Qualification Accredited

OCR
Oxford Cambridge and RSA

A LEVEL

Examiners' report

FURTHER MATHEMATICS B (MEI)

H645

For first teaching in 2017

Y432/01 Summer 2022 series

Contents

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 answers are 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 OCR.

Advance Information for Summer 2022 assessments

To support student revision, advance information was published about the focus of exams for Summer 2022 assessments. Advance information was available for most GCSE, AS and A Level subjects, Core Maths, FSMQ, and Cambridge Nationals Information Technologies. You can find more information on our website.

Paper Y432/01 series overview

This is the Statistics Minor option component of the MEI Further Mathematics course.

This component was accessible to most candidates who typically attempted to answer all questions.

Questions involving calculations using probability distributions (1 (a), 1 (b), 3 (b), 3 (e), 4 (a), 4 (b), 6 (a), 6 (b)) were usually done well. This was also true for calculations of statistics from samples (2 (b), 2 (c), 3 (a), 5 (b)). Together these accounted for just over half of the marks in the component.

For the two questions involving hypothesis tests (3 (d), 5 (c)) candidates demonstrated a good understanding of the required structure. Hypotheses were stated, critical values were found, and comparisons made with test statistics. Conclusions were then usually suitably non-assertive. More successful responses ensured that the conclusion referred to the context of the question. To improve further candidates should ensure that the null hypothesis is chosen correctly and includes just the right amount of detail. They should also take care to know how to identify the correct critical value.

Typically, candidates did not perform so well on 'explain' and 'comment' questions (1 (c), 2 (d), 2 (e), 3 (c), 5 (a), 5 (d), 5 (e)) and on this examination these accounted for one quarter of the marks. Centres are encouraged to ensure that sufficient attention is given to these questions when preparing candidates. Words such as 'accurate', 'reliable' and 'valid' are not interchangeable and candidates need to be aware of which is appropriate under different scenarios.

Candidates should be encouraged to write neatly. A number of responses to the 'explain' and 'comment' questions were challenging to read.

Useful information regarding sampling, bivariate data and modelling assumptions can be found in the specification for this component.

Candidates who did well on this paper generally did the following:	Candidates who did less well on this paper generally did the following:
• consistently carried out calculations accurately, giving answers to a sufficient degree of accuracy • carried out tests accurately stating correct hypotheses, identifying correct critical values, and drawing non-assertive conclusions that were stated in context • showed good understanding of limitations and assumptions associated with various statistical processes.	• did not maintain sufficient accuracy in their working to be able to give answers to the required number of decimal places • did not accurately use formulas that are provided in the formula booklet • found it difficult to answer 'explain' and 'comment' questions using sufficiently clear language and terminology • gave insufficient detail in response to 'show that' questions.

OCR support

Centres may find it helpful to make the Command Words poster available to candidates.

Question 1 (a)

1 In a quiz a contestant is asked up to four questions. The contestant's turn ends once the contestant gets a question wrong or has answered all four questions. The probability that a particular contestant gets any question correct is 0.6, independently of other questions. The discrete random variable X models the number of questions which the contestant gets correct in a turn.

(a) Show that $P(X = 4) = 0.1296$. **[1]**

This was well answered by most candidates.

Unsuccessful responses typically relied on built-in probability distributions on calculators that could not meet the 'show that' requirement of the question.

Question 1 (b)

The probability distribution of X is shown in **Fig. 1.1**.

r	0	1	2	3	4
$P(X = r)$	0.4	0.24	0.144	0.0864	0.1296

Fig. 1.1

(b) Find each of the following.

- $E(X)$
- $Var(X)$ **[2]**

This was answered correctly by most candidates.

When calculating $Var(X)$ candidates sometimes only evaluated $E(X^2)$.

Question 1 (c)

The number of points that a contestant scores is as shown in **Fig. 1.2**.

Number of questions correct	Number of points scored
0 or 1	0
2	2
3	3
4	5

Fig. 1.2

The discrete random variable Y models the number of points which the contestant scores.

(c) Without doing any working, explain whether each of the following will be less than, equal to or greater than the corresponding value for X.

- $\mathrm{E}(Y)$
- $\mathrm{Var}(Y)$ **[3]**

The instruction 'Without doing any working' in this question indicates that explanations should compare the distributions of X and Y without resorting to calculations.

The first mark was awarded for correctly noting that $E(Y)$ is less than $E(X)$ and was gained by many candidates. The awarding of the second mark was much less common as it required a full explanation referring to both the change from one question correct to no points scored and four questions correct to five points scored, together with the associated probabilities. Candidate responses often gave only a partial explanation and so were not credited.

The third mark was awarded for noting that $Var(Y)$ is greater than $Var(X)$ with a comment that the distribution of Y was more spread out.

Exemplar 1

1(c)

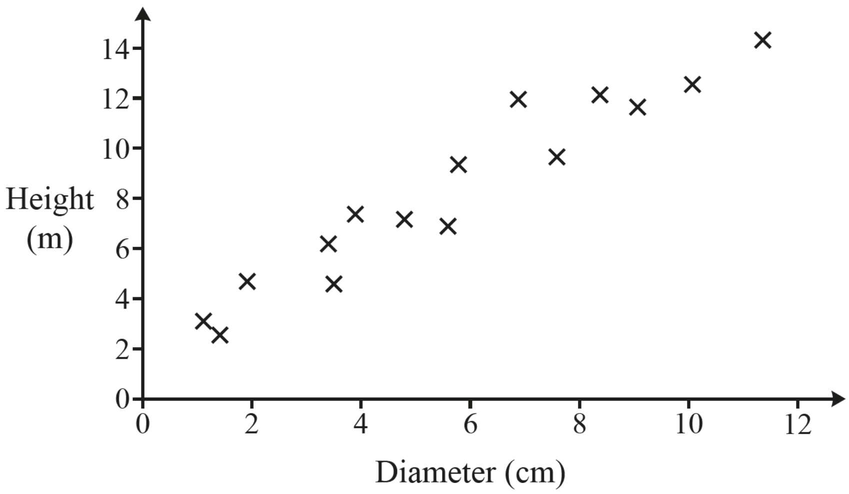
The points value for $X=0$, $X=2$ and $X=3$ are the same as X. The number of points scored for $X=4$ is increased to 5 and the points scored for $X=1$ is decreased to 0. Since $P(X=1) > P(X>4)$, this decrease has a larger effect on the expectation so $E(Y) < E(X)$.

The maximum points score is greater than the number of questions correct maximum as when $X=4$, $Y=5$. Furthermore, the probability of getting the lowest value (0) is greater as $X=1$, $Y=0$. Therefore, $Var(Y) > Var(X)$.

Exemplar 1 shows a comment in the first paragraph that includes the necessary detail to be given the second mark. The second paragraph then implies an understanding of the greater spread and gains the third mark.

Question 2 (a)

2 A forester is investigating the relationship between the diameter and the height of young beech trees. She selects a random sample of 15 young beech trees in a forest and records their diameters, d cm, and their heights, h m. The data are illustrated in the scatter diagram.



(a) State whether either or both of the variables d and h are random variables. **[1]**

Successfully answered by most candidates. Some went on to justify their response though the 'state' command does not require elaboration.

Question 2 (b)

Summary data for the diameters and heights are as follows.

$$n = 15 \quad \sum d = 84.9 \quad \sum h = 124.7 \quad \sum d^2 = 624.55 \quad \sum h^2 = 1230.57 \quad \sum dh = 866.63$$

(b) Find the equation of the regression line of h on d. Give your answer in the form $h = ad + b$, giving the values of a and b correct to **2** decimal places. **[4]**

Most candidates were able to gain some marks on this question. Less successful responses did not correctly calculate the gradient, instead evaluating the product moment correlation coefficient. In a few responses the regression line of d on h was found.

The requirement to give the values of a and b to **2** decimal places was sometimes overlooked. More commonly candidates rounded prematurely once they had calculated the gradient. This led to an incorrect value of 1.97 for b. Without fully detailed working showing the calculation as shown in the mark scheme this would score 2 out of 4. Some candidates used their calculators efficiently to obtain the correct value before rounding their final answer.

Assessment for learning

In both parts (b) and (c) of this question accuracy of working and final answer matters.

Centres can help candidates by drawing attention to the instructions on the front of the question paper:

- Where appropriate, your answer should be supported by working. Marks might be given for using a correct method, even if your answer is wrong.
- Give your final answers to a degree of accuracy that is appropriate to the context.

Question 2 (c)

(c) Use the regression line to predict the heights of beech trees with the following diameters.

- 7.5 cm
- 20.0 cm **[2]**

The context of the question matters. The diameters have been given to 1 decimal place and so it is inappropriate to give the predictions to more than 1 decimal place. As a result, it was common for candidates to gain 1 mark only.

Question 2 (d)

(d) Comment on the reliability of your predictions. **[3]**

Candidates were usually able to gain 2 of the 3 marks available for relating their comments to interpolation (sometimes incorrectly spelt 'intrapolation') and extrapolation.

The most successful candidates also understood that when interpolating there also needs to be strong linear correlation for the prediction to be reliable.

Assessment for learning

A good examination technique here is to note the number of marks available and use this as a prompt for the number of distinct points to make in the response.

Centres can help candidates by including examples of interpolation with weakly correlated variables in their teaching.

Question 2 (e)

(e) There are many mature beech trees with diameter of 60 cm or greater. However, there are no beech trees with a height of more than 50 m.

Comment on this in relation to your regression line. **[2]**

Comments relating to reliability were required in part (d) where there is a lack of certainty about the predictions. In part (e) there is no lack of certainty. The information provided tells us that the regression line is not valid for mature beech trees. Successful responses demonstrated this by candidates using the values provided and using sufficiently strong language in their conclusion.

Question 2 (f)

(f) State the coordinates of the point at which the regression line of d on h meets the line which you calculated in part **(b)**. **[1]**

Candidates who found the equation of the regression line of d on h and then solved simultaneous equations typically lost accuracy in their work and so did not gain the mark.

Noting that the command word is 'state', and that there is only 1 mark available, successful responses gave the mean point. Some candidates did not appear to be aware that this would be the point of intersection.

Question 3 (a)

3 Jane wonders whether the number of wasps entering a wasp's nest per 5 second interval can be modelled by a Poisson distribution with mean μ. She counts the number of wasps entering the nest over 60 randomly selected 5-second intervals. The results are shown in **Fig. 3.1**.

Number of wasps	0	1	2	3	4	5	6	7	8	9	$\geqslant 10$
Frequency	0	2	5	5	12	10	10	11	1	4	0

Fig. 3.1

(a) Show that a suitable estimate for the value of μ is 5.1. **[2]**

Many candidates gained these 2 marks. Those who did not tended to give insufficient detail to satisfy the 'show that' command. For example, simply writing down $\frac{306}{60} = 5.1$ was insufficient as the 306 could have been obtained by multiplying 5.1 by 60, both figures given in the question. There needed to be some evidence of working with the figures in the table.

Question 3 (b)

Fig. 3.2 shows part of a screenshot for a χ^2 test to assess the goodness of fit of a Poisson model. The sample mean has been used as an estimate for the population mean. Some of the values in the spreadsheet have been deliberately omitted.

	A	B	C	D	E
1	Number of wasps	Observed frequency	Poisson probability	Expected frequency	Chi-squared contribution
2	⩽2	7	0.1165	6.9887	0.0000
3	3	5		8.0874	1.1786
4	4	12			0.2765
5	5	10			0.0255
6	6	10	0.1490	8.9400	0.1257
7	7	11	0.1086	6.5134	3.0904
8	⩾8	5	0.1440	8.6414	
9					

Fig. 3.2

(b) Determine the missing values in each of the following cells, giving your answers correct to **4** decimal places.

- C3
- D5
- E8

[4]

This question was generally well answered. Candidates demonstrated good understanding of how to obtain the required values. Some candidates did not give their answers accurately to the required four decimal places.

Question 3 (c)

(c) Explain why some of the frequencies have been combined into the categories $\leqslant 2$ and $\geqslant 8$. **[1]**

Responses could only be credited if they made explicit reference to expected frequencies. Unsuccessful responses typically either referred to non-specific frequencies, or to observed frequencies.

Question 3 (d)

(d) **In this question you must show detailed reasoning.**

Carry out the hypothesis test at the 5% significance level. **[6]**

Typically, candidates demonstrated a good understanding of the structure required for the hypothesis test. Taking each stage in turn:

- Hypotheses were sometimes stated the wrong way round. There were also responses that incorrectly included the mean value, 5.1; this value has been estimated from the data. In the opening sentence of Question 3 the mean is not specified.
- Some candidates incorrectly gave the test statistic as 4.6967 omitting the value from cell E8 calculated in part (b). This error may have been avoided by writing the values from part (b) in the table on the question paper.
- For the degrees of freedom some candidates wrongly worked from the original table rather than the spreadsheet with the combined categories. Candidates sometimes incorrectly calculated $7 - 1 = 6$ leading to an incorrect critical value, perhaps due to not realising that the reliance on the data to generate a value for the mean added a restriction. There were also examples of candidates fortuitously obtaining 5 from incorrect working such as $(7 - 2)(2 - 1)$ or $(7 - 1)(2 - 1) - 1$ suggesting some confusion with contingency tables. Had they not shown the calculation but simply stated 5, examiners would have had to infer correct reasoning.
- Candidates generally compared their test statistic and critical value leading to a suitably non-assertive conclusion. Less successful responses often only made a generic statement whereas more successful candidates ensured the conclusion referred to the context of wasps entering their nest.

Exemplar 2

3(d)

H_0: the data can be modelled as a Poisson ~~dist~~ distribution with $\lambda = 5.1$

H_1: the data cannot be modelled as a Poisson distribution with $\lambda = 5.1$

λ is the mean

$\chi^2 = 1.1786 + 0.2765 + 0.0255 + 0.1257 + 3.0904 + 1.5344$

$= 6.2311$

For $v = (7-1)$

$= 6$

at 5% level, the critical value is 12.59

$6.2311 < 12.59$ so don't reject H_0.

There is insufficient evidence to suggest that the data cannot be modelled by a Poisson distribution with $\lambda = 5.1$

Exemplar 3

3(d)

H_0: Poisson is a good model for measuring the amount of wasps ~~enter~~ entering a nest per 5 second interval

H_1: Poisson isnt a good model for measuring the amount of wasps entering a nest per 5 second interval

χ^2 value: 6.2311

Critical value ~~with~~ with 5% significance level with 5 degrees of freedom is 11.07

6.2311 < 11.07 so not significant

~~thus~~ There is insufficient information to reject H_0 and suggest poisson is not a good model for measuring the amount of wasps entering a nest per 5 second interval

Exemplars 2 and 3 illustrate two responses with similar structures but significantly different scores.

Exemplar 2 was given 2/6. This is an example of a response that incorrectly includes the mean in the hypotheses which may then have led to the incorrect number of degrees of freedom and critical value. There is also no mention of the context in the conclusion.

Exemplar 3 was given 6/6. Reference to the context in the hypotheses, while not necessary to gain the first mark, may have helped ensure that it was included in the conclusion.

Question 3 (e)

(e) Jane also carries out a χ^2 test for the number of wasps leaving another nest. As part of her calculations, she finds that the probability of no wasps leaving the nest in a 5-second period is 0.0053. She finds that a Poisson distribution is also an appropriate model in this case.

Find a suitable estimate for the value of the mean number of wasps leaving the nest per 5-second period. **[2]**

There were many correct responses using the Poisson probability formula accurately. Some candidates used a trial and improvement approach. While this could score both marks it was often unsuccessful.

Assessment for learning

There were several responses that misread 0.0053 as 0.053

To help avoid this slip candidates are encouraged to annotate the value on the question paper by, for example, highlighting or circling.

Question 4 (a)

4 Alex is practising bowling at a cricket wicket. Every time she bowls a ball, she has a 30% chance of hitting the wicket.

(a) Assuming that successive bowls are independent, determine the probability that Alex first hits the wicket on her third attempt. **[2]**

This question was answered correctly by most candidates.

Question 4 (b)

(b) Determine the probability that Alex hits the wicket for the fourth time on her tenth attempt. **[3]**

This question was answered fully correctly by many candidates.

Less successful responses did not recognise the significance of the fourth success occurring on the tenth attempt. This often led to candidates calculating $p(X = 4)$ using $X \sim B(10,0.3)$.

Successful responses recognised the need to find $p(X = 3)$ using $X \sim B(9,0.3)$ and then multiplied by 0.3, sometimes writing this concisely as $\binom{9}{3} \times 0.7^6 \times 0.3^4$. The final mark was sometimes not given due to candidates over-rounding the answer to 0.08.

Assessment for learning

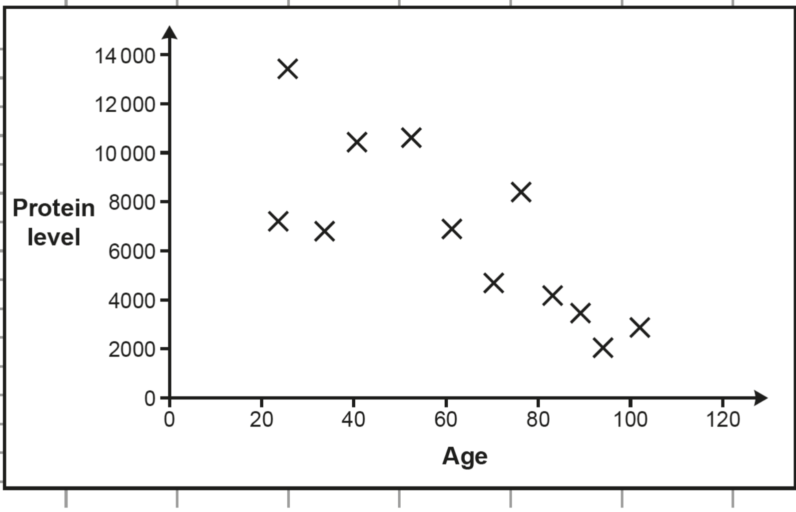
Centres can help candidates by drawing their attention to the subject specific marking instructions in section 2 at the start of mark schemes. Of relevance here is note f:

"When a value is **not given** in the paper accept any answer that agrees with the correct value to **2 s.f.** unless a different level of accuracy has been asked for in the question, or the mark scheme specifies an acceptable range."

Question 5 (a)

5 A medical researcher is investigating whether there is any relationship between the age of a person and the level of a particular protein in the person's blood. She measures the levels of the protein (measured in suitable units) in a random sample of 12 hospital patients of various ages (in years). The spreadsheet shows the values obtained, together with a scatter diagram which illustrates the data.

	A	B
1	**Age**	**Protein level**
2	23	7200
3	25	13400
4	33	6800
5	40	10400
6	52	10600
7	61	6900
8	70	4700
9	76	8400
10	83	4200
11	89	3500
12	94	2100
13	102	2900



(a) The researcher decides that a test based on Pearson's product moment correlation coefficient may **not** be valid. Explain why she comes to this conclusion. **[2]**

The first part of a response to this question should focus on the shape of the cluster of points. Candidates should be using correct terminology, for example "not elliptical" rather than "not an oval". This then enables them to conclude that the sample may not be drawn from a bivariate Normal distribution.

Less successful responses sometimes simply stated that the distribution was not bivariate Normal without explaining why. Other candidates noted the shape and then described the data itself as not being bivariate Normal rather than making it clear, as more successful candidates did, that it is the underlying distribution that may not be bivariate Normal.

Question 5 (b)

(b) Calculate the value of Spearman's rank correlation coefficient. **[3]**

A few candidates did not produce rank lists, and this then hindered them in part (c).

A more common error was to rank the values inconsistently, leading to a positive correlation coefficient.

Generally, this question was answered fully correctly.

Assessment for learning

Centres can help candidates by encouraging them to estimate a correlation coefficient by looking at the scatter of points before starting the calculation. In this case candidates should recognise that a positive value must be wrong. There was evidence that some candidates did do this and were able to correct their initial error.

Question 5 (c)

(c) Carry out a test based on this coefficient at the 5% significance level to investigate whether there is any association between age and protein level. **[5]**

As with Question 3 (d), candidates demonstrated a good understanding of the structure required for the hypothesis test. Taking each stage in turn:

- Hypotheses were sometimes stated the wrong way round. Few candidates were given both marks, usually due to not mentioning the population.
- The correct critical value was identified by most candidates. The most common error was use of the one-tailed value (0.5035). Some candidates wrongly used the table for the product moment correlation coefficient.
- The comparison of test statistic and critical value was usually done well with candidates avoiding potential sign slips.
- Conclusions usually mentioned the context. While some were too assertive, and some incorrectly referred to correlation rather than association, generally this was done well.

Question 5 (d)

(d) Explain why the researcher chose a sample that was random. **[2]**

Many responses to this question concentrated on ideas around bias. While avoiding systematic bias is an advantage of a random sample (and suitable comments were given a mark) it is not the primary reason for the researcher. A few candidates identified proper inference about the population as the key point to make here.

Misconception

A random sample will not, necessarily, be representative of the population.

Question 5 (e)

(e) The researcher had originally intended to use a sample size of 6 rather than the 12 that she actually used.

Explain what advantage there is in using the larger sample size. **[2]**

A common error here was state that the accuracy of calculations and tests is improved when samples are larger. Some candidates discussed improved reliability of conclusions. A good number of candidates were credited for noting that the influence of outliers is reduced as sample size increases.

Few candidates commented that the test statistic will tend to be closer to the population value for larger samples.

Misconception

?

While the study of type I and type II errors lies beyond the scope of this component, some responses to this question highlighted a misconception that it would be helpful to address.

The probability that a null hypothesis, if true, will be rejected is limited by the choice of the level of significance. Increasing the size of the sample does not increase this probability.

Question 6 (a)

6 The random variable X has a uniform distribution over the values $\{1, 4, 7, \ldots, 3n-2\}$, where n is a positive integer.

(a) Determine $\mathrm{Var}(X)$ in terms of n. **[3]**

There was an even split between candidates on this question. Successful responses recognised the relationship between X and the standard discrete uniform distribution for which they are given the variance formula in the booklet provided. Typically, they then obtained the correct expression.

The most common error by far was to substitute $(3n-2)$ directly into the variance formula and this was not given any credit.

Question 6 (b)

(b) Given that $n = 100$, find the probability that X is within one standard deviation of the mean. **[4]**

It was possible for candidates to make good progress with this question even if working with an incorrect expression from part (a).

The most common approach was to find mean and then the upper and lower bounds for X using their variance expression. Some candidates incorrectly used 50.5 as the mean rather than 149.5 and this then limited them to a maximum of the second mark only. Those who found the bounds then either sought to count values of X in the interval or first converted to values for the uniform distribution over $1,2,\ldots,n$ which should have led to $[22,79]$. Some candidates then mistakenly calculated $79 - 22 = 57$ (an error sometimes referred to as 'the fencepost problem') and did not gain the final mark for the probability.

A variation on this approach was to use the symmetry of the uniform distribution, identify the number of values falling below the lower bound, using this to first find the complement of the required probability. Candidates following this strategy were able to avoid the fencepost problem.

Some candidates realised that a more concise approach was to work directly with the uniform distribution over $1,2,\ldots,n$ from the start.

Some candidates sought to work solely with the standard deviation leading to an interval width of 173.2 for X (57.7 over $1,2,\ldots,n$). This approach overlooks the mean and leads to an incorrect probability of 0.57

Candidates making the common error in part (a) worked with an expression equivalent to $\frac{3}{4}n^2 - n + \frac{1}{4}$. Due to the large value of n being used this gave a 'variance' sufficiently close to the correct value that it was possible to obtain the required probability of 0.58. In this case the final accuracy mark was not given.

Copyright information

Supporting you

Post-results services

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.

Keep up-to-date

We send a weekly roundup 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 CPD 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 on the relevant subject page on our website or visit OCR professional development.

Signed up for ExamBuilder?

ExamBuilder is the question builder platform for a range of our GCSE, A Level, Cambridge Nationals and Cambridge Technicals qualifications. Find out more.

ExamBuilder is **free for all OCR centres** with an Interchange account and gives you unlimited users per centre. We need an Interchange username to validate the identity of your centre's first user account for ExamBuilder.

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

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.

It allows you to:

- review and run analysis reports on exam performance
- analyse results at question and/or topic level
- compare your centre with OCR national averages
- identify trends across the centre
- facilitate effective planning and delivery of courses
- identify areas of the curriculum where students excel or struggle
- help pinpoint strengths and weaknesses of students and teaching departments.

Find out more.

Need to get in touch?

If you ever have any questions about OCR qualifications or services (including administration, logistics and teaching) please feel free to get in touch with our customer support centre.

Call us on
01223 553998

Alternatively, you can email us on
support@ocr.org.uk

For more information visit
- **ocr.org.uk/qualifications/resource-finder**
- **ocr.org.uk**
- **/ocrexams**
- **/ocrexams**
- **/company/ocr**
- **/ocrexams**

We really value your feedback

Click to send us an autogenerated email about this resource. Add comments if you want to. Let us know how we can improve this resource or what else you need. Your email address will not be used or shared for any marketing purposes.

I like this **I dislike this**

Please note – web links are correct at date of publication but other websites may change over time. If you have any problems with a link you may want to navigate to that organisation's website for a direct search.