

RESEARCH AND ANALYSIS

# Grading and Admissions Data for England (GRADE) framework

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

**ofqual**

# Contents

|  |           |
|--|-----------|
| <b>Introduction to the GRADE data sharing project .....</b>                    | <b>3</b>  |
| Purpose of the initiative.....   | 3         |
| Principles underpinning the GRADE initiative .....                             | 3         |
| <b>The GRADE data.....</b>   | <b>5</b>  |
| GRADE at a glance.....   | 5         |
| The content of the GRADE data .....  | 7         |
| Data linking.....  | 10        |
| Data limitations.....  | 11        |
| <b>Data governance: accessing the GRADE data .....</b>                         | <b>12</b> |
| Requesting access to the data .....  | 12        |
| The GRADE application review process .....                                     | 15        |
| Timeline .....   | 18        |
| <b>GRADE and data protection.....</b>  | <b>19</b> |
| Data security.....   | 19        |
| Legislation .....  | 21        |
| Data privacy.....  | 22        |
| <b>Promoting the use of the GRADE data .....</b>                               | <b>23</b> |
| Data documentation and research output repository .....                        | 23        |
| Funding opportunities .....  | 23        |
| <b>Further information.....</b>  | <b>25</b> |
| <b>Appendix.....</b>   | <b>26</b> |
| Figure 1: Structure of GRADE data sources – accessible version .....           | 26        |
| Figure 2: The application review process at a glance – accessible version..... | 26        |

# Introduction to the GRADE data sharing project

## Purpose of the initiative

The GRADE (Grading and Admissions Data for England) data sharing project is a joint open data initiative by Ofqual, the Department for Education (DfE), and UCAS. Its main objective is to provide accredited external researchers with access to data to conduct independent research on the educational, assessment and admission systems in England, and to facilitate internal analysis.

The GRADE initiative makes available, for the first time, linked data from Ofqual, DfE and UCAS covering exam results from 2017 to 2019 and grades from summer 2020. In summer 2020, with the closure of schools and cancellation of exams due to the coronavirus (COVID-19) pandemic, alternative awarding arrangements were put in place to allow students to progress to the next stage of education or into the workplace. The initial approach to award standardised grades did not command public confidence and a late decision was taken to change arrangements. It was apparent that trust in the system was damaged.

The GRADE data sharing project aims at rebuilding trust and confidence in the assessment system by showing commitment to transparency and accountability. It will allow external researchers to conduct independent high-quality evaluation of the judgements made in awarding grades in 2020, ensuring that the right lessons are learned. More broadly, the GRADE data will enable research to enhance the quality of the assessment system and produce evidence to inform future education policy, particularly around the fairness of methods for measuring students' attainment, and implications for university admissions processes.

This initiative also intends to strengthen the relationships between policy makers and the statistical, policy and education research communities to ensure that the right research questions are asked. This way, teachers will be better informed about the outcomes of their teaching, and school leavers will have more information and guidance to think about their future.

## Principles underpinning the GRADE initiative

The GRADE initiative is based on the sharing of micro-data: detailed information on each student, the qualification they took, their prior schooling, their socio-economic background, their preferences and choices in relation to progress. Given the wealth of information in the GRADE data, ensuring the security of the data and protecting

the identity of young people was of primary importance and constituted the first underpinning principle of the initiative. This meant designing a data set including only de-identified data, using a secure system to share it and putting in place further controls.

In order to support meaningful research, the GRADE data is rich in nature. To be used effectively, this data requires skills and knowledge of statistical techniques that only trained and experienced researchers possess. In light of that, and as a further means of ensuring appropriate use of the data, the second principle was that only trusted researchers could access the data. Furthermore, researchers with the right credentials will only be allowed to use this data to conduct research for the public good. Operational and commercial uses of the data are not allowed.

It was then necessary to ensure the independence of the research conducted with this data. This was deemed to be a key step to show full transparency and accountability. However, to make sure lessons are learned and that there is a link between research and policy, it was considered essential to ensure that the data owners keep track of how the data is used and, especially, maintain sight of the findings.

Last, but not least, we considered it necessary to engage with the statistical research and education policy communities. This would allow us to understand their needs and define the data in ways that could serve to answer as many key research questions as possible. It was also key to designing the process to grant access to the data, following best practices in the creation and sharing of data for administrative purposes. We therefore established an [advisory committee with key representatives of the statistical research and education policy communities](#).

# The GRADE data

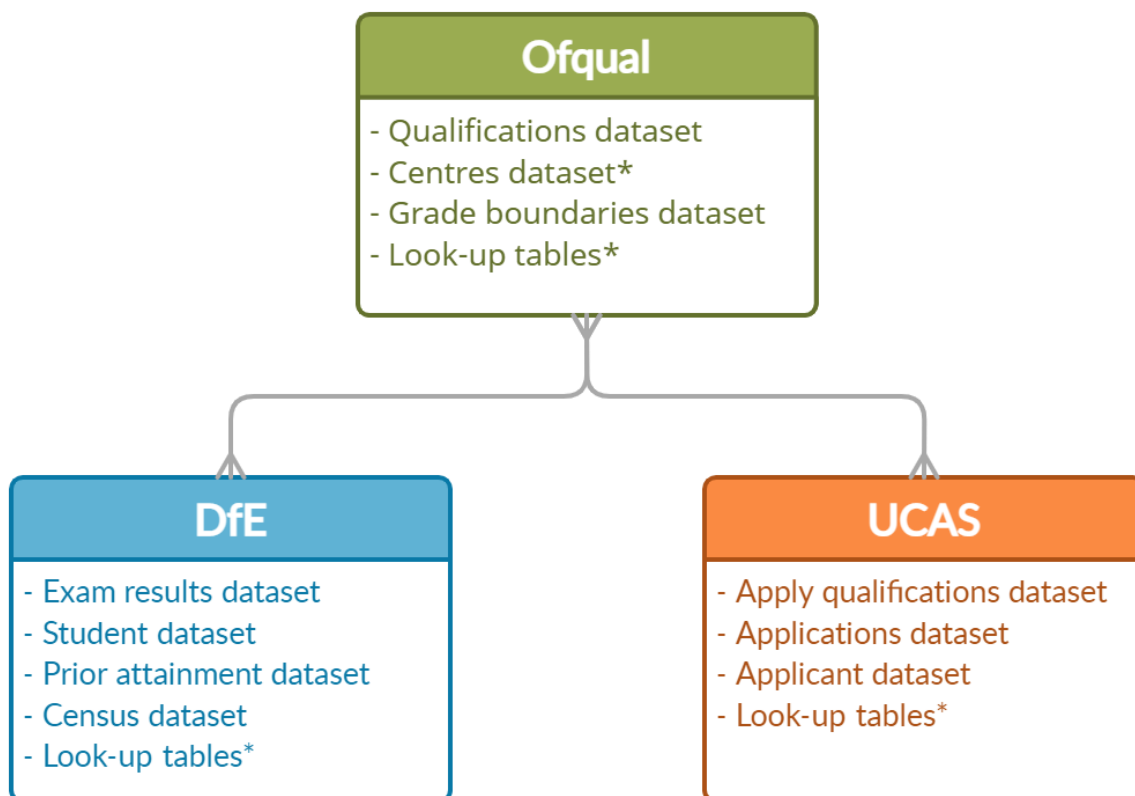
## GRADE at a glance

The data shared through this initiative includes 3 main sources of administrative data:

- Ofqual – including data on GCSE and A level examinations and qualifications collected from awarding organisations.
- DfE – including extracts of the National Pupil Database (NPD) for GCSE and A level students.
- UCAS – including data from the university application process.

Figure 1 shows the structure of the GRADE data at a high level. For each data source, a number of different data tables were created. This structure has been designed around the needs of data users and data manageability as it will be easier for external researchers to work with smaller data sets. This approach has allowed us to maintain the structure of the data sets that are already known by researchers. There are also advantages from a data protection perspective, as this will allow researchers to apply for data needed by their research project, making sure that only the necessary data is actually shared (described further in the section on ‘Data governance: accessing the GRADE data’). The modularity of the data shared will potentially allow adding further data in the future, without having to change its current structure.

Figure 1: The structure of the data at a glance



Data tables marked with \* will not be shared for data protection reasons

The different data tables, within and across data sources, are linked together by common identifiers (see section on ‘Data linking’). This will make possible, for example, for Ofqual data on grading of qualifications to be used alongside the rich set of information on pupils available from the National Pupil Database (NPD) held by DfE. [Research has already been conducted on this data](#) to show its potential and how the data can be used. Additional data, for example on predicted grades, will also be available in the UCAS data and linked to Ofqual and DfE data.

Potential areas of research were used as drivers to define the content of the GRADE data set. Key research questions were identified through a workshop with staff in each of the participating organisations and then developed in conversation with the members of the advisory committee.

For all the 3 data sources, the data covers a number of years. Currently, the GRADE data covers the period 2017 to 2020. This is to ensure that all the data used for the [standardisation model introduced in summer 2020](#) is available to external researchers so that the model can be independently assessed. Further research conducted on standardisation methods and their impact on students will also be possible. This, combined with the fact that the data used to develop and run the

standardisation model in 2020 is augmented with additional data from DfE and UCAS, will allow researchers the access to a much wider set of information on students than those available when the model was developed.

More broadly, in addition to potential reviews of the standardisation model, the GRADE data will allow independent evaluation of summer 2020 grading and of the judgements made in awarding grades. Making data available from 2017, 2018 and 2019, alongside data from 2020, means that the GRADE data can also be used to gain insights from the functioning of the assessment system in years when exams are available. Thanks to the information provided by UCAS, research aimed at informing university and college entrance policy will also be possible.

## The content of the GRADE data

This section presents an overview of the GRADE data. More detail is provided in the [data specifications](#).

### Ofqual data

Ofqual data includes information on each GCSE and A level qualification as provided by awarding organisations and taken by pupils in England, regardless of their age. The backbone of the Ofqual data shared as part of this initiative is the 'summer awarding' data, that is, the data collected for each qualification taken by pupils prior to grades being issued to centres and students in August each year. The Ofqual data will also include the grade awarded to pupils after any type of review.

The qualifications data set includes detailed information on the qualifications taken (such as subject, awarding organisation, tier of entry), as well as granular data on pupils' attainment (such as grades and marks) for the period 2017 to 2019. Information about pupils, such as gender and age, is also contained in this data set. For 2020, given that exams were cancelled and alternative approaches to grading taken, each entry was associated with a Centre Assessment Grade and calculated grade, which were used to determine the awarded grade. These alternative grading approaches will also be included in the qualifications data set, along with details of pupils' prior attainment: an indicator of Key Stage 2 attainment for GCSE pupils, indicators of GCSE attainment for A level students.

As described by Figure 1, in addition to the qualifications data set, Ofqual data features a grade boundaries data set to help to interpret the marks in the main qualifications data set.

## Department for Education data

The NPD is compiled for DfE from data supplied by local authorities, centres and awarding organisations. The NPD is made up of a number of data collections, linked together with a unique identifier, covering all pupils within schools in England up to the age of 19. It constitutes the main source of information for the computation of accountability measures and is widely used for research purposes.

As described by Figure 1, there are 4 main extracts of the NPD shared as part of this initiative. The NPD exam results data sets contain pupil-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. It includes aggregated achievement indicators, such as those related to EBacc, Attainment 8 and Progress 8.

The NPD student data sets contain data on pupils, 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 pupils is also available, covering the type of school or institution, its description, and the centre's admissions policy. Additional information on pupils is also available in the census data sets. This includes further socio-demographic characteristics (such as ethnicity) and socio-economic indicators (such as IDACI score, Income Deprivation Affecting Children Index). The NPD also contains data on pupils' prior attainment including achievement indicators, teacher assessments and test results for Key Stages 1 and 2.

## UCAS data

Universities and Colleges Admissions Service (UCAS) data is based on the information gathered to operate the university application process. This comprises data submitted by applicants to the UCAS undergraduate scheme and by the HE institution receiving perspective students' applications. UCAS data features 3 main data tables.

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



The applications data set contains the data included in the application made by each applicant, the offer made by each HE provider and the response from the applicant. 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

From the descriptions of the data provided it is possible to infer that, given the purposes for which data were initially collected, the coverage of different data sources may be slightly different. Table 1 summarises the provenance and the coverage of each data source, from which some differences become apparent. It is worth noticing that first of all, UCAS data refers to university applicants that are a subset of Key Stage 5 learners (from DfE data) and A level students (from Ofqual data). Furthermore, whereas Ofqual data are collected by qualifications, DfE data is collected by Key Stage. This means that Ofqual data features pupils of any age, while DfE data only features students in Key Stage 4 and 5.

Table 1: The provenance and coverage of each data source

| <b>Data source</b> | <b>Provenance</b>   | <b>Population coverage</b>   | <b>KS4 and GCSE</b> | <b>KS5 and A level</b> |
|--------------------|---|--|---------------------|------------------------|
| <b>Ofqual</b>      | Summer awarding data as routinely collected on qualifications achieved by pupils and other data collections submitted by awarding organisations   | GCSE and A levels achieved by learners in schools in England between 2017 and 2020 | Yes                 | Yes                    |
| <b>DfE</b>         | National Pupil Database - Submission of data on pupils as statutory requirement on schools, and exams results as submitted by exam boards   | School-aged learners in England  | Yes                 | Yes                    |
| <b>UCAS</b>        | 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 providers within the UK | English 18-year-olds for the 2017 to 2020 application cycles                       | No                  | Yes                    |

## Data linking

The data collected by each data owner was not initially intended to be used alongside other data. For this reason, in the source data sets there are no unique personal identifiers that allow pupils within centres to be tracked across data sets from different data owners. Without a unique personal identifier, we rely on comparing other identifying information, such as names and dates of birth. This personal information will not be shared with external researchers but has been used to link records belonging to the same pupils across different data sets. As a result of the linking procedure, two meaningless identifiers, one for each pupil and one for each centre, have been generated for these data sets. These identifiers are not used in any existing operational systems by any of the data owners.

The linking procedure was designed on the basis of common practice used by the 3 data owners to link their data and data from other sources. In brief, the linking procedure was based on the use of identifying information that are common across the data sets held by all data owners, including: first name, middle name, surname,

date of birth and gender. To allow for potential errors in the source data sets a number of linking rounds were performed, starting from exact matches and also 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. The details of the linking procedure and its results will be published in [the user guide that will be available on the GRADE webpage](#).

## Data limitations

The GRADE data is based on administrative records. The data 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. This means that research needs are not generally considered as part of the data collection design. This can clearly lead to limitations of the GRADE data and may result in the lack of potentially useful information. It can also imply the possibility of errors in the data, depending on how the data was recorded. With appropriate controls and safeguards in place, administrative data sources can be a rich source of information for quantitative analysis and evaluation, without imposing an additional burden or risks on data subjects.

The data shared through this initiative includes work to structure data sets for sharing and linking. While the same underlying administrative data sources may be used for other research reports and for Official Statistics, differences are expected due to separate processing. Although we do not anticipate material differences in trends or conclusions, researchers should be aware that analysis carried out using these data sets may not be exactly comparable to other published statistics or research.

It should also be noted that the GRADE data sets may allow to be further linked to additional information, such as information at geographical level. However, being de-identified at student, and school or college level, the GRADE data cannot be linked to any external data source at student, school or college level.

# Data governance: accessing the GRADE data

## Requesting access to the data

From an external researcher's perspective, there are 2 steps needed to access the data in the Office for National Statistics (ONS) Secure Research Service (SRS):

1. become an accredited researcher with ONS
2. apply for access to the specific data required for a research project

## Becoming an accredited researcher

Only accredited researchers can access the data through the ONS SRS. Becoming an accredited researcher allows researchers to carry out analysis and produce outputs on projects in the SRS.

Full [guidance to apply to become an accredited researcher](#) is available from the Office of 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 \(RAS\)](#). Once granted, accredited researcher status is valid for 5 years, after which the researcher will need to apply again.

## Apply for access to the GRADE data

### The research project application form

To request access to the GRADE data, external researchers will have to submit a research project application to the ONS SRS. Accredited researchers must submit a research proposal by completing an application for project accreditation in the [Research Accreditation Service \(RAS\)](#). 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 data (as described in the section on 'Becoming an accredited researcher').

Before submitting the project application form, researchers are advised to use the [research project application example guidance](#). They can also refer to an [exemplar research project application](#), provided by ONS SRS. In addition to the list of those involved in the research, the application form will prompt researchers to provide a description of the research they are proposing and details of the methodology they intend to employ.

Researchers will be requested to provide the specifics of the data required. Researchers are expected to apply for the data they need for their research by selecting the relevant data items as prompted on RAS. This means that they will have to specify which data researchers are requesting access to. This can be:

- Ofqual-only
- Ofqual and DfE
- Ofqual and UCAS
- Ofqual, DfE and UCAS

It should be noted that DfE-only data and UCAS-only data cannot be requested as part of the GRADE initiative. DfE data and UCAS data are already accessible through the ONS SRS.

In completing the project application form, researchers will have to state how and for which purpose each data source will be used, making a clear connection between the aim of the research and how this contributes to the Purpose of the initiative and to each participating organisation's remit (Table 2). This will constitute a key element for the accreditation of the project proposal.

Table 2: Organisations' remit and reasons for sharing data: an overview

| <b>Data source</b> | <b>Status</b>   | <b>Organisation's remit</b>   | <b>Reasons for sharing data</b>  |
|--------------------|---|---|--|
| <b>Ofqual</b>      | 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  |
| <b>DfE</b>         | 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  |
| <b>UCAS</b>        | Charity operating the application process for British 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 supplementary form

Depending on the specifics of the data requested through the project application form, additional controls may be needed and researchers may be required to provide further information in relation to the processing of data, via a supplementary form. Applicants who are required to submit the supplementary form will be notified shortly after the submission of the RAS form to ONS. They will receive the supplementary form from ONS via email along with instructions on how to complete and submit it.

The information provided by the researchers in the supplementary form is necessary to ensure that the data can be shared securely.

## Ethical considerations

Data can only be used for valuable research that delivers clear public benefits. In addition, the use of the data must meet the ethical standards appropriate to the nature and intended use of personal information. This is necessary to show the contribution of the research project to the public good and to identify ethical issues, that is to support an accurate and consistent estimation of the 'ethical risks' and benefits of research proposals.

In order to meet these criteria, researchers will be requested to provide proof of ethical approval of their research proposal. Researchers can seek ethical approval within their institutions, through their ethics committee. Alternatively, researchers can use the [ethical self-assessment tool provided by the UK Statistics Authority and available online](#) along with additional guidance.

## The GRADE application review process

An overview of the application review process is represented by Figure 2. The process is initiated by the submission of the research application process (at the top left of the diagram) by the researcher. Once the research project application is submitted through the ONS SRS, each application will be reviewed by ONS SRS and by each of the organisations providing the data applied for. Details of the review process set up to ensure sound governance of the data shared with external researchers are provided in the ONS SRS review section.

## ONS SRS review

As illustrated by Figure 2, the ONS SRS will conduct 2 types of checks when a project application is submitted:

### First check –conducted by a member of the SRS Data Access team

The SRS Data Access team check the application for the following:

1. Have all the required fields been completed?
2. Has all the necessary information been provided?
3. Has the required level of detail been provided in each section?
4. Has ethical approval or an ethical self-assessment form been submitted, and does this provide the necessary level of detail?

## Feasibility check –typically conducted by a member of the SRS Statistical Support team

The SRS statistical support team are responsible for:

1. Reviewing the methodology of the proposal to ensure that the intended research can actually be pursued with the requested data.
2. Checking that the methodology provides the correct information and level of detail.

The ONS SRS will query any sections of the application which does not pass these checks and provide guidance to the researcher on how to improve the application. Once the application is deemed to be up to the required standard, the ONS SRS will send the application to data owners for their approval.

It is at this stage that, depending on the specifics of the data requested through the project application form, a researcher may be asked to provide a supplementary form. If so the ONS will contact the external researchers and provide them with the supplementary form and instructions on how to complete it and submit it.

The supplementary form will be submitted directly to ONS and will be assessed by data owners alongside the research project application submitted to ONS through RAS. The data owners' review will not start until the supplementary form is submitted.

## Data owners' review

Data owners will review the application (including the supplementary form, if required) in 2 steps.

First, there will be an independent review of each owner of the data applied for. This means that for Ofqual-only data, applications will only be reviewed by Ofqual. If the data applied for includes DfE and/or UCAS data, the relevant organisations will also review the applications independently. The focus of the data owner review will be on the purpose of the proposed research to assess whether this is in line with the organisation's remit and the reasons for which the data is shared (see Table 2 for an overview). Where the researchers have been asked to submit a supplementary form, data owners will also assess the information provided in the supplementary form.

Second, the outcome of each independent review will then be submitted to the GRADE Project Board, where each organisation will be represented by at least one member of each data owning organisation with authority to approve the sharing of data. Each data owner will have a final say regarding the sharing of data they own and the project board will not be allowed to overrule the decision taken by each



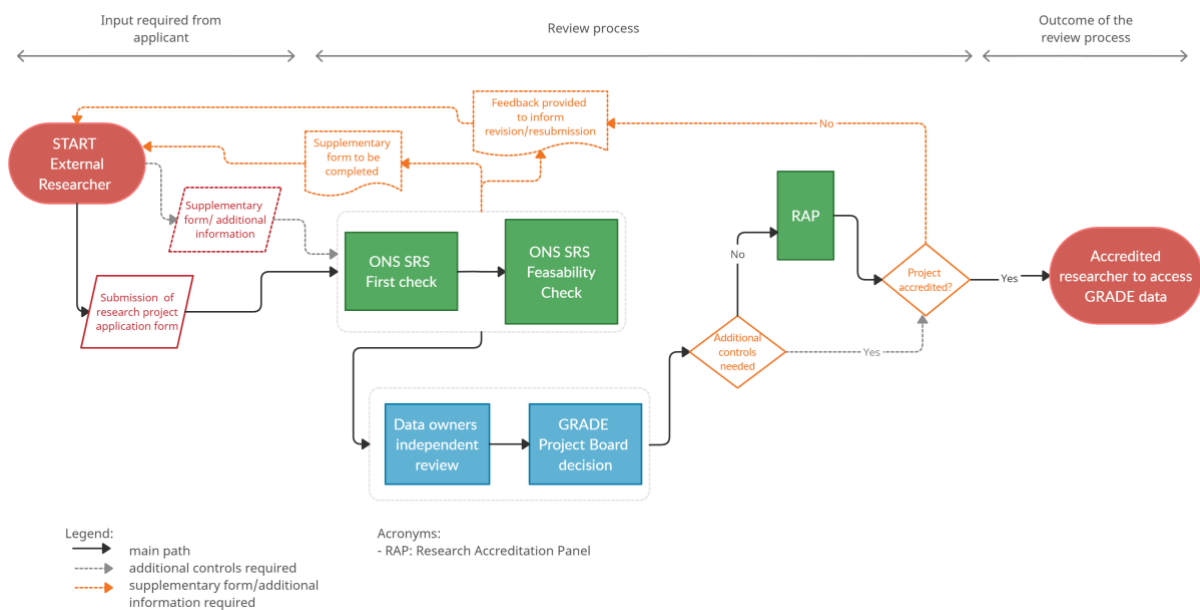
individual organisation. With respect to the project application review process, the GRADE Project Board has 2 key roles:

1. Ensure that all the participating organisations have sight of the decisions taken by other organisations.
2. Maintain visibility of all the research project applications approved by data owners.

The project board will also coordinate the data owners' response to ONS SRS and/or to the accredited researcher.

## The application review process

Figure 2: The application review process at a glance



## The outcome of the review process

Each project proposal will be assessed against the following criteria:

1. Is there public benefit?
2. Is there demonstrable analytical merit?
3. Is the project feasible?
4. Is the data requested appropriate to answer the research questions?
5. Is the statistical method to be used appropriate to answer the research questions?

6. Do any relevant data protection implications arise and are they sufficiently mitigated?
7. Has the project successfully completed a formal ethical review?

Depending on the data requested, access to the data may be granted by the [Research Accreditation Panel \(RAP\)](#) or directly by the data owners. In either case, the researcher will be notified and provided with the documentation to sign to be able to access the data through ONS SRS. Once all the documentation is signed by the researcher, the extracts of the data required can be prepared and the project can start.

## Feedback to researchers

The review process is rigorous, and feedback may be given to researchers at each stage of the process (indicated by dashed connectors in Figure 2). This is to make sure that all the necessary information is provided, the contribution of the research is clear, and the methodology to employ is sound. At each stage of the process, researchers can expect to receive feedback and requests to make minor or major revisions to their proposal. This should minimise the risk that the project is rejected at the end of the review process, therefore reducing the lead time from the submission of the research project application form to the actual access to the data.

## Timeline

Given the process described above, the potential need for additional information, and the possibility that the proposal initially submitted may require revisions, it is difficult to anticipate how long it will take to get a research project accredited. It is, however, possible to provide some indication of the timeline involving the key steps of the review process:

1. The SRS review is performed on a rolling basis as the applications come in.
2. The data owners independent review will also be performed on a rolling basis, although different timelines should be expected depending on the data owner.
3. The GRADE Project Board meeting is scheduled monthly, but it is envisaged that the handling of the data owners' reviews may be conducted electronically to speed up the process.
4. RAP meetings, which used to be 4 times per year (every 3 months), have recently moved to a rolling basis.

# GRADE and data protection

## Data security

The security of data on pupils shared through this initiative is of paramount importance. External researchers will be able to access the data shared through this initiative only through the ONS SRS.

The ONS SRS was accredited as a Digital Economy Act (DEA, 2017) processor by the UK Statistics Authority for the preparation and provision of data for research purposes. The SRS provides a safe setting to access data to conduct research, support rapid policy analysis and complete wider research for the public good. To ensure the safe access and secure use of data, the ONS SRS uses 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 are 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 have been de-identified.

The first 2 Safes are achieved through the accreditation of external researchers and the accreditation of projects (as detailed in the 'Data governance: accessing the GRADE data' section). The controls that are in place to guarantee the privacy and security of the data shared in relation to the three remaining Safes are as follows.

## Safe data: de-identification and pseudonymisation

External researchers will be only able to access data that has been de-identified and pseudonymised. Personal identifiers (including forename, surname, date of birth, school or college attended) are used for linking the data from different organisations. External researchers, however, will not have access to this information.

Personal identifiers were only used to generate a meaningless unique identifier that allows tracking of learners and centres across the different data tables. To add an additional layer of security to the data, a pseudonymisation algorithm was applied to both candidates' identifiers and centres' identifiers. Learners and centres were

pseudonymised using a one-way hashing algorithm, SHA-256 ([U.S. Department of Commerce, 2015](#)). SHA-256 is considered a secure hashing algorithm that produces alphanumeric hashes of intermediate length. The combination of an anonymised identifier with other variables yields input strings of variable length and minimises vulnerabilities to attacks. Further details on the methodology are not shared to ensure the security of the data.

This level of caution was deemed to be necessary given the wealth of information on learners (including some special category related data, for example, learners' ethnicity and special educational needs status) and centres (such as centre type) that will be available to external researchers. Sharing this data adds value to the data to be used for research purposes. This will allow independent evaluation of alternative grading approaches and their impact on different learners. This would not be possible if special category data was not shared.

## Safe settings: the ONS SRS

External researchers will only be able to access the data through the ONS SRS, which is an Accredited Processor under the DEA (2017) for the provision of data to conduct research, support rapid policy analysis and complete wider research for the public good.

The use of SRS ensures a safe setting to access data. Once approval has been granted, external researchers can access the GRADE data through ONS safe rooms or in other safe settings provided by approved organisations. More [details on the access of safe rooms and other safe settings](#) are available on the ONS website.

An alternative way to access the data through SRS is the use of a remote connection. Remote connection to the SRS is available to organisations that achieve certification under the Assured Organisation Connectivity Scheme to ensure that they meet safe setting criteria by using secure technology systems.

In both cases researchers will be able to access the data and analyse it using the systems provided. The use of personal electronic devices when working in the ONS SRS is not allowed. Researchers working in the SRS will not be allowed to copy or export extracts of the data. Taking notes on paper is also not allowed. Researchers will not be 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 ONS.

## Safe outputs: Statistical Disclosure Control

A safe output is an output that is non-disclosive and maintains the confidentiality of the data subjects. The SRS has substantial expertise in the checking of outputs before they leave the centre. All research outputs are checked by the SRS Statistical Support team to ensure that data subjects cannot be in any way identified. This involves the application of a rule-based Statistical Disclosure Control to ensure that all outputs are non-disclosive. SRS uses a threshold rule-of-thumb of 10: any statistical output computed on fewer than 10 data subjects will not be disclosed and will not be exported outside of the SRS. The ONS SRS team will continue to provide guidance and checks for the researcher until publication.

## Legislation

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 DEA, however, the data has to be completely de-identified or functionally anonymised.

Although personal identifiers are not shared, a wide range of information on each individual is available to researchers if they apply for Ofqual data linked to DfE and/or UCAS data. Recognising the greater richness of the data in this situation and greater potential for data subjects to be identifiable, additional controls are put in place in relation to such data. For example, a specific data sharing agreement is entered between parties. External researchers are required not to identify individuals from the de-identified data (and are also reminded that it would be a criminal offence to do so).

Any researcher who wants to access the data through the SRS must be accredited under the DEA (as explained in the 'Becoming an accredited researcher' section) and the research project must have been approved by the data suppliers and/or RAP. 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 project](#).

Each data owning organisation publishes a policy on how it processes and protects personal information. In order to reach as many data subjects as possible, we engaged in a range of communication and engagement activities. To inform data subjects of the GRADE project and of their rights in relation to their data shared as part of this initiative, we worked with a number of stakeholders to reach out to as many students as possible and make them aware of the use of their data.

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, we also jointly issued a [privacy notice for the GRADE project](#). The notice provides students with information including the purposes of the project, how their data is processed, the retention period for the data, and how it will be shared. Work was undertaken to ensure that the privacy notice was accessible to students.

# Promoting the use of the GRADE data

## Data documentation and research output repository

In addition to this framework setting out the high-level structure of the data and how to access it, a range of resources are available to researchers interested in working with the GRADE data. All the resources will be available on [the GRADE page](#). These will include:

- data specifications, with the details (names and descriptions) of all the variables in each data table provided by the three data owners
- a 'read me' user guide, with context and detailed guidance on the GRADE data sets, how the data from different sources has been linked together and some of its key features, including missing data and unique values

This documentation has been developed by data owners based on expert advice and in consultation with users of the data. The engagement carried out also highlighted the potential benefit for researchers of having access to low-fidelity synthetic data<sup>1</sup> outside of the ONS SRS, which is available upon request.

[The GRADE webpage](#) will also host a research output repository, a list of the research and analysis conducted with GRADE data. This is intended to provide researchers with an up-to-date list of findings based on this data to favour knowledge accumulation.

## Funding opportunities

This initiative aims at providing accredited researchers with a key tool to conduct independent research and evaluation on education and educational assessment issues. In order to promote the use of the GRADE data Ofqual, DfE and UCAS are also committed to help researchers overcome funding barriers to the analysis of this data, while ensuring the independence of the research conducted.

---

<sup>1</sup> For low-fidelity synthetic data here we refer to a data set with the exact same format and structure of the real data but containing only made-up information based on simulated data. No real data on pupils is featured in the low-synthetic data. To know more about synthetic data, see [the report published by the Behavioural Insights Team](#). To access the synthetic data for the GRADE data please contact [data.sharing@ofqual.gov.uk](mailto:data.sharing@ofqual.gov.uk)

In conjunction with the release of the data through the ONS SRS, Ofqual, DfE and UCAS have partnered with ADR UK to secure the availability of funding. [A funding call for 2 research fellowships](#) was launched to conduct research and analysis to demonstrate the policy impact potential of the linked, de-identified data sets under the GRADE initiative. The fellowship grant of 12 months in duration amounts to a maximum of £130,000 per annum per each fellowship fully funded by ADR UK. Future funding opportunities will be advertised on [the GRADE webpage](#) if and when they become available, but independent funding could also be sought.



## Further information

A range of further supporting material on the GRADE initiative is available online on [the GRADE webpage](#). This includes data specifications and additional metadata for researchers interested in using the data. It also features additional information on the project for data subjects and the general public. More information about the ONS SRS approved researcher scheme can be found on the [ONS SRS guidance page](#).

For further enquiries you can contact the organisations involved in the project:

- for ONS SRS researcher and project accreditation queries please contact [adrcuration@ons.gov.uk](mailto:adrcuration@ons.gov.uk)
- for Ofqual data specific queries please contact [data.sharing@ofqual.gov.uk](mailto:data.sharing@ofqual.gov.uk)
- for DfE data specific queries please contact [data.sharing@education.gov.uk](mailto:data.sharing@education.gov.uk)
- for UCAS data specific queries please contact [stats@ucas.ac.uk](mailto:stats@ucas.ac.uk)
- for data, synthetic data, or metadata and data feasibility queries please contact [data.sharing@ofqual.gov.uk](mailto:data.sharing@ofqual.gov.uk)
- for queries regarding ADR UK funding opportunities please contact [hub@adruk.org](mailto:hub@adruk.org)

## Appendix

Figure 1: Structure of GRADE data sources – accessible version

### Ofqual data sources

- qualifications dataset
- centres dataset (not shared for data protection reasons)
- grade boundaries dataset
- look-up tables (not shared for data protection reasons)

### Department for Education data sources

- exam results dataset
- student dataset
- prior attainment dataset
- census dataset
- look-up tables (not shared for data protection reasons)

### UCAS data sources

- apply qualifications dataset
- applications dataset
- applicant dataset
- look-up tables (not shared for data protection reasons)

Figure 2: The application review process at a glance – accessible version

1. The researcher submits their research project application form to ONS.
2. ONS SRS perform the first check. At this point ONS will either progress it to the next stage or pass it back to the researcher with feedback to inform revision or resubmission. It also at this point that, if needed, researchers may be asked to submit a supplementary form.
3. ONS SRS perform their feasibility checks. The application is either passed back to the researcher with feedback to inform revision or resubmission, or passed on to the data owners for their review.

4. Each data owner performs their independent review.
5. The Grade Project Board make a joint decision on the research proposal.
6. Depending on whether additional controls are needed, the project may be accredited either by the data owners or the Research Accreditation Panel, or it may be passed back to the researcher with feedback to inform revision or resubmission.



© Crown Copyright 2022

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated.

[View the Open Government Licence at The National Archives website](http://www.nationalarchives.gov.uk/doc/open-government-licence)<http://www.nationalarchives.gov.uk/doc/open-government-licence> or write to: Information Policy Team, The National Archives, Kew, London TW9 4DU

Published by:

**ofqual**

Earlsdon Park  
53-55 Butts Road  
Coventry  
CV1 3BH

0300 303 3344

[public.enquiries@ofqual.gov.uk](mailto:public.enquiries@ofqual.gov.uk)

[www.gov.uk/ofqual](http://www.gov.uk/ofqual)

April 2022