TECHNICAL APPENDICES

Summer 2022 student-level equalities analysis: technical appendices A & D

ofqual

Contents

Appendices B, C and E are provided in the accompanying spreadsheet.

Appendix A – Variables and data sources

Gender

Source: Awarding organisations

Values: Male/Female/Unknown or neither

Reference group: Female

Prior attainment

Source: Awarding organisations

Values: Very low, low, medium, high, very high, missing

Reference group: Medium

Notes: Normalised KS2 or GCSE score originally ranging from 0 – 100, divided into quintiles.

Ethnicity

Source: DfE (NDP and/or ILR)

Values:

White: British, White Irish, any other White background (AOWB)

Asian: Indian, Pakistani, Bangladeshi, Chinese, any other Asian background (AOAB)

Black: African, Caribbean, any other Black, African or Caribbean background (AOBACB)

Mixed or Multiple: White and Black Caribbean, White and Black African, White and Asian, any other Mixed or Multiple ethnic background (AOMMB)

Any Other Ethnic Group (AOEG)

Missing

Reference group: White British

Notes: For the A level analysis, because of their very low numbers of entries, WIRT and WROM were merged with WOTH. For the GCSE analysis, WIRT, WROM and WOTH remained separate categories.

For the VTQ analysis, categorisations were harmonised between the NPD and ILR for consistency and to align with the Government Statistical Service guidance where

possible (some groups had to be merged into one of the 'other' groups, as numbers were too small to include as a separate category – for example, 'Gypsy or Irish Traveller' learners had to be merged into the 'Any Other White Background' group)

Special Educational Needs and Disabilities (SEND)

Source: DfE (NDP and/or ILR)

Values: With SEND, without SEND, missing

Reference group: Without SEND

Free school meals (FSM) eligibility

Source: DfE (NDP and/or ILR)

Values: Eligible for FSM, not eligible for FSM, missing

Reference group: Not eligible for FSM

Income Deprivation Affecting Children Index (IDACI) score

Source: DfE (NDP and/or ILR)

Values: Very low, low, medium, high, very high, missing

Reference group: Medium

Notes: Higher scores indicate higher levels of local deprivation and lower scores indicate lower levels of local deprivation. For the purpose of our modelling, these scores were divided into quintiles

Language

Source: DfE (NDP and/or ILR)

Values: ENG (English), OTH (other than English),missing/unknown

Reference group: English

Notes: Used for GQ analysis only

Region

Source: Awarding organisations

Values: EM (East Midlands), EA (East of England), LD (London), NE (north-east), NW (north-west), SE (south-east), SW (south-west), WM (West Midlands), Y&H (Yorkshire and the Humber)

Reference group: south-east

Notes: Used for GQ analysis only

Centre Type

Source: Awarding organisations

Values: Acad (academy), Free (free school), FurE (further education establishment), Indp (independent school [including city training college]), SecComp (secondary comprehensive or middle school), SecMod (secondary modern or high school), SecSel (secondary selective school), Sixth (sixth form college), Tert (tertiary college), Other (covering college of higher education, university department, tutorial college, language school, special school, pupil referral unit, HM Young Offender Institution, HM Prison, training centre).

Reference group: Academy

Notes: Used for GQ analysis only

Appendix D - Model specifications

Appendix D.1 – GQ Model Specification

All models fitted can be expressed mathematically as:

$y_{ijk} = \alpha + \beta x_{ijk} + u_i + u_j + u_k + e_{ijk}$,

where

y_{ijk} = [in numeric grade analysis] numeric grade awarded for exam entry by candidate *i* in centre *j* in subject *k*; [in grade probability analysis] exam entry by candidate *i* in centre *j* in subject *k* being awarded the target grade or not (1 or 0);

x_{ijk} = a set of background variables about candidate *i* in centre *j* taking the exam entry in subject *k*;

u_i = random intercept of candidate *i*;

u_j = random intercept of centre *j*;

u_k = random intercept of subject *k*;

e_{ijk} = entry level residual.

All analyses included student, centre and subject as random effects, to take account of students taking multiple subjects and students clustering within centres. The subject random effects had the following categories:

- A level: Art & Design: 3D Studies, Art & Design: Art, Craft and Design, Art & Design: Critical and Contextual Studies, Art & Design: Fine Art, Art & Design: Graphics, Art & Design: Photography, Art & Design: Textiles, Biology, Business Studies, Chemistry, Classical Greek, Computing, Dance, Drama & Theatre Studies, Economics, English Language, English Language & Literature, English Literature, French, Geography, German, History, Latin, Music, Physical Education, Physics, Psychology, Religious Studies, Sociology, Spanish
- GCSE: Art & Design: 3D Studies, Art & Design: Art, Craft and Design, Art & Design: Critical and Contextual Studies, Art & Design: Fine Art, Art & Design: Graphics, Art & Design: Photography, Art & Design: Textiles, Biology, Chemistry, Citizenship Studies, Classical Greek, Combined Science, Computing, Dance, Drama, English Language, English Literature, Food Prep and Nutrition, French, Geography, German, History, Latin, Mathematics, Music, Physical Education, Physics, Religious Studies, Spanish

The fixed effects of the models were: (please refer to Appendix A above for keys to the acronyms and abbreviations)

- Gender: male, female (reference category) (the unknown/neither category was omitted in the modelling because of the extremely small number of entries belonging to the category)
- Ethnicity: ABAN, AIND, AOTH, APKN, BAFR, BCRB, BOTH, CHNE, MOTH, MWAS, MWBA, MWBC, WBRI (reference category), WIRI, WIRT (only in GCSE analyses), WOTH (subsuming WIRT and WROM in A level analyses), WROM (only in GCSE analyses), unknown
- Major language: English (reference category), NotEnglish, unknown
- SEND status: NoSEND (reference category), SEND, unknown
- FSM eligibility: NoFSM (reference category), FSM, unknown
- Deprivation: very low, low, medium (reference category), high, very high, unknown
- Prior attainment: very low, low, medium (reference category), high, very high, unknown
- Centre type: Acad (reference category), Free, FurE, Indp, Other, SecComp, SecMod, SecSel, Sixth, Tert
- Region: EM, EA, LD, NE, NW, SE (reference category), SW, WM, Y&H

For the grade probability measures, we also ran logistic regressions in which the dependent variable y_{ijk} was the natural logarithm of the odds of the exam entry by candidate *i* in centre *j* in subject *k* being awarded the target grade. The exponentials of the β coefficients of a fitted logistic regression model are the odds ratios between groups which can be interpreted as estimates of the likelihood of entries by students of a particular group being awarded the target grade relative to entries by students of the reference group after controlling for other variables. For the between-group comparisons examined in our modelling, the pattern of changes in odds ratio across the years in the logistic models was consistent with the pattern of changes in relative outcome difference across the years in the linear probability models. We present results of the linear models for the grade probability measures in this report, as probabilities are more intuitive to interpret than odds ratios.

Appendix D.2 - VTQ model specification

A mixed effect modelling approach was also adopted for VTQ analysis, whereby the learner characteristics (including their prior attainment) were treated as fixed effects, and the learner's centre was treated as a random effect. This is because there may be factors affecting results that only learners from the same centre have in common, such as teaching materials used at their school or college.

The logistic regression specification used for modelling takes the form:

$$log\frac{P(Y_{ij}=1)}{1-P(Y_{ij}=1)} = \alpha + \beta X_i + \gamma Z_i + c_j$$

where:

- $P(Y_{ij} = 1)$ is the dependant variable. It determines the probability of achieving a top grade for candidate *i* at centre *j*
- X_i is a set of independent fixed effect variables summarising the demographic and socioeconomic characteristics of the candidate (see list of variables described above)
- Z_i is an independent fixed effect variable determining the highest level of prior attainment of candidate *i*. It is an additional variable that is controlled for in the logistic regression analysis
- c_j is a random effect determined by centre-level characteristics. It is also an additional variable that is controlled for in the logistic regression analysis
- α, β, γ are the regression coefficients

Appendix D.3 – Confidence intervals

All *β* coefficients had associated standard errors (SEs), which quantified how precisely *β* coefficients had been estimated. The SEs were used to compute the 95% confidence intervals (CIs) of the *β* coefficients, using the formula: 95% CI = *β* ± 1.96*SE. *β*s (taken as estimates of relative outcome differences after controlling for other background variables). The same calculation was used for both GQ and VTQ analyses.

Published by:

ofqual

Earlsdon Park
53-55 Butts Road
Coventry
CV1 3BH

0300 303 3344

public.enquiries@ofqual.gov.uk

www.gov.uk/ofqual

January 2023

Ofqual/22/6981/2