Nuffield Foundation

OXFORD BROOKES UNIVERSITY

APPENDIX 3

Technical notes to accompany the spatial analysis methodology and the qualitative study

The Extended Project Qualification: An Opportunity for All?

Brown, C.
Walkington, H.
Wood, G.
Bradley, S.
Pinot de Moira, A.

https://doi.org/10.24384/n8ge-5t58

Table of contents

Front cover: photo by Ambrose Chua on Unsplash

Report design: Jayne Stuart, Oxford Brookes University

Part 1. Data sources

1.1. EPQ data

All spatial analysis was conducted using publicly available data from the Department for Education (DfE), specifically covering students at the end of 16–18 study in the 2018/19 academic year. Each educational provider was identified using its Unique Reference Number (URN), which enabled consistent matching across multiple DfE datasets.

Key data sources included:

1. "Get Information About Schools" (get-information-schools.service.gov.uk
   This platform provided:
    - Easting and Northing coordinates (used to geolocate providers)
    - Middle Layer Super Output Area (MSOA) code
2. "Compare School and College Performance" (compare-school-performance.service.gov.uk) Two datasets from this source were used:
    i. "england_ks5final" provided:
        - Gender of the sixth form (girls, boys, or mixed)
        - Number of students at the end of 16–18 study
        - Admissions policy (e.g. open, selective)
    ii. "england_ks5underlying" provided:
        - School or college name
        - Provider type
        - Total number of EPQ entries
        - Number of students achieving each grade band (A*, A, B, C, D, E)
        - Count of students recorded as Fail or No Result.

Note: "Fail & No Result" includes students who were disqualified, did not meet the required standard, received no result, were ungraded, or were absent. The DfE suppresses data for providers with ≤5 entries or passes, to protect confidentiality.

To integrate these datasets, the Power Query tool in Microsoft Excel was used to match records by URN and to create a consolidated data table.

The "Provider type" variable was subsequently recoded to align with the six categories described in Section 4 of the main report (defined as academies, community & foundation, free, voluntary aided/controlled, FE/sixth form colleges, independent and other).

This final table was then imported into ArcGIS Pro, where Easting and Northing coordinates were used to geolocate each provider for subsequent spatial analysis.

1.2. Spatial data

1. Ordnance Survey Open data was used, namely:
    - “Boundary-Line” - to provide an outline of England and the Government Regions
    - “OS Open Built-Up Areas” - to clip the modelling outputs to the boundaries of built-up areas (and hence populated zones)
2. The 2019 Index of Multiple Deprivation (IMD) at Middle Layer Super Output Area (MSOA). The IMD is part of a suite of deprivation indices by the Ministry of Housing, Communities and Government. More information can be found at: **https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019** and **research.mysociety.org/sites/imd2019/about/**

Part 2. Data visualisation and analysis

2.1. EPQ uptake: Kernel density estimation

Kernel density estimation (KDE) was used to model the intensity of EPQ uptake across England. The analysis used the locations of educational providers offering the EPQ and the number of student entries at each provider as the input data. For each provider, a smooth curve - known as the kernel function - was applied over the surrounding area. This function assigns greater weight to nearby values and less weight to those farther away, reflecting a distance-decay effect.

A user-defined search radius (or “bandwidth”) determines how far each kernel extends. Within this radius, the density of EPQ uptake is estimated based on the number of student entries and their proximity to the kernel’s centre, tapering to zero at the outer edge. Where kernels overlap - such as in areas with multiple providers in close proximity - their densities are summed, producing a continuous surface that shows variation in uptake per unit area (measured here as students per square kilometre). The results were then clipped to the boundaries of built-up areas to focus the analysis on populated zones.

In this context, high-density areas may indicate clusters of providers offering the EPQ to large student cohorts, while low-density areas may reflect limited access or lower participation rates.

Several search radii were tested, and a 10 km radius was agreed upon with the project Advisory Group as an appropriate approximation of the area from which students might be drawn. A cell resolution of 250 metres was used to maintain a fine level of spatial detail without unduly increasing processing time.

KDE analysis was performed both for (a) state-sector providers only, and (b) all providers combined (i.e. including independent schools).

2.2. Population adjusted EPQ uptake

To calculate a population-adjusted density value for EPQ uptake, it is necessary to account for the underlying density of the potentially eligible population. This approach helps to distinguish between areas where EPQ uptake is high due to large populations and those where uptake is proportionally high relative to the size of the eligible cohort.

In this analysis, the eligible cohort was defined as Year 13 students that entered for at least one advanced (level 3) academic qualification in providers offering 16–18 education in 2018/19, including those that did not offer the EPQ. To enable further comparison, the Year 13 population was disaggregated into (a) state-sector providers only, and (b) all providers combined (i.e. including independent schools).

KDE was then used to model the spatial distribution of the eligible population, using the same parameters as for EPQ uptake i.e. a 10km search radius and a 250m cell resolution.

The Population-Adjusted Density of EPQ Uptake was then calculated using the formula:

$$\text{EPQ Population Adjusted Density} = \frac{\text{EPQ students per km}^2}{\text{Eligible population per km}^2}$$

To support mapping and interpretation, the result was then multiplied by 100, expressing it as a percentage of the eligible population.

2.3. EPQ uptake and relative deprivation

Data on provider catchment areas and admissions policy is not universally available, making it challenging to determine the spatial and socioeconomic characteristics of the EPQ student population. Additionally, the aggregated provider-level dataset does not include information on students' residential locations.

As a proxy, Geographic Information Systems (GIS) overlay analysis was used to determine the Index of Multiple Deprivation (IMD) decile for the MSOA in which the provider is located. IMD deciles range from 1 (most deprived) to 10 (least deprived). MSOAs typically contain between 2,000 and 6,000 households, with a population of 5,000 to 15,000 people, and were used to represent the proximal community served by each provider.

The resulting data table was exported as a CSV file and analysed in Microsoft Excel, using Pivot Tables to aggregate the number of EPQ student entries by IMD decile. The findings were visualised using bar charts and box-and-whisker plots, also produced in Excel.

2.4. EPQ attainment: Kernel density estimation

KDE was used to model the spatial distribution of top-grade attainment (A*/A) for EPQ candidates across 1,232 state-sector providers. Only providers with 6 or more entrants were included in the analysis (data was suppressed for providers with 5 or fewer students).

Two KDE models were created: (a) using the raw number of A*/A grades achieved at each provider, and (b) using the percentage of the provider EPQ cohort achieving an A*/A grade.

As with previous analyses, a search radius of 10km and a cell resolution of 250m were used to generate fine-grained spatial surfaces. Again, the results were clipped to the boundaries of built-up areas to focus the analysis on populated zones.

2.5. EPQ attainment and relative deprivation

To explore the relationship between EPQ attainment and relative deprivation, student outcome data (covering all grades: A*, A, B, C, D, E, and Fail/No Result) for state-sector providers was aggregated by IMD deciles at the MSOA level.

This was done using the same GIS overlay approach described previously, where each provider was assigned the IMD decile of the MSOA in which it is located (decile 1 = most deprived, decile 10 = least deprived).

Part 3. Data communication and engagement: GIS web app

ArcGIS Experience Builder was used to create a **web-based GIS application** that allows users to explore the map outputs from this research in a dynamic and interactive format. The app is structured to allow easy navigation through the underlying maps and datasets.

Tabs located at the top of the screen include:

- EPQ Provision
- EPQ Uptake
- EPQ Attainment

Users can:

- Zoom in and out
- Toggle map layers on and off
- Apply pre-configured filters (e.g. to display state-sector providers only)

A series of interactive widgets on the right-hand side of the screen provide further control options, including:

- Access to the underlying data table
- Selection tools (rectangle, circle, and lasso)
- A swipe tool that reveals map layers beneath others in the drawing sequence

By default, all maps are displayed with a greyscale basemap to provide geographic context. However, users can change the background to suit their preference - for example, to aerial imagery, OpenStreetMap, or a range of standard ESRI basemaps.

Part 4. Methodology for the qualitative study (spatial targeting of educational providers, educator interviews and student focus groups)

GIS was used to target a sample of potential providers for the qualitative research. Using the Indices of Deprivation 2019 at Lower Super Output Area (LSOA) level, the selection criteria were based on two key deprivation indicators:

1. Income Deprivation Affecting Children Index (IDACI) subdomain; and
2. Children and Young People (CYP) subdomain (which includes measures of educational attainment and associated measures).

A Boolean attribute query was then performed to identify state-sector providers that met all of the following conditions:

- Were located within 500 metres of an LSOA in decile 1 (most deprived) for IDACI
- Were located within 500 metres of an LSOA in decile 1 for CYP
- Had 25 or more EPQ entries

This methodology enabled targeted sampling of providers situated in areas with high levels of child poverty and low educational attainment, while also ensuring a sufficient level of EPQ activity for meaningful engagement in the qualitative research.

An example of this spatial targeting is illustrated in Figure 1 (below).

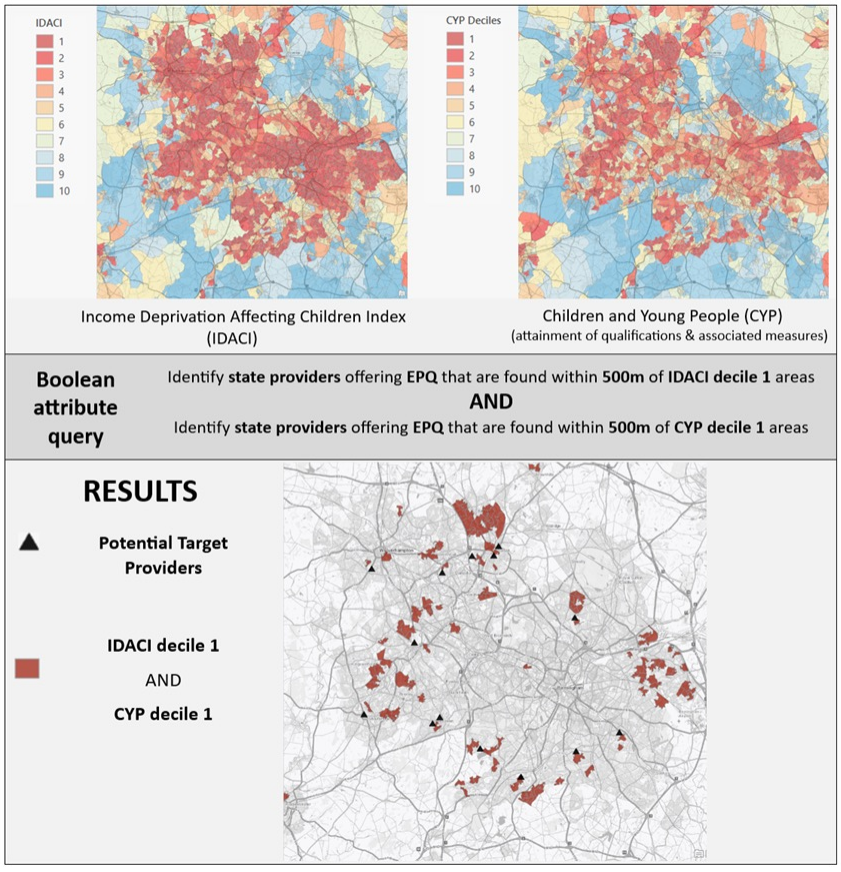

Figure 1 Spatial targeting of potential educational providers for the qualitative methodology

The final list of identified providers was exported to Excel and used as the sampling frame for targeting the educator interviews and student focus groups.

An initial wave (Wave One) of invitations to participate was sent via email to 20 centres identified through the spatial targeting (Sept-Nov 2024). Wave One centres were purposively selected using further sample criteria to target a range of EPQ delivery and engagement experiences. These additional sampling criteria included: geographical location (regional representation); regional characteristics (high levels of deprivation, coastal, rural, inner city, etc.); centres' student profiles (gender, ethnicity, faith); centre type (studio school, university technical college, converter academy, sixth form college, etc.); and centres' EPQ performance and offer (high/low grades, high/low withdrawals, high/low entry numbers, offered alongside A-Level or A-Level and BTEC). Follow up telephone calls and emails were made to non-responders. Two further centres were identified for contact at Wave One to augment these preliminary target centres in terms of offering the EPQ to all students at scale with a catchment that included IMD decile 1 areas.

A secondary wave (Wave Two) of invitations to participate was sent via email to the remaining (n=35) centres identified through the original spatial mapping exercise (Nov 2024 - Jan 2025). Telephone and email follow-ups were made to Wave Two non-responders. One further centre was included after contacting the research team on hearing about the project and its objectives. Student focus groups were recruited through participating centres that agreed to help distribute participant information to EPQ students (who had already completed the qualification) and organise spaces (in person or virtual classrooms) for the focus group discussions.

The educator interviews asked:

- Can you tell us your role in offering the EPQ? What do you think is the best thing about offering the EPQ at your school?
- How is the EPQ offered in your school? Can you describe your practices in terms of recruitment? Are there any reasons that the EPQ is not offered to all your students?
- Have you experienced any barriers to offering the award? Or barriers to some students participating?
- How about barriers to completion or beneficial outcomes?
- Now thinking back to the pandemic - How was the EPQ managed during Covid? Can you comment on any changes to uptake, withdrawal or assessment for the EPQ during the pandemic? Did any new mentoring practices come about as a result of Covid? If so, were these retained?
- How is the EPQ managed in your setting in terms of teaching/supporting students (e.g. practices for retaining students, supervising, teaching, mentoring). Please describe something that you think was effective practice in mentoring students for successful outcomes.

- Please could you give any examples of practices that are particularly effective in supporting the students that you work with?
- What value do you think the EPQ has in your school?
- Do you think there is potential for the EPQ to help 'level up' for students from more disadvantaged backgrounds or areas in your catchment?
- Do you consider your school/college is already using the EPQ (or could more proactively use the award) as a tool for 'levelling up'. Could you describe how?

The student focus groups asked:

- What was your experience of studying a topic of personal interest?
- Why did you take the EPQ?
- Are all students in your school/college offered the opportunity to take the EPQ?
- Were there any challenges or barriers to you taking it? And keeping going?
- What sort of support did you have? (As many examples as you can think of between you… who helped?)
- What has participation in the EPQ specifically allowed you to learn about yourself?
- Do you think participation in the EPQ has changed your aspirations or opportunities available to you?
- Going back to the overall mentoring and support for the EPQ you've had, was there anything offered that was particularly helpful to you?

The interview data was analysed using thematic analysis (Braun & Clarke, 2006), a structured method for identifying, examining, and reporting patterns within qualitative data. This approach follows a series of defined stages, beginning with familiarisation with the material, followed by the development of initial codes. These codes were then used to construct and refine overarching themes in relation to barriers and practices.

4.1. Reference

Braun, V. & Clarke, V. 2006. Using thematic analysis in psychology, Qualitative Research in Psychology, 3:2, 77-101.