School Exclusions, Alternative Provision and Crime in England

Secondary Data Project Analysis Plan

Evaluating institution: CEP, LSE

Principal investigator(s): Matteo Sandi and Stephen Machin

Analysis Plan for YEF Secondary Data Analysis Projects

Project summary

Project title	School Exclusions, Alternative Provision and Crime in England
Research Team	Matteo Sandi, Stephen Machin, Lucas Silva Lopes, and Saandra Nandakumar
Principal investigator	Matteo Sandi
Analysis plan author(s)	Matteo Sandi and Stephen Machin
Overarching research question	How have the post-exclusion destinations of permanently excluded students in England evolved over the last two decades and are these changes associated with variations in children's likelihood of engaging in violence?
Supporting research question(s)	What school-level and local authority-level factors potentially affect the probability of permanent exclusion for a pupil? How have changes in alternative provision attendance (for example, a pupil referral unit vs an alternative provision academy) affected youth crime?
Dataset(s) to be used	DfE-MoJ data linkage 2001-21
Population characteristics	Universe of Pupils in State-Maintained schools in England from 2001-21
Years data spans	2001-21
Geographic coverage	England
Primary outcome(s) investigated	Permanent exclusion, charges for offending (focus on violent offending), as recorded in official school and offending records
Main method(s) to be used or tested	Descriptive unconditional analysis; OLS Regression; Difference-in-Difference; Propensity Score Matching

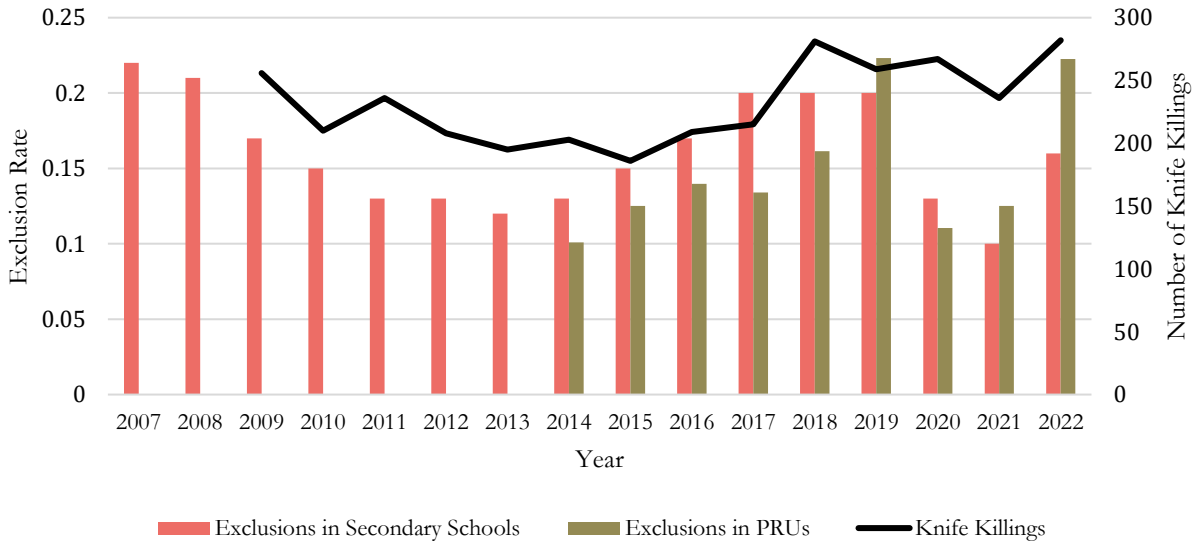
Contents

1. About the project

1.1. Background to the project

In the last decade, England has witnessed a sizeable increase in permanent school exclusions. Following years of declining rates of exclusion, the period between 2013-2019 saw an increase in permanent exclusions of over 70% in state-funded secondary schools, while permanent exclusions more than doubled in pupil referral units (PRUs).[1] During this same period, knife and sharp instrument homicides hit a record high since 1946 (Elkin, 2019). This correlation, illustrated in Figure 1, has stimulated an extensive, mostly non-evidence based, discussion in the media and the political arena (Schraer, 2019). Nevertheless, an empirical investigation that explores how exclusion is linked to youth violence, and the changing educational and crime trajectories of pupils following increasing rates of exclusion is still missing.

Figure 1. Permanent Exclusion Rates and Knife Killings in England



Source: Department for Education, *Permanent Exclusions and Suspensions in England.*

[1] While the pandemic saw the lowest permanent exclusion rates since 2013, owing primarily to school closures (Walker, 2022), by 2022 the permanent exclusion rate had already begun to rise again.[1] The latest statistics from DfE (DfE, 2023c) record a total of 3100 permanent exclusions in the autumn term of 2022/2023, an increase of approximately 1000 compared to the 2097 recorded permanent exclusions from autumn 2021/2022. Data for the summer term of 2022/2023 is still to be released.

A recent report by the UK Department for Education (DfE) and Ministry of Justice (MoJ) shows that 52% of serious violence offenders attended an alternative provision setting (AP) before their first offence, i.e., more than 1 in 2 (DfE & MoJ, 2022). Therefore, developing a better understanding of the AP options available following permanent exclusion, and their impact on excluded pupils has the potential to help keep at-risk juveniles away from crime. This project seeks 1) to build a more complete picture of the AP landscape in England, and how this has changes since 2012 education policy reforms; and 2) to understand the role of exclusions and AP in predicting criminal behaviour in juveniles.

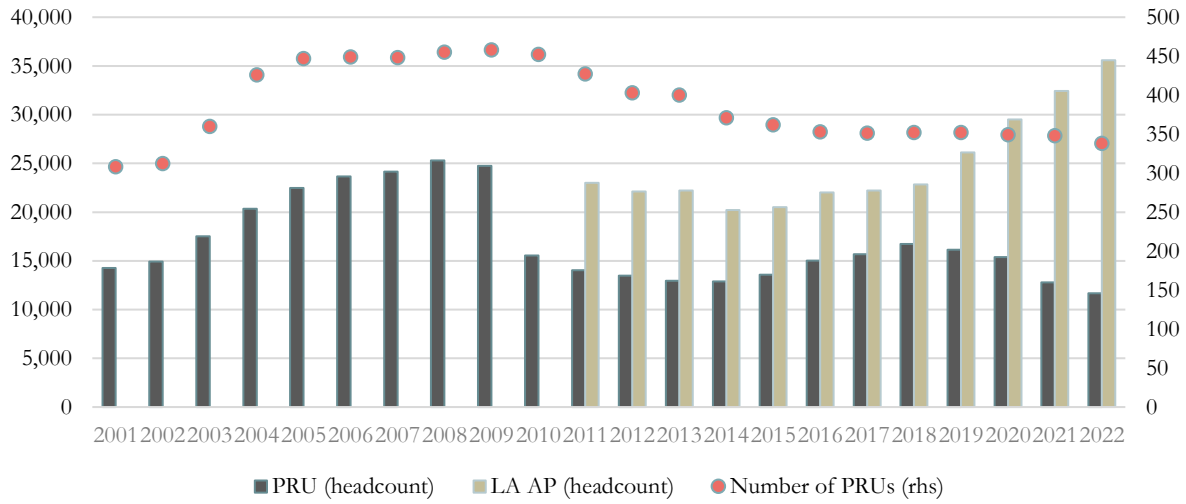
The changing landscape of AP provision

Alternative provision (AP) is defined as any form of education for pupils who are no longer able to attend either mainstream schooling or special education (DCSF, 2008). Children assigned to AP institutions usually fall under one of five different groups. They could be children who have been either permanently excluded or received a fixed-term exclusion for more than five school days; children with emotional and/or behavioural difficulties; children with emotional and/or physical vulnerabilities that prevent them from attending mainstream school (e.g teen mothers, children with school phobia); children that are unable to keep up with mainstream education and thus have been encouraged by their headteachers to seek a different type of education; and finally children who have just recently moved into a new local education authority (LEA) and thus authorities have not yet found them a mainstream school (Ogg & Kaill, 2010). In this sense, AP is a very extensive term, encompassing all the different destinations for a pupil no longer able to attend mainstream schooling such as:

1. Pupil Referral Units (PRUs)
2. AP Academies
3. AP Free Schools
4. Further Education Colleges (FE Colleges)
5. Home Tuition
6. Elective Home Education
7. Alternative Education Initiatives (AEI)
8. In-school alternative provision run by multi-trust academies

PRUs have been the most common form of AP since the early 2000s. As Figure 2 illustrates, the first decade of the 21st century saw a steady rise in the total number of PRUs in England, increasing from 308 to 458 units in a span of eight years (close to a 50% overall increase). Furthermore, the total number of pupils in PRUs also saw a steady increase during the 2000s decade, with a 2008 report by the then-DCSF (now, DfE) estimating that PRUs alone accounted for as many as one third of all alternative provision destinations in the 2007-2008 school year (DCSF, 2008).

Figure 2. Evolution of PRU and LA AP in England, 2001 to 2022



Source: Department for Education (DfE), *Schools, pupils and their characteristics.*

After 2010, however, as illustrated in Figure 2, there was a decline in both the total number of PRUs as well as in the total number of pupils enrolled in PRUs. Two factors might lie behind this. First, following the enactment of the *Education Regulations Amendment of 2012,* high-performing PRUs were allowed to convert into alternative provision academies (AP Academies) and thus gain autonomy from their local education authorities (House of Commons Education Committee, 2018). Second, since 2012 there has been a rise in the variety of AP settings offered to pupils by the local education authority (LEA henceforth). Overall, Figure 2 illustrates that, while remaining one of the most popular destinations for AP assignments, PRUs no longer constitute the absolute majority of AP assignments (as was the case during the 2000s decade).

One particularly notable trend is the rise in the number of AP settings that engage pupils outside the traditional classroom-based setting. These initiatives are usually run by a charity or youth work groups and provide a broader range of activities for students, including work experience or even commercially paid for activities outside the school. Children may attend these sessions either full or part time.

The proliferation of this type of provision started in 2012, when the then-government expert adviser on behaviour, Charlie Taylor, published a report titled “Improving Alternative Provision”. Besides granting high-performing PRUs the right to apply to become academies, this report recommended that the Department for Education (DfE) discontinue keeping a central list of AP providers. The acceptance of this recommendation by the government led to a situation in, which under certain conditions, an institution is no longer required to officially register as a school before starting to operate as an AP institution (House of

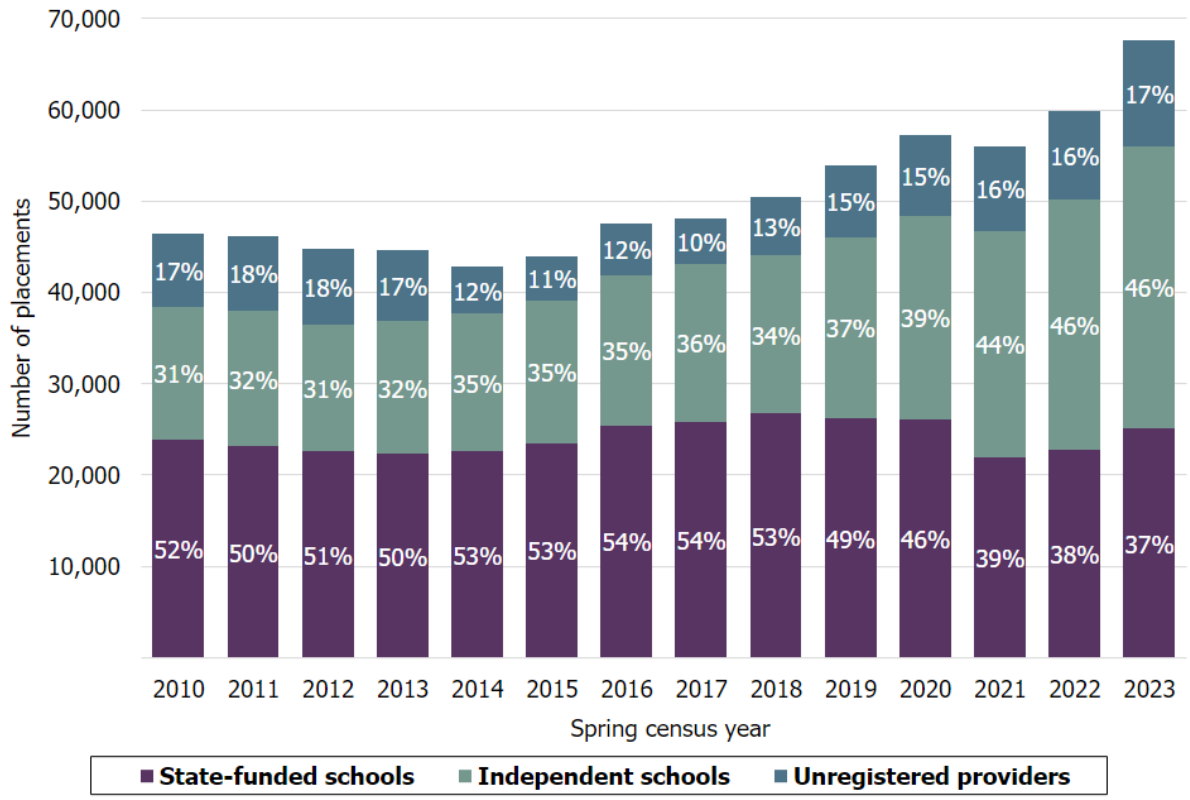
Commons Education Committee, 2018)[2]. This allowed for the proliferation of unregistered AP institutions that are not being routinely inspected by Ofsted. While there is no legal definition for what constitutes an unregistered AP institution, settings that are not state-funded schools or registered independent schools are commonly referred to as unregistered AP institutions (DfE, 2022a; DfE, 2022b; Spielman, 2021). Given the lack of oversight, it is hard to ascertain the quality of the education being provided to pupils in these settings. Meanwhile, with mainstream schools facing increased accountability for attainment, as well as the challenges of poor staffing and budget cuts, the number of pupils being assigned to AP settings is believed to be increasing. Focusing specifically on assignments to unregistered alternative provision, Figure 3 below suggests a steady rise in assignments to unregistered provision as a share of total AP assignments since 2018.

The presence of unregistered AP settings and the extent to which they dominate alternative provision seems to vary considerably by LEA, as illustrated in Figure 4 below. According to Figure 4, whereas in some LEAs unregistered AP accounted for a very small share of the total AP assignments made, in other LEAs such as Essex, Nottinghamshire, and Northumberland; unregistered AP settings accounted for roughly 50% of the total AP assignments made during the 2022-2023 academic year.

Assignment to unregistered AP settings is more common in cases when the pupil is suspended from school. In the case that a pupil is permanently excluded from school and becomes the responsibility of the LEA, unregistered AP settings become the last resort. This is because it then falls onto the LEA to ensure the quality of education and safety of the pupil in the new AP setting. LEAs therefore prefer to have students enrolled either in another mainstream school or in a registered AP institution where the responsibility for the excluded pupil can now be passed on to the school's headteacher. The new headteacher, however, can then decide whether the pupil should be educated offsite. In such cases, the attendance of pupils in these off site settings may be recorded through school absence data, which since September 2022 has included a code for being educated off-site.

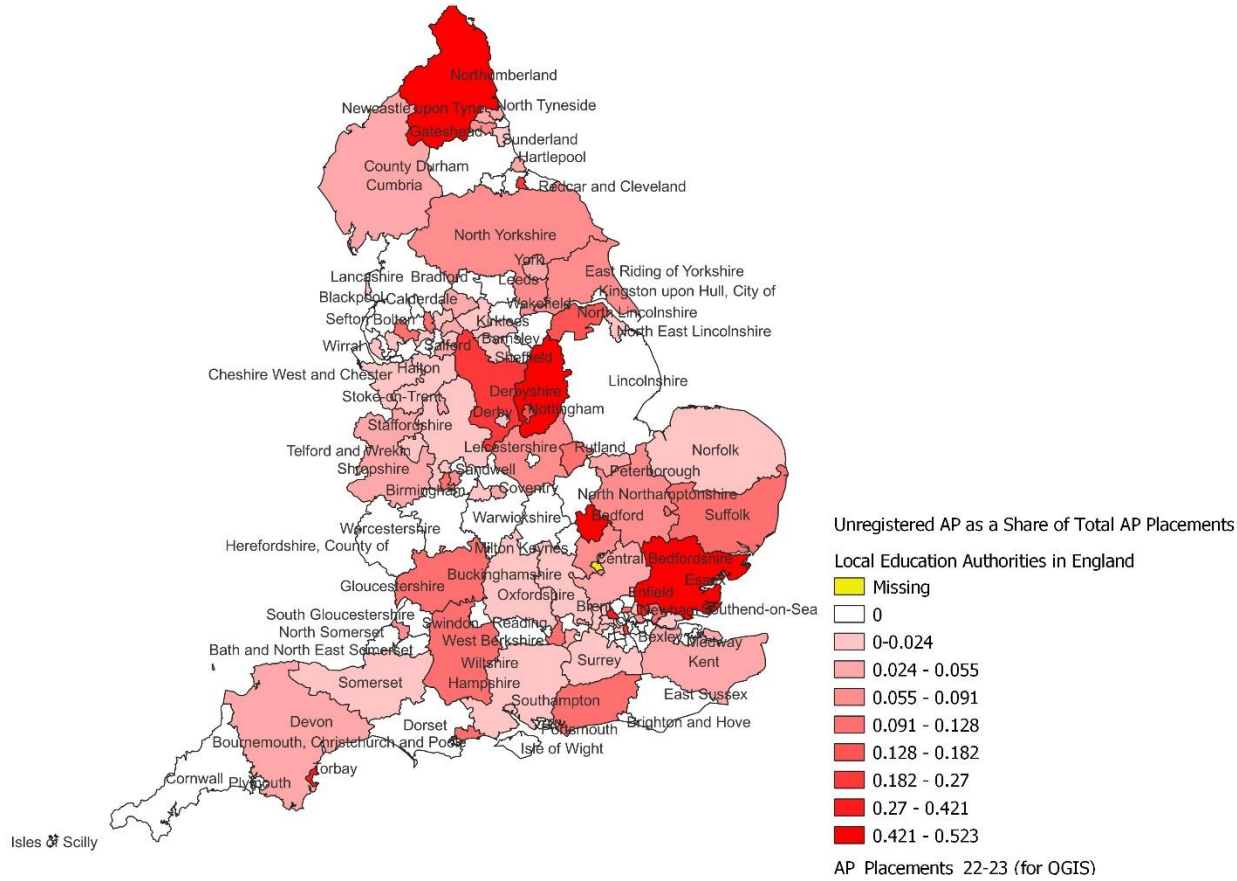
[2] By law, a school is an educational institution that is not a further education or higher education setting and provides either full-time education for children and young people aged (approximately) 5 to 18, or part-time education for children aged 2 to 5. Schools providing full-time education to five or more pupils of compulsory school age or providing education to pupils who are looked-after or have an Education, Health and Care plan and that are not maintained by a local authority or a non-maintained special school, must register as an independent school (DfE, 2022a). Settings that are not state-funded schools or registered independent schools are commonly referred to as unregistered AP institutions (DfE, 2022b).

Figure 3: Number of AP placements in state-funded schools, independent schools and unregistered providers, over time



Source: Ofsted (2024)

Figure 4: Placements in Unregistered AP Institutions as a Share of Total AP Placements by Local Education Authority in the Academic Year 2022-2023



Source: Own elaboration using data from DfE (2023b)

Tracking permanently excluded pupils, who are not (at least) formally enrolled in a registered AP setting, however, is a challenging task. So far, the best attempt to track these pupils has been to compare the number of children registered to GPs in an area with the number of children on school records. This method has suggested that, in recent years, the gap between GP registrations and school records is growing, which is highly concerning given that the missing students are more likely to be vulnerable and at-risk of being exposed to youth violence.

The changing nature of AP and concerns about quality

There are several factors that have led to concerns about the quality of education and care provided in AP settings. The first is funding. Funding for AP comes primarily from a section of the Designated Schools Grant (DSG) called the "high-needs budget" that the Central Government allocates to the local education authorities (LEAs),[3] and the amount that each

[3] AP institutions can receive high-needs funding through three main mechanisms: the core funding, which refers to the annual funding based on the amount per place funding that an AP school receives either from its local authority or from the Education and Skills Funding Agency (ESFA)[3]; top-up funding, which refers to the additional (besides the core funding)

LEA will receive can be unpredictable. One factor that makes AP funding more volatile than regular school funding or special needs education is that in the case of special needs education, the LEA is compelled to fund a certain number of places even if they are unfilled (though this is rarely the case), whereas for AP they are required to fund only the number of filled places. In this sense, given that pupils flow in and out of AP throughout the academic year, this generates a large amount of uncertainty for the AP institution when it comes to budget planning. In a given year, therefore, the budget of an AP institution follows the same cyclic volatility as exclusion rates with a lower budget at the start of the year, followed by a spike towards the end of the autumn term before the School Census in January (at which stage the school in which the student is enrolled is then accountable for their exam performance in that academic year). This leads to issues in hiring for AP institutions: staff tend to be hired on a temporary basis as funding cannot be guaranteed for longer-term contracts. This issue in procurement also contributes to the larger problem that AP institutions face in providing consistent quality education.

The second is that increasing numbers of children attending unregistered AP settings has created oversight problems for the LEA, who have difficulty tracking pupils across these diverse settings and ensuring quality of care. Third, in an attempt to prevent their pupils from landing in unregistered AP settings, large multi-academy trusts have begun to set up AP institutions themselves. In this scenario, excluded or "difficult" pupils from across the different academies in a trust may all be sent to an AP institution run by the trust. This phenomenon started to become more common following the recommendations of the Timpson Review of School Exclusion published in 2019, which among other recommendations, praised and recommended the use of in-school units that provide "some of the benefits of AP" (Timpson, 2019). At the moment, however, very little is known about the quality of education provided in these multi-academy trust run APs.

Overall, the rise in unregistered AP settings over the past decade has added an extra layer to the challenge of evaluating the impact that attendance of alternative provision has on criminal outcomes. Lack of quantitative data and very little qualitative data on what happens inside these settings have made the AP landscape in England even more convoluted. While descriptive evidence suggests that juveniles exposed to AP are more likely to commit crime, to our knowledge no study analyses whether the type of alternative provision the pupil is assigned to plays a role in making the pupil more likely to commit a crime. This is an important knowledge gap in the public debate and this project aims to rectify this omission.

---

funding required for a pupil to participate in education that is paid for either by the local authority or by the mainstream school that commissions each place; and locally negotiated funding for AP services such as outreach, that are independent of the top-up and core funding sources (ESFA, 2021).

1.2. Research question(s)

This research project, conducted by academics from the Centre for Economic Performance (CEP) at the London School of Economics (LSE), sets out to understand the educational and criminal trajectory of children excluded from mainstream education in England and how their exposure to any form of alternative provision (AP), including local authority AP as well as PRUs and academies which are not run by the LA, may affect the risk of school exclusion and youth crime. Specifically, we aim to address the following set of questions:

1. How has the alternative provision landscape in England changed over the past decade: where do the excluded children go after permanent exclusion; what are the types of AP available to them and how has this evolved?
2. Which school-level or local authority-level factors (e.g. supply of different types of alternative provision) predict permanent exclusion?
3. How does attending different forms of alternative provision (for example, a pupil referral unit vs an alternative provision academy) affect the probability of committing a violent offence?

For us to be able to address question 3 in a causal sense, we will need to identify an exogenous source (e.g., a valid instrumental variable) of permanent exclusion resulting in pupils attending different forms of Alternative Provision. In the absence of this instrumental variable, the analysis would only be correlational, and its insights and policy takeaways would be more limited.

This research project proposes to address these questions through the empirical analysis of the Department for Education-Ministry of Justice (DfE-MoJ) data linkage. It will provide descriptive insights on the evolution of alternative provision in England in the last 20 years, i.e., from the early 2000s, both in terms of scale and variety. It will also attempt to provide causal estimates of relevance to policymakers, describing the relationships between exclusions, alternative provision, and youth crime.

Table 1.2. How will the questions be addressed at each stage?

Question Number[4]	Interim report	Final report
1	A descriptive analysis of the AP landscape will be presented. This analysis will include highlighting legislative changes that have impacted AP as well as studying cohorts of excluded students across the last two decades to highlight the differences in their post-exclusion destinations.	Adjustments, edits, and additions to descriptive statistics
2	School-level and local authority-level factors will be investigated through regressions to determine to what extent they predict trends in permanent exclusion in the area. Potential factors are the total availability of different types of AP at the local authority level, or the gender composition at the school-level. Main results will be shown in tables or graphs and briefly summarised. The determinants of school exclusion will be investigated to understand the factors correlated with permanent exclusion rates.	Presentation in graphs or tables and write up in full of results. Final methods, interpretation and discussion of results will be provided.
3	Descriptive results on the types of AP provision in relation to youth crime will be presented through graphs and summary statistics.	Simple OLS regressions will be performed to estimate the correlations between the different types of AP institutions and the probability of committing future violent offences. We will also test potential instrument variables in order to estimate the causal

[4]Question numbers should follow the ordering in the section above.

		effect of exclusion and alternative provision on the propensity of crime.

1.3. Hypotheses

Concerning question 1, we will examine descriptively the evolution of AP and its various forms over the past 20 years. We will describe both the typical pathways of excluded children through the exclusion process and the main legislative changes since the early 2000s that govern this process, such as the *Education Regulations Amendment of 2012,* whereby high performing PRUs were allowed to convert into Alternative Provision Academies (AP Academies) and thus gain autonomy from their local education authorities.

Concerning question 2, we hypothesise that the relationship between the availability of alternative provision (AP) in the local authority (LA) and school exclusion and youth crime will be positive. The reason for this is twofold: on one side, more APs may open in localities/years where there is more demand for places to accommodate a greater number of excluded pupils. On the other hand, the direction of causality may also be the other way around – the greater availability of these alternative forms of education may make school exclusion more likely because an excluded pupil who has not yet reached the dropout age would need to be reinstated within 5 working days in school by the local educational authority. Therefore, having many forms of AP around might potentially induce headteachers to exclude more easily. The idea of a greater supply of AP units in a local area potentially contributing to a higher rate of exclusion within schools is somewhat reinforced by the findings of a report from the Institute of Education (IoE) and the National Foundation for Educational Research (NFER). The findings of the report suggest that prior to excluding a pupil permanently, school staff and headteachers were now deliberating more carefully about where to send the excluded pupil, while also taking the extra step to monitor the pupil's outcomes in their new AP setting (IoE & NFER, 2014). Disentangling the causal channels through which availability of AP affects youth exclusion and crime from these correlations will be a major challenge for our study.

For question 3, we do not have a strong prior on the impact of attending different forms of alternative provision (for example, a pupil referral unit vs an alternative provision academy) on youth crime. This is an open question that is central to this proposal. A large amount of empirical evidence mostly from the UK and the US has documented that school attendance is a protective factor that keeps pupils away from crime (e.g., Bell et. al, 2022; Jacob & Lefgren, 2003; Luallen, 2006; Machin et al, 2011). Therefore, we hypothesize that reinstating a pupil in mainstream schooling would be more effective at keeping the pupil away from crime, compared to assigning the pupil to any kind of alternative provision. However, if the pupil is assigned to an AP, the extent of this difference may vary based on the nature of the AP, for

example, the time commitment required by the AP or the quality of provision available in each setting.

1.4 Key Concepts

Table 1.4 Definitions of key concepts

Permanent Exclusion	*We will use the official records of permanent exclusion from the DfE database.*
Crime	*We will use the official records of criminal offences from the MoJ's Police National Computer database which include charges and subsequent convictions and/or cautions.*
Youth violence	*We will use the official records of charges for violent [5]criminal offences with or without injuries for summary as well as all indictable[6] offences, which are more serious offences that must be tried in the Crown Court, from the MoJ's Police National Computer database.*
Pupil Referral Units (PRU)	*As per Section 19(2) of the Education Act 1996 (Education Act, 1996), any registered school in England that was established to provide full-time education to pupils no longer able to attend mainstream education and that is not a community or foundation school, a community or foundation special school, or a maintained nursery school, is considered a pupil referral unit (PRU). PRUs are run by their local education authorities.*
Converter Academies	*As per the Education Regulations Amendment of 2012 (DfE, 2012), converter academies are schools that* ***have chosen to become academies****. They are funded directly by the government and are run by an academy trust.*
Sponsored Academies	*As per the Education Regulations Amendment of 2012 (DfE, 2012), sponsored academies are schools that were previously underperforming and have been* ***taken out of local authority control through government intervention****. They are run by an academy trust*

[5] "Violent crime" includes violence against the person and sexual offences as defined in the UK Home Office Crime Classification codes.
[6] An indictable offence is a more serious offence which that must be tried in the Crown Court, and usually has more serious punishments (CPS, 2019).

	and are supported by sponsors such as businesses, universities, other schools, faith groups or voluntary groups.
AP Academy	*When a PRU converts to academy status, it becomes an AP academy. This may happen in two ways: one, where they are failing or not delivering expected outcomes, the Education Secretary can start the process of converting them into academies, and two, from September 2012, all PRUs rated outstanding by Ofsted can choose to apply for academy status (Sharma, 2012).*
AP Free Schools	*Free Schools are schools that are funded by the government but not run by the local authority; they are also not run for profit and therefore are often set up by more independent groups such as charities, community and faith groups, and even by parents (UK Government, 2023). AP Free Schools are therefore autonomous Free Schools that cater exclusively to pupils no longer able to attend mainstream education.*
Further Education (FE) Colleges	*The definition of FE College refers to any college that offers academic, vocational, and or technical education to students regardless of their age (Kendall et.al, 2003).*
Special Schools	*A special school is a school destination tailored specifically for pupils with emotional and behavioural difficulties (EBD) (Kendall et.al, 2003).*
Home Tuition	*Home tuition refers to a situation in which the pupil is sent to the offices of her local education authority to receive tutoring by a PRU teacher. Both the pupil and the tutor should be supervised by the senior staff of the local education authority (Kendall et.al, 2003).*
Elective Home Education	*Parents willingly choose to have their child/children educated at home or from home by either hiring private tutors or educating them themselves. Once parents choose this option, the Local Education Authority is no longer responsible for making any educational arrangements for the pupil (DfE, 2019).*

Alternative Education Initiatives (AEI)	*These include partnerships between local education authorities and local organizations in the local area designed to engage permanently excluded pupils in activities that seek to improve the local community (Kendall et.al, 2003).*

2. About the datasets

2.1 Overview of datasets used

We propose to access the Department for Education – Ministry of Justice (DfE-MoJ) linked administrative dataset from 2002-2021 to conduct our research. The dataset contains education and justice data for individuals born from September 1985 and includes detailed, individually linked data from the DfE's National Pupil Database (NPD), School Census, PRU Census and Alternative Provision Census data linked at the individual level with the MoJ's data from the Police National Computer (PNC).

The DfE datasets allow us to track an individual's educational trajectory and attainment, as well as any exposure to alternative provision. The MoJ records allow us to determine if a young person has had contact with the judicial system, recording both the date and type of offence committed as well as the sentence received.

All researchers in the CEP (LSE) team are ONS-accredited, with eligibility to access and analyse sensitive, de-identified data in the Secure Research Service. Both Sandi and Machin are currently accessing the DfE-MoJ data linkage for analysis and are very familiar with these data. Finally, the CEP (LSE) has an Assured Organisational Connectivity Compliance agreement for access to the Secure Research Service in place.

2.2 Secondary data source(s)

Table 2.2.1 Dataset Description – School Census Pupil Level

Name of dataset	***School Census Pupil Level***
Data owner(s)	*Department for Education*
Type of data	*Cross-sectional education census*
Availability of data	*Licence required by the data owners*
Team member(s) who will have access	*Matteo Sandi, Stephen Machin, Lucas Silva Lopes, Saandra Nandakumar*

Population/geographic coverage or sampling frame	*Pupil census for all state-maintained schools in England*
Years covered or survey waves	*2001-2021*
Exclusion criteria	*N/A*
Expected population/sample size (following exclusion criteria)	*This has information on pupils attending maintained schools from 2001/2 on. From Spring 2013/14, this includes local authority (LA) maintained PRUs and alternative provision (AP) academies, including AP Free Schools.*
Documentation	*https://www.find-npd-data.education.gov.uk/datasets/775def61-ecd2-4e9a-8ef9-c168c4f51aac*

Table 2.2.2 Dataset Description – Pupil Referral Unit Census

Name of dataset	***Pupil Referral Unit Census***
Data owner(s)	*Department for Education*
Type of data	*Cross-sectional education census*
Availability of data	*Licence required by the data owners*
Team member(s) who will have access	*Matteo Sandi, Stephen Machin, Lucas Silva Lopes, Saandra Nandakumar*
Population/geographic coverage or sampling frame	*Pupil census for all PRUs in England*
Years covered or survey waves	*2009-2013 (incorporated into the School Census from 2013/14)*
Exclusion criteria	*N/A*
Expected population/sample size (following exclusion criteria)	*This has information on all children attending local authority (LA) maintained PRUs and alternative provision (AP) academies, including AP Free Schools.*
Documentation	*https://www.find-npd-data.education.gov.uk/datasets/36479c85-5dff-42ec-bdf6-492773eccbae*

Table 2.2.3 Dataset Description – Alternative Provision Census

Name of dataset	***Alternative Provision Census***
Data owner(s)	*Department for Education*
Type of data	*Cross-sectional education census*
Availability of data	*Licence required by the data owners*
Team member(s) who will have access	*Matteo Sandi, Stephen Machin, Lucas Silva Lopes, Saandra Nandakumar*
Population/geographic coverage or sampling frame	*Pupil census for students in AP not maintained by the LEA in England*
Years covered or survey waves	*2007-2021*
Exclusion criteria	*N/A*
Expected population/sample size (following exclusion criteria)	*This has information on children in Alternative Provision, i.e. a school not maintained by an LEA but which the authority is paying full tuition fees for.*
Documentation	*https://www.find-npd-data.education.gov.uk/datasets/2f10ee6d-506e-4182-957b-ca88f1a3907c*

Table 2.2.4 Dataset Description – Exclusions

Name of dataset	***Exclusions Default Data***
Data owner(s)	*Department for Education*
Type of data	*Cross-sectional education census*
Availability of data	*Licence required by the data owners*
Team member(s) who will have access	*Matteo Sandi, Stephen Machin, Lucas Silva Lopes, Saandra Nandakumar*
Population/geographic coverage or sampling frame	*All pupil exclusions as collected in the termly School Census (Reason included from 2005-06)*

Years covered or survey waves	*2001-2021*
Exclusion criteria	*N/A*
Expected population/sample size (following exclusion criteria)	*This has information on pupil exclusions as collected in the termly School Census*
Documentation	*https://www.find-npd-data.education.gov.uk/datasets/78f71e9f-856b-43ee-b0b8-749dd7dd2bb5 https://www.find-npd-data.education.gov.uk/datasets/a79b7ee2-b1cf-4a61-b564-9f6401ad4aa2*

Table 2.2.5 Dataset Description – Absences

Name of dataset	***Absences Default Data***
Data owner(s)	*Department for Education*
Type of data	*Cross-sectional education census*
Availability of data	*Licence required by the data owners*
Team member(s) who will have access	*Matteo Sandi, Stephen Machin, Lucas Silva Lopes, Saandra Nandakumar*
Population/geographic coverage or sampling frame	*Absence data for all pupils in state-maintained schools, PRUs and AP academies in England*
Years covered or survey waves	*2005-2021*
Exclusion criteria	*N/A*
Expected population/sample size (following exclusion criteria)	*This has information on pupil absences derived from the termly School Census*
Documentation	*https://www.find-npd-data.education.gov.uk/datasets/9cafe398-67af-4dc6-90f3-a9dec511ba92*

Table 2.2.6 Dataset Description – Educational Attainment

Name of dataset	***KS2 Pupil and Exam Table***
Data owner(s)	*Department for Education*
Type of data	*Cross-sectional education census*
Availability of data	*Licence required by the data owners*
Team member(s) who will have access	*Matteo Sandi, Stephen Machin, Lucas Silva Lopes, Saandra Nandakumar*
Population/geographic coverage or sampling frame	*All learners in England who have completed Year 6*
Years covered or survey waves	*2001-2021*
Exclusion criteria	*N/A*
Expected population/sample size (following exclusion criteria)	*Key stage 2 attainment data. This has information on the assessment of learners by the end of year 6 of schooling.*
Documentation	*https://www.find-npd-data.education.gov.uk/datasets/295d75da-8634-4fa0-9c80-ce3ba2113f7a* *https://www.find-npd-data.education.gov.uk/datasets/d6453111-b401-4420-a68f-7dad865d120f* *https://www.find-npd-data.education.gov.uk/datasets/d7c2aef7-d051-4b07-86c0-a619bcf94b96* *https://www.find-npd-data.education.gov.uk/datasets/82643964-d488-43b2-a50a-0cd4ee3fa2bc*

**We also use similar datasets for KS4 (end of Year 11) and KS5 (post-16 assessment)*

Table 2.2.7 Dataset Description – PLAMS

Name of dataset	***Post-16 Learning Aims (PLAMS) Default Data***
Data owner(s)	*Department for Education*
Type of data	*Cross-sectional education census*
Availability of data	*Licence required by the data owners*
Team member(s) who will have access	*Matteo Sandi, Stephen Machin, Lucas Silva Lopes, Saandra Nandakumar*
Population/geographic coverage or sampling frame	*Post-16 learning aims as collected in the School Census.*
Years covered or survey waves	*2007-2021*
Exclusion criteria	*N/A*
Expected population/sample size (following exclusion criteria)	*This has information on post-16 learning aims as collected in the School Census.*
Documentation	*https://www.find-npd-data.education.gov.uk/datasets/bd036978-12a2-4528-a9f8-b65711e6bc67*

Table 2.2.8 Dataset Description – Crime Records

Name of dataset	***Police National Computer***
Data owner(s)	*Ministry of Justice*
Type of data	*It is used to record convictions, cautions, reprimands and warnings for any offence punishable by imprisonment and any other offence that is specified within the regulations.*[7]
Availability of data	*Licence required by the data owners*

[7] See (Unlock, 2023)

Team member(s) who will have access	*Matteo Sandi, Stephen Machin, Lucas Silva Lopes, Saandra Nandakumar*
Population/geographic coverage or sampling frame	*All linked individuals from Dfe-MoJ dataset*
Years covered or survey waves	*2000-2021*
Exclusion criteria	*N/A*
Expected population/sample size (following exclusion criteria)	*All linked individuals from DfE-MoJ dataset*
Documentation	*https://www.data.gov.uk/dataset/ab2ef0ee-e741-43c7-b939-d88c19eb69b0/moj-extract-of-police-national-computer*

2.3 Primary data collection

This project uses only secondary data, specifically the dataset provided by the DfE-MoJ and constructed datasets from publicly available school-level data, described in the previous section. No primary data collection has been undertaken for this project.

2.4 Linking datasets

The publicly available school-level data on school characteristics and school-level dynamics that we collected will be merged with the DfE-MoJ dataset at the school-level using a school-specific anonymous identifier (URN).

2.5 Access and data protection

The DfE-MoJ dataset will be accessed uniquely via the ONS SRS. Therefore, our use of the data will be subject to the ONS' current regulations in place. We will not need to use any high identifiability data variables (i.e. levels 1 and 2)[8] in our analysis. However, we do need information on the anonymous individual identifier, e.g., the Pupil Matching Reference (PMR) number of pupils in the National Pupil Database (NPD), to be able to merge the different NPD and Ministry of Justice (MoJ) datasets together, e.g., PLASC data with KS4 data and criminal records, at the individual level.

[8] Data items (variables) in secure data is graded based on its identifiability on a scale of 1-5. The lower the number, the more identifiable an individual is because of that information. For example, name would have a value of 1, whereas country of residence would be more likely to have a value of 5.

We are aware of the foremost importance of preserving the confidentiality of the data in the analysis and we have extensive experience in working with highly confidential data in the UK and other countries for research purposes. The data will be stored on a secure server and will be accessed by ONS-accredited researchers within the LSE premises, and no attempt will be made to identify young individuals in the DfE-MoJ dataset. At CEP, we fully comply with the LSE Research Laboratory Security Standards for Sensitive Data that are publicly available on the LSE website at the following link:

LSE Research Laboratory Data Security Policy

LSE also publishes a privacy notice for research subjects that is available at the following link:

Privacy-Notice-for-Research-v1.2.pdf (lse.ac.uk)

Other LSE-wide information on security policies, if required, can be found at the link belowPolicies and procedures (lse.ac.uk)Should further checks of disclosure and conduct for the procedure be necessary, we would be glad to enclose them.

3 About the data

3.1 List of variables

Table 3.1.1 Key Variables in the DfE-MoJ dataset

Variable abbreviation	Variable definition	Variable source
PupilMatchingRefAnonymous	Character: Unique identifier for a pupil	DfE-MoJ: School Census Pupil Level
AgeAtStartOfAcademicYear	Numeric: Age of pupil at start of the academic year (in full years).	DfE-MoJ: School Census Pupil Level
EthnicGroup	Categorical: Pupil's ethnic group based on ethnic code.	DfE-MoJ: School Census Pupil Level
FSMeligible	Binary	DfE-MoJ: School Census Pupil Level

FirstLanguage	Categorical The language to which the child was exposed during early development and continues to use this language in the home or in the community. If a child acquires English after early development, then English is not their first language no matter how proficient in it they become. ENG = English ENB = Not known but believed to be English OTH = Other than English OTB = Not known but believed to be other than English REF = Refused NOT = Information not obtained	DfE-MoJ: School Census Pupil Level
SENprovision	Categorical: Provision types under the SEN Code of Practice. N = No Special Educational Need A = School Action or Early Years Action P = School Action Plus or Early Years Action Plus S = Statement	DfE-MoJ: School Census Pupil Level
LSOA01	National Statistics Postcode Directory Lower Layer Super Output Area derived from the pupil's postcode (based on 2001 Census)	DfE-MoJ: School Census Pupil Level
URN	School unique reference number.	DfE-MoJ: School Census Pupil Level
HomeLA	LA number based on pupil postcode	DfE-MoJ: School Census Pupil Level
StartDate	For each exclusion, exclusion start date	DfE-MoJ: Exclusions Data
PermanentExclusionInd	Binary: Permanent Exclusion Indicator.	DfE-MoJ: Exclusions Data
Reason	Categorical: For each exclusion, reason for exclusion.	DfE-MoJ: Exclusions Data
MoJUID	MoJ non-identifiable unique ID	DfE-MoJ: PNC
CaseID	Identifies individual cases related to each offender. One case may relate to multiple offences.	DfE-MoJ: PNC

OffenceID	Identifies individual offences for an offender in a case	DfE-MoJ: PNC
CourtCautionDate	The date on which the offender was convicted of, or cautioned for, the offence(s).	DfE-MoJ: PNC

Table 3.1.2 Key Variables for Ingested Secondary School Headteacher Dataset

Variable abbreviation	Variable definition	Variable source
urn	Unique Reference Number for the school	Previous Freedom of Information (FOI) requests published on the WhatDoTheyKnow website
year	Year the data about the school was collected.	Previous Freedom of Information (FOI) requests published on the WhatDoTheyKnow website
LAcode	DfE code for the local education authority the school belongs to	Previous Freedom of Information (FOI) requests published on the WhatDoTheyKnow website
LAname	Name of the local education authority the school belongs to	Previous Freedom of Information (FOI) requests published on the WhatDoTheyKnow website
school_number	Unique identification number assigned to the school within its local education authority	Previous Freedom of Information (FOI) requests published on the WhatDoTheyKnow website
LAESTAB	Local authority establishment number unique to each school. It is the amalgamation of the three digits from the “LAcode” variable and the four digits of the “school_number” variable	Previous Freedom of Information (FOI) requests published on the WhatDoTheyKnow website
school_name	Name of the school	Previous Freedom of Information (FOI) requests published on the WhatDoTheyKnow website

school_type	Describes the type of school in question (e.g community school, sponsored academy, academy converter, etc)	Previous Freedom of Information (FOI) requests published on the WhatDoTheyKnow website
gender_pupils	Describes whether the school is boys-only, girls-only, or mixed. 1 = boys-only 2 = girls-only 3 =mixed	Previous Freedom of Information (FOI) requests published on the WhatDoTheyKnow website
education_stage	The education stage the school belongs to. In this dataset, we only kept Secondary Schools and Middle-Deemed Secondary Schools	Previous Freedom of Information (FOI) requests published on the WhatDoTheyKnow website
D1	Binary indicator for whether the school switched headteacher on a given year. 0 = No 1= Yes	Previous Freedom of Information (FOI) requests published on the WhatDoTheyKnow website
DD1	Binary indicator for whether the school switched headteachers between 2006 and 2010. 0 = No 1 = Yes	Previous Freedom of Information (FOI) requests published on the WhatDoTheyKnow website
school_street	Street address for the school	Previous Freedom of Information (FOI) requests published on the WhatDoTheyKnow website
school_town	The town where the school is located	Previous Freedom of Information (FOI) requests published on the WhatDoTheyKnow website
postcode	Postcode for the school's address	Previous Freedom of Information (FOI) requests published on the WhatDoTheyKnow website

trevbal	Total recorded revenue balance (£) for the school on a given year	For the years of 2010 and 2011: The National Archives For the years of 2012-2022: Pupil and school finance data team
B01	Committed revenue balance (£) for the school on a given year	For the years of 2010 and 2011: The National Archives For the years of 2012-2022: Pupil and school finance data team
B02	Uncommitted revenue balance (£) for the school on a given year	For the years of 2010 and 2011: The National Archives For the years of 2012-2022: Pupil and school finance data team
cfextrevbal	Community focused extended school revenue balance (£)	For the years of 2010 and 2011: The National Archives For the years of 2012-2022: Pupil and school finance data team
trevbal_pcttinc	Total revenue balance as a % of total revenue income (£)	For the years of 2010 and 2011: The National Archives For the years of 2012-2022: Pupil and school finance data team
tpupils_sroll	Total number of pupils on school roll (all ages)	Find school and college performance data in England
Lvl2_acc1	share of students in the two that have achieved the Level 2 threshold in their GCSE	Find school and college performance data in England

3.2 Measurement of key concepts

Table 3.2 Measurement of key concepts

Concept	How the concept will be measured and encoded
Alternative Provision	In DfE data, there is a categorical variable that indicates if a pupil is assigned to alternative provision type *i* at time *t*. Therefore, we will be able to observe whether an individual is in an alternative provision setting and which kind of setting it is.
Youth Violence	In the MoJ data, charges as well as convictions, cautions and reprimands are recorded and report a rich array of information on both the offence and the offender which will be used in our analysis.
Exclusion	In the DfE data, permanent and temporary exclusions, as well as reasons for exclusion, are recorded and report a rich array of information on both the type of misconduct that was committed and the pupil which will be used in our analysis.

3.3 Missing data and attrition

We anticipate three missing data problems when using the DfE-MoJ linked dataset.

First, and arguably the biggest threat to the quality of the data used, lies with respect to pupils with frequently changing addresses eventually no longer being tracked by the NPD. To identify a pupil, the NPD makes use of instant pupil identifiers such as the pupil's name and postcode (DfE, 2022c). However, if personal circumstances dictate that a pupil must frequently change addresses over a short span, then it becomes harder for the NPD to accurately track this pupil across different years, until eventually the pupil leaves the datasets altogether. Given that frequent changes of address are more likely among pupils that come from low-income and broken households, it is therefore important to acknowledge that the pupils who have been able to be matched in both the NPD and the PNC datasets are more likely to come from households with more stable socio-economic conditions.

Second, an additional problem to be considered is the fact that individuals who are home-educated are effectively excluded from the DfE-MoJ linked dataset (ADR UK, 2022). Given that home education is also considered a type of AP, removal of home educated pupils from the linked dataset will not allow us to analyse the effect (if any) that this type of AP has on keeping pupils away from crime.

Finally, there is also the problem that individuals who were not able to be matched across the DfE and MoJ datasets have very specific characteristics with respect to gender, ethnicity and age. Regarding gender, the ADR UK (2022) report finds that 75% of the unmatched cases were male. Regarding ethnicity, the report finds that 75% of the unmatched cases belonged to

White Northern Europeans, followed by the general category of "Unknown" ethnicity[9] at 11% of all unmatched cases. Cases pertaining to Black, Asian, Middle Eastern, and Japanese, Chinese or Southeast Asians; the so-called BAME individuals, sum up to a total of 9% of all unmatched cases in the dataset. Finally, regarding age, unmatched cases were more likely to come from the older (initial) cohorts due to the greater probability of the address listed in the justice data matching the address listed in the education data for the younger cohorts. In this sense, we anticipate the population of white, male and older individuals to be under-represented in the MoJ-DfE dataset.

We are aware that the issue of pupils disappearing from the DfE dataset is likely to be biased towards children that are excluded from school. The extent to which this is the case will be tested comparing the rate at which pupils disappear from the dataset whether they are excluded or not. This comparison will be made using regression analysis and controlling for other potential determinants of this attrition in the data (e.g., foreign native language). If a significant fraction of excluded pupils systematically disappeared from the dataset after exclusion, then this could constitute an issue for the design of this study and our findings and policy implications would be limited to the pupils that we are able to observe in the data and study in our analysis.

Apart from these three shortcomings, we do not anticipate any additional gaps in our data. This is because we requested access to the above NPD extract for all pupils in state-maintained schools, pupil referral units and alternative provision in all school years linked at the individual level with the Police National Computer data from 2000 to 2020. From the Police National Computer 2000-2020, we requested to access the list of variables requested above for records at all ages of individuals (i.e., for a linked individual, and to observe the list of variables requested also for criminal offences occurred after a linked individual has reached the compulsory schooling age and, thus, occurred at an older age than the compulsory schooling age). The DfE-MoJ also provides a Match Quality dataset that provides details on how each offender was matched to the NPD: this information would allow us to choose the observations for the analysis better, as well as highlight any potential biases in the matching processing. The ADR UK (2022) report finds that 70% of individuals with a MoJ identifier can be identified to an individual in the DfE data sources.

3.4 Other sources of bias

[9] It is important to keep in mind that in the Police National Computer data, ethnicity of an individual is not self-reported but rather identified by the officer in question. This could potentially explain why the "Unknown" ethnicity category is the second leading category among unmatched cases.

Even though our analysis uses administrative data from DfE and MoJ, the data may be biased as some ethnic groups may be over-represented and some others may be under-represented. Regarding permanent exclusion data, the latest data (for the 2020-2021 school year) on permanent exclusion rates by ethnicity from the Department of Education (DfE) identifies Roma/gypsy pupils as the ethnic group with the highest rate of permanent exclusions, followed by Mixed White and Black Caribbean in second place (DfE, 2023a). However, DfE notes that Roma/gypsy pupils make up a small share of the total amount of pupils within English schools, and therefore their high rates of permanent exclusion should be interpreted with caution.

Regarding criminal activity data, statistics from the Ministry of Justice (MoJ) from 2019 acknowledge that people from BAME ethnic groups (Black, Asian, Mixed, Chinese, and "other") are over-represented in the UK at every single stage of the UK criminal justice system, be it arrest, prosecution, conviction, or imprisonment (Yasin & Sturge, 2020). The authors also explain that in the UK criminal system, pleading "guilty" at the sentencing stage often leads to a sentence length discount of one third. However, the authors also highlight that pleading guilty as early as in the sentencing stage is correlated with a greater degree of trust in the criminal justice system, which is something higher among White than among BAME defendants. As a result, while White defendants have a higher rate of "guilty" pleading, the average sentence length for BAME defendants in 2019 was 27.1 months compared to 19.5 months for White defendants (Yasin & Sturge, 2020). Given this sharp discrepancy in trust with respect to the UK criminal justice system, we therefore expect BAME individuals to be over-represented both in terms of offending and reoffending statistics in the datasets.

4 About the analysis

4.1 Overview of analytical approach

As soon as we receive the permission from the data owners, we will start conducting tests of reliability of the linked DfE-MoJ dataset (henceforth, the data), and core dimensions of data quality (completeness, uniqueness, timeliness, validity, accuracy, and consistency) will be assessed. We will conduct an initial exploration of the linked DfE-MoJ dataset, a descriptive analysis, as well as empirical tests to answer some key questions about the design and methodology, i.e. do we have a valid instrumental variable we could potentially use to estimate causal effects. We will then start the preparation of a descriptive interim report on the evolution of alternative provision in England in the last 20 years, i.e., from the early 2000s.

4.2 Approach to addressing research question(s)

Research question 1: Approach and Methods

Research question	How has the alternative provision landscape in England changed over the past decade: where do the excluded children go after permanent exclusion, what are the types of AP available to them and how has this changed over the past two decades?
Hypothesis, if relevant	-
What will you be able to say by the interim report	By the interim report we will be able to provide a detailed and clear picture of the scale and variety of AP and how prevalent each option is.
Descriptive analysis, if relevant	Describe the evolution of AP in England over the last 20 years, using graphs and tables to understand the introduction and relative prevalence of the different AP options available today. We will describe the typical pathways of pupils who go through the exclusion process: where do they go after exclusion? Does this vary by age of the pupil at the time of the exclusion? How many and which pupils get excluded multiple times? How long does it last before they get excluded again? We will also describe changes in legislation that may have played a role in shaping this landscape. We are also interested in potential factors that affect availability and kinds of AP in an area, such as the level of deprivation, which may play a role in where children go after permanent exclusion.
Models, specifications and statistical techniques used, if relevant	This analysis will be descriptive.
Estimating equation, if relevant	N/A.
What does the approach need to succeed (constraints/assumptions)?	We require data on the type of AP that pupils go to after permanent exclusion, available in the Alternative Provision Census and PRU Census. We require that these students' educational path post-exclusion is tracked in the DfE-MoJ

	dataset to be able to study these questions descriptively as proposed here.
Uncertainty and inference	We do not know yet the extent to which excluded pupils are represented in the DfE-MoJ dataset.
Robustness checks	-
Subgroup you intend to study	Ethnic minorities and pupils excluded from school at different ages.
Changes to the analysis	-

Research question 2: Approach and Methods

Research question	Which school-level or local authority-level factors predict permanent exclusion?
Hypothesis, if relevant	Individual factors (e.g., gender, age, free-school-meal eligibility status), school-level factors (e.g., student body composition, principal switches) and local authority factors (e.g., availability of AP) may affect the probability that a pupil is permanently excluded. The increasing rates of exclusion witnessed in the UK in the last decade may have generated increased demand for AP and thus may have led to the increased offer of formal and informal AP settings in England. On the other hand, AP availability in the locality may positively affect the likelihood of exclusion. We will examine this correlation empirically. We will test for a potential instrumental variable for the diffusion of AP settings, which would allow us to move from a correlational analysis to a more causal analysis of the impact of the provision of AP on permanent exclusion.
What will you be able to say by the interim report	We will be able to provide a descriptive picture of exclusion rates in England in the last 20 years. We will also be able to highlight factors that are correlated with the probability of

	a pupil being permanently excluded, which may allow us to identify potential controls for better estimates.
Descriptive analysis, if relevant	We will produce graphs to show the correlations between school exclusions and school-level and local authority-level factors of interest.
Models, specifications and statistical techniques used, if relevant	We look at factors at three levels that may affect the probability of a pupil getting permanently excluded: the pupil-level, school-level as well as local-authority level. Understanding school-level and local authority-level factors that affect permanent exclusion may have important policy implications, while pupil-level factors may serve as useful controls and can help highlight the differences between groups (male vs female pupils, for example). Apart from looking at simple correlations, we aim to model the probability that a pupil gets excluded as a function of these aforementioned factors. The primary factor of interest is the local authority-level diffusion of AP. Since the outcome of interest is binary (the student is either permanently excluded or not), we can model this relationship as an OLS, probit or logit regression that calculates the probability of a pupil being permanently excluded from school as a function of pupil characteristics, school characteristics, and local authority characteristics, including the local authority-level diffusion of AP. However, this estimate could be biased as there may be unobservable factors that affect both diffusion of AP as well as a pupil's likelihood of exclusion. If we are able to identify a valid instrument for the diffusion of AP, we can estimate the causal impact of the diffusion of AP at the local authority-level on the pupil's likelihood of exclusion. This would allow us to better understand the relationship between exclusions and the AP landscape in England.
Estimating equation, if relevant	$$e_{isl} = \delta_0 + \sum_{a=1}^{A} \beta_a P_a + \sum_{b=1}^{B} \beta_b S_b + \sum_{c=1}^{C} \beta_c L_c + \varepsilon_{isl}$$ where e_{isl} is a binary variable that indicates whether pupil i in school s in local authority l has been permanently excluded (from school s). P refers to the set A of pupil-level

	characteristics, such as gender or FSM-eligibility, S refers to the set *B* of school-level characteristics such as the type of school (maintained school vs academy for example), and L refers to the set *C* of local-authority level characteristics, such as the diffusion of AP or local crime rate. If we are able to identify a valid instrument for the diffusion of AP in the local authority, *D*, we would be able to estimate diffusion of AP in the local authority as: $$\widehat{AP} = \delta_1 + \beta_0 D + \sum_{a=1}^{A} \beta_{a2} P_a + \sum_{b=1}^{B} \beta_{b2} S_b + \sum_{c=1}^{C-1} \beta_{c2} L_c + \varepsilon_1$$ Thereby the likelihood of exclusion could be expressed as: $$e_{isl} = \delta_2 + \sum_{a=1}^{A} \beta_{a2} P_a + \sum_{b=1}^{B} \beta_{b2} S_b + \sum_{c=1}^{C-1} \beta_{c2} L_c + \widehat{AP}_l + \varepsilon_{2isl}$$
What does the approach need to succeed (constraints/assumptions)?	We require access to the DfE data on permanent exclusions as well as the Pupil Census and the School-level data we have collected and merged with the DfE data at the school (URN) level.
Uncertainty and inference	Uncertainty will be reported, and inference will be conducted clustering standard errors at the local authority level.
Robustness checks	We will test robustness of our findings to sampling restrictions and estimation methods.
Subgroup you intend to study	Ethnic minorities, pupils at different ages and children in London vs elsewhere in England.
Changes to the analysis	Setbacks and changes to analysis may occur. Depending on the data availability provided in the DfE-MOJ dataset, we may have to switch the unit of analysis from the pupil level to the school level.

Research question 3: Approach and Methods

Research question	How does attending different forms of alternative provision (for example, a pupil referral unit vs an alternative provision academy) affect youth crime?
Hypothesis, if relevant	Given the recent increase in the forms of AP that are available to excluded pupils, we are keen to study what works best to keep excluded pupils away from crime. Different experiences post-exclusion may increase the risk of crime. However, we do not have a strong prior as to which forms of AP may increase the risk of crime.
What will you be able to say by the interim report	By the interim report, we will be able to highlight correlations between the type of AP an individual is exposed to and their risk of crime.
Descriptive analysis, if relevant	We will describe the correlations observed in the DfE-MoJ data between the type of AP an individual is exposed to and their risk of crime.

Models, specifications and statistical techniques used, if relevant	We will focus on excluded pupils and define a series of difference-in-differences specifications regressing the outcomes of interest for pupil *i* in year *t*, i.e., permanent exclusion or youth crime of pupil *i* in year *t*, on the enrolment (and duration of enrolment) in either a PRU, or an AP academy of pupil *i* in year *t*. In this case, we will first focus on excluded pupils and estimate a set of simple OLS regressions modelling criminality of excluded pupil *i* as a function of the type of AP attended after being excluded. A second approach will be to use IV techniques to instrument enrolment in an AP academy by pupil *i* in year *t*. One potential instrument might be headteacher switches in secondary schools. Preliminary analysis using school-level data shows that the event of a headteacher switch is on average followed by a spike in the number of exclusions in the school. Replicating this analysis using individual level data would result in a lot more statistical power, and thus this is very unlikely not to work. The assumption behind this approach would be that headteacher switches affect subsequent enrolment in different forms of AP and criminality only by raising the risk of permanent exclusion. This seems plausible as school headteachers have no authority over excluded pupils following the exclusion. That being said, however, we will also test for parallel pre-trends in the outcome of interest prior to the principal switch to check the validity of this research design. If pre-trends do not look parallel, a set of alternative empirical approaches (propensity score matching, synthetic control and difference-in-discontinuity) will be used.
Estimating equation, if relevant	$Offend_{ilt} = \lambda_i + \theta_l + \beta_t + \sum_{a=-q}^{-1} \gamma_a \cdot AP_a + \sum_{a=0}^{m} \gamma_a \cdot AP_a + \varepsilon_{ilt}$ We are interested in the probability of a pupil *i* from a local education authority *l* committing a crime in a specific time

period *t* and how being assigned permanently to a type of AP Institution (e.g. PRU, AP Free School, AP Academy) affects that probability. This is captured by the binary variable *Offend* in the equation above. The series of difference-in-differences captured by the equation above attempts to answer this question by observing the pupil both before and after being assigned permanently to an AP institution. Treatment in this case is captured by the variable *AP,* which tells us whether the pupil has been permanently assigned to an AP institution. We intend to run separate analys for permanent assignment to PRUs, AP Free Schools, AP Academies, and all other types of AP institutions.

In this sense, the pre-treatment period consists of the time periods from a=q to a=-1, the last time period before the pupil is sent to the AP institution. In a similar fashion, the post-treatment time period consists of the time periods from a=0 (the first time period when the excluded pupil is assigned permanently to an AP institution) to a=m (the last time period the excluded pupil is observed in the dataset). λ_i refers to individual pupil characteristics that may contribute to the probability of the pupil committing an offence (e.g, previous history with the police). θ_l refers to local education authorities' specific characteristics that may contribute to the probability of the pupil committing an offence. Finally, $\beta_t$refers to time period characteristics that may contribute to the probability of the pupil committing an offence.

For the second approach, which consists of the use of an instrumental variable, the strategy explained in the preceding section also applies here, and is described by the same system of equations

$$\boldsymbol{AP}_{it} = f(s, t)+\theta_{\boldsymbol{4}} ITT_{ia}+ \sum_{j=1}^{J} \beta_{4j} X_{jit}+ \phi_{\boldsymbol{4}} P_{it}+ v_{4it}$$

	$$Y_{it} = f(s, t) + \theta_5 ITT_{ia} + \sum_{j=1}^{J} \beta_{5j} X_{jit} + \phi_5 P_{it} + v_{5it}$$
What does the approach need to succeed (constraints/assumptions)?	It requires parallel pre-treatment trends in the outcome variable that is studied. For the instrumental variable approach to be successful, we also need a strong predictive power of headteacher switch in the first stage regression. Furthermore, for the instrumental variable approach to work, it is paramount that headteacher switches, the proposed IV, affects subsequent enrolment in different forms of AP and criminality only by raising the risk of permanent exclusion for the pupil.
Uncertainty and inference	Uncertainty will be reported, and inference will be conducted clustering standard errors at the LA level.
Robustness checks	We will test robustness of our findings to sampling restrictions and estimation methods.
Subgroup you intend to study	Excluded pupils in England. We will study our outcomes of interest separately for pupils by gender, ethnicity, FSM status, locality and school.
Changes to the analysis	Setbacks and changes to analysis may occur. Alternative empirical strategies may be pursued in that case, such as matching techniques to generate a proper control group for our analysis.

5 Project management

5.1 Risks and Mitigants

Table 5.1 Risks and mitigations

Number	Risk	Likelihood (*Low/Medium/ High*)	Mitigation

1	*Data Reliability*	*Low*	We have extensive experience of assessing data reliability for DfE as well as numerous police forces in the UK. As a recent example, since 2016 we have had access to National Pupil Database (NPD) data linked with HMRC data on individual tax records and DWP data on individual records of benefits receipts. We are also currently examining the database of the West Midlands Police (WMP) and providing analytical support to WMP's operational agenda. We recently produced more than 200 pages of descriptive results and presented this in meetings with WMP's data analysts and senior officials. Our analysis revealed empirical trends that were not known to WMP before. This analysis also exposed anomalies in the data and led to modifications in the production of official statistics and data extraction practices by WMP. We have experience of dealing with missing data and we would be able to detect whether some groups of population are disproportionately represented in a pool of observations that may be missing.
2	*Identifying individuals from the data*	*Low*	We do not need to use any high identifiability data variables (i.e.

			levels 1 and 2) in our analysis. In contrast, we need information on the anonymous individual identifier, e.g., the Pupil Matching Reference (PMR) number of pupils in the National Pupil Database (NPD), to be able to merge the different NPD and Ministry of Justice (MoJ) datasets together, e.g., PLASC data with KS4 data and criminal records, at the individual level. Our analysis of the DfE-MoJ data linkage will strictly comply with the regulations in place by the data owners as well as by the ONS. The DfE-MoJ dataset contains de-identified data for each individual, making it impossible to identify any particular person within the dataset. Furthermore, as part of our data access agreement, we are subject to strict data disclosure protocols, and any observations below a threshold of 10 will be suppressed and removed from any documentation that is prepared for publication.
3	*Data Confidentiality*	*Low*	We are aware of the foremost importance of preserving the confidentiality of the data in the analysis and we have extensive experience in working with highly confidential data in the UK and other countries for research purposes. No

			identifiable information will be revealed to anyone of course, and no attempt will be made to identify young individuals in the DfE-MoJ dataset. At CEP, we fully comply with the LSE Research Laboratory Security Standards for Sensitive Data that are publicly available on the LSE website at the following link: LSE Research Laboratory Data Security Policy LSE also publishes a privacy notice for research subjects that is available at the following link: Privacy-Notice-for-Research-v1.2.pdf (lse.ac.uk) Other LSE-wide information on security policies, if required, can be found at the link below: Policies and procedures (lse.ac.uk) Should further checks of disclosure and conduct for the procedure be necessary, we would be glad to enclose them.
4	*Data Complexity*	*Low*	We have detailed knowledge of the NPD data and we are extremely familiar with its structure. In particular, as we

			have examined the legislation and the dynamics of behavioural outcomes in England (i.e., school absence as well as lunchtime, temporary and permanent exclusions), we would be able to detect anomalies in the data and thus test its reliability very easily. We do not foresee any difficulties in producing a metadata dictionary and giving highly relevant, practical recommendations in analysing the available data.

5.2 Timeline

Using the table template below, please include a project timeline with the main activities, dates and who's in the project team. Where possible include specific dates or date intervals.

Stage	**Task**	**Timing - Staff Responsible**
Project start	Submit application to ONS for access to PRU Census and AP Census datasets.	February 2023 – Sandi and Machin
	Start of hiring process of one or more part-time Research Assistants (RAs) who will be supervised by Matteo Sandi and Stephen Machin.	March-April 2023 – Sandi and Machin
	Start of descriptive interim report on the evolution of alternative provision in England in the last 20 years, i.e., from the early 2000s.	December 2023 – Sandi, Machin, Nandakumar and Silva Lopes

Agree study plan	Discussion between YEF and CEP on study plan	January 2024 – Sandi, Machin, Nandakumar and Silva Lopes
Data Access	Submit application to ONS for access to PRU Census and AP Census datasets.	February 2023 – Sandi and Machin
	Complete data access obtained.	February 2024 – Sandi, Machin, Nandakumar and Silva Lopes
Interim report	Descriptive interim report completed on the evolution of alternative provision in England in the last 20 years, i.e., from the early 2000s. Complete data access obtained.	June 2024 – Sandi, Machin, Silva Lopes, and Nandakumar
Final report	Dissemination of preliminary findings and presentation of the early results of this analysis and collection of feedback from YEF colleagues.	September 2024 – Sandi, Machin, Silva Lopes and Nandakumar
	Respond to comments from YEF and YEF appointed external peer review	October 2024 – Sandi, Machin, Silva Lopes and Nandakumar
	Submit final report	December 2024 – Sandi, Machin, Silva Lopes and Nandakumar

6 References

ADR UK (Administrative Data Research UK). (2022, July). *Ministry of Justice – Department for Education linked dataset Feasibility of evaluating early interventions for violence prevention: Data quality report.* Retrieved from: https://www.adruk.org/fileadmin/uploads/adruk/Documents/Feasibility_study_1_MoJ-DfE_linked_dataset_Data_quality_report.pdf

Bell, B., Costa, R., and Machin, S. (2022). Why Does Education Reduce Crime? *Journal of Political Economy,* 130 (3), pp. 732 – 765. https://doi.org/10.1086/717895

Cole, T., Daniels, H., and Visser, J. (2003). Patterns of Provision for Pupils with Behavioural Difficulties in England: a study of government statistics and behaviour support plan data. Oxford Review of Education, 29(2), pp.187-205. DOI: 10.1080/0305498032000080675

The Crown Prosecution Service. (2019). *Indictable only cases: sending to the Crown Court*. Retrieved from: https://www.cps.gov.uk/legal-guidance/indictable-only-cases-sending-crown-court

Department for Children, Schools and Families. (2008). *Back on Track: A strategy for modernising alternative provision for young people*. Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/328241/Back_on_Track.pdf

Department for Education (2012). *Questions & Answers Pupil Referral Units conversion to Alternative Provision Academies.* Retrieved from: https://www.minutes.haringey.gov.uk/documents/s48844/item%206%20PRU%20academy%20conversion%20FAQ.pdf#:~:text=The%20Education%20%28Pupil%20Referral%20Units%29%20%28Application%20of%20Enactments%29,work%20on%20their%20conversion%20since%2027%20February%202012.

Department for Education. (2019, April). *Elective home education Departmental guidance for local authorities*. Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/791527/Elective_home_education_gudiance_for_LAv2.0.pdf

Department for Education (2022, May). *Children not in School : Schools Bill Factsheet*. Retrieved from : https://assets.publishing.service.gov.uk/media/628bab96e90e071f67b2803d/Children_not_in_School_-_Schools_Bill_Factsheet.pdf

Department for Education (2022, July 11). *Understanding the use of unregistered alternative provision : Call for Evidence*. Retrieved from: https://consult.education.gov.uk/behaviour-attendance-exclusions-and-alternative-provision/unregistered-alternative-provision-call-for-

eviden/supporting_documents/Unregistered%20AP%20Call%20for%20Evidence%20document.pdf

Department for Education (2022, December 22). *Guidance National Pupil Database (NPD): Privacy Notice.* Retrieved from: https://www.gov.uk/government/publications/national-pupil-database-npd-privacy-notice/national-pupil-database-npd-privacy-notice#the-kinds-of-personal-data-we-process-about-you

Department for Education (2023, February 7). *Ethnicity facts and figures: Permanent exclusions.* Retrieved from: https://www.ethnicity-facts-figures.service.gov.uk/education-skills-and-training/absence-and-exclusions/permanent-exclusions/latest#:~:text=Permanent%20exclusion%20rate%2C%20and%20number%20of%20permanent%20exclusions%2C,%20%2020%20%2020%20more%20rows%20

Department for Education (2023, July 20). *Permanent exclusions and suspensions in England, Summer term 2021/22.* Retrieved from*:* https://explore-education-statistics.service.gov.uk/find-statistics/permanent-and-fixed-period-exclusions-in-england/2021-22-summer-term

Department for Education (2023, November 23). *Suspensions and Permanent Exclusions in England: Autumn Term 2022/2023*. Retrieved from: https://explore-education-statistics.service.gov.uk/find-statistics/suspensions-and-permanent-exclusions-in-england

Department for Education, & Ministry of Justice (2022, March 10*) Education, children's social care and offending.* Retrieved from: https://www.gov.uk/government/publications/education-childrens-social-care-and-offending

Education Act 1996, c.56. Retrieved from: https://www.legislation.gov.uk/ukpga/1996/56/section/19

Education and Skills Funding Agency. (2021). High needs funding: 2021 to 2022 Operational guide. Retrieved from: https://assets.publishing.service.gov.uk/media/602b99138fa8f5037d3719e6/High_needs_funding_operational_guide_2021_to_2022.pdf

Elkin, M. (2019). *Homicide in England and Wales - Office for National Statistics.* Retrieved from: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/homicideinenglandandwales/yearendingmarch2018#highest-number-of-sharp-instrument-homicides-seen-in-year-ending-march-2018

House of Commons Education Committee. (July 2018). Forgotten children: alternative provision and the scandal of ever increasing exclusions. House of Commons. Retrieved from: https://publications.parliament.uk/pa/cm201719/cmselect/cmeduc/342/342.pdf

Institute of Education and National Foundation for Educational Research. (2014, July). School exclusion trial evaluation. Department for Education. Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/331795/RR364_-_School_Exclusion_Trial_Final_Report.pdf

Jacob, B. A., and Lefgren, L. (2003). Are Idle Hands the Devil's Workshop? Incapacitation, Concentration, and Juvenile Crime. *American Economic Review*, 93 (5): 1560-1577. DOI: 10.1257/000282803322655446

Kendall, S., Kinder, K., Halsey, K.,Fletcher-Morgan, C., White, R., Brown, C. (2003). An Evaluation of Alternative Education Initiatives. National Foundation for Educational Research. Retrieved from: https://www.nfer.ac.uk/publications/AES01/AES01.pdf

Luallen, J. (2006). School's out… forever: A study of juvenile crime, at-risk youths and teacher strikes. *Journal of Urban Economics*, 59 (1), pp. 75-103. doi:10.1016/j.jue.2005.09.002

Machin, S., Marie, O., and Vujic, S. (2011). The Crime Reducing Effect of Education. *The Economic Journal,* 121 (552), pp.463-484. https://www.jstor.org/stable/41236987

Ofsted. (2011, June). *Alternative Provision.* Retrieved from: https://assets.publishing.service.gov.uk/media/5a81a73eed915d74e62336cf/Alternative_provision.doc

Ofsted. (2024, February 1). *Alternative provision in local areas in England: a thematic review*. Retrieved from: https://www.gov.uk/government/publications/alternative-provision-in-

local-areas-in-england-a-thematic-review/alternative-provision-in-local-areas-in-england-a-thematic-review#Figure-1

Ogg, T., and Kaill, E. (November 2010). *A New Secret Garden? Alternative Provision, Exclusion and Children's Rights*. Civitas. Retrieved from: https://civitas.org.uk/pdf/NewSecretGarden.pdf

Sandi, M., and Boshoff, J. (2022). *Data Explained - Ministry of Justice - Department for Education Linked Dataset*. Retrieved from: https://www.adruk.org/fileadmin/uploads/adruk/Documents/Data_Explained/Data_Explained_Matteo_Sandi.pdf.

Schraer, B. R. (2019, March 8). *Knife crime: Are school exclusions to blame? BBC News*. Retrieved from: https://www.bbc.co.uk/news/uk-47485867

Sharma, N. (2012, July). *Improving Alternative Provision – A London Councils Member Briefing.* Retrieved from: https://www.londoncouncils.gov.uk/node/1299

Spielman, A. (2021, November 5). A loophole is leaving the most vulnerable pupils in England at risk. The Guardian [online]. Retrieved from: https://www.theguardian.com/commentisfree/2021/nov/05/loophole-vulnerable-pupils-england-risk-alternative-provision-registered

Timpson, E. (2019). *Timpson Review of School Exclusion*. Department for Education. Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/807862/Timpson_review.pdf

UK Government. (2023). *Types of School*. Retrieved from: https://www.gov.uk/types-of-school/free-schools

Unlock. (2023). *Recordable offences.* Retrieved from: https://unlock.org.uk/advice/recordable-offences-2/

Walker, A. (2022, September 23). *School suspensions up 13%, but exclusions continue to fall. Schools Week*. Retrieved from: https://schoolsweek.co.uk/school-suspensions-up-pandemic-and-exclusions-continue-to-fall/

Yasin, B. and Sturge, G. (2020, October 02). *Ethnicity and the criminal justice system: What does recent data say on over-representation?* Retrieved from: https://commonslibrary.parliament.uk/ethnicity-and-the-criminal-justice-system-what-does-recent-data-say/