United Nations Economic Commission for Europe

United Nations Children's Fund

Expert meeting on statistics on children

Geneva, Switzerland, 4–6 March 2024
Item 2 of the provisional agenda

Monitoring attainment gaps for students with protected characteristics. Advancements from England

Note by Ofqual (Office of Qualifications and Examinations Regulation, England)*

Abstract

Attainment gaps are a key indicator of educational and societal inequalities. Yet, statistical reporting of educational attainment often takes the form of breakdowns of students' achievement by selected characteristics. These descriptive statistics are potentially misleading, as they fail to account for the interplay between students' characteristics and other factors known for affecting students' achievement. This paper presents the experimental work to improve the statistical monitoring of attainment gaps in England that was conducted by Ofqual and focused on three areas. First, it is based on the creation of a linked administrative dataset at student-level. This combines measures of attainment with a broad set of protected characteristics (gender, ethnicity, language spoken, special educational needs and disabilities status), and indicators of socio-economic background, and other determinants of attainment (most notably, prior attainment). Second, it exploits multivariate regression analysis to retrieve estimates of the association between attainment and each characteristic, once all other observed factors have been held constant. It also shows differences between raw differences and regression results. Third, the statistical output is published as an interactive web dashboard in an accessible format to allow as many users as possible to easily navigate the findings and focus on the results they are interested in.

*Prepared by Nadir Zanini (Associate Director for Research, Data & Analytics – Ofqual, England)

NOTE: The designations employed in this document do not imply the expression of any opinion whatsoever on the part of the Secretariat of the delimitation United Nations concerning the legal status of any country, territory, city or area or of its authorities, or concerning the of its frontiers or boundaries.

I. Introduction

A. Context

1. Education in England

1. In England, education is compulsory for all children between the ages of 5 and 16, though most pupils remain in education up until 18. At both 16 and 18 students take a number of high-stakes examinations. The qualifications and grades achieved in these qualifications play a key role in determining students' progression to the next stage of their life.
2. At 16, to mark the end of the lower secondary school and of the compulsory education, pupils are normally entered for a range of external examinations. Most frequently, these are GCSE (General Certificate of Secondary Education), that are General Qualifications (GQs) available in a range of subjects. Students following an academic track usually take between 10 and 12 GCSEs, whereas those interested in a more applied education take a smaller number of GCSEs (usually including maths and English) alongside Vocational and Technical Qualifications (VTQs).
3. In the upper secondary school, lasting two years, the vast majority of students specialise in a subset of subjects. Those interested in an academic pathway take A levels usually in three subjects. These are highly regarded qualifications used for admissions to university. A range of VTQs are also available to students (often provided by Further Education colleges), who can take them instead of A levels (if interested in a more applied education), or alongside A levels (if interested in pursuing a mixed pathway featuring a combination of academic and applied qualifications).
4. In England, qualifications are offered by a range of Awarding Organisations that operate in a market regulated by the Office of Qualifications and Examinations Regulation (Ofqual), a non-ministerial government department. It is the Awarding Organisations' responsibility to set and mark the exams and provide students with a grade. It is the Regulator's responsibility to set requirements and monitor the system; Ofqual's statutory objectives are:
    i. To regulate for the validity of qualification standards, ensuring comparability of examination results across space as well as over time; and
    ii. To promote public confidence in the system, showing transparency and improving knowledge and understanding of the examinations system.

2. The coronavirus (COVID-19) pandemic and teacher judgement

5. Following the outbreak of the coronavirus (COVID-19) pandemic, in March 2020 schools were closed and exams cancelled. To ensure that students who were supposed to take exams could progress to the next stage of education or into the workplace, a form of teacher judgement was introduced. In summer 2020 and 2021, teachers were asked to submit, for each pupil and for each qualification and subject for which they were entered, the grade they judged the candidate would most likely have received had the exams taken place.

6. The greater subjectivity of teacher judgement, however, is known to be more vulnerable to bias than test-based assessment. Studies of potential bias in teacher assessment suggest that differences between teacher assessment and exam assessment results can sometimes be linked to student characteristics, including: gender, ethnicity, special educational needs, having English as an additional language, and socio-economic background (Lee & Walter, 2020; Lee & Newton, 2021).
7. Awareness of the potential risks to the validity of teacher judgements is, however, not enough to prevent bias against certain groups of students. Ofqual published guidance for teachers involved in making judgements for qualification awarding during the pandemic, to help them make judgements as objectively as possible, and ultimately to promote fairness and minimise bias (Ofqual, 2021).
8. Nevertheless, exams were re-introduced as soon as possible after the public health crisis was under control. In summer 2022 and 2023, exams were available again to students. Stakeholder' concern was that the pandemic might have impacted specific groups of students more than others and that the re-introduction of exams would have disadvantaged those students more than others.

B. Background and motivation

9. The concerns around potential bias in teacher judgement posed a threat to the public confidence in the validity of the grades awarded in absence of exams during the pandemic. Ofqual, therefore, embarked on an experimental project to provide evidence on attainment gaps, defined as differences in qualification results between groups of students with different protected characteristics and socioeconomic status.
10. As in other countries, the UK Department for Education (DfE) provide evidence on educational inequalities in England through their annual Official Statistics publications on attainment of 16-year-olds and 18-year-olds. These are cross-sectional descriptive statistics in the form of bi-variate distributions: simple breakdowns of achievement by some students' socio-demographic characteristics (for 2020, see respectively DfE 2020a and DfE 2020b). Although these are an important source of evidence on educational inequalities, they present two major limitations.
11. The first limitation is that a simple bi-variate analysis may fail to account for:
    i. *The interplay between students' social-demographic characteristics* – This can occur, for example, when two ethnic groups with a very different socio-economic make-up are compared. If differences in results are found, these may be attributed to ethnicity, rather than socio-economic background.
    ii. *Other factors known for affecting students' qualification results* – There is ample evidence for England (and beyond) that the strongest predictor of achievement is prior attainment. Failing to take this into consideration would potentially lead to attribute to students' protected characteristics or socio-economic background a role that they do not have.
12. The second limitation of simple breakdowns of qualification results by students' characteristics is that they may not only reflect potential bias in teacher judgement, but also pre-existing differences in the attainment of different groups. Although this is true at any point in time, retrieving attainment gaps only in 2020 and 2021 when exams were cancelled,

could lead to attribute to teacher judgement differences in educational attainment that would also be apparent in normal years, when qualification grades are based on exams.

C. Analytical approach

13. To tackle the limitations of descriptive statistics regularly published and retrieve more insightful evidence, we designed an analytical approach that was based on three main features:
    i. Move from a bi-variate analysis to a multi-variate approach to measure the impact of each student characteristic once other factors potentially affecting attainment have been held constant;
    ii. Broaden the set of students' characteristics to include not only protected characteristics, but also observable factors knowing to be linked to achievement (most notably, prior attainment);
    iii. Focus on how attainment gaps have changed over time, rather than on the gaps at a specific time, by introducing a set of criteria to identify changes we considered to be 'notable'.

II. Data and variables

14. To broaden the set of students' characteristics included in the analysis we used a linked dataset combining:
    i. Data collected by Ofqual from the Awarding Organisations, containing detailed information on the grades achieved by each student in each qualification awarded and their (current and prior) schooling;
    ii. Records from the National Pupil Database and the Individualised Learner Record held by the UK Department for Education and containing for each student socio-demographic characteristics and measures of socio-economic deprivations.
15. The two datasets were linked using students' names, date of birth and gender. As a result, for each student it was possible to retrieve:
    i. Educational attainment – grades achieved by each student in each qualification taken at 16 and 18, since 2018.
    ii. Schooling – prior attainment measured by qualifications and grades achieved throughout their education; for students taking General Qualifications also the type of education establishment attended and its geographical location.
    iii. Protected characteristics – ethnicity, gender, special education needs and disabilities (SEND), first language spoken at home (only available for students taking General Qualifications);
    iv. Indicators of socio-economic deprivation – Free School Meal eligibility (FSM), and IDACI (Income Deprivation Affecting Children Index), measuring the proportion of

all children aged 0 to 15 living in income deprived families in a certain geographical area.

16. Table 1 below displays the number of grades analysed, separately for the general and vocational qualifications, included in the analysis. Numbers for VTQs are smaller, as they are less frequently taken by students.

Table 1. Number of observations included in the analysis.

		2018	2019	2020	2021	2022	2023
GQ	GCSE	3,385,980	3,491,945	3,587,055	3,671,210	3,673,080	3,737,850
	A level	398,110	415,400	414,135	440,775	460,020	463,435
VTQ	Technical Award	251,005	310,270	386,980	419,080	399,755	403,265
	Technical Certificate	43,540	21,240	16,230	24,870	20,855	17,845
	Applied General	83,180	114,275	140,575	198,075	233,345	252,045
	Tech Level	28,215	40,040	40,225	55,885	54,060	54,345

Note: In interpreting these figures it is important to bear in mind that students are awarded a grade for each qualification and subject taken. Values have been rounded to the nearest 5.

III. Methodology

17. We used a multivariate analytical approach. This allows us to explore the impact on overall results of each feature separately while controlling for other features. This is important because we know that there are relationships between different features (for example, ethnicity and first language) which is necessary to account for. We used regression modelling to estimate differences in results for groups of students after controlling for other variables.

18. We carried out mixed-effects modelling to analyse students' attainment, given the information on the characteristics of students clustered within schools. This is to be able to account for factors affecting results that only students from the same school share, such as admissions policy and teaching quality. The student-level characteristics described above were treated as fixed effects; the student's school was treated as a random effect.

19. For GQ, we considered 3 attainment measures:

    i. Mean numeric grade (for A level, grades A* to E were converted to 6 to 1 respectively and U treated as 0 while for GCSE, grades on 9 to 1 scale were numeric and U was treated as 0);

    ii. Probability of attaining A level grade A and above / GCSE grade 7 and above

    iii. Probability of attaining A level grade C and above / GCSE grade 4 and above

20. For VTQ, given that different grading scales are used across the different types of qualifications considered, we modelled the probability of attaining the top grade available in each qualification.

21. Linear models or generalised models (logistic regressions) were used, depending on the attainment measure being analysed. Regression models were run for each year separately.
22. This approach allowed us to estimate the result for an 'average' entry, that is, an entry by a student who was in the reference category of every one of the background variables. In numeric grade analyses, we estimate the grade an average entry would receive. In grade probability analyses, we estimate the probability that an average entry would be awarded the key grade in question.
23. We can then estimate the size of the difference in outcome between a particular group and the reference group after controlling for effects of other background variables. Variation of these group estimates from the models covering each of the years tells us how that group's results relative to the relevant reference groups have changed over time.

IV. Presentation and interpretation of findings

A. Notable change criteria

24. Findings are presented graphically and in table format. Given the large sample size it is possible that even very small differences between groups/years are flagged as significant. For this reason, we developed a multi-step method for evaluating changes in relative outcome differences between key years of comparison, for example 2020 and 2019 (but more recently also 2023 and 2019, for a pre-post pandemic comparison). The aim is to identify practically significant changes while taking into consideration normal between-year fluctuations.
25. This method is described below:
    i. Step 1 – Identify subgroups of students whose relative outcome differences were not significantly different from zero (at 5% level of significance) in any year and exclude them from further consideration.
    ii. Step 2 – From the subgroups not excluded in Step 1, identify those whose over-time change in relative outcome difference in absolute value was larger than their 2018 to 2019 change in absolute value. This step is aimed at identifying changes that are larger in magnitude than normal year-on-year fluctuations.
    iii. Step 3 – From the subgroups of students identified in Step 2, flag as 'notable' the subgroups whose change exceeded an effect size criterion. We set the effect size criterion at 0.1 grade for the numeric grade measure and 1 percentage point for the grade probability measures (for VTQs this was raised to 5 percentage points in the average marginal effects).
26. Although all the tables and charts are presented, only the notable changes flagged by the above approach are drawn out in the report.

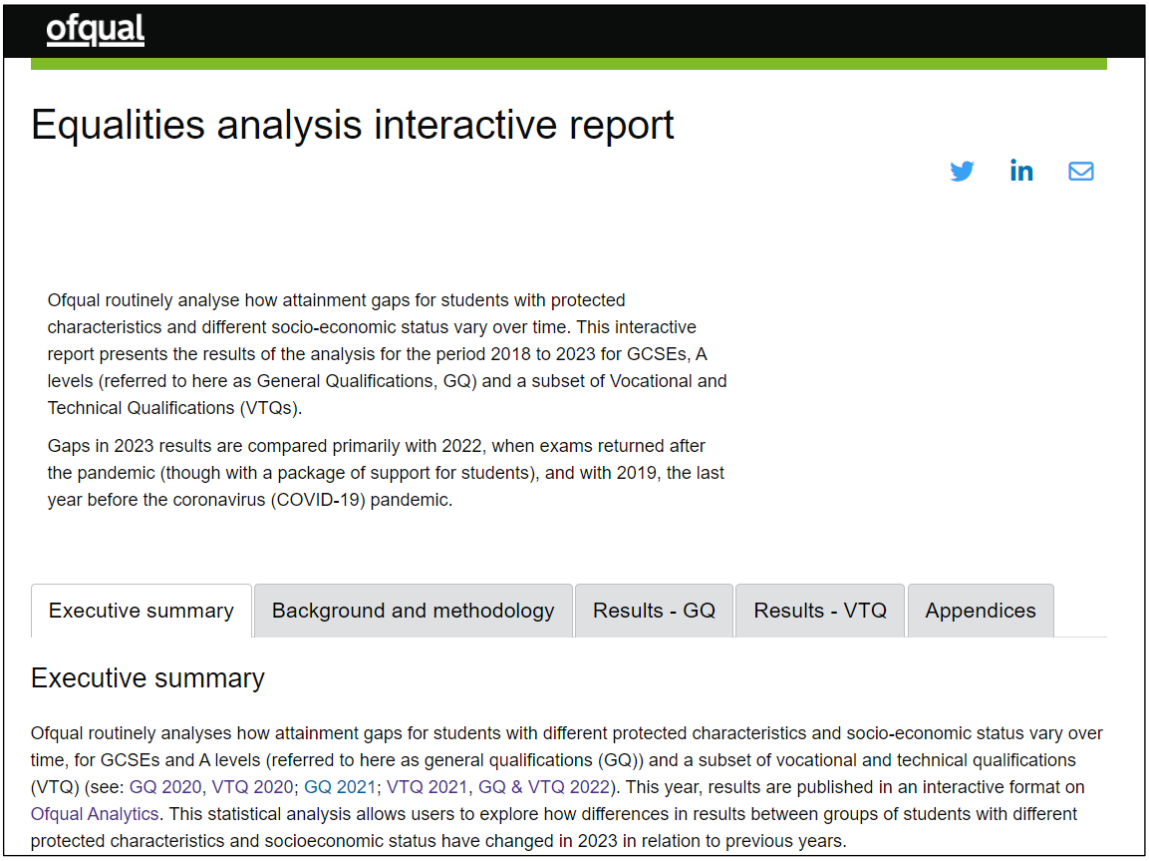
B. Report format and move to a web-based interactive report

27. Ofqual's equalities analyses were initially published as research reports, and separately for GQ (Lee, Stringer & Zanini, 2020; Lee, 2021) and VTQ (Cuff & Rama, 2021). Feedback received on early publications highlighted the need of a more cohesive approach across general and vocational qualifications. From 2022, GQ and VTQ analyses pulled together in the same report, ensuring a more consistent use of definitions and terminology (Lee, Rama & Johnson; 2022). This resulted in a very long report, with a large number of charts and tables.

28. Following user engagement that highlighted the need to make the findings of the analysis more easy to navigate, from 2023 (Lee, Rama & Johnson; 2023) the analysis was made available for the first time in an interactive report format available at this link (https://analytics.ofqual.gov.uk/apps/AllQualifications/Equalities/). In this way users can more easily navigate to the results they are interested in (for example ethnicity), while still being able to find the detailed technical information made available in the previously published reports.

29. The interactive report comprises multiple sections which are navigable via the tabs at the top of the page (see Figure 1). The results are shown in interactive charts in the separate results tabs for GQ and VTQ.

Figure 1. Interactive report format: the landing page.



30. As shown in Figure 2, for the GQ charts, users can select the qualification level, variable and outcome measure that they are interested in. Users can also select whether they would like to see the 'raw results' (ie descriptive statistics based on simple, bi-variate breakdowns) and/or 'modelled' results (ie results which control for the other variables included in the analysis). Users can also choose whether to look at 'relative' results (ie differences from the reference variable) or 'absolute' results (ie overall estimates).
31. For VTQ, the control panel is slightly different (see Figure 3). It allows users to select the qualification level, qualification group and variable that they are interested in. As for GQ, it allows to select the results interested in ('modelled' or 'raw'). Results will show the average marginal effect, that is the difference in average probability of receiving a top grade between the results displayed and the reference group.
32. For both GQ and VTQ results, once a selection has been made, users will be prompted to view the results via a chart, or in a table. Text statements presented underneath the chart highlight which of the findings were identified as notable. Results statements are generated automatically, though carefully reviewed by analysts. Figure 4 shows an example of the findings for modelled and raw results for numeric/average grade in GCSEs for the eligibility to Free School Meals (FSM).

Figure 2. Interactive report format: the control panel for GQ.

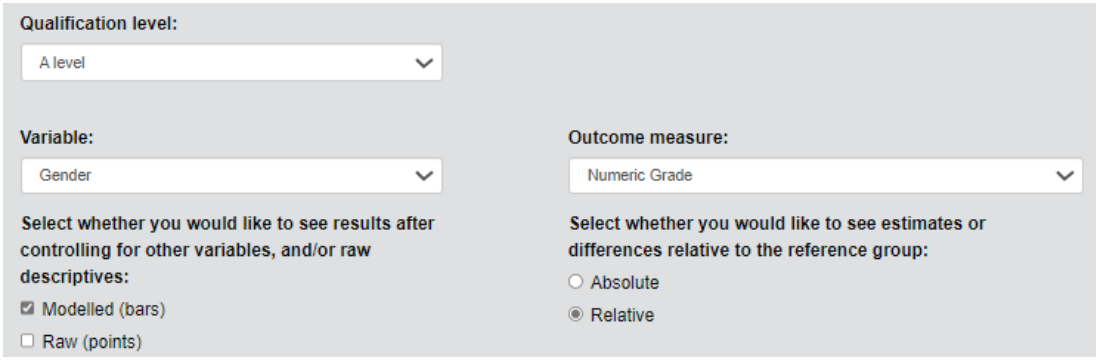


Figure 3. Interactive report format: the control panel for VTQ.

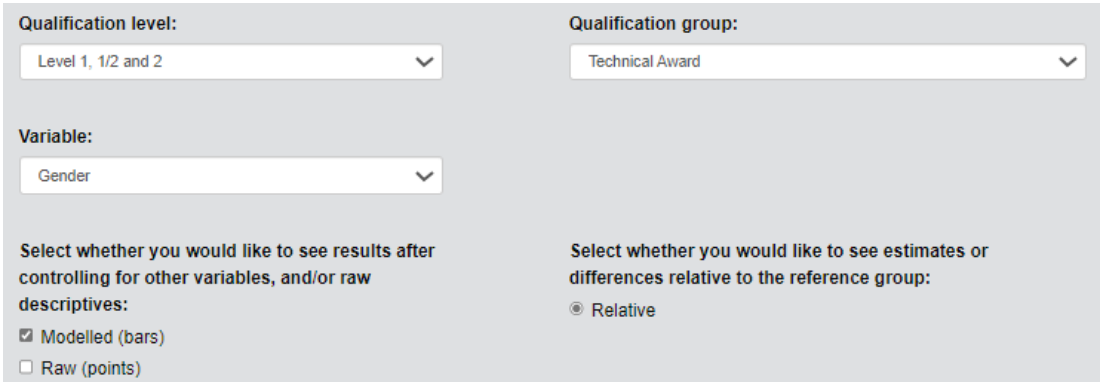
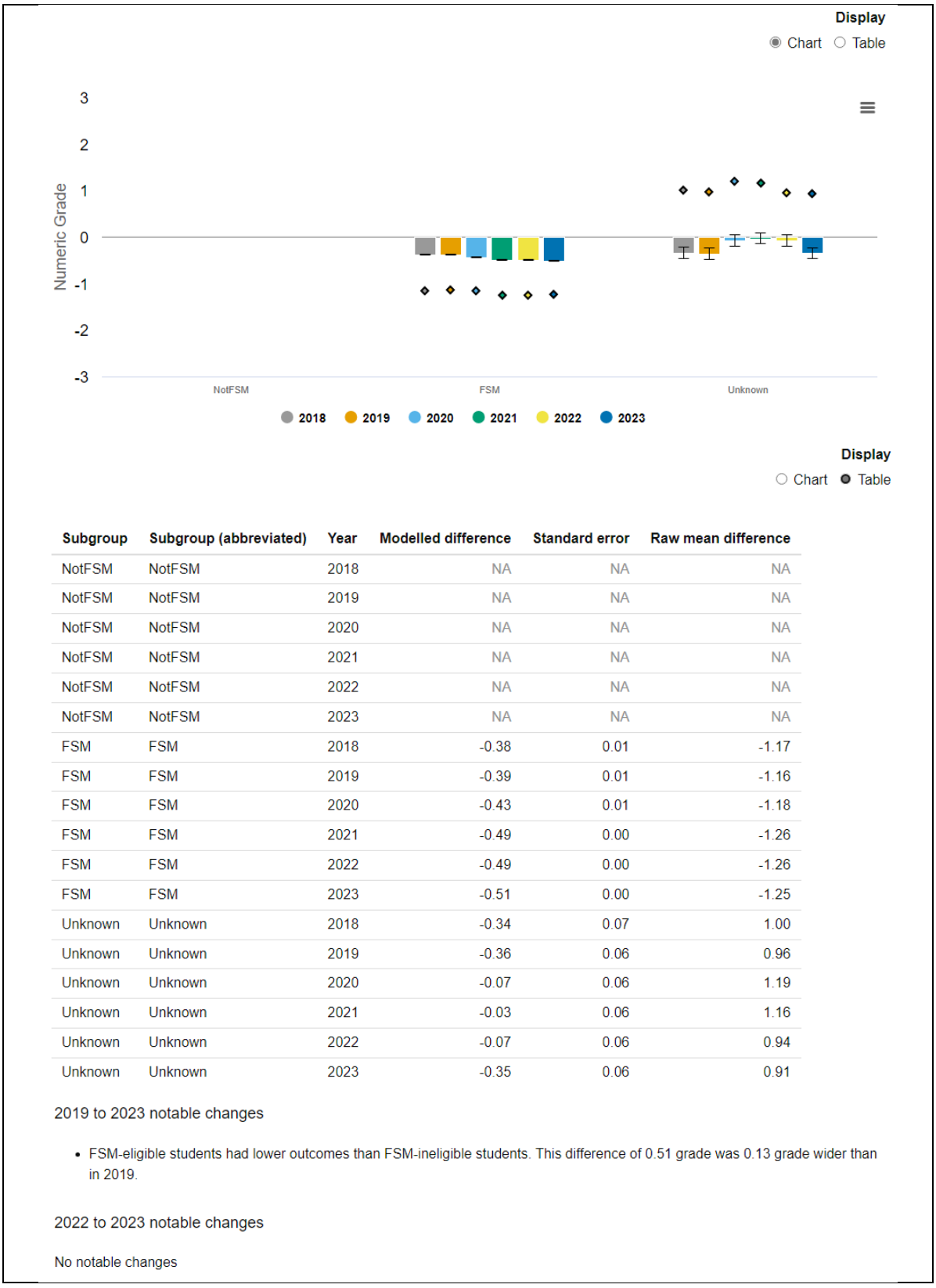


Figure 4. Interactive report format: the results chart, table and results statements.



Subgroup	Subgroup (abbreviated)	Year	Modelled difference	Standard error	Raw mean difference
NotFSM	NotFSM	2018	NA	NA	NA
NotFSM	NotFSM	2019	NA	NA	NA
NotFSM	NotFSM	2020	NA	NA	NA
NotFSM	NotFSM	2021	NA	NA	NA
NotFSM	NotFSM	2022	NA	NA	NA
NotFSM	NotFSM	2023	NA	NA	NA
FSM	FSM	2018	-0.38	0.01	-1.17
FSM	FSM	2019	-0.39	0.01	-1.16
FSM	FSM	2020	-0.43	0.01	-1.18
FSM	FSM	2021	-0.49	0.00	-1.26
FSM	FSM	2022	-0.49	0.00	-1.26
FSM	FSM	2023	-0.51	0.00	-1.25
Unknown	Unknown	2018	-0.34	0.07	1.00
Unknown	Unknown	2019	-0.36	0.06	0.96
Unknown	Unknown	2020	-0.07	0.06	1.19
Unknown	Unknown	2021	-0.03	0.06	1.16
Unknown	Unknown	2022	-0.07	0.06	0.94
Unknown	Unknown	2023	-0.35	0.06	0.91

2019 to 2023 notable changes

- FSM-eligible students had lower outcomes than FSM-ineligible students. This difference of 0.51 grade was 0.13 grade wider than in 2019.

2022 to 2023 notable changes

No notable changes

34. Publishing the analysis as an interactive report allowed us to increase its accessibility. The interactive report can be fully navigated with keyboard alone, including navigating tabs, drop-down menus and even interactive visualisations. All elements of these web applications work with assistive technology. For example, interactive visualisations have descriptive text explaining them in detail, going above and beyond what is available from alt-text alone. Descriptive tags explain what the interactive visualisation shows, the type of visualisation on display, the label and values of the axes, how many data points there are in a series and even each data points value.
35. The interactive report is created using R Shiny. Graphs are crated in Highcharts, which helps ensure they are as accessible as possible, using the WCAG 2.1 standard as a guideline. This allows users, even those with visual impairments, to interact with the visualisation and understand the relationships of all the components of the visualisation to all other components rather than just getting a description of the visualisation.

V. Conclusion

A. The advantages of the analytical approach

36. The attainment gaps highlighted by the multivariate analysis are in most cases different from the raw differences between groups of students. This is an important finding: it indicates that there is an interplay between student characteristics (eg, ethnicity and first language). Hence, it shows how simple descriptive statistics, though informative, can be misleading.
37. More specifically, the modelled attainment gaps tend to be smaller than the raw comparisons. For example, when looking at the disadvantage gap as measured by students' eligibility to free school meals, the simple comparison between GCSE students who are eligible and those who are not eligible indicates more than 1-grade gap, whereas the modelled results show a difference of half a grade.
38. It is not just the size of the difference to be affected, in some instances also the sign (or the direction) of the attainment gap changes. When focussing on the gender gap at A level in 2023, for example, descriptive statistics show that on average male students are outperformed by female students. The multivariate approach, however, shows the opposite result, with boys attaining on average better grades than girls.
39. In terms of over-time changes, it is worth noting that the majority of attainment gaps remain broadly stable across years. In the other cases, the gaps have sometimes narrowed but in some cases widened. The use of the approach described above to flag 'notable' changes, however, allows us to focus on change that are statistically significant, and at the same time that are worthy of note as operationally relevant, given that they are greater than normal year-on-year fluctuations (and of an effect-size criterion).

B. Other lessons learnt

40. There is a wealth of evidence produced with this analysis, helpful from both a methodological and substantive perspective. From a methodological perspective, one result is consistent throughout the analysis: prior attainment is always the strongest predictor of students' achievement. Its predictive power is stronger for general qualifications than vocational and technical qualifications, but always a key factor in determining students' grades. This confirms that failing to take students' prior attainment into account when reporting students' educational attainment may lead to draw the wrong policy recommendations.
41. From a substantive point of view, an interesting result is the one concerning the gender gap and how it changed during the pandemic. Whereas before and after the pandemic grades based on exams were higher for boys, during the pandemic girls outperformed boys. This may be an indication that teacher judgement (the approach used in 2020 and 2021) tends to be biased towards specific groups of students (in this case, girls; for a discussion, see Stratton, Zanini & Noden, 2021) and that objectivity guidance (Ofqual, 2021) may not be enough to avoid bias in grades awarded by teachers.
42. The linking of administrative records from multiple data sources led to the creation of a rich dataset, potentially useful for investigating a range of policy-related research questions by government analysts and academics. To promote quantitative research in the field of education and education assessment, the data set resulting from this analysis was further augmented with information on university admissions and made available to external researchers. GRADE (Grading and Admissions Data for England) data is accessible through the Office for National Statistics Secure Research Service[1] by any accredited researcher in the UK.
43. The presentation of the analysis as an interactive report have promoted transparency and engagement by external stakeholders. This is because users are allowed to select the characteristics they are interested in and how the results are displayed. It has to be considered, however, that many of those interested in these results may not be used at engaging with dashboards or the findings of a regression analysis. Summarising the main findings and describing the analysis without the use of technical language is also very important to ensure that all users can benefit from it. We tried to do so by publishing an authored article (Stockford, 2023) alongside the interactive report.

C. Limitations

44. The picture of attainment gaps and how they change over time is a complex one, especially during the pandemic years. In reporting this, it is important to carefully consider the many factors at play, and how they interact. Societal differences and the disruption to teaching and learning caused by the pandemic may have affected children differently, according to characteristics that we can observe, such as protected characteristics, socio-economic background, or ability profile. But there are also other characteristics that we cannot observe,

[1] More information about the GRADE (Grading and Admissions Data for England) is available at this link: https://www.gov.uk/government/collections/grade-data-sharing-project

such as motivation and parental support. And those effects may change at different points in time. This is why it is important to recognise the complexity of these findings.

45. Although every effort has been made to explain the advantages of using a regression analysis and how to interpret the findings, they can be easily misinterpreted by non-technical audiences. Engagement activity is needed to present the results of such an approach to key stakeholders and improve their presentation.

46. Despite the use of administrative records allowed us to get almost full coverage of the students' population under scrutiny, not all information was available for all student. More specifically, the retrieval of protected characteristics and socio-demographic information is difficult for students in some educational settings. For students in the independent sector, for example, their demographic characteristics are not always available. These students were not excluded from the analysis, but obviously the validity of the results for these students may be limited. Further work and potentially further data linkage is needed to retrieve additional information for these students.

D. Concluding remarks

47. Withing these limitations, however, there are real benefits from the analytical approach used by Ofqual to produce and present equalities analysis, both from a methodological and substantive perspective. Further work is needed to secure completeness of the data and ease the interpretation of the results. Nonetheless, the use of more advanced statistical methodology to estimate indicators to be routinely reported, in education and beyond, should be encouraged as they have the potential to provide more robust evidence to inform policies.

References

Cuff, B., & Rama, D. (2021). *Equalities analysis of grades awarded for VTQs in spring and summer 2021.* Coventry: Ofqual. Available at: https://www.gov.uk/government/publications/analysis-of-results-vtqs-2021/equalities-analysis-of-grades-awarded-for-vtqs-in-spring-and-summer-2021

DfE (2020a). *Key stage 4 performance, Academic year 2019/20*. Explore education statistics – GOV.UK (explore-education-statistics.service.gov.uk)

DfE (2020b). *A level and other 16 to 18 results, Academic year 2019/20*. Explore education statistics – GOV.UK (explore-education-statistics.service.gov.uk)

Lee, M.W. (2021). *Summer 2021 student-level equalities analysis. GCSE and A level.* Coventry: Ofqual. Available at: https://assets.publishing.service.gov.uk/media/61138b5dd3bf7f043c4bae07/6828-3_Student-level_equalities_analysis_for_GCSE_and_A_level_summer_2021.pdf

Lee, M.W., & Newton, P. (2021). *Systematic divergence between teacher and test-based assessment: literature review.* Coventry: Ofqual. Available at:

https://www.gov.uk/government/publications/systematic-divergence-between-teacher-and-test-based-assessment/systematic-divergence-between-teacher-and-test-based-assessment-literature-review#executive-summary

Lee, M.W., Rama, D., & Johnson, K. (2022). *Student-level equalities analysis for GCSE, A level and VTQ - 2022*. Coventry: Ofqual. Available at: Student-level equalities analyses for GCSE, A level and VTQs, 2022 - GOV.UK (www.gov.uk)

Lee, M.W., Rama, D., & Johnson, K. (2023). *Student-level equalities analyses for GCSE, A level and VTQs – 2023.* Coventry: Ofqual. Available at: https://www.gov.uk/government/publications/student-level-equalities-analyses-for-gcse-a-level-and-vtqs-2023

Lee, M.W., Stringer, N., & Zanini, N. (2020). *Student-level equalities analyses for GCSE and A level: summer 2020.* Coventry: Ofqual. Available at: https://assets.publishing.service.gov.uk/media/5fbfbf4ce90e077ee1daf4de/6713_Student-level_equalities_analyses_for_GCSE_and_A_level.pdf

Lee, M.W., & Walter, M. (2020). *Equality impact assessment: literature review*. Coventry: Ofqual. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/879605/Equality_impact_assessment_literature_review_15_April_2020.pdf

Ofqual (2021). *Information for centres about making objective judgements*. Available at: https://www.gov.uk/government/publications/information-for-centres-about-making-objective-judgements

Stockford, I. (2023). *Equalities analysis: what the 2023 results tell us*. Available at: https://www.gov.uk/government/speeches/equalities-analysis-what-the-2023-results-tell-us

Stratton, T., Zanini, N., & P. Noden (2021). *An evaluation of centre assessment grades from summer 2020*. Coventry: Ofqual. Available at: https://www.gov.uk/government/publications/evaluation-of-centre-assessment-grades-and-grading-gaps-in-summer-2020