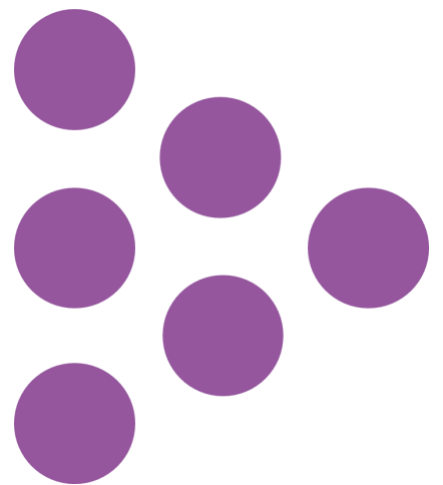


Report

National Reference Test Results Digest 2022

National Foundation for Educational Research (NFER)



National Reference Test Results Digest 2022

Bethan Burge

Louise Benson

Published in August 2022

By the National Foundation for Educational Research,

The Mere, Upton Park, Slough, Berkshire SL1 2DQ

<https://www.nfer.ac.uk/>

© 2022 National Foundation for Educational Research

Registered Charity No. 313392

ISBN: 978-1-912596-64-5

How to cite this publication:

Burge, B. and Benson, L. (2022) *National Reference Test Results Digest 2022*. Slough: NFER

Contents

1. Introduction.....	1
2. The sample.....	2
3. Results for the test booklets in 2022.....	5
4. Performance in English in 2022.....	10
5. Performance in maths in 2022.....	14
6. Appendix A: A brief summary of the NRT.....	19

1. Introduction

Ofqual has contracted the National Foundation for Educational Research (NFER) to develop, administer and analyse the National Reference Test (NRT) in English and maths. The first NRT took place in 2017 and established a baseline from which any future changes in standards can be detected. This report represents an overview of the findings of the 2022 testing process.

The NRT, which consists of a series of test booklets, provides evidence on changes in the performance standards in GCSE English language and maths in England at the end of key stage 4. It does this by testing content taken from the GCSE English and maths curricula. It has been designed to provide additional information to support the awarding of GCSEs in English language and maths and is based on a robust and representative sample of Year 11 students who will, in the relevant year, take their GCSEs.

More information about the NRT can be found in the [NRT document collection](#).

The first live NRT took place in 2017. The outcomes of the 2017 GCSE examinations in English language and maths provided the baseline percentages of students at three grade boundaries and these were mapped to the NRT for 2017 to establish the corresponding proficiency level. The percentages of students achieving those proficiency levels in each subsequent year are calculated and compared.

The NRT structure is intended to remain the same each year. For each of English and maths, there are eight test booklets in use. Each question is used in two booklets, so that effectively all the tests can be analysed together to give a single measure of subject performance. This is similar to other studies that analyse trends in performance over time, for example, international surveys such as PISA and TIMSS.

This report provides summarised information of the key performance outcomes for English and maths in 2022 and provides information on the changes from the baseline standards established in 2017. It also includes data on the achievement of the samples, their representativeness and the performance of the students on the tests. Further information on the nature of the tests, the development process, the survey design and its conduct, and the analysis methods used is provided in the accompanying document: [Background Report: National Reference Test](#).

2. The sample

The NRT took place between 28 February and 11 March 2022. The numbers of participating schools and students are shown in Table 2.1 and Table 2.2. In 2022, the number of schools in the sample has returned to pre-pandemic levels and is now above target.

Table 2.1 Target sample sizes and achieved samples in current and previous years

	NRT Target Sample	Achieved sample 2022	Achieved sample 2021	Achieved sample 2020	Achieved sample 2019	Achieved sample 2018	Achieved sample 2017
English: Number of Schools	330	334	214	332	332	312	339
Maths: Number of Schools	330	334	216	333	331	307	340

The sample was stratified by the historical attainment of schools in GCSE English language and GCSE maths and by school size. In addition, the types of schools were monitored. Checks were made on all three of these variables to ensure that the achieved sample reflects the sampling frame. Students at independent schools have been under-represented in all years of the NRT, although the exact number of independent schools participating each year¹ has varied. As a result of this, students in other types of school such as academy converters are slightly over-represented in the sample. The under-representation of independent schools may have contributed to the achieved sample being slightly lower attaining than the national population but this small difference between the achieved sample and the sampling frame at the top end of the distribution based on previous school GCSE performance has remained broadly consistent across the years.

Table 2.2 shows the number of students in the final sample for whom booklets were dispatched and the number completing the tests for both English and maths. As this shows, just over 80 per cent of students who were selected took part in the tests.

¹ The number of independent schools in each of the NRT administrations were as follows: Seven in 2017 and 2018, nine in 2019, four in 2020, none in 2021 and four in 2022 (the ‘establishment type’ information included in the sampling frame for each school was obtained from Get Information About Schools <https://www.get-information-schools.service.gov.uk/>). Students from independent schools have historically comprised around one per cent of the achieved sample (both weighted and unweighted).

Table 2.2. Completed student test returns for English and maths in all NRT administrations

Year	No. of students: dispatched English tests	No. of students: completed English tests	% of students: completed English tests	No. of students: dispatched maths tests	No. of students: completed maths tests	% of students: completed maths tests
2022	7,969	6,457	81	7,961	6,406	80
2021	5,124	4,030	79	5,152	4,143	80
2020	7,845	6,639	85	7,886	6,756	86
2019	7,928	6,739	85	7,917	6,825	86
2018	7,354	6,193	84	7,320	6,169	84
2017	8,040	7,082	88	8,080	7,144	88

In total, 1,512 students from 334 schools were recorded as non-attendees during the English NRT, which is 19% of the total number of 7,969 sampled students spread across the schools participating in the assessment. A total of 1,555 students from 334 schools were recorded as non-attendees during the maths NRT, which is 20% of the total number of 7,961 sampled students spread across the schools participating in the test.

The pattern of non-attendance is similar in maths to English. The principal reason given for non-attendance was absence due to illness or other authorised reason, which accounted for 63% of non-attendance for both English and maths. Students being absent from the testing session but present in school remains the second most frequently recorded reason, accounting for 14% of non-attendance in both subjects. Of the remaining reasons for non-attendance, around five per cent of students were withdrawn by the headteacher (five per cent for English and six per cent for maths) and another five per cent of students were studying at a different venue.

The percentage of non-attendance in 2022 was higher than that seen in previous cycles of NRT. Student participation rates in 2022 were 81% for English and 80% for maths. Although very slightly higher than the attendance achieved in 2021, this is lower than the student participation rates achieved in earlier years of NRT (between 84% and 86%) but comparable with the response rate required in large scale international studies. In addition, there was no evidence that the pattern of student non-attendance has changed relative to previous years.

2.1. Access arrangements

The NRT offers access arrangements consistent with JCQ requirements (for GCSE examinations) in order to make the test accessible to as many sampled students as possible. Schools were asked to contact NFER in advance of the NRT to indicate whether any of their students required modified test materials or if students' normal working practice was to use a word processor or laptop during examinations. In cases where additional time would be needed for particular students, schools

were asked to discuss this need with the NFER test administrator and ensure that the extra time for the testing session could be accommodated. All requests from schools for access arrangements and the type of arrangement required were recorded. Table 2.3 shows the different types of access arrangements that were provided to students for the 2022 NRT, organised by NFER. This table includes instances where students required more than one access arrangement. These are the access arrangements facilitated by NFER for the NRT in 2022; we do not collect complete data on the permitted arrangements that are organised by the school such as readers, scribes, extra time and examination pens so they are not included in the table below. Overall, the percentage of sampled students receiving access arrangements was similar to 2021.

Table 2.3. Number of access arrangements facilitated by NFER in 2022

Arrangement provided	No. of students English	No. of students maths	Total number of students	% of sampled students
Word processor	280	128	408	2.6
Different colour test paper	122	115	237	1.5
Modified enlarged print and enlarged copies	24	14	38	0.2
Braille	0	0	0	0
Total	426	257	683	4.3%

NB: Due to some students having multiple access arrangements, they will be featured twice in the table.

3. Results for the test booklets in 2022

Details of the analysis procedures are given in the accompanying document: Background Report: National Reference Test Information. The analysis process followed a sequence of steps. Initially, the tests were analysed using Classical Test Theory to establish that they had performed well, with appropriate difficulty and good levels of reliability. The subsequent analyses used Item Response Theory (IRT) techniques to link all the tests and estimate the ability of all the students on a common scale for each subject, independent of the test or items they had taken. These ability estimates were then used for calculating the ability level at the percentiles associated with the GCSE grade boundaries in 2017. From 2018 onwards, the percentages of students achieving above these baseline ability levels are established from the NRT.

3.1. English

The results of the Classical Test Theory analyses are summarised in Table 3.1. This shows the range of the main test performance statistics for the eight English test booklets used.

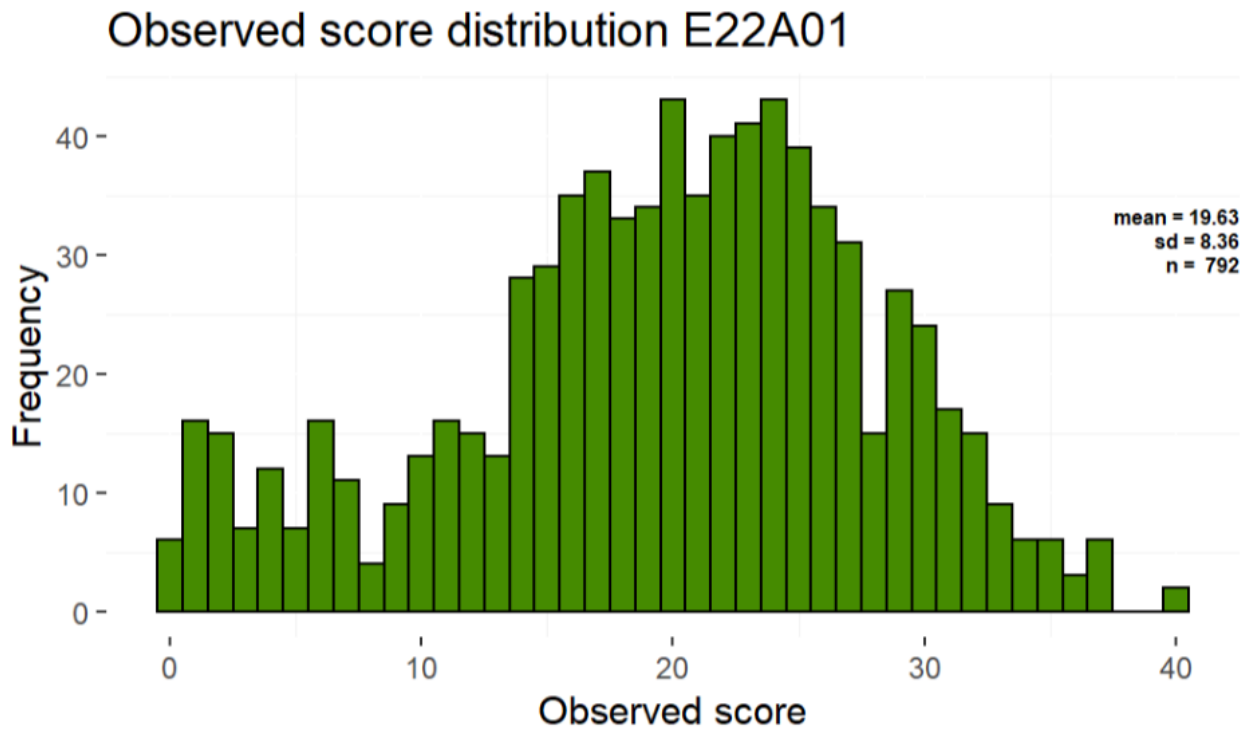
Table 3.1. Range of Classical Test Theory statistics for the English tests in 2022

Classical Test Theory statistic	Minimum	Maximum
Number of students taking each test booklet	792	835
Maximum score attained (out of 50)	40	47
Average score attained	18.62	19.99
Standard deviation of scores attained	8.36	8.97
Reliability of the tests (Coefficient Alpha)	0.77	0.80
Average percentage of students attempting each item (%)	92	95

These results show that the English test booklets functioned well, and similarly to previous years. The booklets were challenging, with few students attaining over 40 marks and average scores somewhat less than half of the available marks. Maximum raw scores ranged from 40 to 47 across the eight booklets, showing a wider range compared to 2021 where maximum raw scores were between 41 and 43. The standard deviation shows that the scores were well spread out, allowing discrimination between the students. This is confirmed by the reliability coefficients which are at a good level for an English test of this length. Finally, the average percentage of students attempting each item was over 90 per cent for all booklets, indicating that the students were engaging with the test and attempting to answer the majority of questions.

These results were confirmed by the distribution of scores students achieved on the tests. This is shown for one of the tests in Figure 3.1. It is an example of one test booklet only but the distributions were similar for the other tests. The figure shows that students were spread across the range, although no students attained the very highest marks.

Figure 3.1. Score distribution for one of the English tests



In addition, a full item analysis was carried out for each test, in which the difficulty of every question and its discrimination were calculated. These indicated that all the questions had functioned either well or, in a small number of cases, adequately and there was no need to remove any items from the analyses. It should be noted that in 2022 one of the reading components was replaced with a reading component from the refresh item bank. The replacement reading component performed well and in line with the existing reading components. Therefore, all items were retained for the IRT analyses. Additionally, an analysis was conducted to establish if any items had performed markedly differently in 2022 compared with the previous years. Where there are such indications, a formal procedure is followed for reviewing the items to establish whether there could be an external reason for the change. In 2022, one reading item was removed from the link between 2017-2021 and 2022. This item was removed based on the evidence provided in the formal review process indicating that there had been a systematic change in marking standard. There was no evidence from the DIF analysis that the adaptations introduced for GCSE English language exams this summer (i.e. advance information) had an impact on performance on the NRT items. This is not unexpected given that advance information was released close to the NRT taking place.

Using the common items, the IRT analyses equated the eight tests. The IRT analyses also used the items common between years to equate the tests over years, allowing ability estimates for students in all six years to be on the same scale. After this had been done, the results showed that the mean ability scores for students were similar for all the tests, confirming that the random allocation to tests had been successful. The results also showed that the level of difficulty of the eight tests was fairly consistent, with only small differences between them.

Both the Classical Test Theory results and the IRT results for the English tests showed that these had functioned well to provide good measures of the ability of students, sufficient for estimating averages for the sample as a whole.

3.2. Maths

The results of the Classical Test Theory analyses are summarised in Table 3.2. This shows the range of the main test performance statistics for the eight maths tests used.

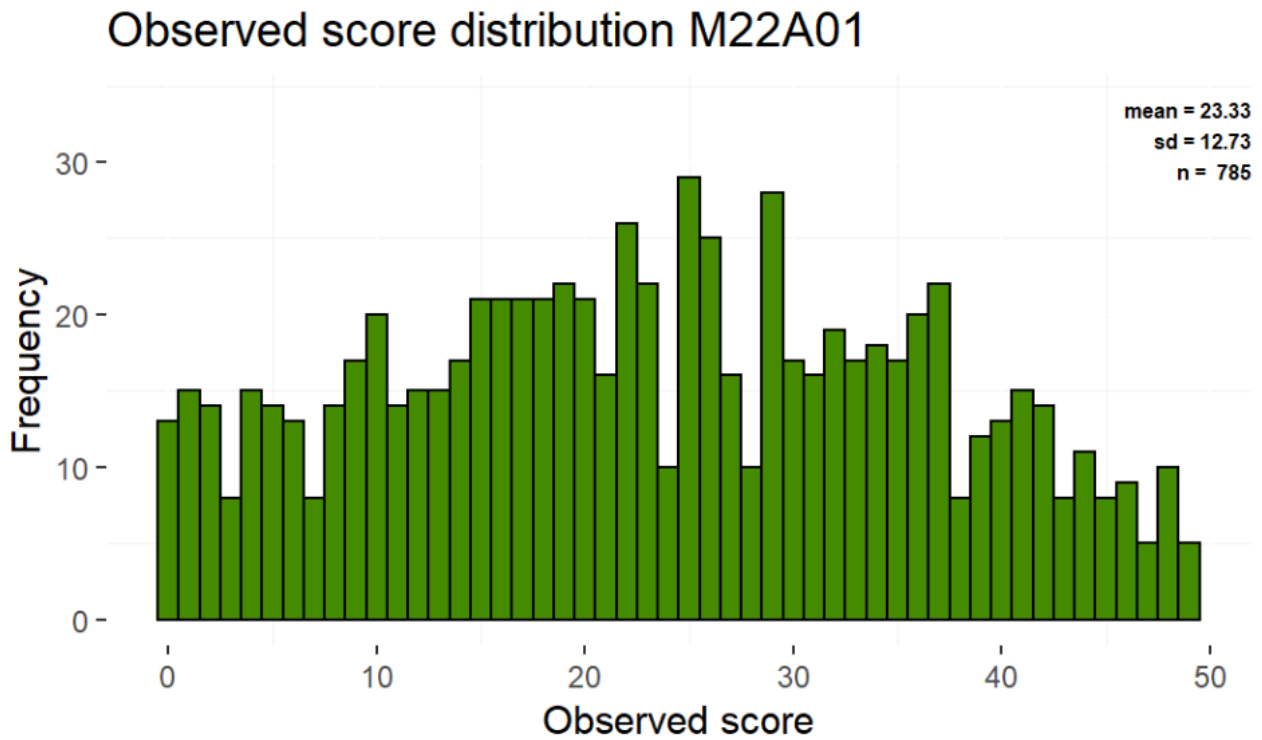
Table 3.2. Range of Classical Test Theory statistics for the maths tests in 2022

Classical Test Theory statistic	Minimum	Maximum
Number of students taking each test booklet	785	824
Maximum score attained (out of 50)	49	50
Average score attained	20.7	23.9
Standard Deviation of scores attained	12.0	14.0
Reliability of the tests (Coefficient Alpha)	0.89	0.91
Average percentage of students attempting each item (%)	84	88

These results show that the maths tests also functioned well. The maximum score, or one mark short of it, was attained on all booklets. The average scores were, again, slightly less than half marks for most booklets which is similar to 2021 but lower than in earlier years of the NRT. The standard deviation shows that the scores were well spread out, allowing discrimination between the students. This is confirmed by the reliability coefficients which are at a very good level for a maths test of this length and higher than for English, which is usual. Finally, the average percentage of students attempting each item (between 84 and 88 per cent) is similar to the percentages seen in 2021, although lower than in 2020. As has been the case in previous years, the average percentage of students attempting each item for maths was also lower than that seen for the English test. However, there are more individual items for students to attempt in the maths test.

These results were confirmed by the distribution of scores achieved on the tests. This is shown for one of the tests in Figure 3.2. The distributions were similar for the other tests. The figure shows that scores were attained over the range of possible marks and that the students were fairly evenly spread over the range.

Figure 3.2. Score distribution for one of the maths tests



In addition, a full item analysis was carried out for each test, in which the difficulty of every question and its discrimination were calculated. These indicated that all the questions had functioned either well or, in a small number of cases, adequately. Therefore all items were retained for the IRT analyses. Additionally, an analysis was conducted to establish if any items had performed markedly differently in 2022 compared with the previous years. Where there are such indications, a formal procedure is followed for reviewing the items to establish whether there could be an external reason for the change and if there is sufficient evidence to remove the item from the link between years. It should be noted that in 2022 the names of characters in four of the items were changed to more ethnically diverse names in order to improve the diversity of the characters referenced in each booklet and across the whole maths NRT. These four items did not function any differently and therefore were treated as common items. In 2022, no items were removed from the link. There was also no evidence from the DIF analysis that the adaptations introduced for GCSE maths exams this summer (i.e. advance information and the provision of formulae sheets) had an impact on performance on the NRT items. This is not unexpected given that advance information was released close to the NRT taking place.

Using the common items, the IRT analysis was used to equate the eight tests. The IRT analysis also used the items common between years to equate the tests over years, allowing ability estimates for students in all six years to be on the same scale. After this had been done, the

results showed that the mean ability scores for students were similar for all the tests, confirming that the random allocation to tests had been successful. The results also showed that the level of difficulty of the eight tests was fairly consistent, with only small differences between them.

Both the Classical Test Theory results and the IRT results for the maths tests showed that these had functioned well to provide good measures of the ability of students, sufficient for estimating averages for the sample as a whole.

Summary

These initial stages of the analyses, the Classical Test Theory evaluation of test functioning and the IRT equating of the tests, indicate that the NRT performed as well in 2022 as it had in previous years. This allowed the final stages of the analysis, the estimation of the percentages of students above the same ability thresholds as in 2017 and the calculation of their precision, to be undertaken with confidence. These are described in Sections 4 and 5 for English and maths respectively.

4. Performance in English in 2022

The objective of the NRT is to get precise estimates of the percentages of students each year achieving at a level equivalent to three key GCSE grades in 2017: these key grades are 4, 5 and 7. For the NRT in 2017, these baseline percentages were established from the 2017 GCSE population percentages. The NRT ability distribution, based on the IRT analysis, was then used to establish the ability thresholds which corresponded to those percentages. From 2018 onwards, the thresholds correspond to the same level of student ability as the thresholds established in 2017, thus allowing us to estimate the percentage of students above each of those thresholds and track performance over time. Alongside this, based on the sample achieved and the reliability of the tests, we can model the level of precision with which the proportion of students achieving the ability thresholds can be measured. The target for the NRT is to achieve a 95% confidence interval of plus or minus no more than 1.5 percentage points from the estimate at each ability threshold.

Ofqual provided the percentages of students at or above the three relevant grades (grades 4, 5 and 7) taken from the 2017 GCSE population. These are shown in Table 4.1. These percentages were mapped to three ability threshold scores in the NRT in 2017.

Table 4.1. English 2017 NRT baseline thresholds

Threshold	Percentage of students above threshold from 2017 GCSE
Grade 7 and above	16.8
Grade 5 and above	53.3
Grade 4 and above	69.9

In 2022, the NRT data for the years 2017 to 2022 were analysed together using IRT modelling techniques. By analysing all the data concurrently, ability distributions could be produced for the samples for each year on the same scale. The percentages of students at each of the three GCSE grade boundaries, fixed on the 2017 distribution, could then be mapped onto the distributions for the subsequent years to produce estimates of the percentage of students at the same level of ability in those years. For example, the percentage of students at the ‘Grade 4 and above’ threshold in the 2017 GCSE population was 69.9 per cent. This was mapped onto the 2017 distribution to read off an ability value at that grade boundary. The same ability value on the distributions for all other years can then be found, and the percentage of students at this threshold or above in those years can be established. In this way, we are able to estimate the percentage of students at the same level of ability as represented in the 2017 GCSE population for each year of the NRT going forward. The precision of these estimates is dependent on both the sample achieved and the reliability of the tests as measures.

Table 4.2 presents the percentages of students achieving above the specified grade boundaries for the years 2017 to 2022. Confidence intervals for percentages are provided in brackets alongside the estimates. This is important as it shows that, although there have been changes in

performance, these are often within the confidence intervals. The statistical interpretation of the differences is discussed below.

Table 4.2. Estimated percentages at grade boundaries in English

Year	Estimated percentages at Grade 4 and above	Estimated percentages at Grade 5 and above	Estimated percentages at Grade 7 and above
2017	70.0 (67.8-72.2)	53.3 (51.2-55.5)	16.8 (15.3-18.3)
2018	69.2 (67.0-71.4)	53.2 (51.3-55.1)	17.3 (15.8-18.7)
2019	66.1 (64.2-68.0)	50.3 (48.1-52.4)	16.7 (15.1-18.3)
2020	67.7 (65.7-69.7)	52.1 (50.3-54.0)	18.1 (16.4-19.8)
2021	67.1 (64.7-69.6)	52.7 (50.2-55.1)	19.3 (17.4-21.2)
2022	65.3 (63.2-67.4)	49.4 (47.6-51.1)	16.9 (15.5-18.3)

The 2017 figures in the table above are based on the NRT study, rather than the 2017 GCSE percentages. Note that, because of the way in which they have been computed, they match closely with the GCSE percentages. The confidence intervals for them reflect the fact that the NRT 2017 outcomes carry the statistical error inherent in a sample survey, as per the subsequent years.

In each year of the NRT, the percentages for previous years are re-estimated due to the concurrent calibration approach which analyses all of the data together in a single IRT model. Some degree of variation is therefore expected with the addition of more data, and the differences seen are generally small.

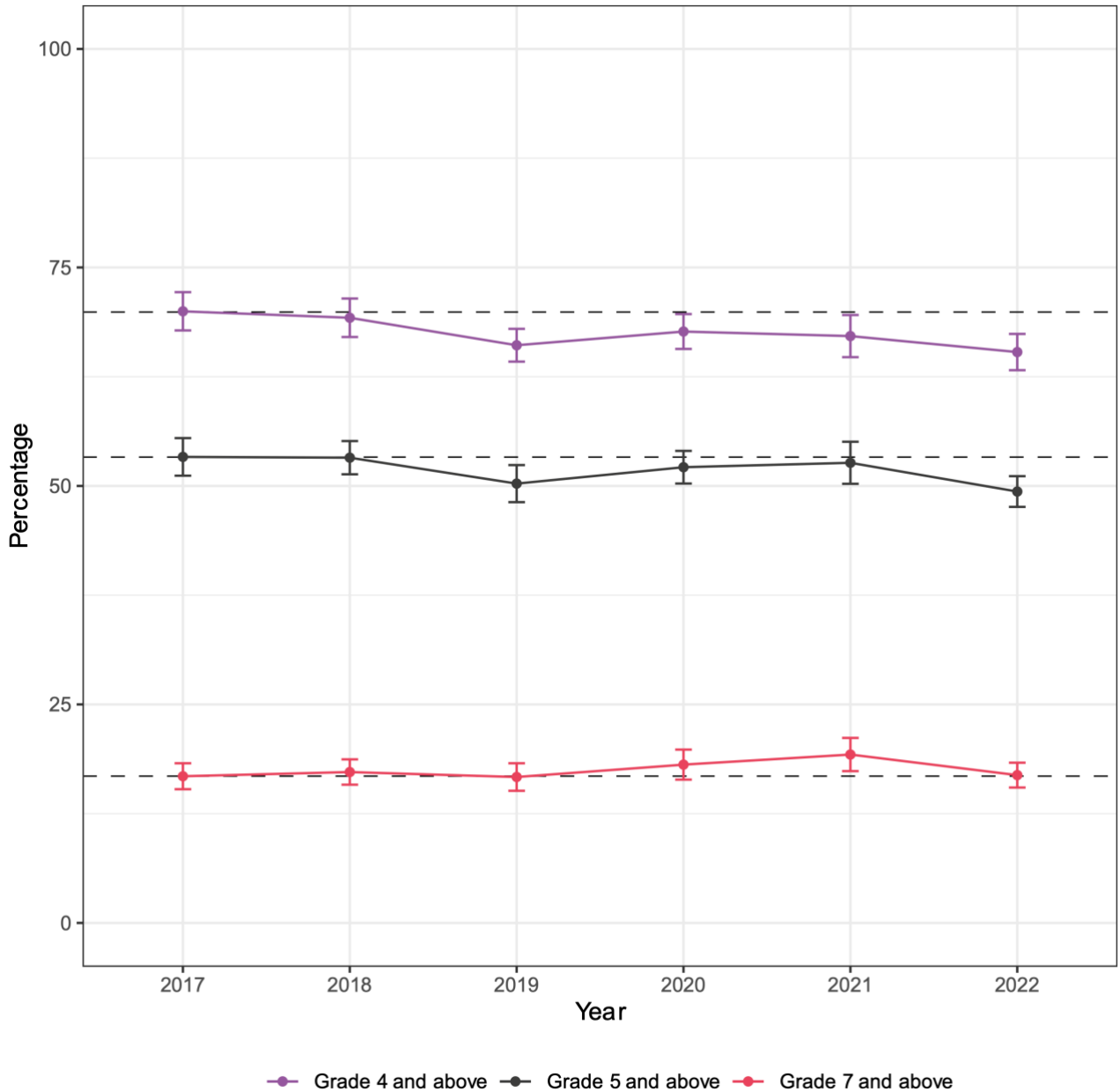
Table 4.3 shows the half widths of the confidence intervals. The confidence intervals increased in 2021, reflecting the smaller sample size in that year, and have returned this year to similar levels as seen in earlier years of the NRT.

Table 4.3. English NRT half width of confidence intervals each year

Year	Half width of confidence intervals: Grade 4 and above	Half width of confidence intervals: Grade 5 and above	Half width of confidence intervals: Grade 7 and above
2017	2.2	2.1	1.5
2018	2.2	1.9	1.4
2019	1.9	2.1	1.6
2020	2.0	1.9	1.7
2021	2.4	2.4	1.9
2022	2.1	1.8	1.4

Figure 4.1 presents 95% confidence intervals around the percentages achieving at least the specified grade boundary in 2022, as compared with previous years and the 2017 population baseline percentages. The 2017 population percentages are represented as dotted lines and the trend lines across years as solid lines. This format has been used to encourage the reader to compare the point estimate confidence bands for each year with the 2017 baseline population percentages, bearing in mind the confidence intervals.

Figure 4.1. Long term changes in NRT English over time from 2017 baseline



The chart suggests that performance in English has declined relative to 2017. There had previously been a small decline in the percentage of students achieving at-or-above both grades 4 and 5 from the baseline in 2017 to 2019, but 2020 had seen an upturn in performance, bringing performance much closer to that seen in 2017. This performance then remained stable in 2021, despite the impact of school closures due to the pandemic, but appears to have declined in 2022 back to similar levels of performance as seen in 2019. At grade 7 and above, performance has been relatively consistent across the years, with a slight improvement in 2020 and 2021 and now a drop back to the level seen in the earlier years of the NRT.

A key question arising for the NRT results in a given year is to determine if differences in outcomes across the years are statistically significant. For the NRT, several comparisons could be made between different pairs of years at different grade boundaries, and this gives rise to the possibility that changes arising by chance may seem real. Hence, the criteria for significance that have been used are adjusted for multiple comparisons. For more information, see Appendix A.

The research question NFER was asked to address is to compare the performance in 2022 with the performance in 2020 at each of the three grade boundaries. Adjusting for three comparisons, the NRT English data shows that there are no statistically significant differences in performance between 2020 and 2022 at any of the three grade boundaries.²

5. Performance in maths in 2022

The objective of the NRT is to get precise estimates of the percentages of students each year achieving at a level equivalent to three key GCSE grades in 2017: these key grades are 4, 5 and 7. For the NRT in 2017, these baseline percentages were established from the 2017 GCSE population percentages. The NRT ability distribution, based on the IRT analysis, was then used to establish the ability scores which corresponded to those percentages. From 2018 onwards, the thresholds correspond to the same level of student ability as the thresholds established in 2017, thus allowing us to estimate the percentage of students above each of those thresholds and track performance over time. Alongside this, based on the sample achieved and the reliability of the tests, we are able to model the level of precision with which the proportion of students achieving the ability scores can be measured. The target for the NRT is to achieve a 95% confidence interval of plus or minus no more than 1.5 percentage points from the estimate at each ability threshold.

Ofqual provided the percentages of students at or above three relevant grades (grades 4, 5 and 7) taken from the 2017 GCSE population. These are shown in Table 5.1. These percentages were mapped to three ability threshold scores in the NRT in 2017.

² The results of a given year's NRT can be compared with the NRT results from a previous year (both are sample surveys, and the statistical error is therefore reflected in confidence intervals for each administration) or with the GCSE percentages of 2017, regarded as external constants. The *2018 Results Digest* reported comparisons with the GCSE 2017 population percentages. However, in order to make ongoing comparisons from year to year it was decided, for 2019 onwards, that comparing the outcomes between NRT studies (e.g. making statistical comparisons with the 2017 NRT study, rather than 2017 GCSE percentages) would be more informative.

Table 5.1. Maths 2017 NRT baseline thresholds

Threshold	Percentage of students above threshold from 2017 GCSE
Grade 7 and above	19.9
Grade 5 and above	49.7
Grade 4 and above	70.7

In 2022, the NRT data for the years 2017 to 2022 were analysed together using IRT modelling techniques. By analysing all the data concurrently, ability distributions could be produced for the samples for each year on the same scale. The percentages of students at each of the three GCSE grade boundaries, fixed on the 2017 distribution, could then be mapped onto the distributions for the subsequent years to produce estimates of the percentage of students at the same level of ability in those years. For example, the percentage of students at the ‘Grade 4 and above’ threshold in the 2017 GCSE population was 70.7 per cent. This was mapped onto the 2017 distribution to read off an ability value equivalent to that grade boundary. The same ability value on the distributions for all other years can then be found, and the percentage of students at this threshold or above in those years can be established. In this way, we are able to estimate the percentage of students at the same level of ability as represented in the 2017 GCSE population for each year of the NRT going forward. The precision of these estimates is dependent on both the sample achieved and the reliability of the tests as measures.

Table 5.2 presents the percentages of students achieving above the specified grade boundaries for the years 2017 to 2022. Confidence intervals for percentages are provided in brackets alongside the estimates. This is important as it shows that although there have been changes in performance, these are often within the confidence intervals. The statistical interpretation of the differences is discussed below.

Table 5.2. Estimated percentages at grade boundaries in maths

Year	Estimated percentages at Grade 4 and above	Estimated percentages at Grade 5 and above	Estimated percentages at Grade 7 and above
2017	70.7 (69.3-72.1)	49.7 (48.1-51.3)	19.9 (18.6-21.3)
2018	73.4 (71.9-74.9)	52.5 (50.8-54.1)	21.7 (20.3-23.0)
2019	73.2 (71.8-74.7)	51.9 (50.2-53.6)	23.0 (21.6-24.3)
2020	74.1 (72.6-75.5)	54.3 (52.9-55.8)	24.1 (22.7-25.4)
2021	69.6 (67.7-71.5)	49.0 (46.8-51.1)	21.1 (19.4-22.9)
2022	71.4 (69.9-72.8)	49.9 (48.4-51.4)	21.0 (19.7-22.4)

The 2017 figures in the table above are based on the NRT study, rather than the 2017 GCSE percentages. Note that, because of the way in which they have been computed, they match closely with the GCSE percentages. The confidence intervals for them reflect the fact that the NRT 2017 outcomes carry the statistical error inherent in a sample survey, as per the subsequent years.

Since the percentages for previous years have been re-estimated following the concurrent calibration with the 2022 data, these figures differ slightly from those reported in previous years. Some degree of variation is expected given the addition of more data, and the differences seen are small.

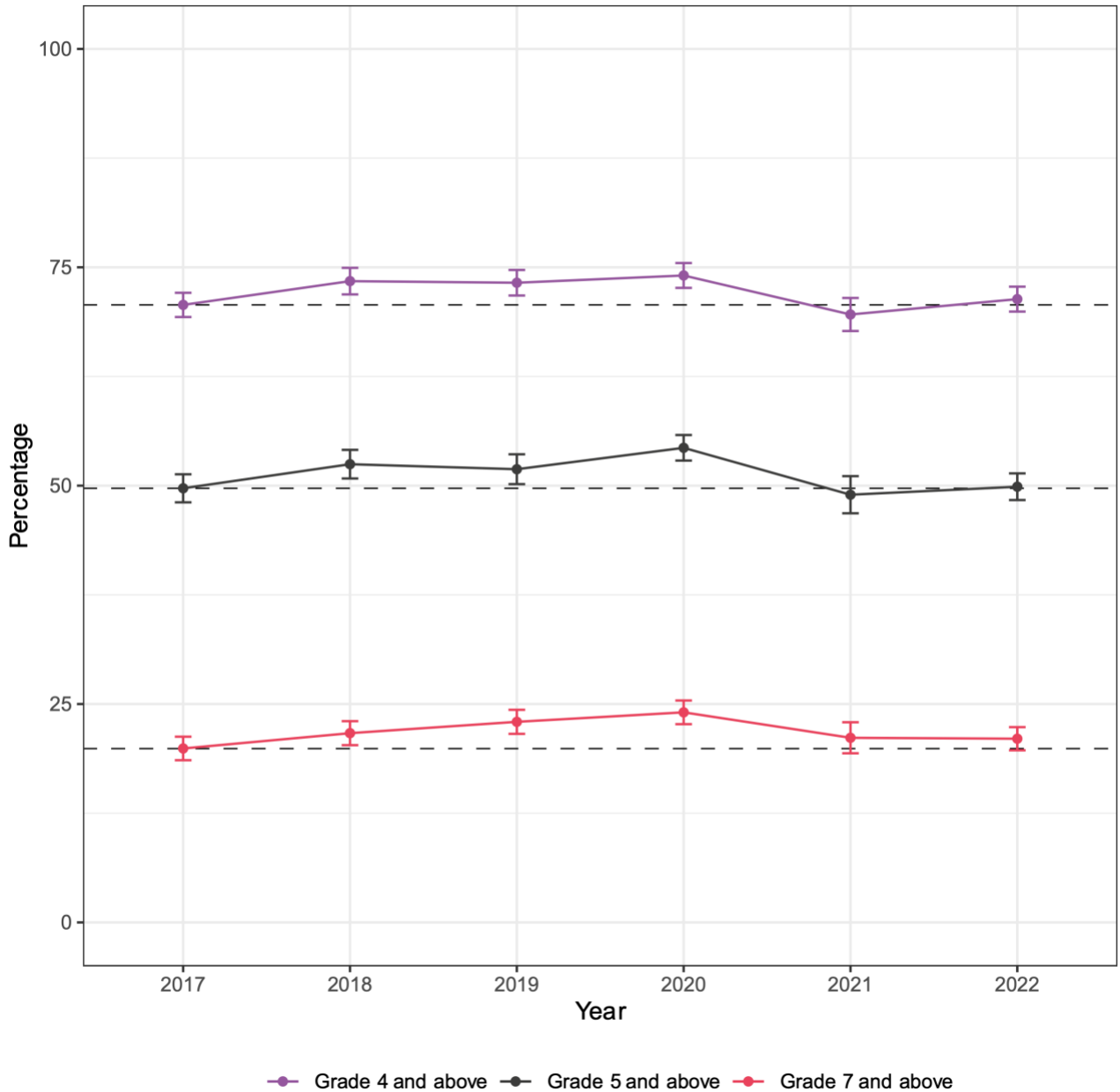
Table 5.3 shows the half widths of the confidence intervals. The confidence intervals for 2021 were wider than for previous years, reflecting the smaller sample size in that year. The precision this year is in line with earlier years prior to 2021. Any changes in the estimates of precision for the earlier years of the NRT due to the addition of 2022 data are minimal.

Table 5.3. Maths NRT half width of confidence intervals each year

Year	Half width of confidence intervals Grade 4 and above	Half width of confidence intervals Grade 5 and above	Half width of confidence intervals Grade 7 and above
2017	1.4	1.6	1.3
2018	1.5	1.6	1.4
2019	1.5	1.7	1.4
2020	1.4	1.5	1.4
2021	1.9	2.1	1.8
2022	1.4	1.5	1.3

Figure 5.1 presents 95% confidence intervals around the percentages achieving at least the specified grade boundary in 2022, as compared to previous years and the 2017 population baseline percentages. The 2017 population percentages are represented as dotted lines and the trend lines across years as solid lines. This format has been used to encourage the reader to compare the point estimate confidence bands for each year with the 2017 baseline population percentages, bearing in mind the confidence intervals.

Figure 5.1. Long term changes in NRT maths over time from 2017 baseline



The chart shows a relatively steady increase in the percentage of students achieving at-or-above all three grade boundaries from 2017 to 2020, followed by a sharp drop in 2021, back to around the 2017 level of performance. In 2022 we see a slight improvement at grades 4 and 5, suggesting some recovery, while performance at grade 7 and above is stable. A key question arising for the NRT results in a given year is to determine if differences in outcomes across the years are statistically significant. For the NRT, several comparisons could be made and this gives rise to the possibility that changes arising by chance may seem real. Hence, the criteria for significance that have been used are adjusted for multiple comparisons. For more information, see Appendix A.

The research question NFER was asked to address is to compare the performance in 2022 with the performance in 2020 at each of the three grade boundaries. Adjusting for three comparisons, the NRT maths data shows that there has been a statistically significant drop in performance between 2020 and 2022 at all three grade boundaries. The differences at grades 5 and 7 and above are significant at the 1% level of significance, whereas the difference at grade 4 and above is significant at the 5% level.³

³ The results of a given year's NRT can be compared with the NRT results from a previous year (both are sample surveys, and the statistical error is therefore reflected in confidence intervals for each administration) or with the GCSE percentages of 2017, regarded as external constants. The *2018 Results Digest* reported comparisons with the GCSE 2017 population percentages. However, in order to make ongoing comparisons from year to year it was decided for 2019 onwards that comparing the outcomes between NRT studies (e.g. making statistical comparisons with the 2017 NRT study, rather than 2017 GCSE percentages) would be more informative.

6. Appendix A: A brief summary of the NRT

English

The English test takes one hour to administer and follows the curriculum for the reformed GCSE in English language. In each of the eight English test booklets, there are two components; the first is a reading test and the second a writing test. Each component carries 25 marks and students are advised to spend broadly equal time on each component.

The reading test is based on an extract from a longer prose text, or two shorter extracts from different texts. Students are asked five, six or seven questions that refer to the extract(s). Some questions of one to four marks require short responses or require the student to select a response from options provided. In each booklet, the reading test also includes a 6-mark question and a 10-mark question, where longer, more in-depth responses need to be given. These focus on analysis and evaluation of aspects of the text or a comparison between texts.

The writing test is a single, 25-mark task. This is an extended piece of writing, responding to a stimulus. For example, students may be asked to describe, narrate, give and respond to information, argue, explain or instruct.

Maths

For maths, a separate sample of students is also given one hour to complete the test. The test includes questions on number, algebra, geometry and measures, ratio and proportion, and statistics and probability – the same curriculum as the reformed GCSE. Each of the eight test booklets has 13 or 14 questions with a total of 50 marks and each student takes just one of the test booklets.

Analysis

The analysis process followed a sequence of steps. Initially, the tests were analysed using Classical Test Theory to establish that they had performed well, with appropriate difficulty and good levels of reliability. The subsequent analyses used Item Response Theory techniques to link all the tests together from 2017 to 2022 and estimate the ability of all the students on a common scale for each subject for each year, independent of the test or items they had taken. These ability estimates were then used for calculating the ability level at the percentiles associated with the GCSE grade boundaries in 2017 and mapping these onto the distributions for subsequent years to generate percentile estimates for those years.

Multiple Comparisons

The statistical significance of the difference between two percentages estimated in two years, say 2020 and 2022, may be approached with a two-sample t -statistic. Because of the huge number of degrees of freedom, the value can be compared with the standard normal distribution rather than the t -distribution. For a comparison of two percentages, say the percentage of students at grade 4 or higher between two years, the critical value at a confidence level of 0.05 (5%) would usually be

1.96. However, since there are three grade thresholds across multiple years, there are several comparisons which could be made (up to 45 if all pairs of years were compared across all three grade boundaries). As the number of simultaneous comparisons grows, the probability that some of them are significant by chance rapidly increases. To guarantee that the chosen level of significance is guaranteed overall, we have implemented an adjustment for multiple comparisons.

Evidence for excellence in education

Public

© National Foundation for Educational Research 2022

All rights reserved. No part of this document may be reproduced or transmitted in any form or by any means, electronic, mechanical, photocopying, or otherwise, without prior written permission of NFER.

The Mere, Upton Park, Slough, Berks SL1 2DQ

T: +44 (0)1753 574123 • F: +44 (0)1753 691632 • enquiries@nfer.ac.uk

www.nfer.ac.uk

NFER ref. OFMT

ISBN. 978-1-912596-64-5