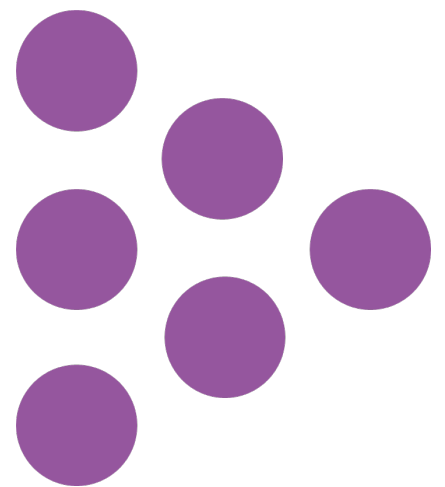# Report

National Reference Test 2021: Performance Across Content Areas

Research into the possible effects of Covid-19 disruption across different areas of the curriculum

National Foundation for Educational Research (NFER)

National Reference Test 2021: Performance Across Content Areas

Bethan Burge
Louise Benson
Alex Cutbill
Sarah Gibb
David Thomas
Jasper Wouda

Contents

Executive Summary

The NRT is uniquely placed, as a Key Stage 4 assessment which directly measures performance over time, to contribute to the evidence on the impact disruption due to the Covid-19 pandemic. The 2020 NRT administration was largely unaffected by the pandemic, as the testing window ended before schools were closed for the first national lockdown. We can assume, therefore, that the outcomes up to, and including, 2020 for the NRT provide a relatively precise and stable measure of performance of the population over time. Accordingly, the NRT outcomes in 2021 have the potential to provide useful insights into the way in which student performance may have been impacted by the pandemic.

The main NRT outcomes, reported in the 2021 Results Digest, show overall performance on the NRT in English and mathematics over time indicating how the subjects may have been impacted as a whole. This study provides additional analysis of the NRT 2020 and 2021 data to explore whether different areas of the curriculum have been differentially impacted by the disruption to teaching as a result of the Covid-19 pandemic. This may highlight areas that were more difficult to learn at a distance, were given a lower priority than others or required more explicit teaching of concepts and/or skills.

The 2021 NRT results indicated that there was no significant change in the overall performance in English for the 2021 cohort of students compared with the 2020 cohort. The absence of a global difference could mask relative changes for different parts of the curriculum (for example, improved performance in writing with decreased performance in reading). The analyses in this study compared relative performance between reading and writing and their composite assessment objectives. No evidence of any clear differences across the curriculum areas was found, suggesting that performance on the NRT in 2021 was comparable to that of 2020 across all areas of the English curriculum assessed in the NRT.

The 2021 NRT results indicated that there was a significant drop in performance in mathematics for the 2021 cohort of students compared with the 2020 cohort. This study investigates further whether this drop in performance is a result of larger decreases in performance in specific areas of the mathematics curriculum, or whether it reflects smaller decreases in performance across all curriculum areas. The evidence from these analyses indicates that performance fell by a similar amount for all areas of the curriculum in mathematics between 2020 and 2021 and that the disruption caused by the pandemic therefore did not have a differential impact across the mathematics curriculum.

As the extent of Covid-19 pandemic-related disruption was variable at the school, class and even student level, then the absence of a clear pattern in changes in performance at the curriculum level for English and mathematics is perhaps unsurprising.

1. Background

As an independent measure of performance over time of a nationally representative sample at GCSE level, the NRT outcomes have the potential to provide some useful insights on the impact of disruption to education due to the Covid-19 pandemic.

The NRT was designed to routinely report outcomes each year in terms of the proportions of students achieving at or above three key GCSE grades, defined as being equivalent to the ability required to achieve those grades in 2017 (the baseline year). The 2020 NRT administration was largely unaffected by Covid-19, as the testing window ended before schools were closed for the first national lockdown[1]. We can assume, therefore, that the outcomes up to, and including, 2020 for the NRT provide a relatively precise and stable measure of performance of the population over time. This should allow us to identify significant changes in performance in 2021 which are likely to be attributable to the impact of the Covid-19 pandemic[2]. The eight test booklets that make up the NRT had the same content in 2020 and 2021, which means that direct comparisons between the performances of the two representative samples are possible.

We might have expected that the performance of the 2021 sample would be lower than for the previous years given the possible loss of learning due to the Covid-19 pandemic. While the term 'learning loss' has attracted some criticism for its negative focus, it has been used to reflect the impact of the pandemic on students' learning and there is a commonly shared understanding of what it means. Here, it is used to refer specifically to a reduction in the level of attainment in different areas of the English and mathematics curriculum that are attributable to both direct and indirect impacts of the pandemic. This definition is consistent with that used by Ofqual in its review of literature on learning during the pandemic, published last year.

As reported in the National Reference Test 2021 Results Digest, there were, as expected, significant drops in the estimated proportions at or above the ability levels equivalent to grade 4, grade 5 and grade 7 for mathematics. For English, the outcomes were perhaps more surprising: performance was very similar, on average, in 2021 and 2020, with no significant differences at any of the three grade boundaries. Though not statistically significant, there was a slightly higher proportion of students at grade 7 and above in 2021 than in 2020 and a slightly lower proportion achieving grade 4 and above, indicating a slight widening of the ability distribution relative to previous years.

---

[1] NRT 2020 was held in February 2020, before the Covid-related closure of schools in late March 2020.
[2] NRT 2021 was held in April 2021, after schools returned following school closures in spring 2021.

Overview of NRT

The NRT is a short test which reflects the sorts of questions that are in the GCSE examinations for English language and mathematics, and the questions largely remain the same each year. It is taken by a representative sample of students in year 11 who are taking their GCSEs in the same academic year. The NRT is designed to be accessible to all eligible students. The questions in the English and mathematics tests are aimed mostly at grades 3 to 7 but also include some questions aimed at grades 1 and 2 and grades 8–9 to help make them accessible and / or challenging to all students.

For each subject, the questions are arranged into eight test booklets of 50 marks. Test questions are grouped into blocks and each test booklet is made up of two blocks of questions. The blocks are organised in an overlapping design such that each block, and therefore each question, appears in two test booklets. This design provides a broad coverage of the curriculum and good measurement of the national population, while minimising the time each student is tested. Each student has one hour to complete either one English or one mathematics booklet.

The English NRT comprises a reading and a writing component. Each component is marked out of 25 marks and students are advised to spend broadly equal time on each. The majority of the marks for reading and all of the marks for writing are for extended response questions. The remaining marks on each reading test are for short-answer responses. The English NRT covers all of the GCSE assessment objectives for reading and writing (excluding those for spoken language).

The mathematics NRT covers the full range of subject content for the GCSE specification (number, algebra, ratio and proportion, geometry and measures, and statistics and probability) and the assessment objectives (using and applying standard techniques; reason, interpret and communicate mathematically; solve problems within mathematics and in other contexts). It is not a tiered assessment. Most of the questions are based on the GCSE content that all GCSE mathematics students have to study and includes some more straightforward questions at the start to help encourage students. A few questions drawing on higher tier content are at the end of each test booklet.

More information on the NRT can be found in the [General Report](#)

The NRT was constructed as a test to report on the performance of a student cohort, in English language and mathematics, between years. During test construction, the curriculum characteristics for every question in the NRT were recorded to ensure that the curriculum balance of the NRT matched that of the GCSEs to allow for a valid comparison to GCSE performance. For the purposes of this study, we can use these designations to explore changes in performance across the different areas of English and mathematics between 2020 and 2021. Specifically, these are the different assessment objectives relating to reading and writing for the English curriculum and the subject content areas and assessment objectives for the mathematics curriculum.

It should be recognised that for the NRT 2021 cohort, both years of Key Stage 4 have been defined by high levels of disruption, including long periods of school closures during the two national lockdowns as well as periods where Covid-19 cases in bubbles or schools resulted in some students spending additional periods of time away from the classroom. However, it is also

important to recognise that since the start of the Covid-19 pandemic this particular cohort, compared with other year groups, have had extended periods of time in school participating in face-to-face lessons and potentially catching up on learning disrupted earlier in the pandemic. This complex picture highlights the fact that individual students, cohorts and schools may have experienced different levels of disruption caused by the pandemic.

As an 'exam' year, the NRT 2021 cohort were one of the priority year groups to return to school at the end of the first national lockdown in June 2020. Although this meant that these students may have had an additional six weeks of face-to-face teaching compared with other year groups, provision of lessons and student attendance may not have been consistent across the cohort. The 2020/21 academic year began relatively normally with all students returning to school at the start of the autumn 2020 term. This normality was short-lived in some regions however, as many students experienced further disruption during this period with bubble and year group closures, as well as time away from schools for self-isolation due to high Covid-19 rates. The second lockdown, which began in January 2021, saw most students experiencing another long period of remote learning and, for the NRT 2021 cohort, the announcement that formal GCSE examinations would not take place that summer. The vast majority of students returned to school in March 2021 and therefore the NRT students spent four weeks participating in face-to-face lessons before the NRT 2021 was administered. However, the cancellation of formal GCSE exams may have meant that these students were working in a different way or completing different sorts of assignments (for example internal assessments) in the weeks preceding the NRT administration. Therefore, in some respects, these students may have been more 'test ready' compared with the students who participated in the NRT in 2020. However, they may also have been less motivated as they were not preparing for formal examinations, as would normally be the case. This, as well as the two-month delay to the administration of the NRT 2021, may have impacted student performance.

It is not possible to say definitively how students' learning experiences during this period may help to explain the findings from this additional analysis. However, Ofqual have undertaken further analysis of the contextual information collected from schools and students during the administration of the 2021 NRT. The findings from Ofqual's analysis may provide some useful context in the interpretation of the findings from this study.

Interpreting the data

The following points should be borne in mind when comparing student performance on the NRT in 2020 and 2021 and interpreting how these differences reflect Covid-19 disruption:

- The administration of the NRT in 2021 took place approximately two months later than would usually be the case, due to the closure of schools to most students in the spring term of 2021.
- The administration period was relatively soon after the return to school for most students, so some students may still have been adjusting to being back in school.
- GCSE examinations had been cancelled prior to the administration of the NRT in 2021; this was not the case for the 2020 administration, even though they were later cancelled.
- The administration took place at a time when teachers were making judgements for their teacher-assessed grades for GCSEs, meaning some students may have been more 'test ready' than they would ordinarily be.

2. Are different areas of the curriculum impacted differentially?

The aim of this research study was to see whether there is evidence from the NRT that different areas of the curriculum have been impacted by the pandemic.

The main NRT outcomes show the overall performance for the NRT over time in English and mathematics and gives some indication as to how performance has changed on each subject as a whole. The additional analysis in this report aims to show if there are areas of the curriculum which have been impacted more by the pandemic than others. This may highlight areas that were more difficult to learn at a distance, were given a lower priority than others or required more explicit teaching of concepts and/or skills.

The NRT has been designed to provide a global quantitative measure of performance in both English and mathematics. In order to provide a valid comparison with the GCSEs, the NRT has been constructed to provide a similar distribution of assessment objectives and subject content. Therefore, within the NRT we can analyse the performance of students in the different specified areas of the curriculum included in the NRT and compare these to the performance in the 2020 NRT administration. Given that the NRT in 2021 had the same questions as the NRT in 2020, it provides a valid measure of differential performance between cohorts for the different curriculum areas.

The English NRT is divided into the two main subject areas of reading and writing. Each of these subject areas is worth 50 per cent of the marks of the English NRT. An initial comparison can be made between the performance of the 2020 cohort and of the 2021 cohort in these two subject areas. This comparison will test the hypothesis that the students' performance in reading and writing has been impacted by the pandemic.

Within these subject areas, more granular detail of potential differences can be derived from looking at the different assessment objectives. The English curriculum is divided into six assessment objectives (AOs). Four of these (AO1–4) refer to reading skills and two to writing skills (AO5–6). In both cases there is a progression of skills (see Table 2.1) through these AOs which means that analysis of them may identify relative changes in the performance of students related to the level of demand of the items.

Each AO for English is represented by at least 20 marks in the NRT so there are sufficient marks for differences in performance of all AOs to be identified. An analysis comparing the change in performance of the English AOs in the 2021 cohort will test the hypothesis that some of the skills may have been differentially impacted by the pandemic. For example, it is plausible that the available time had been focussed on identifying and explaining language in texts and proportionally less time to the higher order skills of comparing and evaluating.

Table 2.1   Assessment objective coverage for English

Assessment objective	Description	Number of marks in the NRT
AO1	Identify and interpret explicit and implicit information and ideas. Select and synthesise evidence from different texts.	20
AO2	Explain, comment on and analyse how writers use language and structure to achieve effects and influence readers, using relevant subject terminology to support their views.	40
AO3	Compare writers' ideas and perspectives, as well as how these are conveyed, across two or more texts.	20
AO4	Evaluate texts critically and support this with appropriate textual references.	20
AO5	Communicate clearly, effectively and imaginatively, selecting and adapting tone, style and register for different forms, purposes and audiences. Organise information and ideas, using structural and grammatical features to support coherence and cohesion of texts.	60
AO6	Candidates must use a range of vocabulary and sentence structures for clarity, purpose and effect, with accurate spelling and punctuation.	40

The mathematics curriculum, and the structure of the NRT, covers five main subject areas as shown in Table 2.2. It is possible that, as a result of the disruption to teaching caused by the pandemic, there may have been a change in emphasis given to the teaching of different subject areas compared with previous years. For example, there may have been an emphasis on the core subject area of number at the expense of a more involved subject area such as algebra. The NRT data allows us to investigate the hypothesis that disruption has differentially impacted these subject areas in mathematics.

Table 2.2   Subject content area coverage for mathematics[3]

Mathematics subject content area	Number of marks associated with items for each subject content area in the NRT	Number of individual marks in the NRT
Number	39	46
Algebra	53	49
Ratio, proportion and rates of change	34	38
Geometry and measures	46	39
Probability and statistics[4]	28	28

As with English, the mathematics skills are also assessed using an AO for both GCSE and NRT. The distribution of marks awarded for each AO in the NRT is presented in Table 2.3.

Table 2.3   Assessment objective coverage for mathematics

Assessment objective	Description	Number of marks in the NRT
AO1	Use and apply standard techniques	85
AO2	Reason, interpret and communicate mathematically	59
AO3	Solve problems within mathematics and in other contexts	56

---

[3] Each item is written with a specific subject intent, for example, to cover algebra. This distribution is reported as the number of marks associated with items for each subject area. However, each mark that makes up an item will also have a specific subject statement allocated for the specific requirement of that step in the question and this may differ from the subject area for the item. This is reported as the number of individual marks in the NRT. The latter provides a more accurate picture of the balance of subject content across the NRT as a whole.

[4] Probability and statistics are separate areas in the curriculum but are considered together for the purposes of the subject content coverage. In some of the analyses in this study the performance of probability and statistics are considered separately.

3. Sample

The achieved sample sizes for the NRT for the years in this study are shown in Table 3.1. Fewer schools took part in the NRT in 2021 than in other years following the postponement of the administration period to the summer term.

Table 3.4   School and student achieved samples NRT 2020 and 2021

Achieved sample	2020	2021
English achieved school sample	332	214
English achieved student sample	6,639	4,030
Mathematics achieved school sample	333	216
Mathematics achieved student sample	6,756	4,143

A bias analysis was carried out to check that the reduced sample was still representative in terms of the stratifier used to select the NRT sample: GCSE performance at school level in 2019. A number of other school factors which could have affected performance due to the impact of the pandemic were also investigated. Therefore, the sample has been broken down further by the proportion of students eligible for free school meals (FSM), geographical region and absence rates. While the data shows that the NRT samples do not perfectly represent schools in terms of the proportion of students with FSM and geographical region, they do indicate that there has not been a major change compared with 2020. Analysis of the absence rates during the autumn term 2020 found that there were no significant differences in the absence rates during this term between those schools that participated in NRT 2021 and those schools which were invited but were unable to participate due to the pandemic. Therefore, the evidence suggests that despite the smaller sample size, comparing performance in 2020 and 2021 can still be considered valid. However, when interpreting the findings of this study it should be borne in mind that there is no direct measure available of how individual schools coped with the disruption of the pandemic, and therefore we cannot rule out the possibility that the schools that struggled more with the impact may have been more likely to withdraw from the NRT.

The number of students completing each of the NRT booklets for English and mathematics is presented in the tables 3.2 and 3.3. The fewer completed booklets in 2021 reflects the fact that fewer schools participated in the NRT in 2021. However, the design of the assessments means that the number of students completing each booklet is similar within each year.

Table 3.5   Number of completed test booklets in English

Booklet	Number of students completed tests NRT 2020	Number of students completed tests NRT 2021
Booklet 1	834	492
Booklet 2	838	499
Booklet 3	827	495
Booklet 4	835	507
Booklet 5	840	505
Booklet 6	832	507
Booklet 7	819	518
Booklet 8	814	507
Total	**6639**	**4030**

Table 3.6   Number of completed test booklets in mathematics

Booklet	Number of students completed tests NRT 2020	Number of students completed tests NRT 2021
Booklet 1	844	513
Booklet 2	832	518
Booklet 3	841	513
Booklet 4	866	520
Booklet 5	841	526
Booklet 6	839	514
Booklet 7	857	513
Booklet 8	836	526
Total	**6756**	**4143**

4. Methodology

All of the data presented in this analysis are derived from Classical Test Theory (CTT) analysis which is based on the raw scores achieved by students from items for the different curriculum areas. For the main analysis of the NRT, Item Response Theory (IRT) is the preferred methodology for comparing the performance of different cohorts of students between years. Given the connected design of the NRT booklets, IRT allows the performance of all students to be compared in a meaningful way, irrespective of the specific items that they sat. For the performance of items themselves, CTT provides a more direct comparison and is the method used to compare item performance between years. As the items in the booklets were the same in 2020 and 2021 then they can be considered as parallel tests and as such, CTT analysis provides a valid comparison between years. Given the validity of this approach and the more straightforward interpretation of CTT statistics for item performance, this approach has been taken in this study. Further information on the technical terms included in the methodology can be found in the Glossary of this report.

A summary of the analyses presented in this study is provided below.

1. **Profile plots**. Profile plots (Bechger and Partchev, 2021) are very useful for investigating whether two groups of students obtain the same test score in the same way. The profile plot maps the most likely combination of scores on one curriculum area compared with the rest of the test for each overall total score on the booklet. The plots are presented at the booklet level so that data for each curriculum area based on the same set of students (that is, the ones that completed that booklet) are not complicated by student specific factors. The profile plots are presented together for both the 2020 and 2021 cohorts. Any large differences in performance in a curriculum area would be noted as a deviation of the plots for the two years. A consistent deviation would be noted as occurring for the same curriculum area in the different booklets.

2. **Booklet scores**. These graphs present the proportion of the total booklet scores allocated to each curriculum area. The curriculum areas are presented in different colours, and graphs are presented as cumulative bar graphs. As with the profile plots, the graphs are presented at the booklet level so that student-specific factors do not form part of the analysis. Note also that the graphs are presented proportionally so that differences in the total scores per booklet between years are not presented. This format allows for a comparison of the relative proportion of scores allocated to each curriculum area between years.

3. **Facilities at the item level**. The facilities show the proportion of students who answered a particular item correctly. These graphs present the facility for each curriculum area in each booklet between 2020 and 2021. This provides a more obvious way of comparing the direction of change in facilities for curriculum areas than the booklet score graphs.

4. **Item facility plots for English**. These graphs present the facilities for items in the analysis in 2020 compared with 2021. The items are separated with each curriculum area plotted on a separate graph. This analysis presents global information on the changes in facilities between these years rather than just at the booklet level. The graphs include a diagonal line indicating equivalence of scores between the years. Items where the facility increased from 2020 to 2021 are plotted above this line and items where the facility decreased are plotted below the line. The data from these plots are provided in Appendix B1 Table B1 and were used for statistical

t-testing for significant differences in assessment objective performance between 2020 and 2021 (Table 5.1).

5. **Sub-item facilities for mathematics**. For mathematics, the items are marked based on achievement of students on each individual mark (of the 200-mark total). As each mark has a designated subject content area and assessment objective, then analysis at the sub-item level allows a more accurate analysis of the different curriculum areas than at the item level. This is because most items in mathematics include a mix of different subject areas and assessment objectives, which does not occur for English. In order to analyse performance across the different curriculum areas between the years, the change in facility between 2020 and 2021 is averaged for all subject areas. To account for changes in the relative difficulty of items in different curriculum areas, proportional changes to facilities (change in proportion correct divided by 2020 proportion correct) are also compared. The data used for these analyses are provided in Appendix B2 Table B4 and the results of the analysis are presented in Table 5.2. These data were also used for statistical t-testing for significant differences in assessment objective performance between 2020 and 2021 (Table 5.3).

6. **Sub-item facility rankings for mathematics**. The sub-items for 2020 and 2021 were ranked in order from the lowest facility (most difficult item) to the highest facility (easiest item). This provides 200 data points for each year. The change in ranking for each sub-item was calculated and these were averaged across the different areas of the curriculum. This allows the relative change in performance for each area of the curriculum to be measured against changes in the other areas. A negative number indicates that the curriculum area has performed comparatively less well than the other curriculum areas. The analysis of facility rankings is provided in Table 5.2.

5. Outcomes

5.1. English

Reading versus writing: profile plots

Initial analysis for English was to look for any evidence of specific changes in relative performance for items targeting reading skills compared with items that target writing skills. This represents the main division in English booklets and is a valid comparison to make as each component makes up half of the marks for each booklet.

The profile plots map the most likely scores for the reading component of each booklet (horizontal axis) against the most likely scores for the writing component (vertical axis) for students with the same total score (the diagonal lines). The reading vs writing profile plot for booklet 1 is shown in Figure 1. The 2020 cohort is presented by the red line and the 2021 cohort by the blue line. We can see that, for example, for a total score of ten (the diagonal line ending with a small number 10), the 2020 sample were most likely to score four marks on reading and six marks on writing; whereas for students in the 2021 sample also scoring ten in total, the most likely combination of scores was six on reading and four on writing. The equivalent profile plots for all booklets are presented in Appendix A1 Figure A1. The plots for some of the booklets show a similar pattern of performance as that seen in the booklet 1 plot, that is for the lower half of the score range the 2020 population seems to favour writing and 2021 seems to favour reading. However, these differences are small and they are not consistent across the booklets. Therefore, there is no evidence of a consistent difference in performance between reading and writing in the NRT between 2020 and 2021.

Figure 1   Profile plot of scores for reading versus writing scores for booklet 1
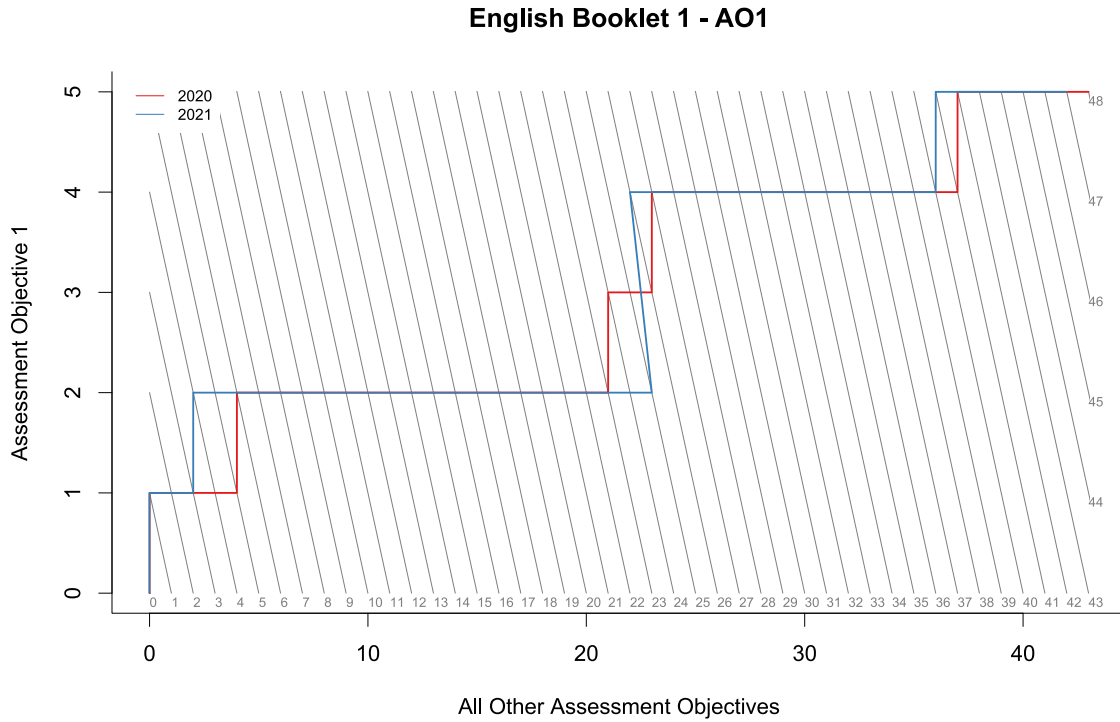
English Booklet 1
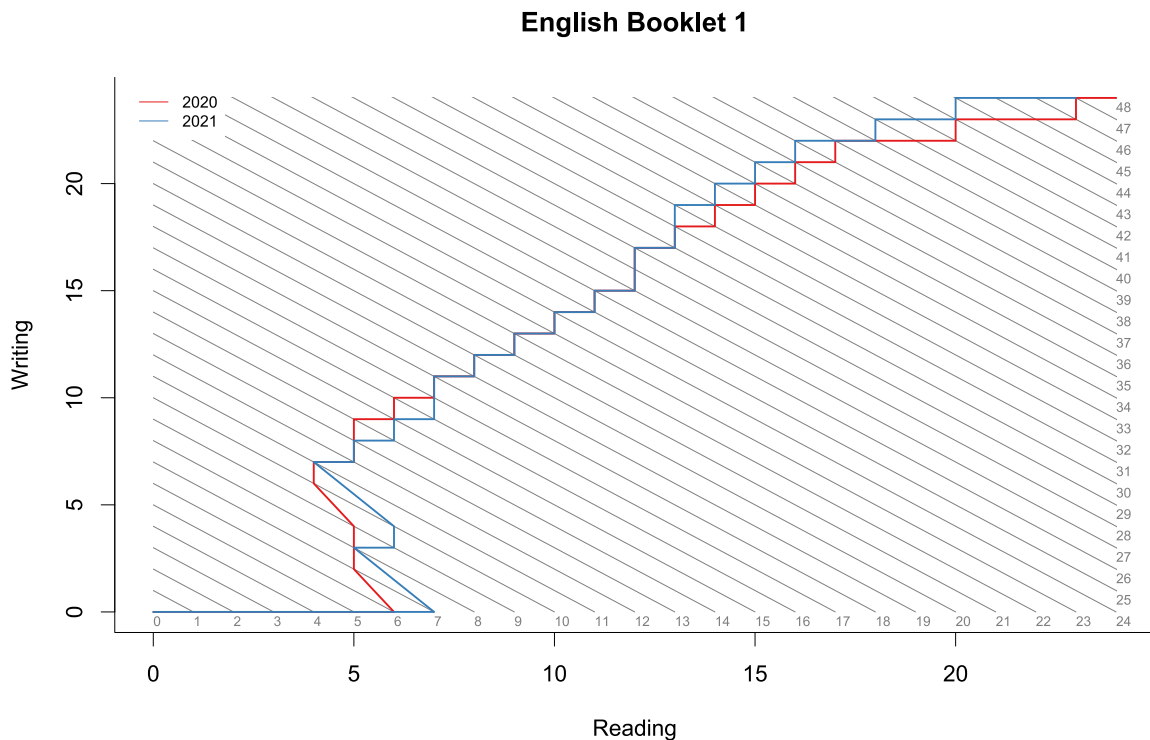
Assessment objectives: profile plots

The performance in reading and writing for English can be further broken down by AO with AO 1–4 denoting reading skills and AO 5–6 denoting writing skills (Table 2.1). Profile plot analysis was also performed on the different AOs with the score for each AO plotted against the score for the remainder of the booklet.

The profile plots for the reading AOs for booklet 1 are shown in Figure 2. The profile plots for the writing AOs for booklet 1 are shown in Figure 3. The equivalent profile plots for all booklets are presented in Appendix A1 Figures A2 to A7

Although there are minor differences in the profile plots across different AOs, these differences are small and not at all consistent across booklets, suggesting that there are no clear patterns for differences in the way that students obtained these scores between 2020 and 2021. This observation applies to all the English AOs. As such there is no evidence of a clear change in performance of the English AOs from 2020 to 2021.

Figure 2   Profile plot for reading assessment objectives for booklet 1[5]

English Booklet 1 - AO1

English Booklet 1 - AO2

---
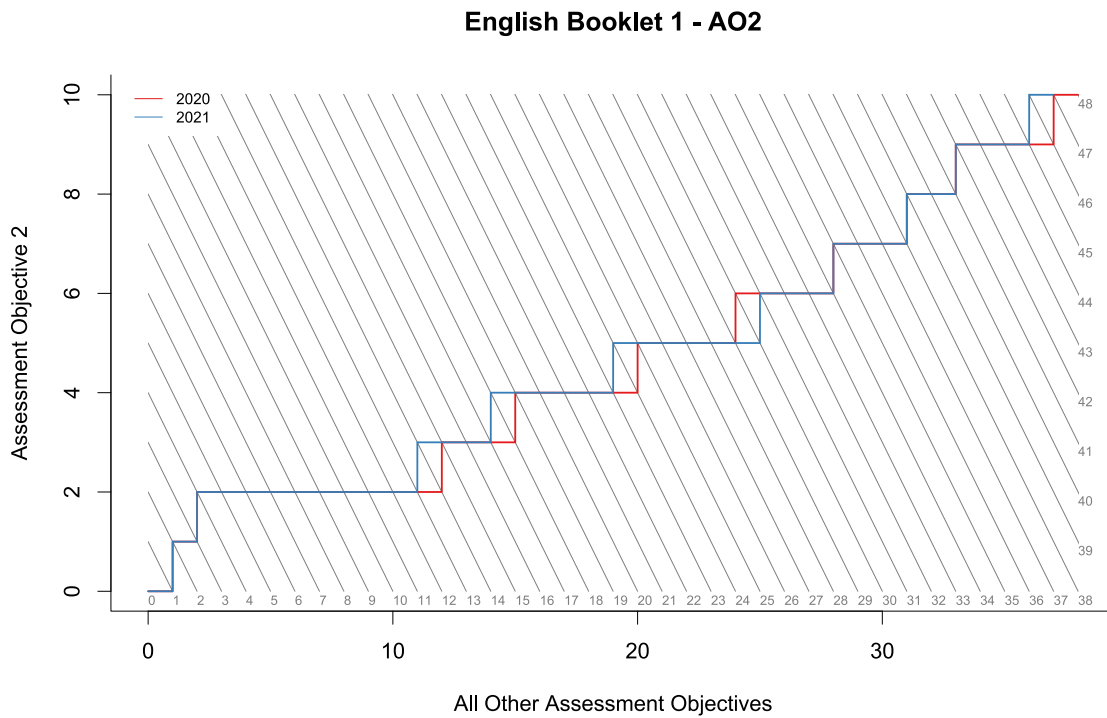
[5] All booklets contain marks for AO1 and AO2. Booklets 1,6,7 and 8 include marks that cover AO3 skills. Booklets 2,3,4 and 5 include marks that cover AO4 skills.

English Booklet 1 - AO3
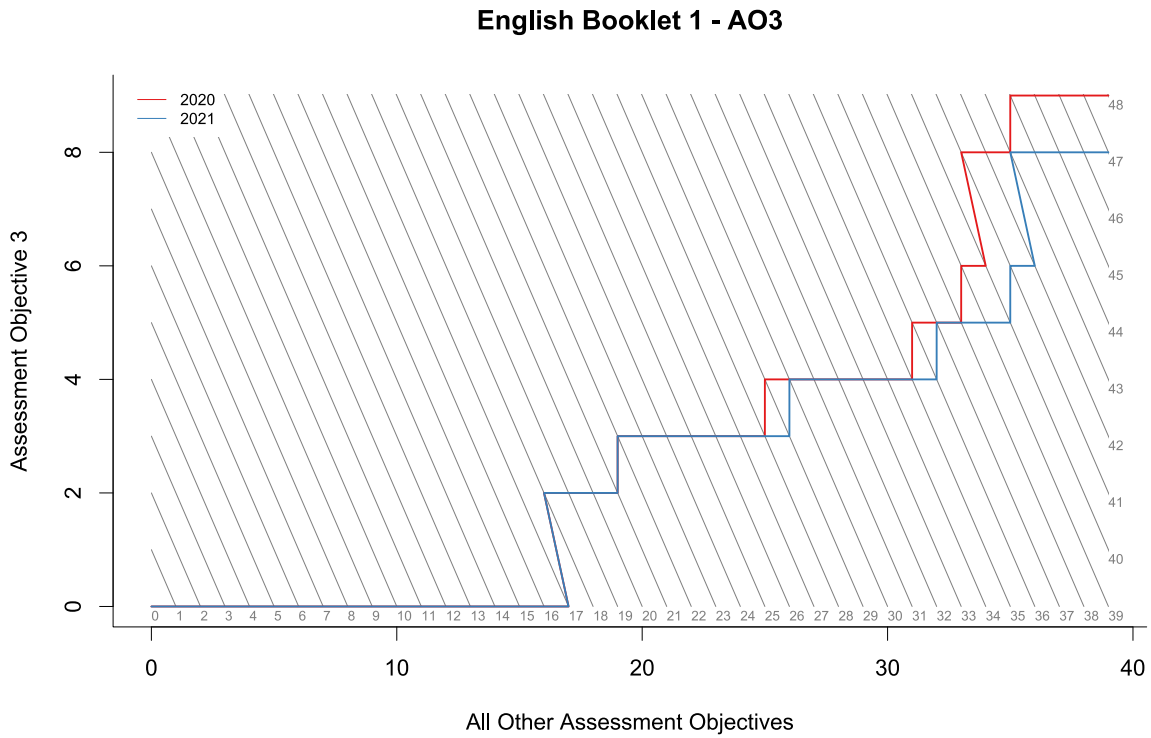
Figure 3   Profile plot for writing assessment objectives for booklet 1

English Booklet 1 - AO5
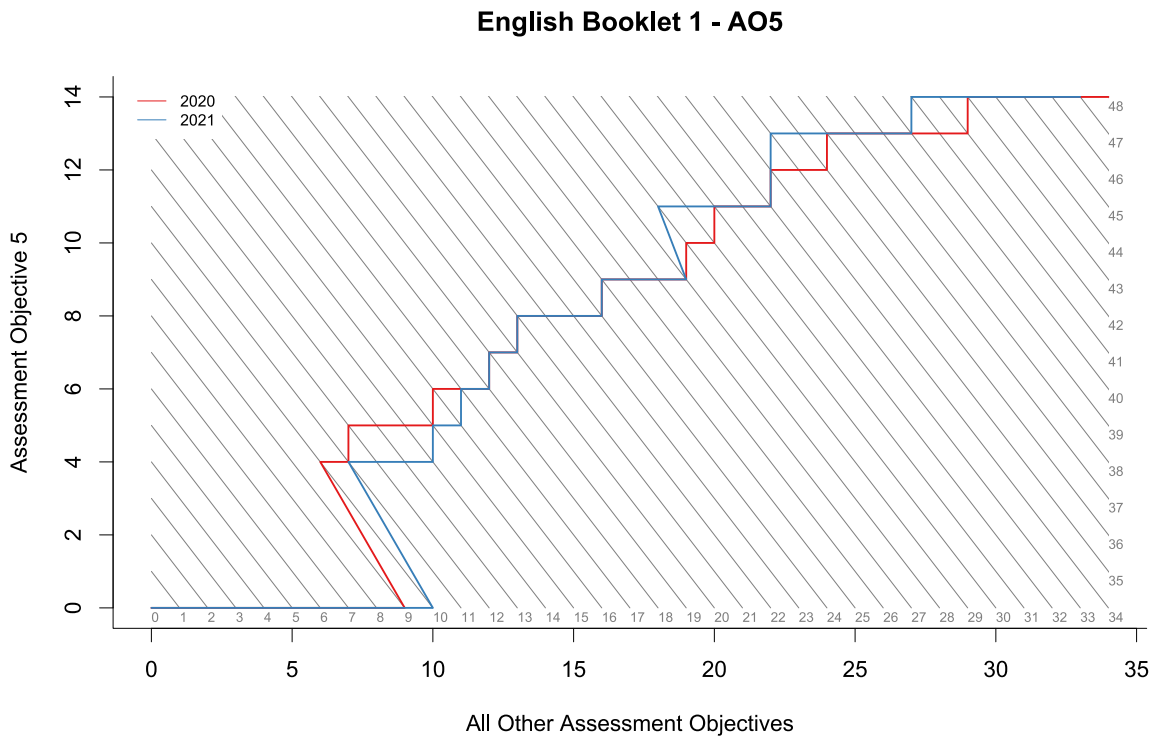
English Booklet 1 - AO6
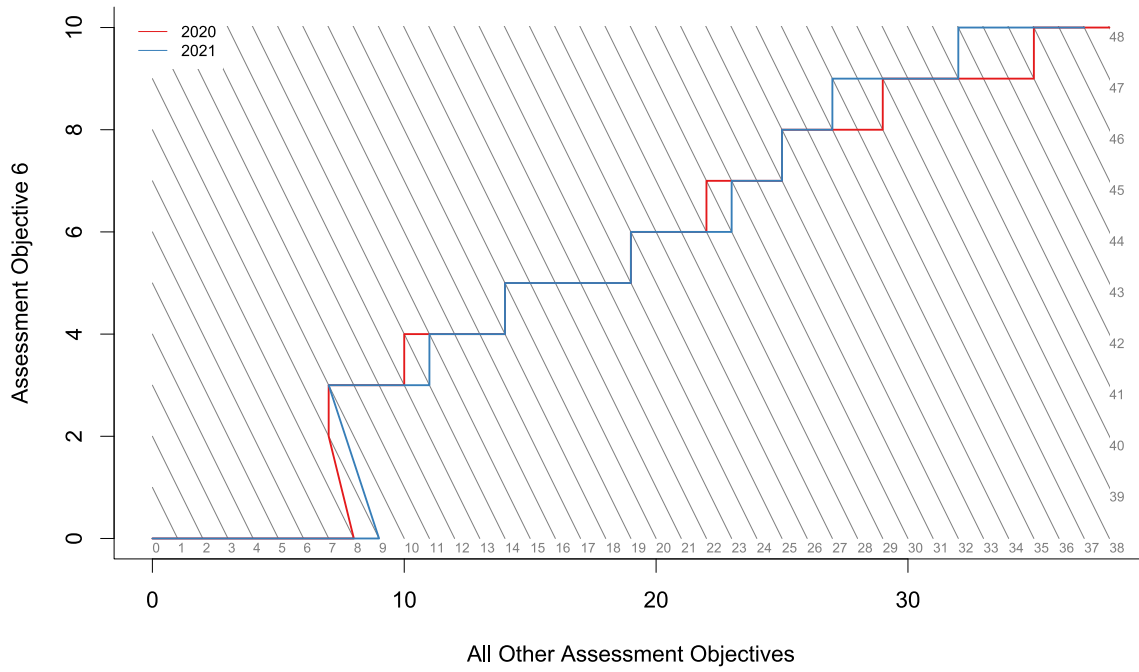
Assessment objectives: proportions, facilities and mean scores

As the booklets for 2020 and 2021 contained the same items, we can compare the scores within those booklets directly. Figure 4 shows small differences for some proportions of certain AOs between 2020 and 2021. However, these differences are not substantial and never consistently favour a certain year.

Figure 4 Proportions of booklet scores by assessment objectives for 2020 and 2021
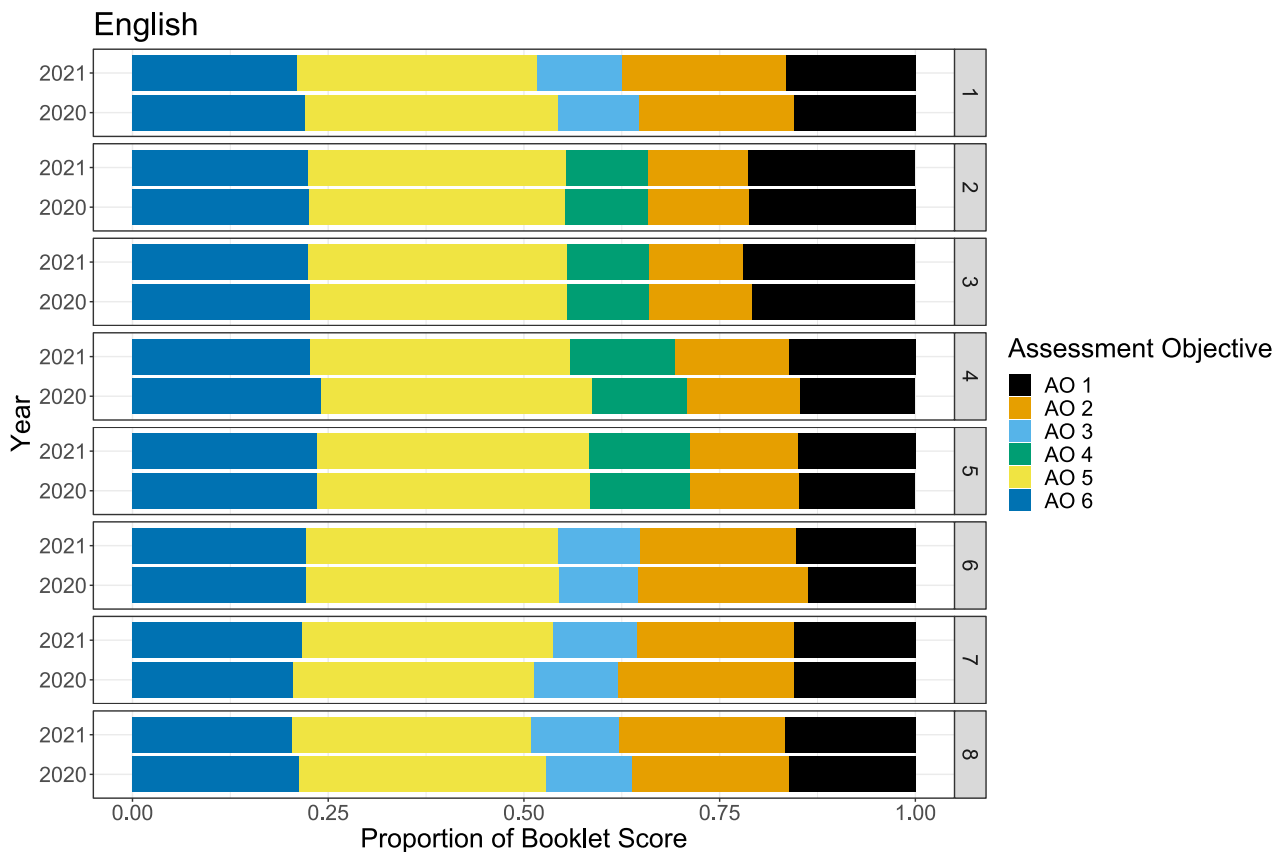
This same observation is true for the mean facilities of the different AOs in 2020 and 2021. The full overview of the graphs is in Appendix B1 Figure B1, but we can give an example in Figure 5. For instance, the data for booklet 1 show a slight increase for AO2, but for booklet 3 there was a decrease for AO2.

Figure 5   Mean facilities for assessment objectives for booklets 1 and 3

Booklet 1 English
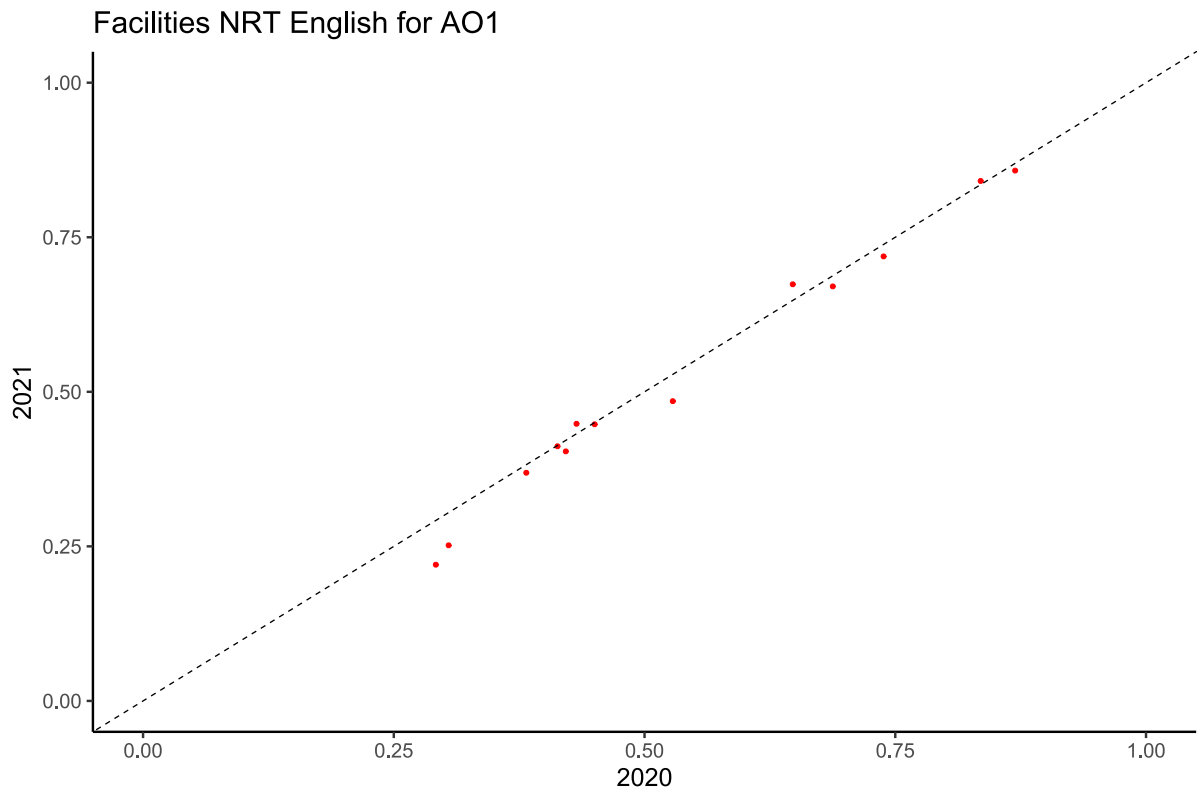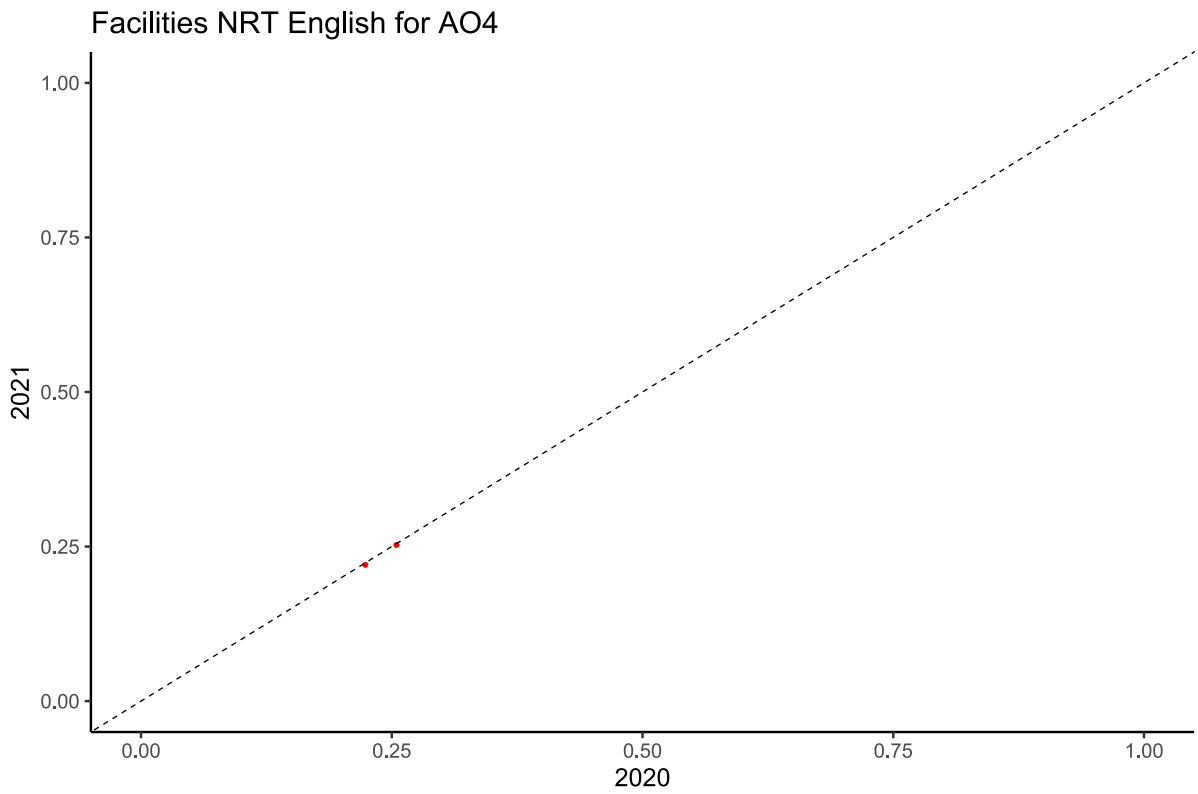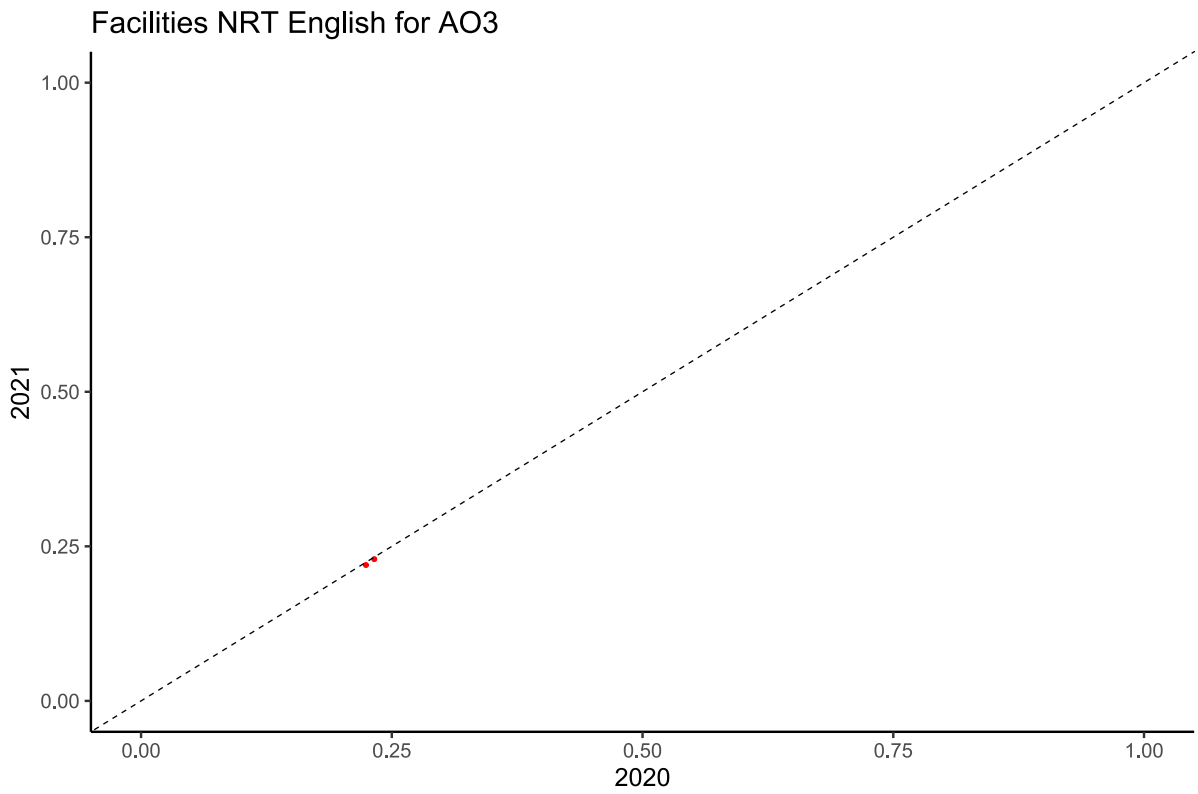

Booklet 3 English
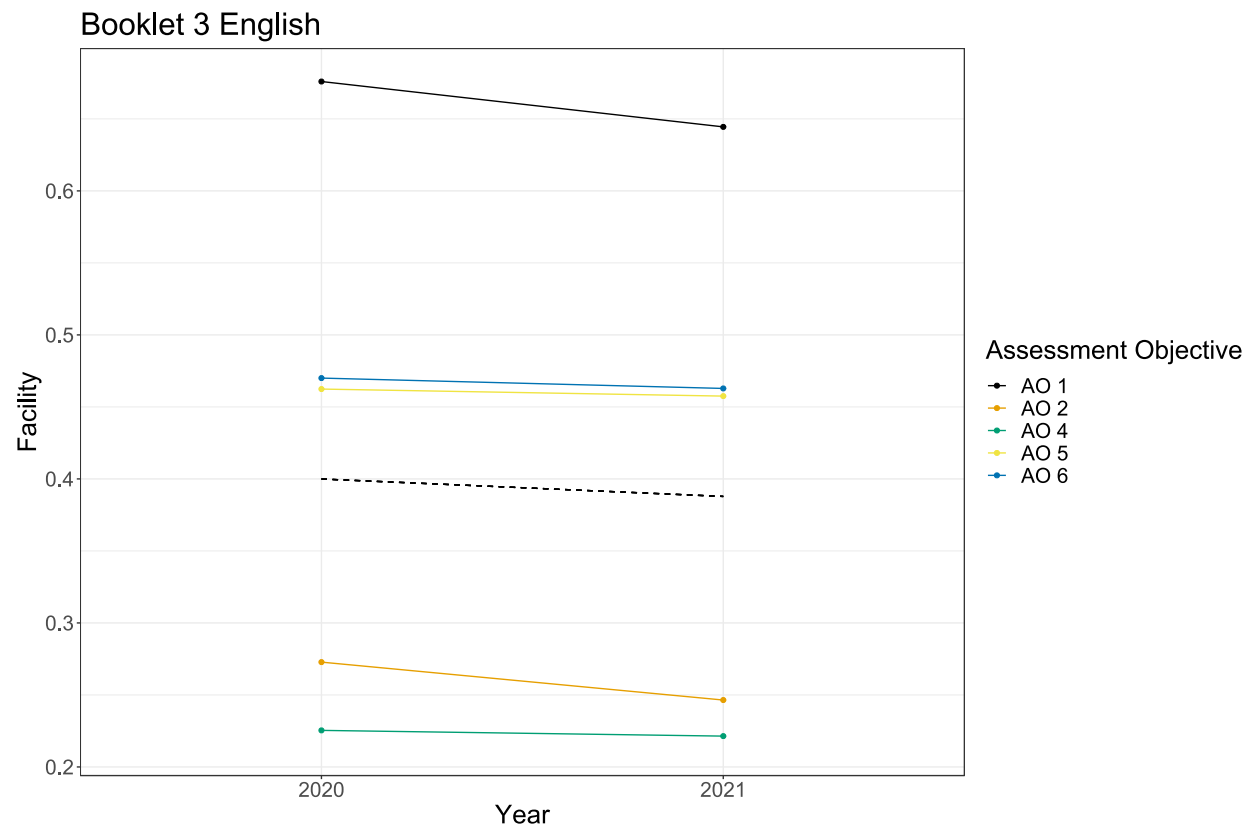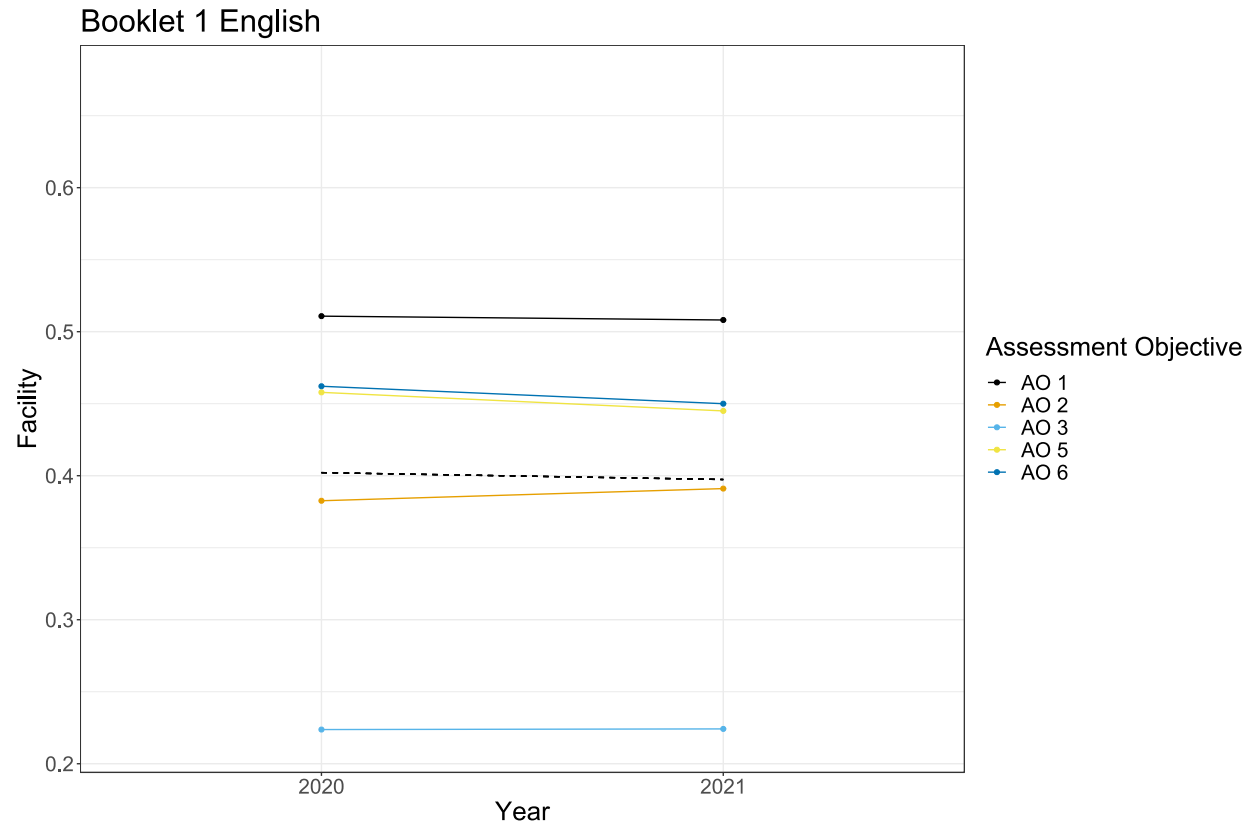
In addition, Table B1 in Appendix B1, with mean scores and standard deviations for the different AOs, does not show substantial differences.
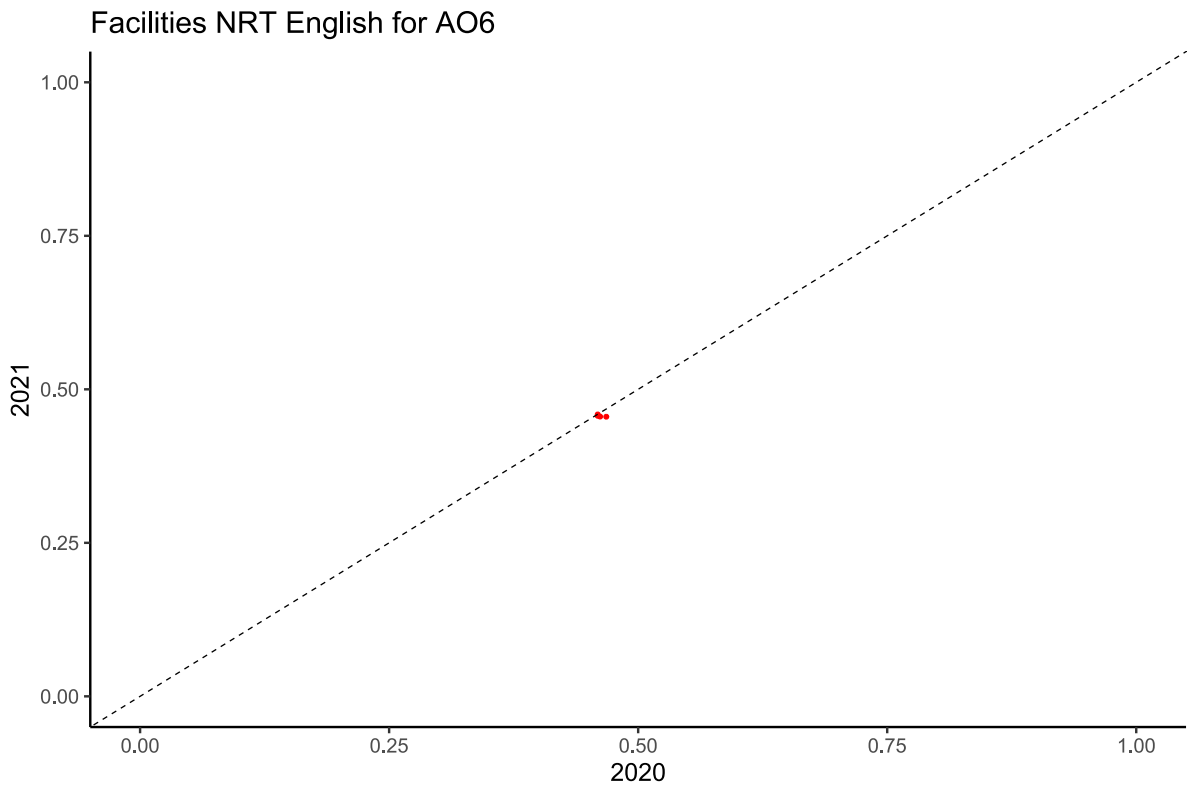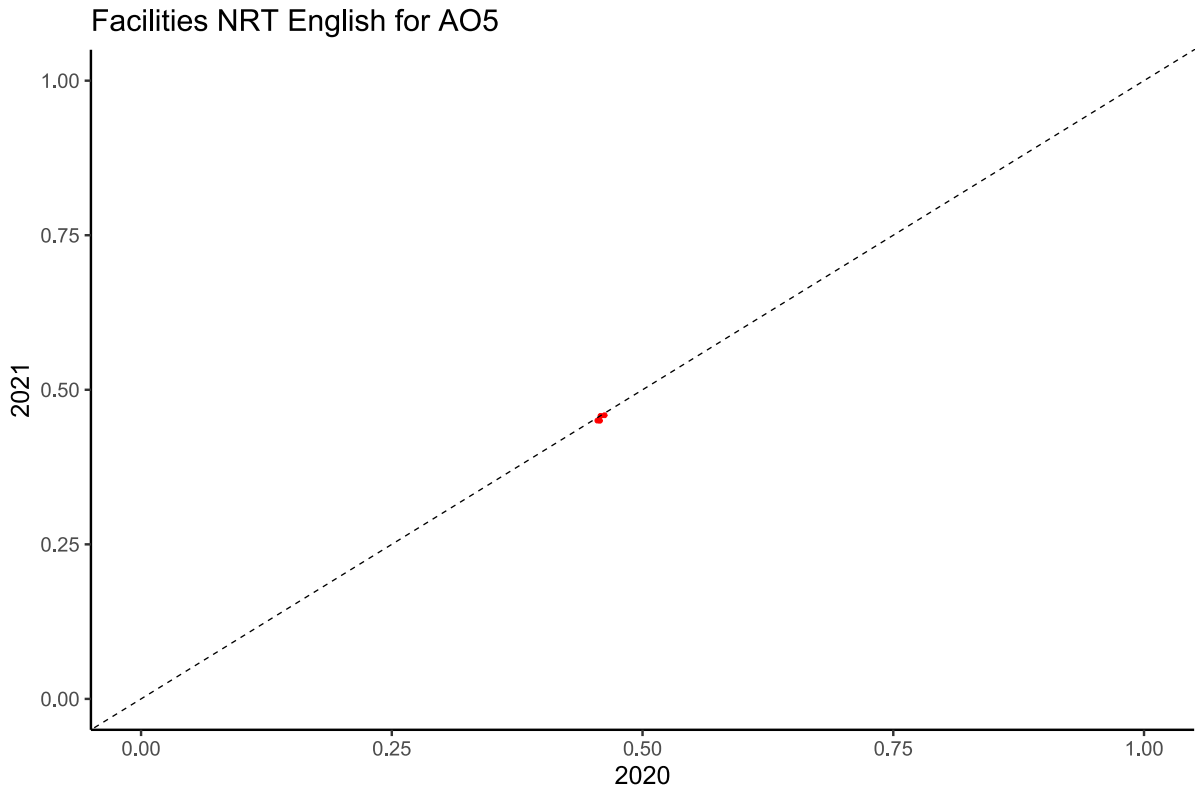
Item facilities comparison plots

Another way to compare the relative performance of AOs between years is to look at the changes in facility for each item covering that AO between 2020 and 2021. This provides a more global analysis than the profile plots as it includes all of the items from the different booklets together. This data is presented as scatter plots in Figure 6.

Figure 6   Item facilities by assessment objective for English comparing 2020 to 2021

Facilities NRT English for AO1

Facilities NRT English for AO2

Facilities NRT English for AO3

Facilities NRT English for AO4

Facilities NRT English for AO5

Facilities NRT English for AO6
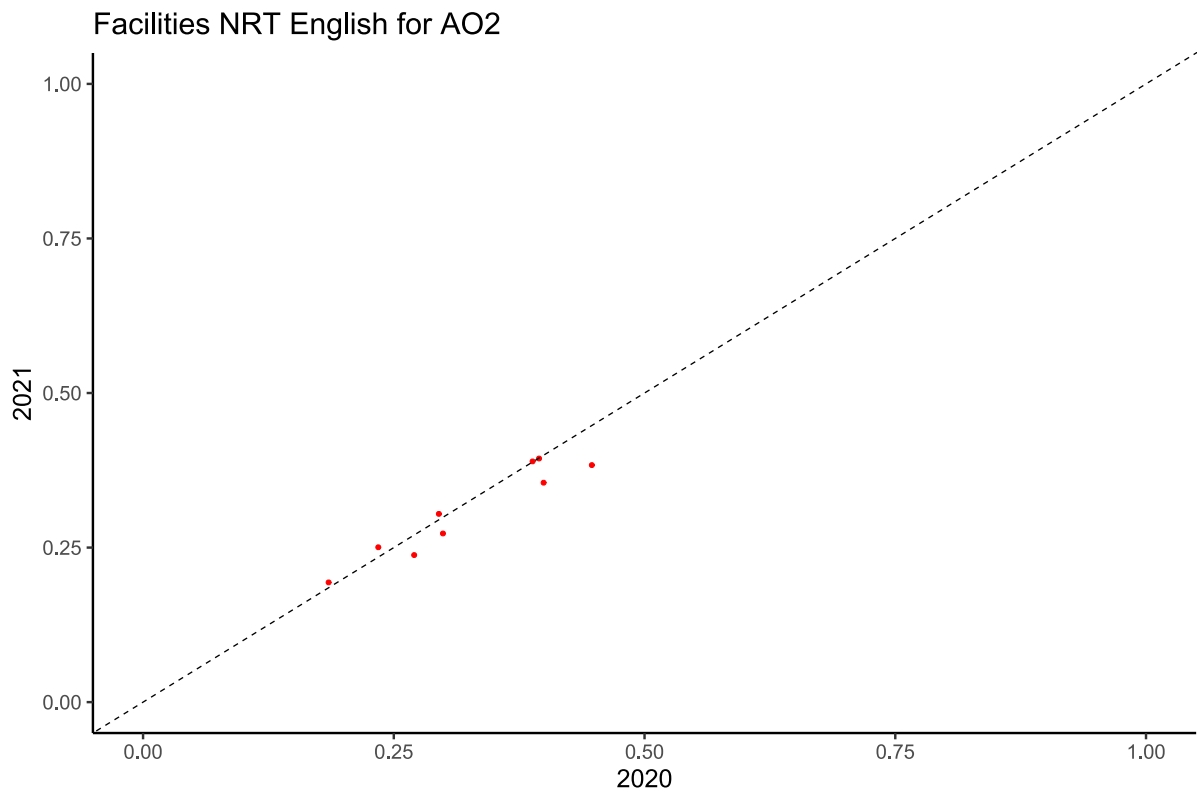
This analysis is consistent with the overall outcome of the English NRT in that there is no significant change between the two years. However, there is a suggestion of a small drop in performance for items covering AO1 and AO2, as the majority of these items are placed below the diagonal line. In contrast, all of the items for AO3 to AO6 lie on, or very close to, the diagonal line indicating that they performed similarly between the two years.

A t-test was used to compare the item facilities for AO1 and AO2 between 2020 and 2021, as shown in Table 5.1. In both cases the average difference in facilities was small and not statistically significant ($p=0.79$ for AO1 and $p=0.65$ for AO2). This supports the evidence from the profile plots that there were no significant differences in performance for any of the English NRT AOs between 2020 and 2021.

Table 5.1   Comparison of mean facility for AO1 and AO2 items

AO1 items	Mean facility	Standard deviation of facility
2020	0.54	1.45
2021	0.52	1.44

AO2 items	Mean facility	Standard deviation of facility
2020	0.32	0.82
2021	0.31	1.44

Gender

This analysis also explored whether the impact of the disruption of school closures due to the pandemic was the same for the boys and the girls in the NRT sample in terms of their relative performance across the curriculum areas. Whilst there are substantial differences between the performance of boys and girls in the NRT English assessment – girls tend to perform better than boys – there appear to be no notable changes in the performance of boys and girls across the different assessment objectives (AOs) between 2020 and 2021. This is perhaps unsurprising given that across the whole sample very few changes in performance on the different AOs were seen. The figures for this analysis are presented in Appendix C Figure C1.

Conclusion for English

The NRT results from 2021 indicate that there was no significant change in performance in English compared with 2020. This study compared the relative performance by assessment objective in English to see whether any relative changes in performance in particular areas of

the curriculum could be discerned that may identify specific English skills that have proved to be difficult to teach or learn, or given different levels of priority as a result of the disruption to the 2021 cohort caused by the Covid-19 pandemic. The analyses of profile plots, booklet scoring and gender factors provide no evidence of clear differences in the performance at the booklet level that might indicate any consistent change in performance across the curriculum. The facility plots across all booklets also show that there were no significant differences in the scores between different AOs.

5.2. Mathematics

Subject content areas: profile plots

The NRT for mathematics is divided into five subject content areas and three assessment objectives as set out in Tables 2.2 and 2.3. Figure 7 shows the representative profile plots for all of the subject content areas for mathematics booklet 1. The equivalent information for all of the booklets is presented in Appendix A2 Figures A8 to A13.

Because of the way that the mathematics curriculum is divided there are many more profile plot comparisons to make. There are few differences between the profile plots for mathematics and, as with the English, there are no consistent differences visible on the profile plots. As such, there is no evidence that performance in the mathematics subject areas differed between 2020 and 2021.

Figure 7 Profile plot of scores for the mathematics subject content areas for booklet 1[6]
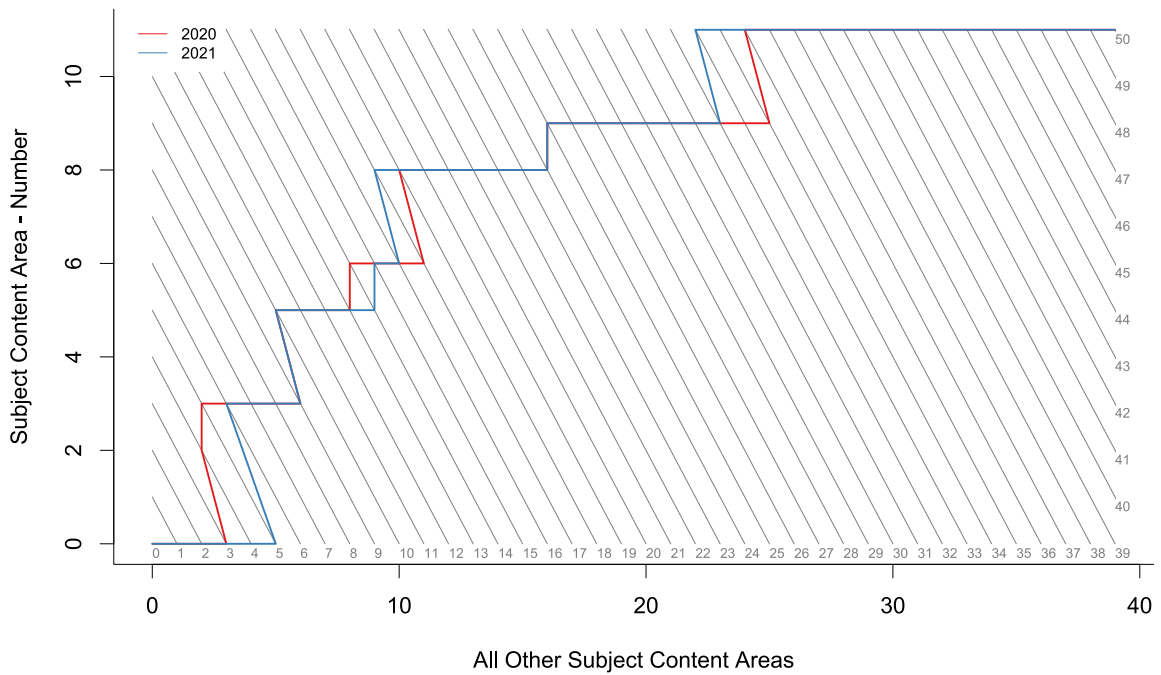

Mathematics Booklet 1 - Algebra

---
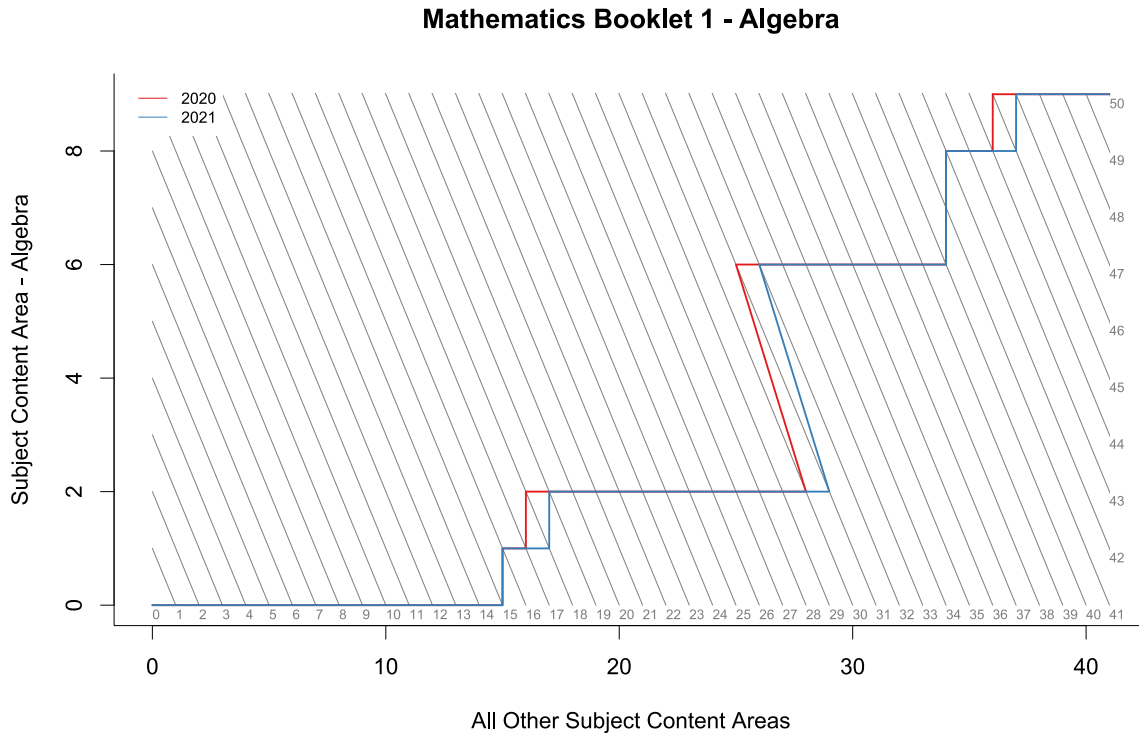
[6] Subject content areas on Figure 5.7: A=algebra; G=geometry and measures; N= number; R=ratio and proportion; P=probability; S=statistics. In the NRT subject content distributions probability and statistics are counted together.
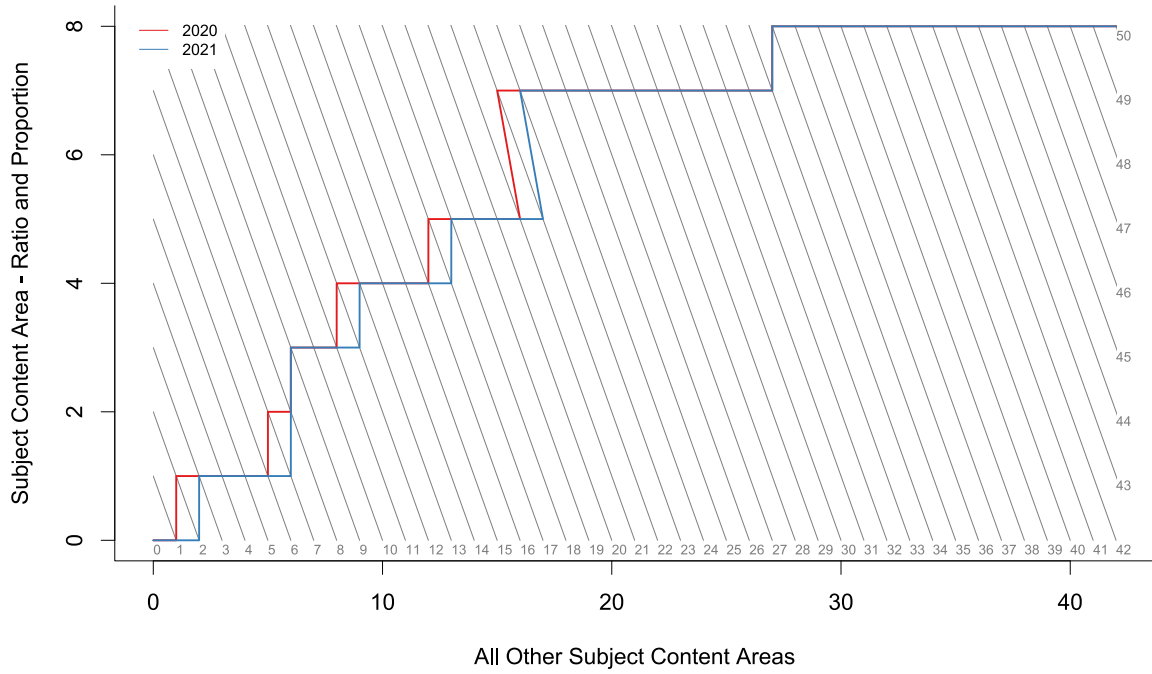
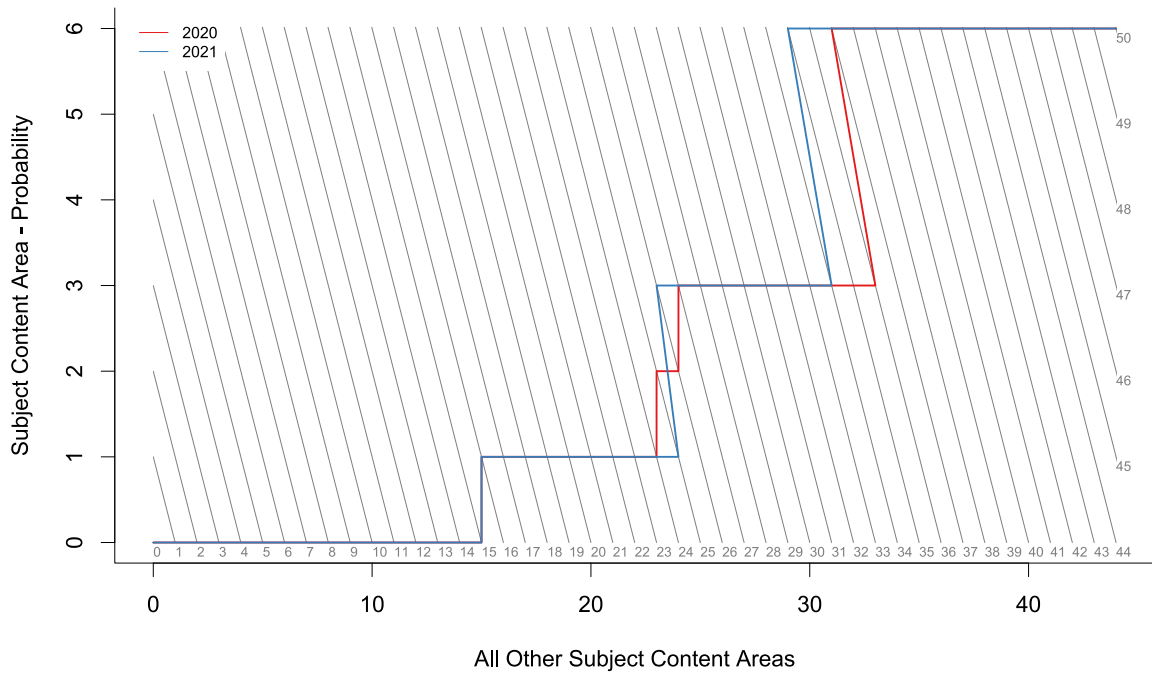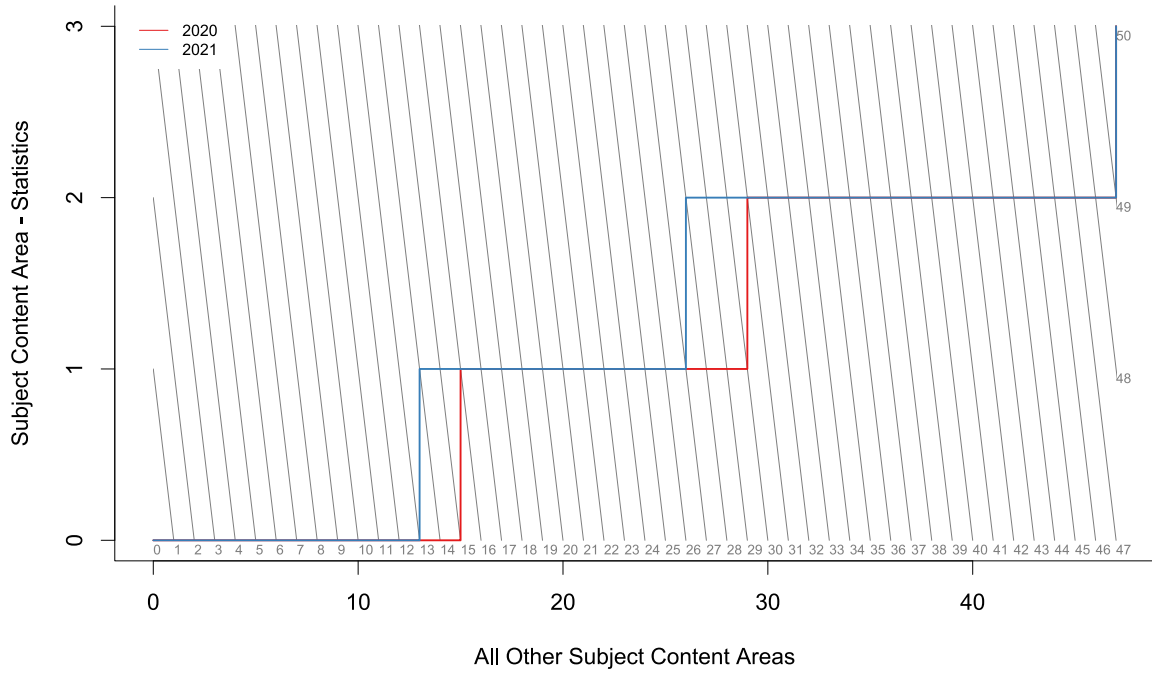## Mathematics Booklet 1 - Geometry and Measures
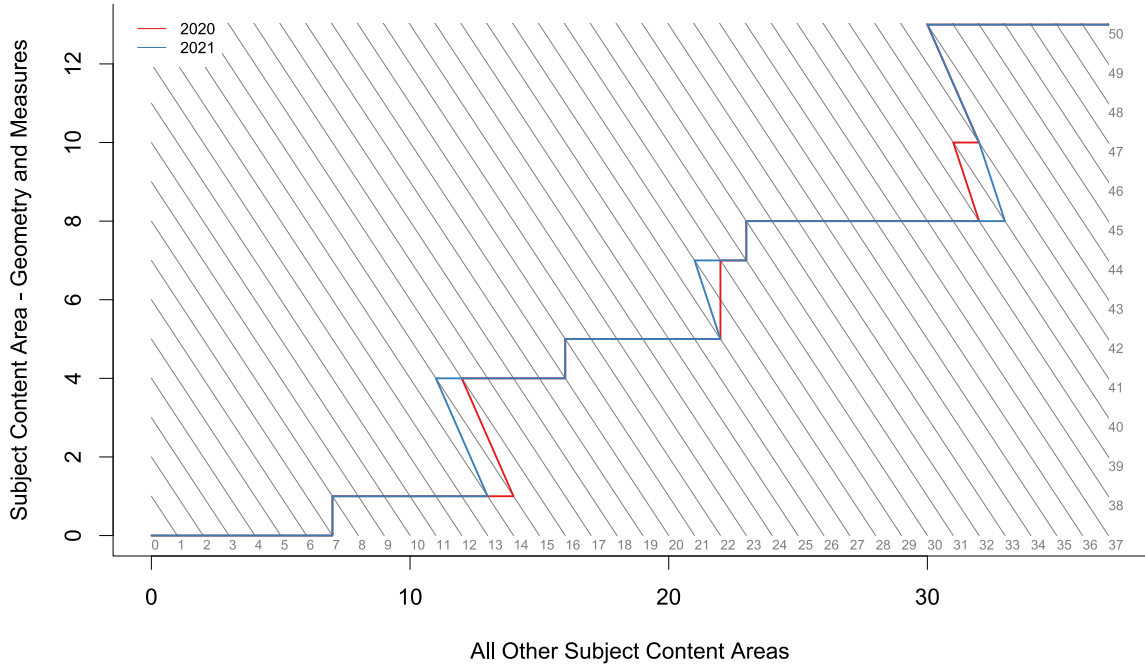
Mathematics Booklet 1 - Number

Mathematics Booklet 1 - Ratio and Proportion

## Mathematics Booklet 1 - Probability

Mathematics Booklet 1 - Statistics

## Subject content areas: proportions, facilities and mean scores

As the booklets for 2020 and 2021 for mathematics contained the same items, we can compare the scores within those booklets directly. Figure 8 shows only small differences in the proportions of certain subject content areas in mathematics between 2020 and 2021. However, these differences are not substantial and never consistently favour a specific year.

Figure 8   Proportions of mathematics booklet scores of subject content areas for 2020 and 2021
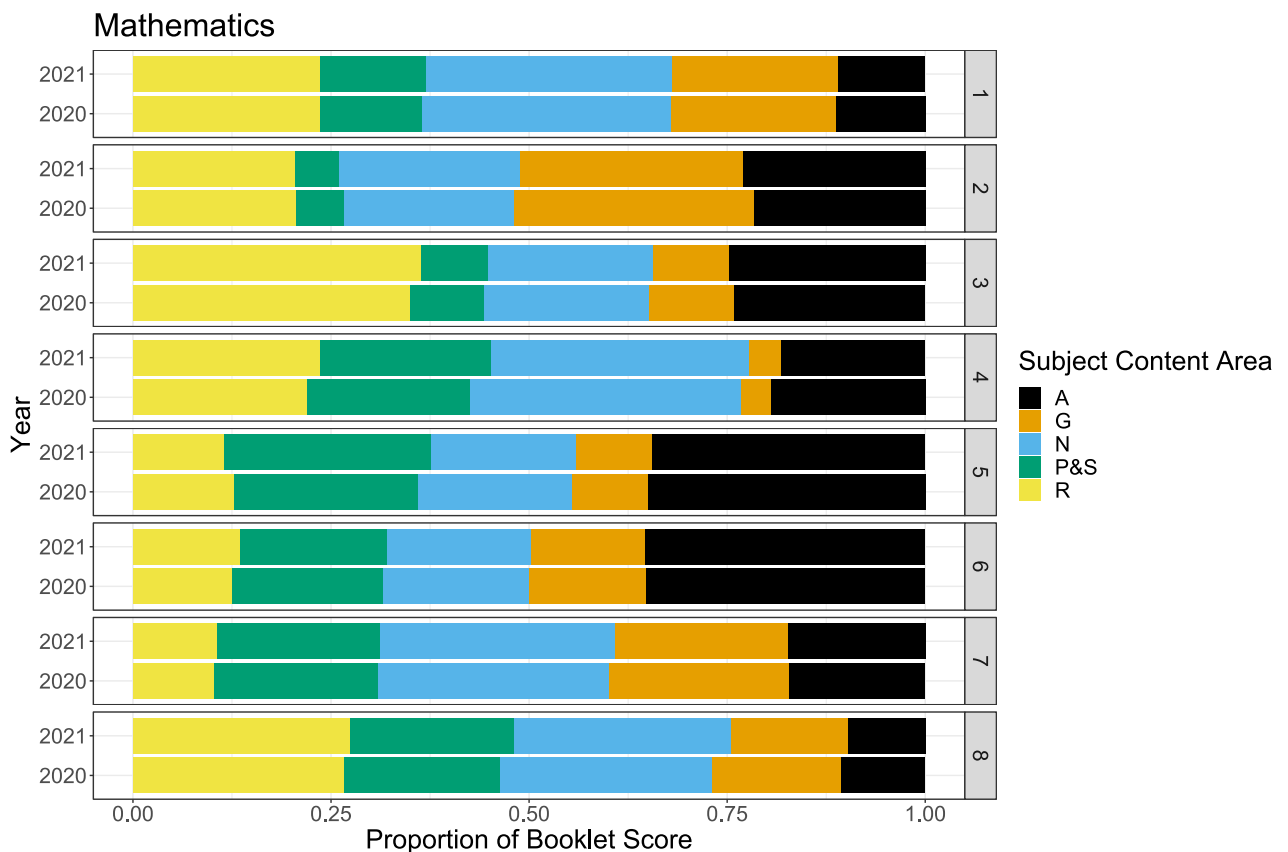
If we take a look at the mean facilities (proportion correct) of the different subject content areas, we can find out whether the performance of the 2021 population is lower than the 2020 population in each of the subject content areas. For example, Figure 9 shows that for booklet 1 and booklet 4 the performance of the 2021 population is lower than that of the 2020 population. Although the size of the drop varies, this drop in performance is seen in almost all subject content areas and booklets (as shown in Appendix B Figure B2). The same holds for the mean scores as shown in Appendix B2 Table B2.

Figure 9   Mean facilities for booklet 1 and booklet 4 for all mathematics subject content areas in 2020 and 2021

Booklet 1 Mathematics

Booklet 4 Mathematics
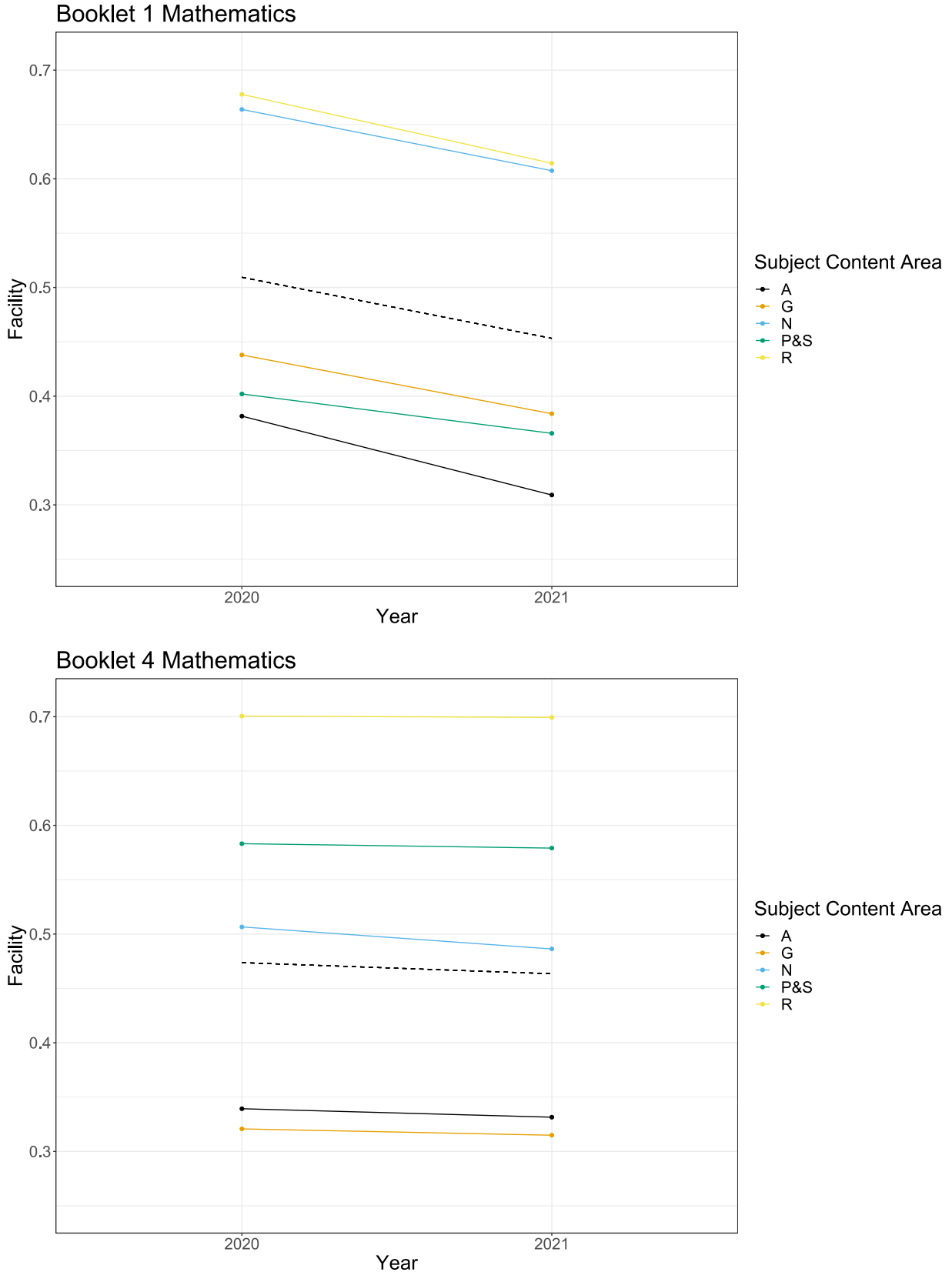
Performance on assessment objectives for mathematics

Mathematics items can also be divided by assessment objectives. Figure 10 plots the proportions of total booklet scores achieved for each of the mathematics AOs in 2020 and in 2021. As with the English AOs, this analysis indicates no clear, consistent changes in performance between the AOs from 2020 to 2021. This is supported by the mean facility plots by assessment objectives, which show a similar trend across all assessment objectives and booklets (Appendix B2 Figure B3).

Figure 10 Proportions of mathematics booklet scores by assessment objective for 2020 and 2021
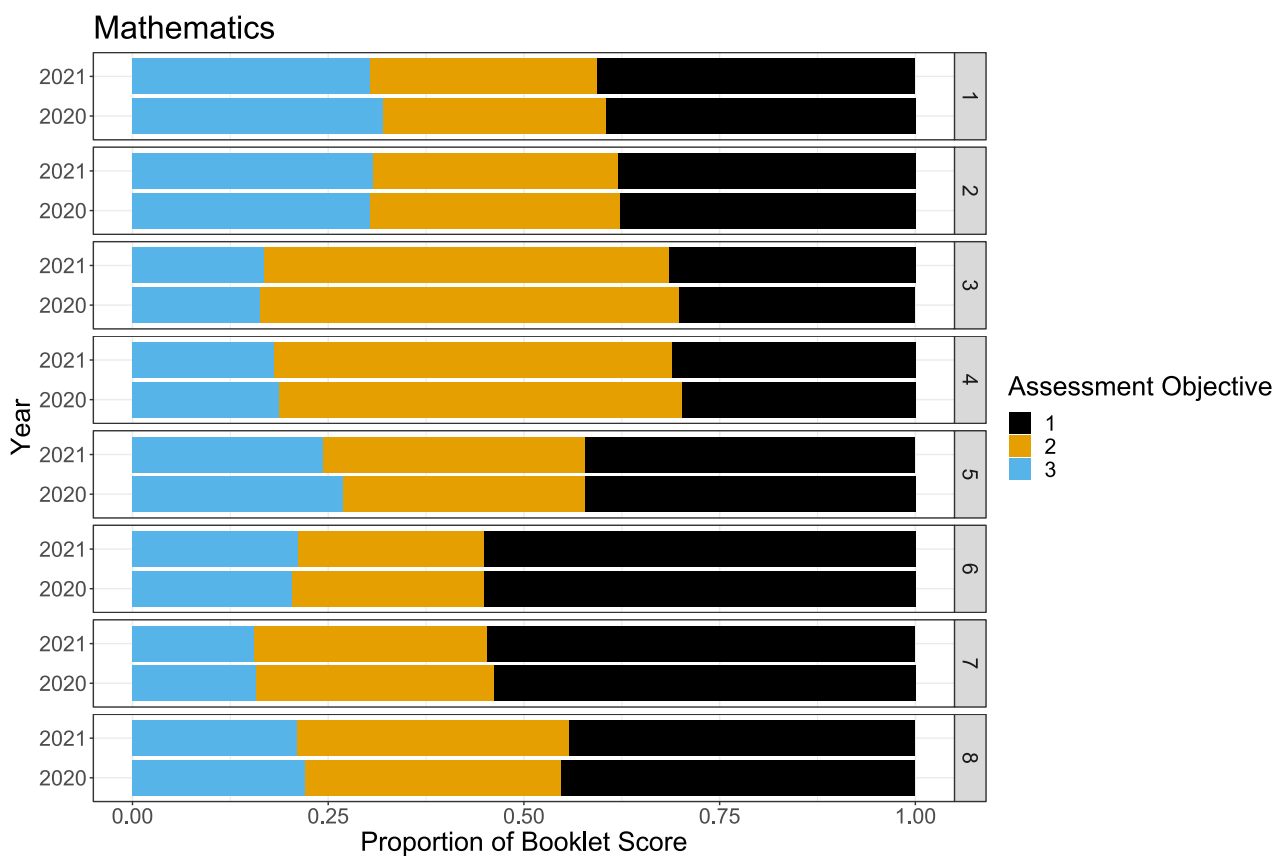
Sub-item level analysis for mathematics

The marking of the mathematics items in the NRT differs from that of English, in that each individual mark (out of the 200 marks in the NRT) is assigned to a specific subject area and assessment objective (AO) and then marked individually as to whether the student has met (1 mark) or not met (0 marks) the specific marking criteria. Each of these marking points is called a sub-item, and this allows us to analyse performance within each subject content area and AO at a more granular level[7].

---

[7] A five-mark item may contain more than one subject content area and more than one assessment objective, thereby merging performance over these different areas or skills. Analysis at the sub-item level, in theory removes this merging effect so analysis will look at performance solely on the specific subject content area or assessment objective skill.

The facility data at the sub-item level is contained in Appendix B2 Table B4. Based on this data we have looked for any evidence of any larger changes in the average performance of sub-items in any particular area of the curriculum (subject content area and AO). There are different ways to look at such changes but the analyses performed were as follows:

- Average changes in facility from 2020 to 2021.
- Average changes in relative facility from 2020 to 2021. This compares the proportional change in facility, taking into account that subsets of sub-items may tend to be harder or easier depending on the skill involved.
- Changes to ranking of facilities from 2020 to 2021. This highlights changes in facility for a subset of sub-items compared with the changes for all other sub-items. The data from this analysis is presented in Table 5.2.

Table 5.2   Sub-item level analysis of mathematics subject content areas.

Mathematics subject content area	Number of sub-items	Average change in facility	Average relative change in facility	Average change in rank order
Algebra	49	-0.031	-0.068	-0.45
Geometry and measures	39	-0.042	-0.125	-2.31
Number	46	-0.037	-0.084	-0.09
Probability and statistics	28	-0.027	-0.071	2.04
Ratio and proportion	38	-0.029	-0.065	1.55
Grand Total	200	-0.034	-0.082	0.00

These three measures produced differing results as they measure changes in performance in different ways. However, the aim is to identify whether any areas of the curriculum had consistent changes in performance across all three measures, suggestive of a meaningful change in item performance relative to the other groups.

It is noticeable that geometry and measures sub-items showed the greatest change by all three measures, suggesting that geometry and measures items may have performed relatively less well in 2021 compared with 2020. A t-test was used to compare the sub-item facilities for this subject content area between 2020 and 2021, as shown in Table 5.3. Although the mean facility was slightly lower in 2021, this difference was not statistically significant ($p=0.60$). Similar analysis by AO for mathematics produced no clear picture (data not shown).

Table 5.3   Comparison of mean facility for geometry and measures subitems

Geometry and measures subitems	Mean facility	Standard deviation of facility
2020	0.43	2.03
2021	0.38	2.00

Gender

As noted above, this analysis also explored whether boys' and girls' performance on the mathematics subject content areas was differentially impacted by the Covid-19 pandemic. Although there have been no significant differences in gender performance in the mathematics NRT, girls tend to do less well than boys. However, with respect to the different curriculum areas, there appear to be no notable differences in the performance of boys and girls across the different curriculum areas between 2020 and 2021. As was the case for English, this finding reflects the fact that across the whole sample very few changes in the performance in the different content areas were observed. Appendix C Figure C2 shows that the proportion of each subject content area contributing to the total score does not change noticeably between the years.

Conclusion for mathematics

The NRT results from 2021 indicate that there was a significant drop in performance in mathematics compared with 2020. This study compares the relative performance by subject content area and assessment objective in mathematics to see whether any relative changes in performance in different areas of the curriculum could be identified that may indicate specific mathematics skills that have proven to be difficult to teach or learn or given different levels of priority as a result of the disruption to the 2021 cohort caused by the Covid-19 pandemic.

The analysis of profile plots, booklet scoring and gender factors shows there were no clear differences in the performance at the booklet level that might indicate any consistent change in curriculum coverage. The inclusion of sub-item data for mathematics provides a method for analysis of potential curriculum effects across the whole of the NRT. Although this analysis suggests that there may be relatively lower performance in marks testing geometry and measures, these differences are not statistically significant.

Comparison of the performance of the different curriculum areas, by item and sub-item analysis, suggests that performance fell for all areas of the curriculum. Therefore, it is reasonable to conclude that disruption due to the Covid-19 pandemic impacted all areas of the mathematics curriculum to a similar degree. As the extent of Covid-related disruption was variable at the school, class and even student level then the absence of a clear pattern in changes in performance at the curriculum level is perhaps unsurprising.

6. References

Bechger, T. and Partchev, I. (2021). *How Members of Different Groups Obtain the Same Test Score: Profile Plots in Dexter* [online].

Glossary of technical terms

Block interlaced design
In these designs, a booklet contains two blocks of items. The first block of items overlaps with the second block of items of the previous booklet. The second block of items of the last booklet overlaps with the first block of items of the first booklet.

Booklet
A booklet is the unique collection of items allocated to a test version. Each student will see only one booklet. Booklets are designed to be approximately equivalent in coverage and difficultly so that the experience of each student is comparable.

Classical Test Theory
Theory that postulates that the observed score is composed of a true score and a measurement error component. CTT measures the difficulty of an item as its facility for the population of students that have completed the item.

IRT
Item response theory. Theory that specifies the relationship between ability and the probability of answering an item correctly using a mathematical model. IRT measures the difficulty of an item on a theoretical ability scale that is valid for all students in the cohort, not just those that completed that item.

Item Facility
In Classical Test Theory, simply the proportion of correct answers to an item from the population of students that have completed that item.

Measurement error
The difference between the reported and true values.

Standard deviation
A measure of the variation of a survey statistic.

Evidence for excellence in education

Public