Qualification Accredited

OCR
Oxford Cambridge and RSA

Examiners' report

AS LEVEL

FURTHER MATHEMATICS B (MEI)

OCR Level 3 Advanced Subsidiary GCE in Further Mathematics B (MEI)

H635

For first teaching in 2017

Y412/01 Summer 2025 series

Version 1

ocr.org.uk/maths

Contents

Introduction 3
Paper Y412/01 series overview 4
Question 1 (a) 5
Question 1 (b) 5
Question 2 (a) 5
Question 2 (b) 5
Question 2 (c) 6
Question 2 (d) 6
Question 3 (a) 7
Question 3 (b) 7
Question 3 (c) 7
Question 3 (d) 9
Question 4 (a) 9
Question 4 (b) 10
Question 5 (a) 10
Question 5 (b) 11
Question 5 (c) 11
Question 6 (a) (i) 12
Question 6 (a) (ii) 12
Question 6 (a) (iii) 13
Question 6 (b) 13
Question 6 (c) 13
Question 6 (d) 13
Question 7 (a) 14
Question 7 (b) 14
Question 7 (c) 15
Question 7 (d) 15
Question 7 (e) 16
Question 7 (f) 16

Introduction

Our examiners' reports are produced to offer constructive feedback on candidates' performance in the examinations. They provide useful guidance for future candidates.

The reports will include a general commentary on candidates' performance, identify technical aspects examined in the questions and highlight good performance and where performance could be improved. A selection of candidate responses is also provided. The reports will also explain aspects which caused difficulty and why the difficulties arose, whether through a lack of knowledge, poor examination technique, or any other identifiable and explainable reason.

Where overall performance on a question/question part was considered good, with no particular areas to highlight, these questions have not been included in the report.

A full copy of the question paper and the mark scheme can be downloaded from Teach Cambridge.

Paper Y412/01 series overview

This is a calculator paper in which situations are modelled by discrete random variables; the suitability of models is tested using chi-squared tests. Bivariate data are investigated, with tests for correlation and association, as well as modelling using regression.

To do well on this paper, candidates must demonstrate their understanding of the contents from the specification. Candidates should be able to choose and apply appropriate models and be able to provide structured responses to questions involving hypothesis tests. Candidates should be aware that questions asking for 'detailed reasoning' to be given or containing command words such as 'show that' and 'determine' indicate that final answers must be supported by relevant working.

Once again, the overall performance was very good this year. It was pleasing to again see suitable structure in hypothesis tests with appropriate, non-assertive conclusions provided. It was also pleasing that candidates could explain the key features of, and use, appropriate probability distributions.

Candidates could have done better in answering the questions where a deeper understanding of the situation was required to calculate a requested probability. Some candidates fell short in their responses where explanation was required.

Candidates who did well on this paper generally:	Candidates who did less well on this paper generally:
• chose and quoted an appropriate probability distribution and stated its parameters • provided working to justify answers • showed good understanding of different probability distributions and their means • provided concise, focused explanations using context when needed • provided correct answers to the more difficult questions involving probability • showed good understanding of hypothesis testing • answered questions fully.	• chose inappropriate probability distributions • provided insufficient working to support correct answers • were unable to distinguish between different probability distributions and use their means • gave explanations of a general nature rather than focusing on statistical features • had limited success with probability calculations where a range of values was involved • showed some understanding of hypothesis testing • did not answer questions fully.

Question 1 (a)

1 A student is investigating what level of exercise is given to dogs in the UK by their owners.

(a) When trying to collect suitable information, give **two** reasons why it might be advantageous to use a sample of dog owners, rather than a census of dog owners. **[2]**

Some candidates appeared to be mixing up 'advantageous' with 'disadvantageous'. Those candidates giving disadvantages of using a census needed to make it clear that their comments referred to a census.

Misconception

Several candidates believed that 'a census' as stated in the question referred to the National Census carried out once every ten years.

Question 1 (b)

(b) Explain why it is preferable that the size of the sample is large. **[1]**

There were many mentions of a larger sample being 'more accurate', 'more reliable' or 'less biased' rather than the required 'more representative of the population'.

Question 2 (a)

2 The probability distribution of a discrete random variable X is given in the table.

x	0	1	2	3
$\mathrm{P}(X = x)$	$\frac{1}{6}$	$\frac{1}{3}$	a	$\frac{1}{10}$

(a) Find the value of a. **[1]**

Question 2 (b)

The probability function of another discrete random variable, Y, is defined below.

$$\mathrm{P}(Y = y) = \frac{y}{595} \text{ for } y = 1, 2, ..., n,$$

where n is a positive integer.

(b) Determine the value of n. **[3]**

Many candidates provided enough information to fully 'determine' the required value of n.

Question 2 (c)

(c) Find $\mathrm{P}(Y < n)$. **[1]**

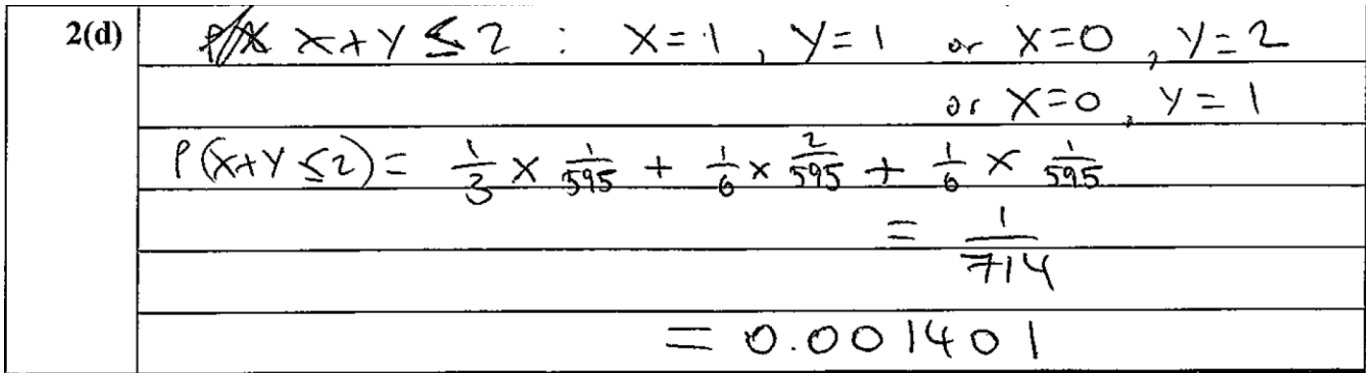
Most candidates used the given probability function and got this mark. A common error was to use a discrete uniform distribution over the values 1, 2, …, n.

Question 2 (d)

(d) Given that X and Y are independent, determine $\mathrm{P}(X+Y \leqslant 2)$. **[2]**

This question was generally not answered well. Most candidates did not work out that there were just three possible combinations of X and Y such that $X+Y \leq 2$.

Exemplar 1



In the successful attempts, as seen in this exemplar, candidates wrote down the three appropriate pairs of values then showed the calculation required to get the answer of $\frac{1}{714}$.

Question 3 (a)

3 The discrete random variable Q has the distribution Geo(0.3).

(a) Find $P(Q = 2)$. **[1]**

Question 3 (b)

(b) Determine $P(Q \leqslant 5)$. **[2]**

Many candidates simply wrote their answer without showing the method used to determine it. It was pleasing to see many candidates using $1 - 0.7^5$ as a more efficient method for determining the answer than using $0.3 + 0.7 \times 0.3 + \ldots + 0.7^4 \times 0.3$.

Assessment for learning

In questions involving the command word 'determine', the method used must be made clear if all the available marks are to be given.

Question 3 (c)

The discrete random variable R has a uniform distribution on $\{4, 5, ..., 8\}$.

The discrete random variable S has the distribution $B(10, 0.1)$.

Q, R and S are mutually independent of each other.

(c) Find $P(Q + R + S = 5)$ **[3]**

Very few candidates showed understanding of what is required here. Many candidates thought $Q + R + S = 5$ meant that $Q = 5$, $R = 5$ and $S = 5$. Many candidates used a sum of probabilities rather than a product. Successful candidates realised that there was just one combination to consider.

Exemplar 2

3(c)

Q can take values ≥ 1

$Q + R + S = 5$

1, 4, 0

This is the only way of obtaining 5 since $R \geq 4$ and $Q \geq 1$.

$P(Q+R+S=5) = 0.3 \times 0.2 \times 0.3486784$

$= 0.0209$

This exemplar shows a clear, complete response. Although this was not a 'determine' question, candidates who presented a complete method tended to be the successful ones.

Question 3 (d)

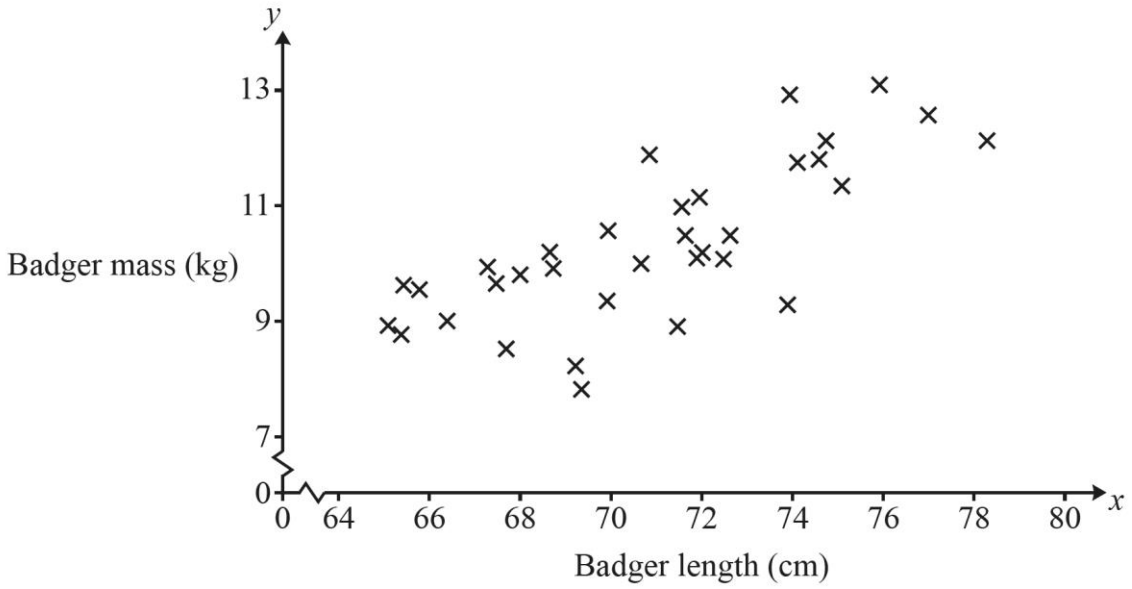
The discrete random variable T is defined by $T = R - Q + 2S$

(d) Determine the value of $\mathrm{E}(T)$. **[4]**

Most candidates answered this successfully. A common error was to calculate and use an incorrect value for $\mathrm{E}(R)$. For many candidates, that was the only error made in this part.

Question 4 (a)

4 A scientist measures the length, x cm, and mass, y kg, of a random sample of 34 adult female badgers. A scatter diagram for the data is shown in the diagram.



(a) Give **two** reasons why it would be appropriate for a hypothesis test based on the product-moment correlation coefficient to be carried out in this situation. **[2]**

As in previous years, many candidates who remembered to mention bivariate normality did not make it clear that this applied to the underlying population.

Question 4 (b)

The scientist analyses the collected data and produces the following summary statistics.

$n = 34 \quad \Sigma x = 2409.4 \quad \Sigma y = 350.87 \quad \Sigma x^2 = 171140.28 \quad \Sigma y^2 = 3681.851 \quad \Sigma xy = 24981.486$

(b) Find and use the equation of a suitable regression line to determine an estimate of the length of an adult female badger of mass 10 kg. **[4]**

Many candidates got 3 marks here. The most common mistake was to give a final answer to an inappropriate level of accuracy, sometimes with as many as ten significant figures given.

Question 5 (a)

5 Two judges, Judge A and Judge B, independently rank 8 entries in a jam-making competition, awarding each jam a rank from 1 (best) to 8 (worst). There are no tied ranks. The ranks are shown in the table below but unfortunately two of the rankings awarded by Judge B have been lost.

	Jam 1	Jam 2	Jam 3	Jam 4	Jam 5	Jam 6	Jam 7	Jam 8
Judge A	4	8	7	1	5	6	3	2
Judge B	4		5	2	6		1	3

(a) Explain which jam is likely to be the best. **[1]**

There was some confusion from candidates here despite the given information that 1 is the best score. Jam 2 was often, incorrectly, stated as the best. Candidates were required to give some explanation why Jam 4 is likely to be the best; an easy way to do this was to state that the sum of Jam 2's ranks was the smallest.

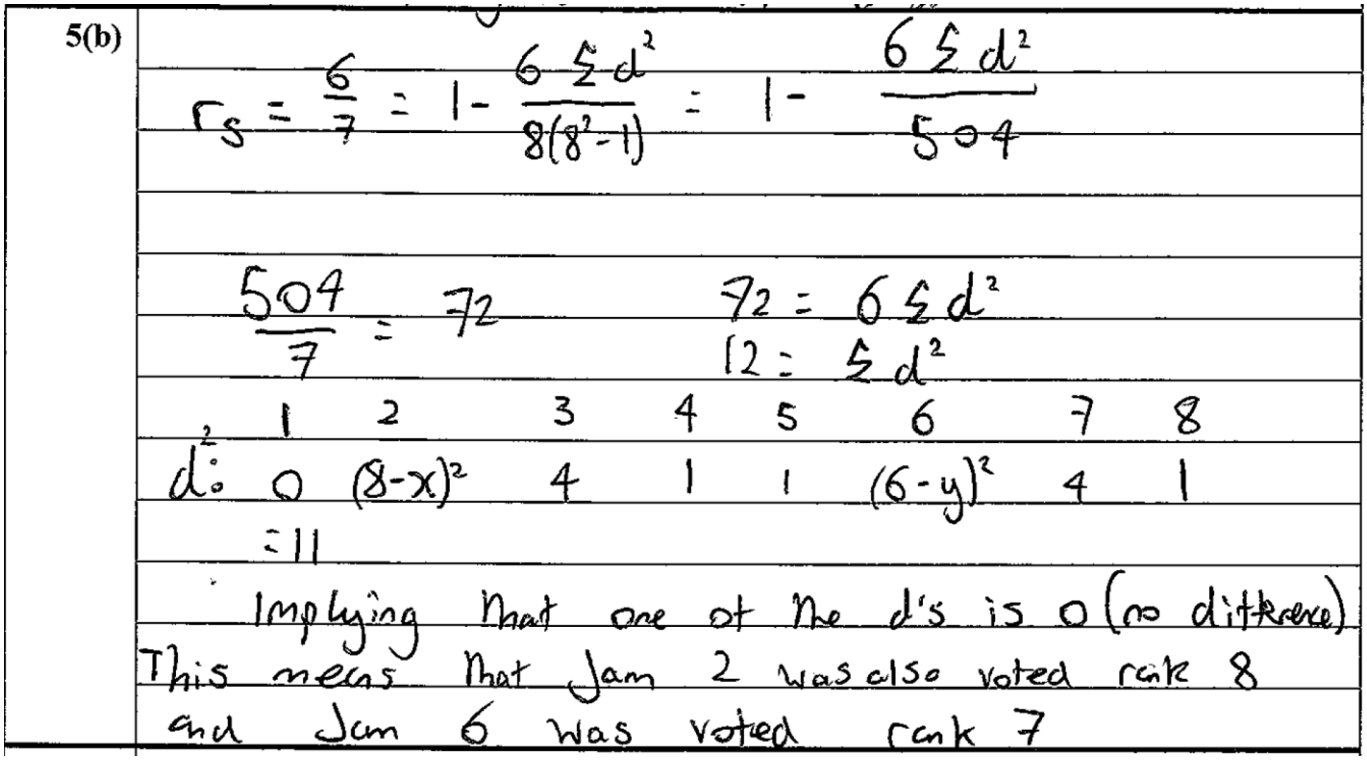
Question 5 (b)

Before Judge B's ranks were lost, Spearman's rank correlation coefficient was calculated for all the data and found to be $\frac{6}{7}$.

(b) Determine the missing ranks that Judge B gave, specifying which rank was given to which jam. **[3]**

This was a 'determine' question which requires evidence of the method used. Most candidates did well here.

Exemplar 3



This exemplar shows a complete and easy-to-follow method.

Question 5 (c)

The two judges want to assess whether there is positive association between their rankings of jam.

(c) Carry out a hypothesis test to make the assessment using a significance level of 1%. **[5]**

It was pleasing to see many fully correct responses here. Candidates showed good understanding of the structure of this hypothesis test. One error that was seen often was stating 'no positive association' in the null hypothesis. Otherwise, most candidates found this to be straightforward.

Question 6 (a) (i)

6 The number of misprints in every 100 000 words of a book is denoted by the random variable X. The probability of any given word being misprinted is 0.000 11. You may assume that misprints occur randomly and independently of each other.

(a) (i) Explain why a certain binomial distribution could be used to model the distribution of X. **[2]**

Compared to the last time such a question appeared, candidates were much more successful in showing an understanding of why, in this case, a binomial distribution could be used.

Assessment for learning

When giving explanations, candidates need to give clear and unambiguous descriptions, ideally using the context of the question. For example, simply stating 'the probability is constant' is not enough. Acceptable versions of this statement would be 'the probability of success is constant' or, even better, 'the probability of a misprint is constant'. Similarly, stating that 'misprints occur independently' is preferable to stating 'trials occur independently', whereas 'events occur independently' is not clear enough.

Question 6 (a) (ii)

(ii) Hence explain why X could also be modelled by a certain Poisson distribution. **[1]**

Many candidates gave reasons for X following a Poisson distribution based on the information given at the start of the question rather than explaining why it could be used as an approximation to the 'certain binomial distribution' referred to in part (a) (i). Those giving descriptions often fell short in detail, stating 'the probability' rather than 'the probability of success' when describing the parameter p.

Question 6 (a) (iii)

(iii) State the mean of this Poisson distribution. **[1]**

Question 6 (b)

For the remainder of this question, you should use a **Poisson** distribution to model the number of misprints in a book.

(b) For a book containing exactly 100 000 words, find the probability that it contains exactly 9 misprints. **[1]**

Question 6 (c)

A proof-reader is paid £20 for each misprint identified in this book. You may assume that the proof-reader identifies every misprint in the book. The payment in pounds made to the proof-reader is denoted by the random variable P.

(c) By considering the mean and variance of P, show that P does **not** follow a Poisson distribution. **[2]**

Many candidates calculated $\mathrm{E}(P)$ and $\mathrm{Var}(P)$ correctly, then went on to explain that their values were not close and that P could not follow a Poisson distribution. Some candidates used calculations suitable for a binomial distribution despite the instruction to use a Poisson distribution for all parts of the question starting from part (b).

Question 6 (d)

The proof-reader is paid for proof-reading another book containing exactly 50 000 words, again identifying every misprint in the book.

(d) Determine the probability that the proof-reader is paid a total amount which is more than £300 and less than £500 for proof-reading **both** books. **[3]**

Many candidates realised that the given information required them to find $\mathrm{P}(15 < Y < 25)$ but did not succeed in finding its value. Mistakes included not acknowledging use of a Poisson distribution with a mean of 16.5, using a Poisson distribution with a mean other than 16.5 and not being able to calculate the probability for this range of values. Typical errors of those close to using a correct method included finding $\mathrm{P}(Y \leq 24) - \mathrm{P}(Y \leq 14)$ or finding $\mathrm{P}(Y \leq 25) - \mathrm{P}(Y \leq 15)$.

Question 7 (a)

7 The manager of a Health Service Trust receives data from two hospitals, Hospital A and Hospital B, about the lengths of stays in hospital for patients with a particular condition. The data are in the form of a list of the number of days that each patient with that condition has spent at each hospital in one particular year.

The manager decides to categorise the length of each stay according to **Table 7.1** before putting the data into a contingency table.

Table 7.1

Length of Stay (days)	Category
1–3	Short Stay
4–6	Medium Stay
⩾ 7	Long Stay

(a) State **one** advantage and **one** disadvantage of categorising the data in this way. **[2]**

Many candidates found it difficult to provide suitable responses here, with the responses for the 'disadvantage' being less successful. Statements about 'accuracy', 'reliability' or what could be done with the data did not gain credit. More successful responses focused on advantages/disadvantages of 'categorising the data in this way' and did not speculate about the nature of the data or what happened in the hospitals.

Question 7 (b)

The contingency table created by the manager is shown in **Table 7.2**.

Table 7.2

	Short Stay	Medium Stay	Long Stay	Total
Hospital A	9	20	15	44
Hospital B	29	25	40	94
Total	38	45	55	138

The manager wishes to use a chi-squared test to assess whether there is any association between the length of stay of the patient at the hospital and which hospital is treating the patient. You may assume that the lengths of stay are independent of each other and the data is a random sample from the underlying population of patients with this condition.

(b) Write down suitable null and alternative hypotheses for the test. **[2]**

Most candidates provided correct hypotheses. Some candidates got the hypotheses the wrong way round.

Question 7 (c)

Table 7.3 shows some of the expected frequencies. The table is incomplete.

Table 7.3

Expected frequency	Short Stay	Medium Stay	Long Stay
Hospital A	12.116		
Hospital B			37.464

(c) Complete the copy of **Table 7.3** in the **Printed Answer Booklet**. Your entries should be given to **3** decimal places. **[2]**

Most candidates answered this correctly.

Question 7 (d)

Table 7.4 shows some of the chi-squared contributions. Entries are given to **4** decimal places. The table is incomplete.

Table 7.4

Chi-squared contribution	Short Stay	Medium Stay	Long Stay
Hospital A			
Hospital B	0.3751		0.1717

(d) Complete the copy of **Table 7.4** in the **Printed Answer Booklet** and carry out the test at the 10% significance level. You do not need to restate your hypotheses. **[6]**

Fully correct answers were common for this question. Some candidates gave an incorrect/inaccurate test statistic or used the wrong number of degrees of freedom. It was pleasing to see correctly worded conclusions in most cases. In some cases, candidates stated their contextual summary before stating 'Reject H_0'.

Question 7 (e)

(e) Use the chi-squared contributions found in part **(d)** to explain which length stay, at which hospital, gives most evidence of any association. **[1]**

Most candidates answered this question correctly. Many candidates opted, incorrectly, for 'Long Stay at Hospital B', perhaps thinking that the lowest contribution to the test statistic indicated the strongest evidence of association.

Question 7 (f)

The manager receives a new instruction that all hypothesis tests carried out by the trust should actually have a significance level of 5%.

(f) Explain whether following this new instruction would have affected the result of the test. **[2]**

Many correct responses were seen here. Some candidates did not finish off their explanations, so did not get the final mark. After providing the necessary detail regarding the different significance level, they did not comment that the result would have been affected by this change.

Supporting you

Teach Cambridge

Make sure you visit our secure website **Teach Cambridge** to find the full range of resources and support for the subjects you teach. This includes secure materials such as set assignments and exemplars, online and on-demand training.

Don't have access? If your school or college teaches any OCR qualifications, please contact your exams officer. You can forward them this link to help get you started.

Reviews of marking

If any of your students' results are not as expected, you may wish to consider one of our post-results services. For full information about the options available visit the OCR website.

Access to Scripts

We've made it easier for Exams Officers to download copies of your candidates' completed papers or 'scripts'. Your centre can use these scripts to decide whether to request a review of marking and to support teaching and learning.

Our free, on-demand service, Access to Scripts is available via our single sign-on service, My Cambridge. Step-by-step instructions are on our website.

Keep up-to-date

We send a monthly bulletin to tell you about important updates. You can also sign up for your subject specific updates. If you haven't already, sign up here.

OCR Professional Development

Attend one of our popular professional development courses to hear directly from a senior assessor or drop in to a Q&A session. Most of our courses are delivered live via an online platform, so you can attend from any location.

Please find details for all our courses for your subject on **Teach Cambridge**. You'll also find links to our online courses on NEA marking and support.

Signed up for ExamBuilder?

ExamBuilder is the exam paper builder platform for a range of our qualifications. Find out more.

Free for all OCR centres with an Interchange account, it allows unlimited users per centre. We require an Interchange username to verify your centre's users for ExamBuilder.

If you do not have an Interchange account, please contact your centre administrator (usually the Exams Officer) to request a username.

Once you have an Interchange username, use the form on this page to sign up for ExamBuilder, or ask an existing user at your centre to create an ExamBuilder account for you.

Active Results

Review students' exam performance with our free online results analysis tool. It is available for all GCSEs, AS and A Levels and Cambridge Nationals (examined units only).

Find out more.

Tell us what you think

Your feedback plays an important role in how we develop, market, support and resource qualifications now and into the future. We want you and your students to enjoy and get the best out of our qualifications and resources, but to do that we need your honest opinions to tell us whether we're on the right track or not.

You can email your thoughts to resources.feedback@ocr.org.uk or visit our feedback page to learn more about how you can help us improve our resources.

Designing and testing in collaboration with you and your students

Helping young people develop an ethical view of the world

Equality, diversity, inclusion and belonging (EDIB) are part of everything we do

---

Contact the team at:

01223 553998

ocr.org.uk

To stay up to date with all the relevant news about our qualifications, register for email updates at ocr.org.uk/updates

Visit our Online Support Centre at support.ocr.org.uk

CAMBRIDGE
UNIVERSITY PRESS & ASSESSMENT