = 14 out of 21

How the candidate could have improved their answer

(a) A more precise description of 'n' would have been 'the number of plants in each species' and 'N' 'the total number of plants in *all* species'.

(b) The candidate could have described a better method for placing quadrats, for example marking an area with tapes and using a suitable method of randomising, such as a random number generator or app, to identify co-ordinates. The answers also showed some misconceptions about standardising variables and the use of a control in field studies which were also irrelevant to the question asked.

(d) (iv) The candidate should have stated that confidence intervals are used for setting the certainty of data, in this case the calculated mean, and should have used the actual values in the question to describe the statistical significance of the difference in the mean values.

(e) Only two answers were required here; the other two answers should not have been included and the time saved used to improve other answers.

Mark awarded = **(a) 0/2,**
Mark awarded = **(b) 5/8,**
Mark awarded = **(c) 1/1,**
Mark awarded = **(d) (i) 1/1, (ii) 2/2, (iii) 2/2, (iv) 0/2**
Mark awarded = **(e) 2/2,**
Mark awarded = **(f) 1/1**

Total marks awarded = 14 out of 21

Example candidate response – low

1 Grassland is an important breeding habitat for some birds. These birds feed on plant material and invertebrates. Biodiversity of the habitat is maintained by domestic herbivores, such as sheep, cows and goats, grazing on growing plant material.

A group of students investigated the effect of grazing by domestic herbivores on the plant biodiversity of a grassland as measured by Simpson's Index of Diversity. They investigated two areas. One area was grazed by herbivores and the other area was not grazed for many years because it was surrounded by a fence to keep out the herbivores.

(a) State the data that the students would have collected from the grazed and ungrazed areas to calculate Simpson's Index of Diversity.

Total number of species in the grazed and ungrazed area.
Number of organisms of each species in both grazed and ungrazed areas.
This information is required to calculate Simpson's Index of Diversity. [2]

(b) Describe a random (unbiased) method which the students could have used to collect the data needed to calculate the biodiversity of the plant species in the two areas.

The description of your method should be detailed enough for another person to follow.

The person must follow the method of random sampling.
First, take a quadrat and place it anywhere in the area randomnly so that the results are not biased and represent the entire area. Count the different number of species present in the quadrat. Also count how many organisms of that same species is present in that quadrat. These Value must be plotted in a table as follows.

No of species	Number of each species

Readings for Quadrat used in grazed area.

Specie Number	Number of organisms in that specie.

Simpson's Index of Diversity can be used to find the species diversity which will represent the biodiversity of that area.

Examiner comments

1 The candidate is awarded a mark for their second sentence, but does not make it clear in their first sentence that it is the total number in all of the species that is needed.

Mark for (a) = 1/2

2 The candidate is awarded a mark for using an appropriate piece of apparatus. Although they refer to random sampling, there is not enough information about how the randomising will be done to gain any further marks.

3 The candidate describes an acceptable method of collecting data from a quadrat.

Example candidate response – low, continued

Simpson's Index of Diversity = $1 - \left(\frac{\Sigma n}{N}\right)$ 4

where,

N is the total number of organisms in all the species.

n is the number of species in any particular specie.

- Divide number of ~~species~~ organisms for each species by the total number of organisms, N.
- Add all of them up and subtract the value obtained by 1.

The value must be between 0 and 1. More the value closer to 1, more is the species diversity and Hence more is the biodiversity.

Species diversity depends on two things : Age abundance of each specie and Total Number of species. More the number of species and more equally their abundances are, more would be the biodiversity of that area.

- Readings for ungrazed area should be taken in exactly the same way as that for grazed area. Quadrat shall be 5 replaced randomly so that the results are not biased.

All over again, Simpson's Index of diversity can be used to find a value. 6

These values indicate how much the biodiversity of that area is.

These value, calculated using Simpson's Index of Diversity can also be compared to get an idea which area has more Biodiversity.

~~These value, values calculated for species diversity can be used because~~

Test crosses must also be done between the same species of plant as more alleles (more genetic variation) also represents an increases in biodiversity 7 [8]

Examiner comments

4 The candidate includes a great deal of irrelevant information about how to calculate Simpson's Index of Diversity on the first half of this page.

5 The candidate gains credit for stating that sampling is carried out in the same way in both areas.

6 The remaining part of this answer is irrelevant as it describes how to *use* information; the question is about how to *obtain* suitable information.

7 In this part of their answer, the candidate appears to have lost sight of the fact that this is field study.

Mark for (b) = 3/8

Example candidate response – low, continued

The students also investigated the effect grazing had on the height of one particular species of plant. Their hypothesis was:

The mean height of the plant is greater in the ungrazed grassland than the grazed grassland.

(c) State the independent and the dependent variables in this investigation.

independent variable ....grazing.....

dependent variable ....mean height of the plant.... [1]

(d) Table 1.1 shows the results of their investigation.

Table 1.1

sample number	height of plant/mm	
	grazed area	ungrazed area
1	586	858
2	549	873
3	526	864
4	589	901
5	545	847
6	538	862
7	573	864
8	549	879
9	604	864
10	611	888
mean	567	870
mode	549	864
median	561	864

(i) Complete Table 1.1 by writing the values of the mode and median for the ungrazed area. [1]

Examiner comments

8 This answer is not awarded any marks as the independent variable is not precise enough.

Mark for (c) = 0/1

9 The candidate correctly calculates both values.

Mark for (d) (i) = 1/1

Example candidate response – low, continued

(ii) Use the information and formula below to calculate the standard error for these results.

Give your answers to 3 significant figures.

$S_M = \frac{s}{\sqrt{n}}$

S_M = standard error

s = standard deviation

n = sample size (number of observations)

grazed area: $s = 29.5$

ungrazed area: $s = 15.7$

standard error, grazed area = 9.33 10

standard error, ungrazed area = 4.96 [2]

Standard error is used to calculate 95% Confidence Intervals (CI).

The values for the grazed area are 548.3 mm to 585.7 mm.

(iii) Use the formula below to calculate the confidence intervals for the **ungrazed** area.

95% CI = mean ± 2 S_M

Show your working.

95% CI = 567 ± 2 × 4.96

= 567 ± 9.92.

– 567 + 4.96

– 567 – 4.96

11

ungrazed area 571.96 mm to 562.04 mm [2]

(iv) State what information is gained by calculating the confidence intervals.

The information gained by calculating the confidence intervals tell us that we are 95% sure that plants with heights 571.96 – 562.04 were found in ungrazed and their height has not been effected by grazing. 12

[2]

Examiner comments

10 The candidate correctly calculates both values.

Mark for (d) (ii) = 2/2

11 The candidate is not awarded any marks as they have calculated the confidence intervals using the mean value of the grazed area, rather than that of the ungrazed area.

Mark for (d) (iii) = 0/2

12 The candidate's answer suggests they are aware that confidence limits are used to express a degree of certainty, but has not linked them to the mean value, which is the parameter being assessed. The final statement, however, suggests that the candidate does not have a clear understanding of how confidence intervals are used to make comparisons or to assess statistical significance.

Mark for (d) (iv) = 0/2

Example candidate response – low, continued

(e) The students used the mark-release-recapture method to estimate the population of an invertebrate animal found living on the grassland. They used the formula:

$$\frac{\text{number of animals marked in the first sample} \times \text{total number of animals in the second sample}}{\text{number of marked animals in the second sample}}$$

State two precautions the students should have taken to ensure that the results they obtained were valid.

1. The should have used a non-toxic waterproof paint to 13 mark the animals so that each one marked, remains marked until the recapture.

2. They should give enough time to the organisms to randomly spread in their habitat so that the results are not biased and represent the entire area being investigated. [2]

(f) The population of an invertebrate that feeds on seeds was estimated in both the grazed and ungrazed areas. Predict which area would have the greatest population and give a reason for your choice.

choice ungrazed area.

reason More plants have been eaten so more availability of seeds as the seeds have been exposed when the plant was eaten as seeds can not be digested by grazing 14 animals and so are left behind. [1]

[Total: 21]

Examiner comments

13 The candidate gains both marks in the first line as 'non-toxic' and 'waterproof' are separate features of any marker used. The answer given in number 2 is also correct.

Mark for (e) = 2/2

14 This is an invalid reason. Most domestic grazing animals eat any seeds along with the plant being consumed.

Mark for (f) = 0/1

Total marks awarded = 9 out of 21

How the candidate could have improved their answer

(a) The candidate could have stated more clearly that the second piece of information needed is the total number of all the individuals present.

(b) The candidate should have included much more detail about how to randomise the quadrats and how the sampling on the two different areas would be standardised, for example, the area, the number and the size of the quadrats. The information given in the answers about how to calculate Simpson's Index of Diversity was irrelevant and should have been omitted.

(c) When identifying an independent variable, the candidate should have taken into account all of the information given. In this case, as two areas are being compared, the answer should have included both grazed and non-grazed grassland.

(d) The information in the question needed more careful reading to avoid making an error in calculation. A clearer understanding of confidence intervals was also needed.

(f) More thought about the effect of grazing on the ability of plants to reproduce might have helped the candidate to reason that since growing shoots are removed continuously, the plants have less chance to produce seeds.

Mark awarded = **(a) 1/2**
Mark awarded = **(b) 3/8**
Mark awarded = **(c) 0/1,**
Mark awarded = **(d) (i) 1/1, (ii) 2/2, (iii) 0/2, (iv) 0/2**
Mark awarded = **(e) 2/2**
Mark awarded = **(f) 0/1,**

Total marks awarded = 9 out of 21

Common mistakes candidates made in this question

(a) Some candidates gave imprecise or inappropriate descriptions of the data collected for Simpson's Index of Diversity, for example, frequency, percentage cover and density.

(b) Some candidates used the term 'quadrat' to describe a large area that is measured to use for sampling.
Some suggested using transect lines for random sampling in uniform areas and described how to standardise external variables in a field investigation.
Some candidates suggested random placing of quadrats by 'throwing' while taking care not to choose 'interesting areas'.
The instruction asking candidates to describe a method that could be used by another person was not followed. Lists of the different variables were given without any clear method.

(d) (iv) Some candidates confused confidence intervals with standard error, standard deviation and probability. Some candidates gave generalised answers that did not use the data in the question.

Question 2

Example candidate response – high

2 Medical researchers carried out an investigation into the effect of smoking in a country. A group of male volunteers had their peak expiratory flow rate (PEFR) measured as shown in Fig. 2.1.

Fig. 2.1

PEFR measures the maximum speed of airflow through the bronchi during breathing out in dm^3 per minute ($dm^3 min^{-1}$). Peak flow readings are lower when the airways are constricted.

The volunteers were grouped according to the number of packets of cigarettes that they smoked per year. Each packet contains 20 cigarettes.

Table 2.1 shows the results of the investigation.

Table 2.1

group	1	2	3	4	5
number of packets of cigarettes smoked per year	0	1–50	51–100	101–150	151–230
mean number of packets smoked per group ± *s*	0	30.61 ± 10.47	73.80 ± 16.52	127.27 ± 9.66	189.22 ± 27.51
mean age of volunteers ± *s* /years	26.42 ± 5.61	22.82 ± 3.28	26.66 ± 3.59	28.90 ± 4.20	36.22 ± 3.21
mean PEFR ± *s* /$dm^3 min^{-1}$	513.43 ± 87.58	494.70 ± 79.22	443.33 ± 45.14	350.90 ± 32.38	300.00 ± 46.90
number of volunteers tested	64	14	15	12	8

s = standard deviation

Examiner comments

Example candidate response – high, continued

(a) State **three** variables which **should have been** standardised in this investigation.

~~The age of the males~~

The ethenicity of the males.

~~How long they have been smoking for~~

Their condition whilst taking the test, for example everyone should be rested/sitting down.

The number of hours they do not smoke before the test, for example 24 hours. 1

[3]

(b) The medical researchers made two conclusions based on the data shown in Table 2.1.

1. An increase in the number of packets smoked decreases the PEFR measurement.

2. The number of packets smoked increases with age.

State how the results from Table 2.1 support these conclusions and how they do not support these conclusions.

support

For conclusion one, it does support because Group 1's mean PEFR is 513.43, and Group 3's is 443.33 and group 5's is the lowest with 300.00, as the mean number of packs smoked increase 2

For conclusion 2 it does support because from group 2 to 5 the age increases from 22.82 to 36.22, as the packs smoked also go up 3

do not support

for conclusion one, it doesn't support, because, the standard deviation for Group 1 and 2 overlap significantly. As well as group 2 and 3 (for mean PEFR) 4

For conclusion 2, Group 1's mean age (0 cigarette) is higher than Group 2's mean age 1-50 cigarette [3]

5 26.42 > 22.82

Examiner comments

1 The candidate is awarded marks for each of their answers. The example of a time in the last line has been ignored.

Mark for (a) = 3/3

2 The candidate gains marking point 2 for quoting suitable PEFR figures linked to an increase in the mean number of packets of cigarettes smoked.

3 The candidate gains marking point 5 for selecting appropriate smoking groups and quoting the relevant figures as part of a description of a trend.

4 As the candidate includes the reference to PEFR in brackets, they gain marking point 7, although their answer does not make a clear reference to the number of packets increasing from group 2 to group 3.

5 The candidate is awarded marking point 9 for a correct comparison of the mean ages of groups 1 and 2 in relation to the number of cigarettes smoked.

Mark for (b) = 3/3

Example candidate response – high, continued	Examiner comments
(c) (i) State a null hypothesis for a statistical test to find out whether the data in Table 2.1 supports the conclusion that: An increase in the number of packets smoked decreases the PEFR measurement. There is no significant relationship between the increase ~~in the~~ number of packets smoked and decrease in PEFR measurement. 6 [1] (ii) State **two** ways in which the data for **group 5** is less trustworthy compared with the data for the other groups. Number of volunteers tested is less. It has the largest standard deviation in the mean number of packs smoked. ±2.75 7 [2] [Total: 9]	6 This answer has all the correct elements for a null hypothesis about a correlation between two factors. Mark for (c) (i) = 1/1 7 The candidate is awarded marks for both answers. Mark for (c) (ii) = 2/2 **Total marks awarded = 9 out of 9**

How the candidate could have improved their answer

Although the candidate gained maximum marks, their answers contained crossings out and included important information in brackets. Some information was omitted. A little more time spent thinking might have resulted in fuller and more carefully structured answers.

Mark awarded = **(a) 3/3,**
Mark awarded = **(b) 3/3,**
Mark awarded = **(c) (i) 1/1, (ii) 2/2**

Total marks awarded = 9 out of 9

Example candidate response – middle	Examiner comments

2 Medical researchers carried out an investigation into the effect of smoking in a country. A group of male volunteers had their peak expiratory flow rate (PEFR) measured as shown in Fig. 2.1.

Fig. 2.1

PEFR measures the maximum speed of airflow through the bronchi during breathing out in dm^3 per minute ($dm^3 min^{-1}$). Peak flow readings are lower when the airways are constricted.

The volunteers were grouped according to the number of packets of cigarettes that they smoked per year. Each packet contains 20 cigarettes.

Table 2.1 shows the results of the investigation.

Table 2.1

group	1	2	3	4	5
number of packets of cigarettes smoked per year	0	1–50	51–100	101–150	151–230
mean number of packets smoked per group ± *s*	0	30.61 ± 10.47	73.80 ± 16.52	127.27 ± 9.66	189.22 ± 27.51
mean age of volunteers ± *s* /years	26.42 ± 5.61	22.82 ± 3.28	26.66 ± 3.59	28.90 ± 4.20	36.22 ± 3.21
mean PEFR ± *s* /$dm^3 min^{-1}$	513.43 ± 87.58	494.70 ± 79.22	443.33 ± 45.14	350.90 ± 32.38	300.00 ± 46.90
number of volunteers tested	64	14	15	12	8

s = standard deviation

Example candidate response – middle, continued

(a) State **three** variables which **should have been** standardised in this investigation.

- the mean age of the volunteers, with same standard deviation. 1
- the number of volunteers tested in each group. 2
- the interval within the number of packets of cigarettes smoked per year. 3

[3]

(b) The medical researchers made two conclusions based on the data shown in Table 2.1.

1. An increase in the number of packets smoked decreases the PEFR measurement.
2. The number of packets smoked increases with age.

State how the results from Table 2.1 support these conclusions and how they do not support these conclusions.

support

for statement 1, the mean PEFR decreases as the number of packets smoked increases, from 513 to 300. 4

for statement 2, the mean number of packet smoked increase with mean age increases, from 26.42 to 36.22. 5

do not support → The overlapping of standard deviation is too large.

- for statment 1, for example, group 4, and 5, group 4 PEFR is in range 317.62 – 382.28 while in group 5 PETR range is 253.1 – 346.90, So Some volunteer in who smokes more packets have higher PEFR than th who smoke fewer packets. 6

[3]

- For statment 2, comparing group 3 and 4, people with age about (26.06+3.59) 30 smoke fewer packets than those who age is about ~~34~~ 25 in group 4. 7

Examiner comments

1 This statement does not gain credit as a mean age cannot be standardised.

2 This answer is credited.

3 This answer is not credited as the investigation showed some standardisation of this value.

Mark for (a) = 1/3

4 This answer scores marking point 2 as the candidate refers to the mean PEFR decreasing and quotes the correct range of increase in number of cigarette packets.

5 This answer scores marking point 3 and the candidate quotes the trend using mean values for age and number of cigarette packets.

6 The first four lines of this answer earn marking point 7 for noting that although the number of cigarettes smoked increases from group 4 to group 5, the standard deviations of the PEFR overlap, so there is no clear decrease.

7 If the maximum mark had not already been achieved by this point, a mark for statement 2 might have been given as 'benefit of the doubt', although the description of the overlap in age is not clearly stated.

Mark for (b) = 3/3

Example candidate response – middle, continued	Examiner comments
(c) (i) State a ~~null hypothesis for~~ a statistical test to find out whether the data in Table 2.1 supports the conclusion that: An increase in the number of packets smoked decreases the PEFR measurement. • there is no significant ~~differences~~ Correlation between increases in the number of packets smoked and decrease in PEFR measurement. 8 [1] **(ii)** State **two** ways in which the data for **group 5** is less trustworthy compared with the data for the other groups. • the interval for number of packets of cigarettes smoked per year is not the same as the other group. 9 ~~• the standard deviation of mean number~~ • the number of volunteers in Group 5 is the smallest. 10 [2] [Total: 9]	8 This answer has all the elements for a null hypothesis for a correlation. Mark for (c) (i) = 1/1 9 The first answer is not precise enough. The answer is either that the interval was greater than the others or that the interval of group 5 was 80 and that of the other groups was 50. 10 A correct answer. Mark for (c) (ii) = 1/2 **Total marks awarded = 6 out of 9**

How the candidate could have improved their answer

(a) The candidate needed to read the summary of the investigation more carefully to realise that there were two variables being changed, so that the actual age should be the same for all groups. Other aspects of the investigation, such as how the PEFR was carried out, could have been considered.

(b) Although maximum marks were awarded, some of the answers in 'do not support' would have been improved by reducing the number of words used.

(c) (ii) Stating that there is a difference, without identifying what that difference is, was not awarded marks.

Mark awarded = **(a) 1/3**
Mark awarded = **(b) 3/3**
Mark awarded = **(c) (i) 1/1, (ii) 1/2,**

Total marks awarded = 6 out of 9

Example candidate response – low	Examiner comments

2 Medical researchers carried out an investigation into the effect of smoking in a country. A group of male volunteers had their peak expiratory flow rate (PEFR) measured as shown in Fig. 2.1.

Fig. 2.1

PEFR measures the maximum speed of airflow through the bronchi during breathing out in dm^3 per minute ($dm^3 min^{-1}$). Peak flow readings are lower when the airways are constricted.

The volunteers were grouped according to the number of packets of cigarettes that they smoked per year. Each packet contains 20 cigarettes.

Table 2.1 shows the results of the investigation.

Table 2.1

group	1	2	3	4	5
number of packets of cigarettes smoked per year	0	1–50	51–100	101–150	151–230
mean number of packets smoked per group ± *s*	0	30.61 ± 10.47	73.80 ± 16.52	127.27 ± 9.66	189.22 ± 27.51
mean age of volunteers ± *s* /years	26.42 ± 5.61	22.82 ± 3.28	26.66 ± 3.59	28.90 ± 4.20	36.22 ± 3.21
mean PEFR ± *s* /$dm^3 min^{-1}$	513.43 ± 87.58	494.70 ± 79.22	443.33 ± 45.14	350.90 ± 32.38	300.00 ± 46.90
number of volunteers tested	64	14	15	12	8

s = standard deviation

Example candidate response – low, continued

(a) State **three** variables which **should have been** standardised in this investigation.

- The number of volunteers tested should be same in all groups. 1
- The number of packets of ~~cigs~~ cigarettes smoked per year in all groups should be the same 2
- Use uncertainty instead of standard deviation. 3

[3]

(b) The medical researchers made two conclusions based on the data shown in Table 2.1.

1. An increase in the number of packets smoked decreases the PEFR measurement.
2. The number of packets smoked increases with age.

State how the results from Table 2.1 support these conclusions and how they do not support these conclusions.

support

- ~~At~~ g from group 3 to 5, ~~does~~ ~~stop~~ as the number of packets smoked increases, the mean age of volunteers also increases. 4
- from group 1 to 5, mean PEFR decrease from 513.43 to 300.00 as number of smoked ~~int~~ cigarettes increase. 5

do not support

- from group 1 to 2, mean age of volunteers decreases as number of packets smoked increases. 6
-

[3]

Examiner comments

1 The candidate is awarded a mark for this answer

2 This answer is incorrect as the investigation does include a way of standardising this variable.

3 This statement is not relevant and suggests that this candidate does not understand that standard deviation is one way of showing uncertainty.

Mark for (a) = 1/3

4 This statement does not earn marking point 4 or 5 because the candidate omits 'mean' from the number of packets and does not quote the figures to show the increase in the numbers of packets.

5 The candidate gains marking point 2 as they quote relevant figures for the decrease in PEFR linked to an increase in the number of packets of cigarettes.

6 This answer scores marking point 9 as the candidate makes a link between a decrease in mean age and increase in number of cigarettes.

Mark for (b) = 2/3

Example candidate response – low, continued

(c) (i) State a null hypothesis for a statistical test to find out whether the data in Table 2.1 supports the conclusion that:

An increase in the number of packets smoked decreases the PEFR measurement.

Number of packets smoked and PEFR measurement is related and inverse to one another. 7

..... [1]

(ii) State **two** ways in which the data for **group 5** is less trustworthy compared with the data for the other groups.

- Mean age of volunteers is above 30 where as the other groups are below 30. 8

- Number of volunteers tested is the least amongst all other groups. 9

..... [2]

[Total: 9]

Examiner comments

7 This answer is a weak description of a negative correlation which is true but which is not a null hypothesis. A null hypothesis should state that there is no significant correlation between the changes in the two parameters being assessed, in this case, the decrease in PEFR and the increase in smoking.

Mark for (c) (i) = 0/1

8 This answer is not relevant. The difference in ages does not affect validity, which is more concerned with the reliability of the data for this group.

9 The candidate gains credit for this answer.

Mark for (c) (ii) = 1/3

Total marks awarded = 4 out of 9

How the candidate could have improved their answer

(a) The candidate needed to be clearer about the methods used in the investigation so that they could consider for which variables there had been some attempt to standardise. A more useful way of approaching this would have been to think about the way in which PEFR measurements are made and choose variables that would influence this. For example, chest size and lung capacity is influenced by body mass, physical fitness and lung diseases.

(c) (ii) It would have been better if the candidate had thought more carefully about the factors, other than group size, that affect reliability of data, in particular the size of the standard deviation. As a general principle, the greater the range of any measures of uncertainty, the less reliable the parameter being assessed. The candidate's second answer to **(a)** would have been appropriate here, as the method of standardising this variable has changed for group 5.

Mark awarded = **(a) 1/3,**
Mark awarded = **(b) 2/3**
Mark awarded = **(c) (i) 0/1, (ii) 1/3,**

Total marks awarded = 4 out of 9

Common mistakes candidates made in this question

(a) Many candidates did not consider how the PEFR test was carried out and so missed obvious issues, such as lung diseases and the time interval between smoking and taking the test. Some candidates also missed that there were two variables being changed, so the focus of the investigation should have been only on the increase in smoking.

(b) Some candidates restated the hypotheses in the question without referring to means of selecting appropriate data from two different groups.

(c) (i) Some candidates gave a null hypothesis suited to a *t*-test rather than a correlation.

(c) (ii) Some candidates stated that the age difference was significant. They also stated that the standard deviation was too large without linking this to the number of packets of cigarettes.

Cambridge International Examinations
1 Hills Road, Cambridge, CB1 2EU, United Kingdom
t: +44 1223 553554 f: +44 1223 553558
e: **info@cie.org.uk** **www.cie.org.uk**

Version 1.0

Cambridge Assessment