Grading and Admissions Data for England (GRADE) user guide

A data sharing framework for linked administrative data on education, assessment, and university admissions

ofqual

Contents

Introduction to the GRADE data sharing project

GRADE (Grading and Admissions Data for England) makes available linked micro-data from Ofqual, the Department for Education (DfE) and the Universities and Colleges Admissions Service (UCAS) covering the period from 2017 to 2022. The GRADE data set includes detailed information on students' progression from school to universities and colleges. It includes information on students' qualifications, prior schooling, socio-economic background and their university and college applications and admissions.

This user guide introduces researchers to the GRADE data sharing initiative and the GRADE data set. It provides an overview of the purpose and principles of making available linked data for independent research on the educational, assessment and admission systems in England. It also outlines the content of the GRADE data set and provides an overview of the data and linking procedures used for its production.

This user guide is published by Ofqual. It was updated in January 2025 alongside the second release of the GRADE data set.

Aim of GRADE

GRADE is a joint open data initiative by Ofqual, the Department for Education (DfE), and UCAS. The main aim is to make data available for independent research that serves the public good. Table 1 gives an overview of the data owning organisations' remit and reasons for sharing data.

The initial aim of the GRADE data sharing project was to allow external researchers to conduct independent high-quality evaluation of the judgements made in awarding grades in 2020. Now, GRADE aims to more broadly enable research to enhance the quality of the educational, assessment and university admission system in England and produce evidence to inform future education policy.

Organisations' remit and reasons for sharing data

Table 1. Organisations' remit and reasons for sharing data: an overview

Data owner	Status	Organisation's remit	Reasons for sharing data
Ofqual	Non-ministerial government department with jurisdiction over regulated qualifications provided in England	Regulate for the validity of qualifications, ensure fairness to learners in England and promote public confidence in the system	Facilitate the carrying out of programmes of research and retrieving evidence for purposes in line with its remit
Department for Education	UK government department	Responsible for education, children's services, higher and further education policy, apprenticeships, and wider skills in England, and equalities	Promote research and analysis to provide guidance or advice on education and/or well-being of children in England
UCAS	Charity operating the application process for UK universities	Provide evidence into higher education access and outcomes (Higher Education Research Act 2017, Section 79)	Promote more comprehensive statistical analysis to allow the performance of tasks carried out in the public interest and support efforts to promote research under the Digital Economy Act

The GRADE data set

GRADE makes available a data set of linked micro-data from Ofqual, DfE and UCAS covering the period between 2017 to 2022. The GRADE data set includes detailed information on students' progression from school to universities and colleges. It includes information on students' qualifications, prior schooling, socio-economic background and their university and college applications and admissions.

Research potential of the GRADE data set

The GRADE data set can be used for research on the educational, assessment and higher education admission systems in England. It provides opportunity to investigate linked data on students' granular attainment at GCSE and A level, their prior attainment, detailed socio-demographic characteristics and their applications to university and college. Covering alternative arrangements for grading in 2020 and 2021, it also provides a unique opportunity to compare and evaluate assessment methods.

Using the GRADE data set, researchers can for instance investigate questions around:
- equality and the effects of socio-demographic factors on education and university admission
- educational trajectories and the factors influencing students' choices at A level and applications to university
- fairness of assessment
- longitudinal trends in the educational, assessment and admission systems, particularly the effects of the COVID-19 pandemic

The GRADE research output repository provides researchers an up-to-date list of published research and analysis using the GRADE data set.

Structure of the GRADE data set

The GRADE data set includes 3 sources of administrative data. Each data source includes multiple data tables.

- **Ofqual data** – data on GCSE and A level qualifications and examinations collected from awarding organisations
  - awarding tables
  - grade boundaries table

- **DfE data** – extracts of the National Pupil Database (NPD) for key stage 4 and key stage 5 students
  - exam results tables
  - student tables
  - prior attainment tables
  - census tables

- **UCAS data** – data from the university and college application and admission process
  - apply qualifications tables
  - applications tables
  - applicant tables

The next section on the Content of the GRADE data set provides an overview of the information contained in the 3 data sources and each table. The data tables, within and across data sources, are linked together by common identifiers, see section on Data linking. More detail on the variables included in each data table of each dataset is provided in the data specifications.

Content of the GRADE data set

This section presents an overview of the information contained in the GRADE data set. More detail is provided in the data specifications.

The GRADE data set as a whole covers the period from 2017 to 2022. It covers 6 cohorts of students, including those who were awarded GCSE and A level grades in 2020 and 2021 when schools were closed, and exams were cancelled because of the COVID-19 pandemic.

Ofqual data

Ofqual data includes information on GCSE and A level qualifications and examinations taken by students of all ages in England.

The awarding tables include detailed information on the qualifications taken (such as subject and tier of entry) and on students' attainment. The grade boundaries table can be linked to the awarding table to help with interpreting marks and grades.

The awarding tables and grade boundaries table are based on Summer awarding data. Summer awarding data is provided to Ofqual by awarding organisations prior to grades being issued to centres and students in August each year. The grades and qualifications included in the summer awarding data are provisional. The awarding tables included in the GRADE data set enrich the summer awarding data with: information on final grades awarded after all types of reviews concluded, an indicator of Key Stage 2 attainment for GCSE students and indicators of GCSE attainment for A level students, and demographic information about students, such as gender and age. The grade boundaries table can be linked to the awarding table of the

corresponding year and qualification level using variables SpecCodeNoP and CertCodeNoP for linkage. The grade boundaries table is only available for 2022.

For the periods 2017 to 2019 and 2022 when exams were taken normally, the awarding tables include qualification grades and marks. In 2020 and 2021, exams were cancelled and alternative approaches to grading were taken. For 2020, the awarding tables include Centre Assessment Grades and calculated grades from the standardisation model, which were used to determine the awarded grade. For 2021, the awarding tables include Teacher Assessment Grades. For summer 2022, exams and other formal assessments went ahead with some planned adaptations intended to recognise the disruption to education caused by the COVID-19 pandemic, and grading was at a midpoint between summer 2021 and 2019.

Department for Education data

Data from the Department for Education is based on the National Pupil Database (NPD). The NPD is compiled by the Department for Education (DfE) from data supplied by schools, local authorities, centres and awarding organisations. The NPD constitutes the main source of information for the computation of accountability measures and is widely used for research purposes.

There are 4 main extracts of the NPD shared as part of GRADE. The NPD exam results tables contain student-level results data by qualification taken for Key Stages 4 and 5, covering A levels, GCSE and other general and vocational or technical qualifications, as provided by awarding organisations. The exam results tables include aggregated achievement indicators, such as those related to EBacc, Attainment 8 and Progress 8.

The NPD student tables contain data on students, including demographic and protected characteristics, such as gender, age, language spoken, free school meal eligibility and special educational needs. Information on the centre attended by students is also available, covering the type of school or institution, its description, and the centre's admissions policy.

The census tables contain additional information on students, including further socio-demographic characteristics (such as ethnicity) and socio-economic indicators such as the Income Deprivation Affecting Children Index (IDACI).

The prior attainment tables contain data on students' prior attainment including achievement indicators, teacher assessments and test results for Key Stages 1 and 2.

UCAS data

UCAS data is based on the information gathered to operate the higher education application process. This comprises data submitted by applicants to the UCAS undergraduate scheme and by the HE institution receiving prospective students' applications. UCAS data features three main data tables.

The applicant table contains information on applicants to the UCAS undergraduate scheme. For each applicant, this data set includes data on demographic characteristics (such as gender, age, geographical region) and socio-economic characteristics (such as ethnicity, socio-economic background, deprivation index).

The apply qualifications table is at qualification level and contains information on qualifications declared by the applicant during their application. This includes the A level grades predicted by teachers and submitted as part of an applicant's application to higher education.

The applications table contains the data included in the application made by each applicant, the kind of offer made by each HE provider and the response from the applicant. It does not contain the specific offer letters for individual students. This allows researchers to have access to a wealth of data, including: the applications that did not receive an offer, if an offer was received which kind of offer (unconditional, conditional) was made, whether each offer was accepted as either firm or insurance.

Data coverage

The coverage of the 3 data sources is slightly different. Table 2 summarises the provenance and the coverage of each data source. Ofqual data is collected by qualifications, DfE data is collected by Key Stage and UCAS data is collected by applications to the undergraduate scheme. This means that Ofqual data includes students of any age, while DfE data only includes students in Key Stage 4 and 5. UCAS data includes all university and college applicants which is a subset of Key Stage 5 students (from DfE data) and A level students (from Ofqual data).

Table 2: The provenance and coverage of each data source

Data source	Provenance	Population coverage	KS4 and GCSE included	KS5 and A level included
Ofqual	Summer awarding data as routinely collected on qualifications achieved by students and other data collections submitted by awarding organisations	GCSE and A levels achieved by learners in schools in England between 2017 and 2022	Yes	Yes
DfE	National Pupil Database - Submission of data on students as statutory requirement on schools, and exams results as submitted by exam boards	School-aged learners in England	Yes	Yes
UCAS	University application process - Data obtained as part of UCAS' position as the provider of a central admissions service for full-time undergraduate courses for higher education	English 18-year-olds for the 2017 to 2022 application cycles	No	Yes

	providers within the UK			

Data linking

Data from Ofqual, DfE and UCAS does not originally share any unique identifiers for students or schools. We produced such shared unique identifiers by linking data between the 3 data sources.

The linking procedure was based on the use of identifying information that is shared across the three data sources, including information about schools and students' personal information (first name, middle name, surname, date of birth and gender). Students' personal information was only used for data linking and will never be shared with external researchers.

Several linking rounds were performed, starting from exact matches and resuming with some lower-quality matching. A variable indicating the quality of the matching will be available to researchers who will therefore have the option to rely on different levels of linking quality. As a result of the linking procedure, 2 anonymised identifiers, one for each student and one for each school, were generated. These identifiers are not used in any existing operational systems by any of the data owners

Data linking was done by Ofqual. Any questions about data linking can be addressed to data.sharing@ofqual.gov.uk.

Data limitations

Data included in the GRADE data set was originally collected for non-statistical reasons, such as for the delivery of a public programme or service, the delivery of qualifications, or for maintaining school records. Research needs are not generally considered as part of the data collection design. As a result, the GRADE data set
- may lack potentially useful information
- may contain errors and systematic inaccuracies related to the data collection procedures

GRADE includes administrative data that is also used for official statistics or in other research reports published by the data owners. Minor differences between the GRADE data set and those other published statistics are expected due to separate processing.

Researchers are not allowed to link the GRADE data set to additional data on individual students, schools, exam boards or HE providers that may be available from external sources.

Releases and updates to the GRADE data set

The GRADE data set was first released in 2021. This first release covered the time period between 2017 and 2020. In 2024, the GRADE data set was re-released and updated to additionally cover 2021 and 2022.

Data specifications of this second release of the GRADE data set differs slightly from the first release. In the current GRADE data set, all body corporate identifiers (such as names of exam boards, schools and universities) have been pseudonymised. This includes body corporate identifiers which had not been pseudonymised in the first release of GRADE data.

Related data sets

- National Pupil Database (NPD)
- UCAS Undergraduate Data Release Archive
- Longitudinal Education Outcomes (LEO)
- Growing Up in England (GUiE)

Data governance: accessing the GRADE data set

Requesting access

The GRADE data set is shared by the Office for National Statistics under the Digital Economy Act 2017 and can be accessed under the Research Accreditation Framework via the Secure Research Service (SRS) and/or the Integrated Data Service (IDS).

To request access to the GRADE data set, researchers need to complete 2 steps:
- Become an accredited researcher with the Office for National Statistics.
- Complete an application for the specific research project

Becoming an accredited researcher

The GRADE data set is only shared with accredited researchers, who have demonstrated the skills and knowledge needed to use the data appropriately. Becoming an accredited researcher allows researchers to gain access to data stored in the Secure Research Service (SRS) and/or the Integrated Data Service (IDS) and to carry out analyses in the SRS or IDS environment.

 Full guidance to apply to become an accredited researcher is available from the Office for National Statistics. In brief, to be an accredited researcher, applicants must have an undergraduate degree (or higher) including a significant proportion of maths or statistics or be able to demonstrate at least 3 years of quantitative research experience. Successful completion of a 'Safe Researcher' training course is also required as part of the accreditation process.

Researchers can apply for accreditation through the Research Accreditation Service. Once granted, accredited researcher status is valid for 5 years, after which the researcher will need to apply again.

Applying for a specific research project

To request access to the GRADE data set, researchers have to submit a research project application to the Office for National Statistics Research Accreditation Service. The research project application includes a list of all people involved in the research, a detailed research proposal and a check for ethical approval of the research.

Researchers should apply only for the data they need for their research project. In their application, they should select the relevant data sources, which may be:
- Ofqual-only
- Ofqual and DfE
- Ofqual and UCAS
- Ofqual, DfE and UCAS

DfE-only data and UCAS-only data cannot be requested as part of GRADE.

In the project application form, researchers must state how and for which purpose each data source will be used. Researchers should clearly state the connection between the aim of their research project, the Aim of GRADE and the data owner's remit. This constitutes a key element for the accreditation of the project.

Researchers are advised to use the research project application example guidance. They can also refer to an example research project application, which has a similar structure and format to GRADE research project applications.

Researchers who are yet to become accredited researchers can still submit a research project application, but they will have to be accredited before being able to access the GRADE data set (as described in the section on Becoming an accredited researcher.

Ethical considerations

The GRADE data set can only be used for research that is legal, ethical and feasible. Research projects must demonstrate a clear public benefit. Research outcomes must be published. Detailed criteria can be found on the UKSA website.

The GRADE application review process

The application review process has 3 stages:
- **Stage 1.** Applications are received and checked by the Office for National Statistics Research Accreditation Service.
- **Stage 2.** Applications that pass the initial check, and feasibility assessment are reviewed by Ofqual and the Department for Education as data owners. UCAS chose to be informed about applications but not to conduct a data owner review.
- **Stage 3.** Applications that are approved by the data owners are then scrutinised and accredited or rejected by the Research Accreditation Panel

At each stage of the process, researchers can expect to receive feedback and requests to make minor or major revisions to their application.

Data security

To ensure the safe access and secure use of the data, the Office for National Statistics Secure Research Service (SRS) and Integrated Data Service (IDS) use the Five Safes Framework. The 5 safes are:
1. **Safe People** – only trained and accredited researchers are trusted to use data appropriately.
2. **Safe Projects** – data is only used for valuable, ethical research that delivers clear public benefits.
3. **Safe Settings** – access to data is only possible using secure technology systems.

4. **Safe Outputs** – all research outputs are checked to ensure they cannot identify data subjects.
5. **Safe Data** – researchers can only use data that has been de-identified.

Safe People and Safe Projects: accreditation of researchers and projects

The first 2 safes are achieved through the accreditation of researchers and the accreditation of projects (as detailed in [Data governance: accessing the GRADE data set](#)).

Safe settings: Office for National Statistics Secure Research Service (SRS) and Integrated Data Service (IDS)

External researchers access the data through the Office for National Statistics [Secure Research Service](#) (SRS) and/or [Integrated Data Service](#) (IDS) which are Accredited Processors under the [Digital Economy Act (2017)](#).

The use of the SRS and IDS requires a [safe setting](#) to access data. Once a project has been accredited, external researchers can access the GRADE data through Office for National Statistics [safe rooms/ SafePod](#) or safe settings provided by approved Assured Organisational Connectivity organisations. The most popular route to use the SRS or IDS is under the [Assured Organisational Connectivity scheme](#) which enables researchers from approved organisations to access projects directly from their organisational office or remotely from home. The organisation must achieve certification under the scheme which ensures that they meet the safe setting criteria by using secure technology and systems.

Safe outputs: Statistical Disclosure Control

A safe output is an output that is non-disclosive and maintains the confidentiality of the data subjects. All research outputs are checked by the SRS or IDS Statistical Support team to ensure that data subjects cannot be in any way identified. This involves the application of Statistical Disclosure Control methods to ensure that all outputs are non-disclosive. Any statistical output computed on fewer than 10 data subjects will not be disclosed and cannot be exported outside of the SRS or IDS. The SRS or IDS Statistical Support team provide guidance and checks until the publication of research outcomes.

Researchers working in the SRS and/or IDS are not allowed to copy or export extracts of the data. Researchers are not allowed to report or discuss their findings until their output is checked. This is to ensure that no information regarding the data can leave the system without being cleared by the SRS or IDS Statistical Support team.

Safe data: de-identification and pseudonymisation

External researchers will be only able to access data that has been de-identified and pseudonymised. Researchers are required not to identify individuals from the de-identified data and are reminded that it would be a criminal offence to do so. An Individual Declaration Form is required to be signed by each accredited researcher before access is given.

Personal identifiers (including first name, surname, date of birth, school or college attended) are used for linking the data across data sources (Data linking) but will never be shared with researchers. Instead, meaningless identifiers created for each student and each school. A pseudonymisation algorithm was applied to both identifiers.

Legislation

Legal gateways to data sharing and processing

Ofqual, DfE and UCAS maintain ownership of their respective data within the GRADE dataset. Ofqual, DfE and UCAS act as joint controllers of the GRADE dataset.

Ofqual may co-operate or work jointly with another public authority where it is appropriate to do so for the efficient and effective performance of its qualification functions under section 156 of the Apprenticeships, Skills, Children and Learning Act 2009. This encompasses the sharing of data.

DfE is able to onwardly share "individual pupil information" collected from schools through Section 537A(4) of the Education Act 1996 with "Prescribed Persons". The Education (Individual Pupil Information) (Prescribed Persons) (England) Regulations 2009 go on to prescribe which persons may be provided with "individual pupil information" under section 537A(4) of the Education Act 1996, although this was subsequently amended and updated in the Education (Individual Pupil Information)

(Prescribed Persons) (England) (Amendment) Regulations 2013.(ii) Legislation within section 253A of the Apprenticeships, Skills, Children and Learning Act 2009 covers the sharing of Key Stage 4, Key Stage 5 for learners in FE colleges with named bodies and third parties as prescribed in the Education (Student Information) (England) Regulations 2015.

UCAS data sharing is in accordance with Article 6(1)(f) of UK GDPR, and related Data Protection Act provisions, for the purposes of legitimate interests pursued by the controller or by a third party where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child. This sharing actively supports efforts to meet government aims to encourage fairness and scrutiny of the higher education application process and promotes research in this field in accordance with the aims of the Higher Education and Research Act 2017.

The UK GDPR allows the processing of personal data by an organisation where "processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller" (Article 6(1)(e)) and states that special categories of data may be used for historical research or statistical purposes (Article 9(2)(j)).

Article 6(1)(e) of the UK GDPR allows DfE and Ofqual to process personal data for their public task and Article 6(1)(f) of the UK GDPR allows UCAS to process personal data for the purposes of legitimate interests.

In addition, Article 9(2)(j) UK GDPR allows all organisations to process special category data where it is necessary for the purposes of research (such processing to be proportionate to the aims pursued and subject to safeguards). Special categories of data may be processed for research purposes when (i) this is in the public interest, (ii) the processing is not likely to cause substantial damage or distress to a data subject, and (iii) the processing is not carried out to take measures or make decisions with respect to a particular data subject.

Also, Article 9(2)(g) UK GDPR allows the processing of special categories of data for reasons of substantial public interest to ensure equality of opportunity or treatment with regards to certain categories of data concerning specific groups of people.

Legal gateways to accessing data for research

The Five Safes Framework established under the Digital Economy Act (2017) enables governmental bodies to share administrative data for the purposes of research. Only researchers accredited by the UK Statistics Authority can have access to the data to conduct research projects also accredited by the UK Statistics

Authority. To be shared under the Digital Economy Act, however, the data has to be completely de-identified or functionally anonymised.

Any researcher who wants to access the data must be accredited (see Becoming an accredited researcher) and the research project must have been approved (see The GRADE application review process). In order for research projects to be approved they must comply with the Research Code of Practice and Accreditation Criteria which was approved by the UK Parliament in July 2018.

Data processing involved as part of this initiative is compliant with all applicable data protection legislation, including the UK General Data Protection Regulation (UK GDPR) and Data Protection Act 2018.

Data privacy

A key transparency requirement under the UK GDPR is that individuals have the right to be informed about the collection and use of their data. The intent to enable independent research by making data available to external researcher was made public from the beginning of the GRADE initiative.

Each data owning organisation publishes a privacy notice on how it processes and protects personal information:
- Ofqual's personal data protection policy
- Department for Education's NPD privacy notice
- Department for Education's personal information charter
- UCAS privacy notice

Given that this project is based on the linking of data sets from the different organisations which will broaden the set of information available for each student, all data owners also jointly issued:
- Privacy notice for the initial release of the GRADE data set
- Privacy notice for the second release of the GRADE data set with additional data for 2021 and 2022

The privacy notices provide students with information on why their data is shared, how their data is processed, the retention period for the data, and how the data will be shared.

Resources

A range of useful resources are available to researchers interested in working with the GRADE data set.
- data specifications, with the details (names and descriptions) of all the variables in each data table for all three data sources

- a low-fidelity synthetic GRADE data set which is available upon request from Ofqual data.sharing@ofqual.gov.uk[1]
- a research output repository, which provides an up-to-date list of the research and analysis conducted with the GRADE data set

Funding opportunities

Research projects using the GRADE data set should be independently funded. To promote the use of the GRADE data set, Ofqual, DfE and UCAS have partnered with Administrative Data Research UK to help researchers overcome funding barriers.

In 2021, ADR UK launched a funding call for 2 GRADE research fellowships with funding for up to 12 months in duration and a maximum of £130,000.

In 2024, GRADE was included in the ADR UK Research Fellowship call for using ADR England flagship datasets, with funding up to 18 months in duration and up to 80% of a maximum of £200,000 at full economic cost.

Future funding opportunities will be advertised here if and when they become available.

Contact

For further information about accessing the GRADE data set and the application and accreditation process:
- Office for National Statistics the Secure Research Service (SRS) srs.customer.support@ons.gov.uk
- Office for National Statistics Integrated Data Service (IDS) ids.customer.support@ons.gov.uk

For further information about the GRADE data set, related resources or for research feasibility queries or feedback about the GRADE data set:
- Ofqual data.sharing@ofqual.gov.uk

---

[1] Low-fidelity synthetic data refers to a data set with a similar format and structure to the real data set but containing only made-up information based on simulated data. No real data on students is featured in the synthetic data. To know more about synthetic data, see the report published by the Behavioural Insights Team. To access the low-fidelity synthetic GRADE data set please contact data.sharing@ofqual.gov.uk.

For queries specific to the individual data sources in the GRADE data set or for information about organisational remit and data governance, please contact the data owner:

- Ofqual data.sharing@ofqual.gov.uk
- Department for Education data.sharing@education.gov.uk
- UCAS stats@ucas.ac.uk

For queries regarding ADR UK funding opportunities:

- ADR UK - hub@adruk.org

January 2025   **Ofqual/25/7198**