

# Evaluating Teacher Recruitment Strategies

## Evaluation Study Plan

Evaluator (institution): Education Intelligence

Principal investigator(s): Professor Becky Allen



|                                      |  |
|--------------------------------------|--|
| <b>PROJECT TITLE</b>                 | Evaluating teacher recruitment strategies  |
| <b>EVALUATOR (INSTITUTION)</b>       | Education Intelligence Ltd (Teacher Tapp)  |
| <b>PRINCIPAL INVESTIGATOR(S)</b>     | Professor Becky Allen  |
| <b>STUDY PLAN AUTHOR(S)</b>          | Becky Allen, Burak Sonmez  |
| <b>STUDY DESIGN</b>                  | Fully randomised teacher paired conjoint experiments (a type of discrete choice experiment)  |
| <b>PUPIL AGE RANGE AND KEY STAGE</b> | KS1-KS4  |
| <b>NUMBER OF SCHOOLS</b>             | Over 5,000   |
| <b>NUMBER OF PUPILS</b>              | Around 8,000 teachers  |
| <b>OUTCOME MEASURE AND SOURCE</b>    | <u>Teacher conjoint experiment</u><br>Job preference for schools, given all attributes (expressed as willingness to apply in terms of % salary change)<br>Source: Teacher responses                        |
|                                      | <u>Headteacher conjoint experiment</u><br>Preference for teachers by school leaders, given all attributes (expressed as willingness to pay in terms of % salary change)<br>Source: School leader responses |
|                                      | <u>Teaching in disadvantaged schools</u><br>Willingness of teachers to take a job at another local school given a particular salary (expressed as required % salary change)<br>Source: Teacher responses   |

## Study Plan version history

| VERSION | DATE       | REASON FOR REVISION |
|---------|------------|---------------------|
| 1.0     | 13/12/2024 | Original version    |

## **Table of contents**

|   |    |
|---|----|
| <i>Background and study rationale</i> .....                                   | 4  |
| <i>Intervention – teacher preferences for job and school attributes</i> ..... | 5  |
| <i>Evaluation design</i> .....  | 7  |
| <i>Inference</i> .....  | 19 |
| <i>Implementation and process evaluation (IPE)</i> .....                      | 20 |
| <i>Cost evaluation</i> .....  | 21 |
| <i>Ethics</i> .....   | 21 |
| <i>Data protection statement</i> .....  | 22 |
| <i>Personnel</i> .....  | 24 |
| <i>Risks</i> .....  | 25 |
| <i>Timeline</i> .....   | 28 |
| <i>References</i> .....   | 29 |

## Background and study rationale

England faces significant challenges in teacher recruitment and retention, especially in secondary education shortage subjects amid rising pupil numbers (Allen et al., 2023). Previous research on teacher job choices has largely focused on surveying or analysing enacted job choices. Salary is a key retention factor (Worth et al., 2018), yet many teachers prioritise a quality working environment over pay rises (See et al., 2020). Over half report unmanageable workloads (Jerrim and Sims, 2019), a major reason for leaving the profession (Perryman and Calvert, 2020). Many teachers highly value flexible or part-time work (Worth, 2023; Sharp et al., 2019), but this is challenging to implement.

This research study uses paired conjoint experiments, alongside other quantitatively descriptive analysis, to estimate how much teachers value specific job attributes and how much headteachers feel able to accommodate these job characteristics in their school. We do this via three research questions that together answer the question:

***[RQ1] Which working conditions or benefits do teachers find most attractive in job adverts?***

***[RQ2] What is headteachers' willingness to pay (in salary terms) for job attributes that are costly for schools to provide in terms of financial burden or organisational flexibility?***

***[RQ3] What salary increase or job benefits are necessary to encourage teachers to accept positions at schools with a greater number of students receiving free school meals?***

A fully randomised paired conjoint experiment methodology will be employed to investigate the first two of these questions, with a non-experimental approach providing new evidence on the third question. Specifically, our study employs a paired conjoined experiment, to study how teachers make choices between complex job adverts that involve multiple attributes. Our sample size will be sufficient to examine variations in job attribute preferences across different demographic groups such as subject, region and family circumstances, which is crucial given the uneven pattern of current teacher shortages.

Paired conjoint experiments allow us to quantify teacher's relative preferences for different job attributes by requiring them to make trade-offs between possible policy options. We consider the most relevant studies in this area to be Burge et al. (2021) in the UK and Lovison and Hyunjung Mo (2024) in the US. Burge et al. (2021) particularly focuses on the value of pecuniary compensation, pension variations, bonuses and pay progression and their core findings will not be replicated in this study. Whilst some job attributes in Lovison and Hyunjung Mo (2024) are US-specific, we aim to replicate others.

The theoretical starting point of the research is that teachers make informed job selections based on the job and school attributes that maximise their expected utility or satisfaction. Equally, school leaders select between job candidates based on maximising the outputs of the school. We assume that jobs can be described as a bundle of these attributes. Since most of these job and school attributes are non-pecuniary, there may be efficiency gains from offering teachers particular employment arrangements that all sides find it hard to straightforwardly value in financial terms.

The paired conjoined experiment allows us to estimate the implicit pecuniary value of factors that are hard to value, such as disutility of workload, stressful experiences during the school day, and job flexibilities. To complement the experiment, qualitative research will both inform

and support the interpretation of findings. We will conduct focus groups prior to the experiment to identify a set of relevant job attributes to test. During the experiment itself, open-ended questions will follow each choice task, prompting participants to explain their choice. Then, following the experiment, additional focus groups with school leaders will explore ways to implement changes based on the experiment's insights into what teachers' value most. This qualitative work will help us probe teacher's credibility of the job adverts presented, enhance our understanding of how stated preferences align with a utility maximisation model, and provide insights to support practical implementation.

## **Intervention – teacher preferences for job and school attributes**

### *Research questions*

School working environment is a complex construct, encompassing attributes like workload, leadership, and professional development, and so poses a recruitment challenge for school leaders seeking to articulate these aspects compellingly in job adverts.

Therefore, our study aims to directly support their challenges by asking:

#### ***[RQ1] Which working conditions or benefits do teachers find most attractive in job adverts?***

Identifying the job attributes that teachers value most can inform school leader's recruitment efforts. However, many of these job attributes entail considerable costs for schools, either in direct financial outlay or organisational flexibility. It is difficult to calculate the financial cost of potential job attributes, as they can affect organisational costs and flexibilities in quite complex ways.

The literature on how leaders make decisions and what they favour tends to be descriptive or correlational and does not estimate their relative willingness to accommodate different types of policies within their school (e.g. Lynch et al., 2024; Robinson and Gray, 2019). For example, there exists only descriptive research on school leaders' feelings about providing job flexibility (Allen et al., 2023; Harland et al., 2023). Therefore, to provide informed recommendations regarding which job attributes should be prioritised, it is essential to understand their implicit cost-effectiveness.

Thus, our second research question is:

#### ***[RQ2] What is headteachers' willingness to pay (in salary terms) for job attributes that are costly for schools to provide in terms of financial burden or organisational flexibility?***

Whilst it is well-documented that schools serving disadvantaged communities are disproportionately affected by teacher shortages, with resultant greater use of non-specialist teachers, inexperienced teachers and higher teacher turnover (Allen and Sims, 2018; Allen and McInerney, 2019), less is known about why these schools find it so difficult to attract staff. Given that schools serving disadvantaged pupils are given greater funding, both through the Dedicated Schools Grant (DSG) and the pupil premium, this research could make a valuable policy contribution by measuring the adequacy of these resources to cover the aforementioned strategies to attract well-qualified teachers.

Hence, our third research question is:

***[RQ3] What salary increase or job benefits are necessary to encourage teachers to accept positions at schools with a greater number of students receiving free school meals?***

### **Summary of survey approach**

We propose a three-stage research methodology, all to be carried out via the Teacher Tapp survey app during the 2024/25 academic year. We provide further details of each research question's design below, alongside an in-depth consideration of the sample. But briefly, Teacher Tapp is a daily survey app that surveys over 10,000 teachers in England. It is a convenience sample in the sense that any teacher is free to download and use the survey app, with post-stratification weights applied to ensure the sample represents the population of teachers on known demographic characteristics.

**Figure 1: A mock-up of a paired conjoint experiment on a mobile screen**

Which job advert do you prefer?

A school offering:

- Comprehensive CPD offer including one day per half-term for self-directed professional development
- School supports staff health and wellbeing including through time off for appointments and a wellbeing helpline
- Planning, Preparation and Administration (PPA) time can be taken at home if scheduling allows
- Salary 5% higher than your current job

A school offering:

- The school values and supports the importance of CPD
- School supports staff health and wellbeing
- Planning, Preparation and Administration (PPA) time given as one day working from home per week
- Same salary as your current job

Figure 1 shows a mock-up of the types of choices teachers will be asked to make within the main paired conjoint experiments of this study. The Teacher Tapp survey dashboard allows full randomisation of the order of the choices presented, the levels of the job attributes, and the order in which the job attributes are listed. Since Teacher Tapp holds full school and demographic information on all panellists, this background information will not need to be re-asked alongside the experiment questions, minimising total surveying time to just a few minutes. RQ 3 exploits the fact that Teacher Tapp knows the name and location of the school currently taught at, and so asks the teacher questions about willingness to work at other named local schools.

## Evaluation design

### Design RQ1

In the main phase of our study, we present teachers with a fully randomised paired conjoint experiment based on a fictional scenario where they are asked to imagine that they need to seek employment at a new school. The opening invitation will state: *“You have moved house so need to find a new school. Take a look at these job adverts and see which ones you prefer.”*

Teachers are then presented with two short lists of four or five job attributes, accompanied by the following statement: *“If two schools that were identical in every other way advertised the following jobs, which would you prefer?”*

With careful design to minimise choice complexity, paired conjoint experiments have been shown to be valid and reliable methods for eliciting policy-relevant preferences (Brutger et al, 2023; Zhirkov, 2022). A fully randomised paired conjoint experiment has three main advantages over other methods of learning teachers' job preferences. First, by introducing random variation in school and job attributes, we break the link between school characteristics and unobservable factors that also correlate with teachers' job selection decisions. This is an important advantage over observational studies since many 'soft' job characteristics such as working conditions are correlated with the socio-demographic profile of students in the school. Second, by asking teachers to choose a school following a hypothetical house move, they are unable to bring to mind any unstated features of the schools in the pair. Third, we can simultaneously estimate treatment effects for each of the policy choices examined in the study, yielding strong efficiency advantages over quasi-experimental studies that focus on a single treatment at a time (Hainmueller et al., 2014).

Typically, in Discrete Choice Experiments (DCEs), attribute levels are not purely randomised but structured to enhance statistical efficiency and meaningful data collection, often relying on model-dependent approaches (e.g. the McFadden model). This study departs from the traditional DCE design found in the labour economics literature, opting instead for a fully randomised approach where attribute levels are assigned without constraints across profiles. This methodology simplifies the statistical estimation of stated preferences using OLS regression, as the design automatically ensures variation in the data, eliminating the need for predefined designs structures.

The following section outlines the advantages of fully randomised paired conjoint experimental designs, particularly in contrast to typical conjoint and DCEs. Overall, with sufficiently large sample sizes, as in this study, full randomisation ensures orthogonality that each attribute level is equally likely to appear, reducing concerns about biases introduced by design choices in DCEs. Moreover, it allows for the testing of interactions between attributes without predefining constraints (Hainmueller, Hopkins, and Yamamoto 2014; Egami and Imai 2019; Leeper, Hobolt, and Tilley 2020; Bansak et al. 2021b; Abramson, Koçak, and Magazinnik 2022). Hence, in identifying the multidimensional trade-offs of choosing a workplace or hiring strategies, we believe that this fully randomised paired conjoint experimental design should be adopted rather than other factorial designs. This well-established design integrates typical conjoint analysis with the potential outcomes framework for causal inference. Its new causal estimand (Average Marginal Component Effect - AMCE) shows that the causal effect of each level of attributes can be nonparametrically identified and easily estimated from conjoint data using a fully randomised design.

The use of AMCEs and marginal means as a primary estimation framework aligns with our focus on causal interpretation and preference descriptives. AMCEs provide clear and interpretable estimates of how changes in attribute levels impact choice probabilities,

averaged across the population. This approach complements our study's goal of deriving actionable, policy-relevant insights in a straightforward and replicable manner. AMCEs isolate the causal effect of changing one attribute level while holding all other attributes constant, providing precise estimates within the same profile. In contrast, marginal means describe the overall average probability of selection for each attribute level, offering a descriptive measure of preferences without relying on a reference category.

**Table 1: Comparisons of Different Factorial Designs**

| Feature                             | Traditional Conjoint               | DCE                               | Fully Randomised Paired Conjoint                      |
|-------------------------------------|------------------------------------|-----------------------------------|---|
| <b>Design Methodology</b>           | Fractional factorial or orthogonal | D-efficient or Bayesian designs   | Full randomisation                                    |
| <b>Attribute-Level Combinations</b> | Constrained                        | Constrained                       | Unconstrained (within plausible ranges)               |
| <b>Flexibility</b>                  | Limited by design constraints      | Limited by design constraints     | High  |
| <b>Bias Control</b>                 | Structured designs ensure balance  | Structured designs ensure balance | Balance achieved via randomisation over large samples |
| <b>Statistical Model</b>            | Part-worth utilities (OLS, logit)  | Utility-based models (logit)      | AMCEs using linear regression                         |

Altogether, this design is empirically tractable under several credible assumptions. First, we assume there is no carryover effect, meaning that responses in one round of assessing profiles do not influence responses in subsequent rounds. To validate this, we will empirically test whether observations from each respondent can be treated as independent. Second, we will control whether the randomisation ensures that our within-subjects design is balanced in the study. In other words, we will show whether each level of conjoint attributes is uniformly distributed across covariates. Third, we will show whether the positionality of the profiles (e.g. up or down) affects our estimates. Finally, we will display the frequency distribution of each attribute level across all profiles to confirm that the randomisation process worked as intended.

Traditional approaches like mixed logit models are well-suited for estimating heterogeneity in preferences, but often rely on pre-constructed designs that might miss complex, unanticipated interaction effects. In contrast, full randomisation ensures that the full space of possible attribute combinations is explored, which enhances the model's ability to identify and test interactions naturally without prior assumptions. Additionally, while the mixed logit model directly addresses the Independence of Irrelevant Alternatives (IIA) assumption<sup>1</sup>, it is important to note that in our fully randomised paired forced-choice design, the issue of IIA is inherently minimised. This is because each fully randomised choice set includes only two alternatives (profiles), ensuring that there are no "irrelevant" third options to bias the choice probabilities.

---

<sup>1</sup> This assumption means that adding or removing an unrelated option in a choice set should not affect the relative probabilities of choosing between the other options. For example, introducing a third, irrelevant school should not change the preference between two existing schools.

While paired conjoint experiments offer robust insights, they also have limitations. Designing choice scenarios to be both realistic and manageable can be challenging, as overly complex choices may overwhelm participants, affecting response quality. Additionally, hypothetical scenarios may not capture all the nuances of real-world decision-making, potentially leading to hypothetical bias. Lastly, paired conjoint experiments rely on certain assumptions, such as the IIA assumption, which may not always hold in practice, limiting the extent to which findings can be generalised to actual job selection behaviour. All three of these disadvantages are particularly relevant when studying teachers evaluating job advertisements, as the real-life process is likely to be slower and more deliberative compared to the way choices are made within the experiment.

The list of job attributes to be tested as part of the main paired conjoint experiment are developed through a literature review, focus groups and pilot questions. For each type of attribute, we plan to include up to three levels, including one statement that aims to roughly reflect “normal practice” to aid interpretation of the value of the attribute. Our current hypothesised list of attributes includes:

- **Salary:** Same as present; 5% increase; 15% increase
- **Flexible working requests,** up to a guarantee to meet all employee requests
- **Childcare benefits,** up to onsite subsidised provision for ages 6 months to 13 years
- **Healthcare and wellbeing benefits,** up to private health insurance
- **Planning, Preparation and Administration (PPA) time,** with the highest level being one day a fortnight at home
- **Professional development,** up to a paid one term research sabbatical after a period in post
- **Workload arrangements,** up to a 35-hour week guarantee through credible workload reducing commitments (e.g. centralised planning, behaviour, marking etc)
- **Behaviour management,** as established by the preliminary stage
- **School environment** such as curriculum and culture, as established by the preliminary stage

In addition to the conjoint profiles, we will always finish with an open-ended question inviting teachers to explain the choices they made. Their free text responses are likely to be helpful in assessing the reasons for any unexpected choices made.

Within the experiment, salary options are presented as "same salary as present," "5% increase," and "15% increase," rather than using nominal amounts such as "£5,000 increase". This approach aligns with the methodology used in other teacher job choice DCEs and is based on the challenge of creating a nominal scale that is meaningful across a wide salary range—from early-career teachers earning around £27,000 to senior teachers earning £76,000 or more. Using percentages allows the salary options to remain proportionate to each respondent's current earnings, ensuring the choices are relevant and reflective of individual contexts.

We acknowledge that this approach may introduce some cognitive burden, as respondents need to calculate percentage changes to their salary during the task. This could potentially cause confusion, for example, about whether the percentage applies to gross or take-home pay. However, we believe the benefits of using percentages outweigh the drawbacks. Presenting nominal amounts would risk introducing irrelevant or unrealistic options for respondents at different salary levels. For instance, a £5,000 increase may be highly meaningful for a new teacher but less so for a senior leader, while a £15,000 increase could seem implausible for early-career teachers.

To mitigate cognitive burden, the instructions explicitly clarify that the percentages apply to respondents' current gross salary and include only simple calculations that are straightforward to process. Our approach ensures that salary choices remain equitable and contextually relevant across the diverse range of teaching roles, thereby enhancing the realism and validity of the decision-making scenarios. While this trade-off may slightly increase cognitive effort, it is consistent with best practices in similar DCEs and is essential for maintaining the realism and fairness of the choice tasks across respondents with vastly different salary baselines.

This experiment will employ a forced conjoint design, requiring participants to choose between presented options without the possibility of opting out. A key advantage of this approach is that it compels respondents to make trade-offs, potentially revealing relative preferences with greater clarity. By removing the option to "opt out" or select "none of the above," this design can encourage participants to carefully consider and prioritise attributes, providing insights into the comparative importance of features. It has been found to encourage respondents to more carefully consider the information about the profiles and increase their engagement with the task, while matching with the behavioural benchmarks remarkably well (Hainmueller, Hopkins, and Yamamoto 2014).

However, not having the option to opt out may lead to upwardly biased estimates, as such scenarios may not fully reflect real-world decision-making contexts where individuals often have the option to decline all presented alternatives or maintain the status quo. It is worth noting that neither school presented within RQ1 represents the typical status quo of remaining at their current school, which is why the hypothetical scenario of a house move is presented where forced choice without the status quo is the normal decision-making framework. Nevertheless, lack of the option to decline the choices could lead to hypothetical bias, as respondents might make choices in the experiment that they would not make in reality (Carson et al., 1994; Hensher, 2010). Additionally, the absence of an opt-out option can introduce challenges related to incentive compatibility and may inadvertently prompt less thoughtful or stochastic responses in some cases. The fully randomised conjoint design can also reduce the social desirability bias for the AMCEs of the sensitive attributes (Horiuchi, Y., Markovich, Z., & Yamamoto 2022).

Despite these drawbacks, forced choice is particularly suited for studies aiming to rank or prioritise features rather than assess absolute acceptability. It may also reduce certain biases, such as an over-reliance on status quo or default options, and can yield richer data on preference rankings. To address potential concerns, we have designed the choice tasks to closely mimic realistic trade-off scenarios based around a house move, aiming to minimise hypothetical bias and maximise the validity of the findings.

One important question for us to consider is how many attributes to include with conjoint profiles before survey satisficing leads to unacceptable declines in response quality. There exists a trade-off between masking and satisficing that is likely to be context specific, thus limiting insights from existing studies (Bansak et al., 2019). On the one hand, including too few attributes will make it difficult to interpret the substantive meaning of AMCEs, since respondents might associate an attribute with another that is omitted from the design. Such a perceived association between an attribute included in the design and another omitted attribute muddies the interpretation of the AMCE of the former as it may represent the effects of both attributes (i.e. masking). On the other hand, including too many attributes increases the risk of survey satisficing, where respondents, faced with cognitive overload and/or response fatigue, may answer less thoughtfully to complete the survey more quickly.

Whilst Bansak et al. (2019; 2021) has made an important contribution to this question, showing only modest increases in survey satisficing when respondents face large numbers of

attributes, this research was carried out on computers rather than smaller mobile phone screens. Given our research will be administered via mobile phones, we will need to account for potential differences in respondent experience due to the smaller screen size. Therefore, during the pilot phase of the project, we plan to evaluate likely response reliability alongside the app development team, testing 4 versus 5 attributes on the mobile phone screen to assess potential masking and satisficing issues.

### *Design RQ2*

This part of our study aims to quantify school leaders' willingness to allocate budgetary resources for certain job attributes. These job attributes will roughly match those used in RQ1, focusing particularly on those that teachers value most in salary terms. One difficulty is that some of these job attributes can be offered to an individual teacher in the school, without providing it for all other teachers, whereas others realistically cannot.

As much as possible we will focus on the job attributes that can be provided to an individual teacher and will present headteachers with a series of paired conjoint questions in which they must select between two potential teaching candidates. These candidates will differ in terms of salary demands and various attributes drawn from the list outlined in RQ1.

We believe that this will be the first paired conjoint experiment to examine this issue, in England and internationally, and will give unique insights into the implicit financial costs (in salary terms) of providing job attributes that may not have straightforward budgetary implications. Given that many of these job attributes affect the nature and flexibility of educational provision, it is important to understand whether the preferences of teachers are palatable options for their employers. We believe that the findings from our conjoint analyses from the experiments for RQ1 and RQ2 may present interesting mismatches between what teachers want and value, and what headteachers are willing to pay for (discussed in the cost evaluation below).

### *Design RQ3*

RQ3 aims to understanding their willingness to transition to jobs at named local schools, in relation to the relative levels of deprivation at these schools. The Teacher Tapp survey app holds information on the unique reference number of their current school, alongside information about their current teaching position. Our goal is to contribute new quantitative evidence to the policy question: *"What salary uplift should higher FSM schools offer to ensure they have as large a pool of job applicants as lower FSM schools?"*

A survey will prompt teachers to choose between their current role and a comparable position at a geographically proximate school, along with various potential salary adjustments. These "real-world" scenarios have high ecological validity but pose interpretative challenges — such as the hesitance towards job changes, the intrinsic value of familiar versus unfamiliar roles, and the impact of unmeasured characteristics — yet they provide a unique avenue to infer teachers' preferences for different school types indirectly.

Our methodology is designed to sidestep social desirability bias as far as possible by determining how the necessary salary adjustment to motivate a teacher to move from their current school to another nearby school correlates with the free school meals (FSM) disparity between the two schools. All school attributes are bundled, so FSM does not represent the demographic of the pupils but rather all the school attributes that are correlated with the demographic of the pupils.

It is important to note that, as teachers are not randomly allocated to their current schools, the overall willingness to relocate estimates cannot be straightforwardly interpreted due to the confounding effect of initial placement. For example, a teacher who finds managing behaviour

difficult is likely to seek out schools where this is less of an issue (and research suggests these are more affluent schools - Allen et al., 2024). Therefore, they are likely to demand a higher salary to transfer to a more disadvantaged school, relative to those who already choose to teach in such schools. This means that the most insightful estimates will be average willingness to transfer between schools in particular demographic brackets, given the starting demographic bracket of the teacher. These estimates will be policy-relevant, given that the aim is to encourage those who have so far sought jobs in more affluent schools to consider positions in more disadvantaged settings.

The research design varies the salary of the named local school, but not other attributes, therefore we can never make causal inferences about the marginal value of other attributes of the school and must consider them bundled goods.

The survey approach for RQ3 is as follows:

1. We will verify their current school's name and invite them to name the nearest school of the same phase to their current one.
2. We will ask them how well they know this school they have named, excluding them from further questions if they say "Not at all"
3. We will ask if they would consider applying for a position at this school, assuming it offers the same role and salary as their current job.
4. We will pose the same question up to four additional times, adjusting the proposed salary by increments/decrements of 5%, 15%, 25%, and 35%.
5. Participants will be asked to evaluate whether they perceive any of the following aspects to be better (and then worse) than at their current school: behaviour, leadership, working hours, culture, collegiality, curriculum, and travel time.
6. An open-ended question will seek their rationale behind the responses they give.

This dataset will yield a number of key insights. Firstly, we will be able to gauge the overall 'stickiness' within the labour market, as indicated by the salary increase required to encourage a teacher to move to another local school with comparable characteristics. Secondly, we will estimate the relationship between the required salary increase and the disparity in free school FSM rate between the two schools. A preference for schools with certain FSM profiles should not be interpreted as a causal preference for particular student demographics but rather as recognition of the set of school attributes often associated with such demographics, including workforce composition, school culture and leadership, Ofsted ratings, curriculum offerings, and governance structures. Thirdly, we will investigate variations in the willingness to switch schools based on teachers' personal circumstances, employing similar demographic categorisations to those used in earlier parts of the project. The variation by subject might prove especially interesting, as some subjects provide a wide range of employment options, while others are more competitive.

Table 22 below summarises the key characteristics of the research project overall.

**Table 2: Summary of the Design Characteristics of the Study**

|   |                            |
|---|----------------------------|
| <b>Design</b>                                     | Paired Conjoint Experiment |
| <b>Unit of analysis<br/>(school, pupils)</b>      | Teachers                   |
| <b>Number of Units to be included in analysis</b> | Around 8,000               |

|                         |   |   |
|-------------------------|---|---|
| <b>Outcome measures</b> | <b>variable</b>                           | Job preference for schools, given all attributes (from which we estimate their preference for job characteristics at the expense of salary increase through interacting attributes with the pay attribute – willingness to apply) |
|                         | measure<br>(instrument, scale, source)    | % change in salary (for inferred estimates)   |
|                         | <b>variable(s)</b>                        | (1) Preference for teachers by school leaders, given all attributes (willingness to pay)<br>(2) Willingness of teachers to take a job at another local school, given a particular salary  |
|                         | measure(s)<br>(instrument, scale, source) | Both in terms of % salary change (for inferred estimates)   |

### Scoping phase

Developing a set of possible job attributes for teachers to choose between, and a set of teacher characteristics for headteachers to choose between, is complex. The greatest implementation challenge we face is establishing appropriate wording for ‘soft’ job attributes that relate to the school culture, behaviour, workload, professional development, colleagues and so on. Chong et al. (2024) shows that these types of statements were frequently occurring in advertisements and were also felt by leaders to be important. However, we think teachers’ responses to these ‘soft’ statements will reflect both how much they value the attribute and whether they believe the statement is credible. The attribute descriptions are developed over the first few months of the project, as follows:

1. Desk research to review existing discrete choice and paired conjoint experiments as well as survey literature to establish which job attributes appear to be valued by teachers, particularly in the UK and US.
2. Four sets of focus groups with teachers in both primary and secondary phases to review and give feedback on job attributes – both individual statements and statements embedded within lists that are similar to those in the paired conjoint experiments, as described below.
3. Pilot survey questions to clarify how teachers value some job attributes, where several specifications of the attribute are possible
4. A technical check of the paired conjoint experiments on ineligible teachers such as those working outside England or with incomplete demographic data returns. We do not pilot on eligible teachers because this would reduce the target sample for the main stage of the experiment. Note that whilst these teachers are ineligible by virtue of their location, they have very similar job characteristics to the main sample. This technical check includes an open-ended question for teachers to express where language is unclear.

We consider the scoping phase, particularly the focus groups, to be a key strength of our study design. Pretesting is a critical step in enhancing the “validity, reliability, and relevance of the survey” while minimising “sources of bias, burden, and error.” It also supports hypothesis generation and aids interpretation (Campoamor et al., 2024, p. 111). To this end, we will conduct four focus groups before piloting our survey to refine the research instruments and

support subsequent analysis. As Campoamor et al. (2024) highlight, despite pretesting being “an essential stage” in designing DCEs, “fewer than one-fifth of DCE studies report including pretesting in their development” (2024, p. 109, p. 116).

In recent DCEs exploring teacher preferences, pretesting has often been overlooked. For instance, Burge et al. (2021) report modifying their study design based on piloting but do not mention pretesting. Similarly, Horng (2009) describes conducting consultations and piloting for her conjoint study of teacher preferences but does not detail a pretesting phase.

To maximise efficiency within a limited timeframe, we choose to use online focus groups instead of the more typical individual interviews often used for pretesting. This approach enables us to gather input from a larger group of teachers while maintaining a high level of detail and relevance. Teachers were recruited for the four focus groups via an advert on Teacher Tapp, targeting classroom teachers in mainstream state schools. This ensured the discussions focused on aspects most relevant to the majority of teachers’ work. To improve clarity, the groups were split by phase, with separate groups for primary and secondary teachers. The target size of each focus group is between 4 and 8 teachers.

The focus group scripts are designed to address four key domains:

- **Content:** Assessing the relevance and comprehensiveness of the attributes, ensuring they capture the key factors teachers consider when selecting roles.
- **Comprehension:** Evaluating whether teachers can envision the proposed scenarios and decision contexts.
- **Presentation:** Identifying the most effective forms of presentation and wording for the attributes.
- **Elicitation:** Exploring trade-offs and the heuristics underlying decision-making.

A semi-structured protocol will guide the focus groups, incorporating visual stimuli (e.g., attribute lists or sample job adverts) and prompts addressing areas such as credibility, interpretation, and understanding. When issues arise, participants will be encouraged to suggest alternatives, which will be reviewed by the research team.

The first two focus groups (one for primary and one for secondary teachers) will cognitively pretest the proposed attributes and statements, focusing particularly on content and comprehension. We will prioritise testing ‘soft attributes,’ such as school culture and ethos, over ‘hard attributes,’ like salary, since the latter lends itself more effectively to the on-app piloting phase. The second two focus groups will review revised statements and explore the design of the choice task that will be presented in the app.

By integrating these focus groups into the pretesting phase, we aim to ensure that the survey is both robust and reflective of teachers’ real-world decision-making processes.

### **Participants**

All three parts of the research will take place on the Teacher Tapp survey app, a mobile platform that surveys over 10,000 teachers in both state and private sectors in England. Once teacher download the app and verify their credentials, they receive daily notifications at 3:30 pm to participate in surveys. These surveys use various questions format, including multiple-choice, single-response, and open-ended text response questions.

As the country’s largest teacher survey platform, Teacher Tapp is exceptionally well-positioned to conduct paired conjoint experiments, especially given their demanding sample size requirements. Although Teacher Tapp does not constitute a randomly selected sample, Jerrim (2023) notes that neither is any other survey of teachers since randomisation is

invalidated by low response rates, which typically do not exceed 10%. Furthermore, given the widespread use of smartphones among teachers (Laricchia, 2024), we consider the app-based method to be no less inclusive than other data collection approaches.

Teacher Tapp employs several strategies to ensure representativeness by:

- Limiting the sample to those providing a valid school name and job information, relevant to our study's focus.
- Applying post-stratification weights based on observable characteristics such as educational phase, funding type, gender, age, and job position.
- Regularly cross-checking the panel against key findings from the OECD TALIS 2018 survey to verify its representativeness.
- Using Teacher Tapp offers several advantages: The platform's extensive daily reach allows for a substantial sample collection on any given day, and the flexibility for further surveying of the same respondents to clarify unexpected findings as the project progresses.
- The large sample size also affords precise estimations for specific demographic groups, including details about their current job roles, years of experience, etc.
- Teacher Tapp continually holds and updates teacher and school demographic information, thus avoiding the priming effect of pre-treatment questions and reducing response fatigue caused by asking demographic questions within the paired conjoint experiments.
- As a longitudinal platform, there exists the possibility of returning to the same sample to test within-teacher changes in job attribute preferences in the future.

Currently, Teacher Tapp has over daily 10,000 respondents. However, once the sample is restricted to those who work as a teacher in schools in England and who agree to answer these extra questions, it is prudent to assume a response rate of around 8,000. Respondents who live in other parts of the UK or who are otherwise ineligible will be used for technical pilot tests of the survey.

The three parts of the research will use slightly different samples, as summarised in the table below.

**Table 3: Sample for survey and primary analysis of outcomes**

|   | <b>Targeted for survey</b>   | <b>Included in primary analysis of outcomes</b>   |
|---|--|---|
| <b>RQ1</b><br><b>Teacher paired conjoint experiment</b>     | Teachers working in a state or private school in England<br>Excluding headteachers | Teachers working in a state or private school in England<br>With full demographic information available<br>Excluding all senior leaders |
| <b>RQ2</b><br><b>Headteacher paired conjoint experiment</b> | Headteachers working in a state or private school in England                       | Headteachers working in a state or private school in England  |
| <b>RQ3</b><br><b>Teaching in disadvantaged schools</b>      | All teachers working in a state or private school in England                       | Teachers and headteachers working in  |

|  |   |   |
|--|---|---|
|  | With a school URN that is consistent with their demographic information | a state or private school in England<br>With a school URN that is consistent with their demographic information |
|--|---|---|

### Sample size calculations

Statistical power in fully randomised conjoint experiments is a function of the number of repeated trials performed by each respondent, the number of levels of an attribute, and the size of the measured effect in the population. Stefanelli and Lukac (2020) show that effect sizes in conjoint experiments are often small despite large sample sizes. This tends to be the case because researchers either overload their design with a high number of experimental conditions or have limited trials.

The app “Cjpowr” ([https://m-freitag.github.io/cjpowR\\_shiny/](https://m-freitag.github.io/cjpowR_shiny/)) for conjoint experiments has enabled us to calculate power and determine the minimum required sample sizes in forced-choice conjoint experiments (Schuessler and Freitag, 2020). Recognising that testing for interaction effects between attributes requires substantially larger sample sizes than testing for average effects, we focus our evaluation of statistical power on a conjoint design involving interacted attributes, each with up to three levels, and two profiles over five trials.

For RQ1, our focus is on treatment probabilities where the main interest lies in detecting causal interaction effects. In doing so, we employ uniform randomisation, which ensures that each of the four possible interactions combinations is presented equally across tasks, each with a probability of 0.25. Based on this design, our calculations with two profiles, each including maximum three levels for each attribute over three trials, suggest that a sample size of at least  $N_{\text{individuals}} = 4,680$  and  $N_{\text{observations}} = 28,079$  is necessary to achieve the conventional power threshold of 0.8 or higher with a 0.5 alpha. This sample size ensures a satisfactory level of power for detecting interaction effects in the experiment. This information is summarised in the table below.

Given that we are likely to achieve sample responses well over twice this size, we are confident that we can identify even quite small differences in values placed on different job attributes. More importantly, we will be able to identify effects in all the important demographic subgroups, even where the group is proportionately small.

For RQ2 we base our evaluation of statistical power on the AMCE (Average Marginal Component Effect), rather than the Average Component Interaction Effects (ACIE). This estimand allows us to understand how the change in each attribute, relative to a baseline, would affect the probability that headteachers would prefer a certain candidate. Given that we will have a maximum number of three levels in each attribute with two potential teacher profiles over six trials, our calculations suggest that a sample size of at least  $N_{\text{individuals}} = 391$  and  $N_{\text{observations}} = 4,698$  is necessary to achieve the conventional power threshold.

**Table 4: Sample size calculations for RQ1 and RQ2**

| Input criteria                           | RQ1 Input value | RQ2 Input value |
|--|-----------------|-----------------|
| Power                                    | 0.8             | 0.8             |
| AMCE (Average marginal component effect) | 0.05            | 0.05            |

|  |       |      |
|--|-------|------|
| Levels of attributes (max)   | 3     | 3    |
| Alpha (P rejecting null when true)   | 0.05  | 0.05 |
| Number of tasks  | 3     | 6    |
| Minimum sample size for estimating Average Marginal Component Effect (AMCE)              | 783   | 391  |
| Minimum sample size for estimating Average Marginal Component Interaction Effect (AMCIE) | 4,680 | N/R  |

### Outcome measures and other data

#### Primary outcome

The primary outcome variable is based on the question in the conjoint task *“If two schools that were identical in every other way had the following job adverts, which would you prefer?”* This will let us measure the job preferences based on the randomly assigned attributes.

From this primary outcome, we infer their preference for job characteristics at the expense of salary increase through interacting attributes with the salary attribute.

#### Secondary outcomes

The first secondary outcome (for RQ2) variable is based on the question in the conjoint task *“If two teachers who were identical in every other way had the following profiles, which one of them would you employ?”* This will let us measure the candidate preferences based on the randomly assigned attributes.

The second secondary outcome (for RQ3) variable is the average increase in salary required to induce a job move to a local school for each additional percentage FSM.

#### Other data

We will use the Teacher Tapp demographic information to explore how estimates vary across sub-groups but have no specific prior hypotheses about what we might find. These demographics include:

- Teacher personal characteristics: age, gender, age of youngest child at home
- Teacher job characteristics: years of experience, seniority of post, subject taught
- School characteristics: phase, funding, Ofsted, free school meals, region

### Primary analysis

In the analysis for RQ1 and RQ2, to assess the impact of the attributes on school/teacher preference, we will regress the school/teacher choice on all attribute levels (through survey::svyglm() function in R - Lumley et al., 2024) with clustered standard errors to account for the fact that school/teacher choices are nested within respondents (Hainmueller et al., 2014).

Specifically, we will regress the binary choice outcome,  $Y$ , for teacher,  $i$ , who is viewing profile,  $j$ , for task,  $k$ , on a vector of indicator variables for each job attribute using the following model:

$$Y_{ijk} = \beta X_{jk} + \epsilon_{ijk}$$

The choice outcome variable,  $Y_{ijk}$ , is a binary variable equal to 1 if the teacher rated the profile as preferred, and 0 otherwise.  $X$  is a vector of indicator variables for each attribute. Because the attribute values are randomly assigned to profiles, and profiles are randomly assigned to teachers, the vector of coefficients  $\beta$  capture the independent effect of each attribute on the probability a teacher preferred that school, average over the randomisation distribution of all other school attributes.

The coefficients will be interpreted as AMCEs, which indicate the average change in the probability of choosing a school/teacher when one of its attributes is switched from the reference category to a specific attribute level.

AMCEs are empirically tractable and practical to estimate using simple statistical methods like differences-in-means or OLS. By focusing on changes within the same profile for a specific attribute, rather than contrasting attributes across profiles, AMCEs isolate the causal effect of each attribute level.

Additionally, to identify how the change in salary level attribute interacts with the change in other attributes in explaining school choice, we will estimate a two-way causal interaction effect, called the ACIE, which captures the additional causal impact of an attribute-level (treatment) combination that exceeds the sum of the individual effects of each attribute-level when considered separately. This ACIE shows effect sizes for the different groups.

AMCEs, however, are relative statements and depend on the selected reference category, which can make cross-group comparisons challenging. Hence, to be able to identify general preferences without reference categories, we will also calculate marginal means, which describes the level of favourability towards schools/teachers with a particular attribute-level when ignoring all other attributes (Leeper et al., 2020).

In other words, the AMCE and ACIE are contingent on the selected reference category, which is often arbitrary and can skew interpretations when comparing across subgroups. For more reliable insights, it is also essential to incorporate marginal means, as they provide estimates that are independent of the reference category, avoiding potential biases. In order to calculate marginal means, we will use `mm()` function from `cregg` package in R. Marginal means are conditional averages and not regression coefficients. They are descriptive estimands/quantities of interest and they do not rely on any reference category or baseline level.

As for RQ3, the analysis of the survey will proceed as follows:

1. We will use the survey responses to indirectly infer a reservation price for each teacher. This is the minimum salary change at which they become willing to consider applying for a job at the named local school. We will observe how this reservation price varies by teacher demographic characteristics and estimate overall 'stickiness' or unwillingness to move in general.
2. We will test our key hypothesis that reservation price depends on the target local school FSM, given the existing school FSM. Our main analysis will be a visual representation of this relationship, but we will also fit regression models to explore how these relationships vary by type of teacher and initial school conditions.
3. We will explore how the relationship between reservation price and target school FSM varies across teacher demographics. Whilst we think it is plausible that there will be demographic variation in preferences for schools, we do not believe the existing literature provides us with good hypotheses as to what they might be. Therefore, this analysis is purely exploratory.

4. We plan to look at whether the reservation prices signal any clear preferences for particular type of governance arrangements, e.g. Catholic schools or schools in large MATs. However, we need to be careful about causal interpretation here since these characteristics are correlated with school population demographics.
5. We will use the final survey questions to explore whether knowledge of a school affects the relationship between reservation price and FSM.

### Inference

AMCE and marginal means estimands for RQ1 and RQ2 will be interpreted as follows:

**Table 5: Interpretation of AMCEs and marginal means**

|                       | RQ1 – Teacher’s willingness to apply   | RQ2 – Headteacher’s willingness to pay  |
|-----------------------|--|---|
| <b>AMCE</b>           | Changing a job’s salary implication from the same level to 10% raise increases the probability of selection by X percentage points.          | Offering a specific job attribute (e.g., flexible working hours) is valued by headteachers as equivalent to a salary adjustment of X percentage points.   |
| <b>Marginal means</b> | Jobs with 10% salary increase are supported by X% of respondents or 10% salary increases are X percentage points more likely to be selected. | Attributes such as flexible working hours or reduced workload are supported by X% of headteachers, reflecting the proportion who value these attributes highly, regardless of other attributes. |

Additionally, following the guidance of Hainmueller et al. (2014), the results will be presented using coefficient plots. In these plots, point estimates will be represented by dots, with 95% confidence intervals displayed to illustrate the level of uncertainty surrounding each estimate. Standard errors will be clustered within respondents to account for within-subject dependencies.

To facilitate our discussion of costs and benefits, we will convert the AMCEs to salary equivalents by dividing the coefficient on each attribute by the coefficient on the salary increase attribute. These are commonly referred to in the literature as willingness to pay estimates.

### Robustness checks

We will evaluate the robustness of our main results in multiple ways. First, to ensure that inferences about the causal effects of our conjoint experimental design are credible, we will test whether assumptions hold:

- **Independence of responses:** In a within-subject design, respondents make multiple choices, which we assume are independent from each other. This means that each choice is made without influence from previous choices, avoiding carryover effects. We can test for this by analysing response patterns in their first task versus subsequent tasks.
- **Balanced design:** We will control whether the within-subjects design is balanced. To ensure that each level of attributes is uniformly distributed, we will compare a covariate across different levels of attributes. As further robustness check, we will control for participant’s demographic characteristics in the regression models.
- **Linearity of preferences for salary increases.** The willingness-to-pay estimates rely on the assumption that teachers’ preferences for salary increases are linear from 0 to 15%. We can use the survey data to explore the plausibility of this assumption.

## Further analyses

### Subgroup analyses

The Teacher Tapp panel is sufficiently large to identify differences in job attribute preferences by demographic groups. The existing literature suggests that we are likely to find some differences, for example by phase and by age of teacher (Burge et al., 2021; Lovison and Hyunjung Mo, 2024).

To get the correct demographic based differences in preferences for schools/teachers, we will calculate conditional marginal means. As recommended by Leeper et al. (2020), in order to descriptively characterise differences in preference level between subgroups, we will directly estimate the subgroup difference using conditional marginal means and differences between conditional marginal means rather than relying on the difference-in-AMCEs.

## Implementation and process evaluation (IPE)

If the research study yields a set of clear findings about what types of job attributes teachers value, and about what types of job attributes headteachers are willing to consider, we believe there is value in supporting the profession through detailed, concrete advice about how to enact change in schools. Thus, we propose hosting a series of online focus groups with school leaders and school business managers to explore the barriers that might hinder the adoption of new working practices.

### Purpose and value of final focus groups

The primary objective of these focus groups is to bridge the gap between research findings and practical implementation. Assuming the research reveals clear insights into teacher-valued job attributes and headteacher perspectives on accommodating those attributes, these focus groups will provide a platform to explore the practical challenges of enacting change within schools. Specifically, the focus groups will centre around understanding the potential barriers that might prevent schools from adopting the new working practices suggested by the research findings. This information will be essential in developing tailored guidance and recommendations for schools.

### Strategic sampling

The focus groups will employ a purposive sampling method to ensure a balanced and insightful discussion. This means that participants will be carefully chosen based on specific criteria, aiming to include a mix of school leaders and business managers from:

- **Schools at the forefront:** Institutions already implementing innovative working practices, offering insights into their successes, challenges, and lessons learned.
- **Schools facing barriers:** Particularly those in disadvantaged settings that might face more significant hurdles in adopting new approaches. These participants can shed light on their specific concerns, resource constraints, and the support needed to overcome these challenges.

### Structure and format

The project proposes around four to six phase-specific online focus groups, each with no more than 12 attendees, to facilitate focused discussions. These online focus groups are likely to be run using Google Meet and recorded.

### Recruitment and incentives

Leveraging the extensive reach of the Teacher Tapp platform, recruitment for these focus groups will be conducted through targeted advertisements within the app. This approach allows for directly reaching the desired audience of school leaders and business managers.

There is no need to provide incentives for participation in focus groups like this, since school leaders are usually keen to access and discuss relevant research findings. Since the event will also provide an opportunity for networking between school leaders, this acts as a further incentive to participate.

### **Analysis and dissemination of focus groups**

The primary goal of the focus groups is to facilitate the dissemination of emerging ideas to school leaders and to gather feedback that informs the project's recommendations. The team will carry out some thematic analysis of the focus groups to help identify common barriers and facilitators to implementation. These insights will be crucial for refining practical guidance, strengthening policy recommendations, and ensuring that the project outcomes are grounded in the realities of what school leadership can influence.

### **Cost evaluation**

One policy goal of this project is to compare the types of job attributes that teachers value, in salary terms, against the implicit cost of providing such job attributes, in salary terms. We will bring together the findings from RQ1 and RQ2 to provide an overall assessment of which job attributes school leaders are most likely to find it advantageous to provide.

RQ1 will yield a teacher average maximum willingness to forego a salary (in £ terms) for every job attribute. Estimation of the costs of providing each job attribute are more complex, for two reasons. First, some job attributes can be given to individual teachers and not to others, whilst other job attributes must be given to all teachers across a school, even if a teacher does not value it. Second, some job attributes have a direct financial cost whereas others have an implicit cost to the school in terms of flexibility of organisation. These implicit costs will be estimated in RQ2 as average minimum implicit salary costs (in £ terms) for each job attribute we evaluate. We will collate these implicit costs with direct financial costs to compile the costs of providing job attributes, either for an individual teacher or for all staff, as relevant.

### **Ethics**

Ethical approval for this research project was granted in July 2024 by the ethical review board at UCL. The review process was comprehensive and encompassed the following key aspects:

- **Focus group materials:** The ethics board thoroughly reviewed the draft Participant Information Sheets and Consent Forms that will be used for the focus groups to be conducted with both teachers and school leaders. These documents detail the purpose of the focus groups, the voluntary nature of participation, data handling procedures, and the rights of participants to withdraw. The review ensured these documents meet ethical standards for informed consent and participant protection.
- **Survey questions and existing protocols:** Recognising that Teacher Tapp is an established research platform with existing ethical protocols, the review process included an evaluation of how the proposed survey questions align with [Teacher Tapp's standard terms and conditions](#). This step ensured that the specific questions used in this study adhered to the platform's established framework for ethical data collection.
- **Teacher Tapp privacy statement:** The ethics board also reviewed [Teacher Tapp's privacy statement](#), specifically focusing on its provisions for safeguarding the privacy of panellists. This review ensured that the platform's broader data protection measures adequately protect the privacy of teachers participating in the research.

- **Data transfer and security:** The process of transferring data from Teacher Tapp's secure servers to UCL's secure servers for analysis was also scrutinised. The review focused on ensuring that robust data transfer protocols and security measures are in place to maintain data confidentiality and integrity throughout the research process. This included an examination of the anonymisation procedures to safeguard against the identification of individual teachers or schools.

The ethics review process confirmed that the proposed research adheres to the ethical guidelines set by UCL and relevant professional bodies. It also addressed specific ethical considerations arising from using the Teacher Tapp platform and conducting research with teachers and school leaders.

This study was registered on the Open Science Framework on October 28<sup>th</sup>, 2024. For greater detail, follow this link: <https://osf.io/958zx>. The registration will be updated upon the completion of the study.

## Data protection statement

This project is committed to ensuring that all data handling and processing activities comply with the General Data Protection Regulation (GDPR) and the Data Protection Act 2018. All data collected will be managed to uphold confidentiality, integrity, and security, ensuring the protection of individual privacy throughout the project lifecycle.

**In this research, no personally identifiable information (PII) will be collected.** Survey responses are separated from personal data such as users' email addresses and passwords using Google's Firebase System. All data for this project will be stored in a GDPR-compliant manner on Education Intelligence Ltd servers, which are designed to securely handle and protect data.

Teacher Tapp employs anonymised data at all times, with app login and password data held on an entirely separate system. The school Unique Reference Number (URN) is the most sensitive piece of data held by Education Intelligence Ltd as part of their routine work as a survey company. However, for this study, the school URN will **not** be passed to the UCL research team, ensuring that the data remains non-identifiable.

### Legal basis for processing personal data:

The legal basis for processing personal data in this project is GDPR Article 6(1)(f), i.e. *"Processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject..."*

### Legal basis for processing special category data:

No special category data, as defined under GDPR Article 9, will be processed in this project. Therefore, no legal basis under Article 9 is required.

### Rationale for legal bases selected:

The legitimate interests pursued by Education Intelligence Ltd and UCL in this project are:

- **Advancement of Educational Research:** Conducting research to evaluate educational interventions contributes to academic knowledge and evidence-based practices.
- **Improvement of Educational Practices:** Findings from the study aim to enhance teaching methodologies and student outcomes.
- **Societal Benefits:** The research has the potential to positively impact society by improving education systems.

Processing data is necessary to achieve these objectives. Given that no PII or special category data will be processed and that all data will be anonymised, the rights and freedoms of data subjects are unlikely to be adversely affected.

### **Approach to demonstrating GDPR compliance**

To ensure compliance with GDPR, the project will implement the following measures:

- **Transparency:** Participants will receive clear information about data processing activities through information sheets and privacy notices, outlining the purpose of data collection, handling procedures, and their rights.
- **Data Minimisation:** Only data essential for the research objectives will be collected, avoiding unnecessary processing of personal data.
- **Anonymisation:** Data will be anonymised to prevent the identification of individuals. Unique identifiers will be removed, and data will be aggregated where appropriate.
- **Security Measures:** Data will be stored securely on Education Intelligence Ltd servers, compliant with GDPR standards, with survey responses separated from personal data using Google's Firebase System and with limited transfer to anonymised data to UCL's secure servers.
- **Access Control:** Access to data will be restricted to authorised personnel directly involved in the project. The UCL team will not receive any PII or school URNs.
- **Data Subject Rights:** Even though PII is not collected, the project respects data subjects' rights under GDPR, including rights to access, rectify, or erase their data.
- **Personal Data:** Any personal data, if collected, will be deleted as soon as it is no longer necessary for the project.
- **Anonymised Data:** Will be retained indefinitely for replication and open science purposes, aligning with best practices in scientific research.
- **Third-Party Compliance:** At project completion, anonymised data will be stored in the Open Science Framework (OSF) repository, which complies with GDPR and has robust data protection measures.

### **Purposes for data processing**

Data will be processed solely for:

- **Conducting Educational Research:** Evaluating the effectiveness of educational interventions through analysis of anonymised survey responses and focus group data.
- **Advancing Knowledge:** Contributing findings to the academic community and informing future educational policies and practices.

### **Parties with access to data and reasons**

**Education Intelligence Ltd:** As the data controller, they collect and securely store the data necessary for the project.

**UCL Research Team:** Receives only anonymised data for analysis, acting under the instructions of the data controller.

**Open Science Framework (OSF):** Hosts the anonymised dataset post-project for public access, supporting transparency and reproducibility in research.

### **Protection of individual data subjects' rights**

While no PII is processed, the project ensures:

- **Right to Access:** Participants can request access to any data held about them.
- **Right to Rectification and Erasure:** Participants can request corrections or deletion of their data if applicable.

- **Right to Object:** Participants can object to data processing activities.
- **Contact Information:** Participants will be provided with contact details to exercise their rights.

This data protection plan ensures compliance with GDPR and the Data Protection Act 2018 while facilitating valuable educational research. By prioritising data minimisation, anonymisation, and secure data handling practices, the project safeguards individual privacy and upholds ethical research standards.

## Personnel

Leading the team is **Professor Becky Allen**, Chief Analyst at Teacher Tapp. Becky is an expert in surveying teachers, large scale data collection and analysis, and education policy. She has published widely on teacher recruitment and retention, and she served as an expert advisor to the RAND-led discrete choice experiment on teacher job preferences (Burge et al., 2021). She is responsible for all aspects of the project.

**John Jerrim**, Professor of Education and Social Statistics at the UCL Institute of Education, brings his expertise in leading substantial projects across the field of education policy research, including the Nuffield-funded study on teacher mental health and wellbeing. He will oversee the statistical analysis of the principal paired conjoint experiments in this study.

**Dr. Sam Sims**, Associate Professor at the Centre for Education Policy and Equalising Opportunities at UCL Institute of Education, is renowned for his research on teachers. Currently engaged in designing discrete choice experiments for the ESRC Centre for Early Mathematics Learning at Loughborough, his expertise will be instrumental in guiding the design of the paired conjoint experiments for the study.

**Dr. Burak Sonmez**, a Lecturer in Quantitative Social Science at UCL Institute of Education, specialises in the design and execution of discrete choice experiments. He will be the principal analyst for the paired conjoint experiments.

**Loic Menzies**, a researcher and policy specialist with a distinguished background as the founder and former CEO of the think-tank 'The Centre for Education and Youth' and now a visiting fellow at Sheffield Institute of Education, brings a wealth of knowledge on teachers. His role will focus on leading the qualitative analysis of job attribute preferences and focus groups.

In addition to the co-investigators listed, the following members of the Teacher Tapp team will contribute the project:

- **Iain Ford** is Head of Data and will lead the set-up and allocation of the sample, the calculation of sample weights, fuzzy matching of school names to URNs, and the monitoring of response rates.
- **Karen Wespeiser** is Chief Operating Officer and will assist with focus groups and overall project management as needed.
- **Darren Royle** is a software engineer who will make technical changes needed to facilitate the paired conjoint experiments on the Teacher Tapp survey app.
- **Grainne Hallahan** is Head of the Teacher Tapp Community and she will ensure the teachers who are participating in the project are informed about its purpose, what to expect and the findings.

## Risks

| Risk   | Type                   | Phase | Description  | Likelihood | Impact | Mitigation Strategy   |
|--|------------------------|-------|--|------------|--------|---|
| <b>Timing conflicts with academic calendar</b>   | Data collection        | All   | We need RQ3 data collection to take place whilst teachers are very familiar with their school (i.e. not September or early October). Focus groups are far easier to run in school holidays.  | Medium     | Low    | Schedule alternative data collection points if the preferred timetable slips.   |
| <b>Challenges in analysing qualitative data from open-ended questions and focus groups</b> | Analysis               | All   | Interpreting qualitative responses may be subjective and resource intensive.   | Medium     | Medium | Develop a clear plan for exactly what we do / don't plan to learn from qualitative analysis. Use robust qualitative analysis software and involve multiple researchers in coding and analysing open-ended responses.  |
| <b>Miscommunication of project aims to teachers</b>  | Stakeholder engagement | All   | Teachers may not fully understand the purpose and importance of the study, leading to disengagement.   | Low        | Low    | Clearly communicate project aims through regular updates, informative blogs, and direct engagement with Teacher Tapp users.   |
| <b>Technical issues with Teacher Tapp app</b>  | Operational            | All   | Technical glitches during data collection could disrupt survey completion and data integrity.  | Low        | Medium | Conduct extensive testing of the app before launch and have technical support on standby during data collection periods. Leave time in timetable for a second attempt at data collection if there is a failure on day one.  |
| <b>Timeline slippage due to unforeseen delays</b>  | Operational            | All   | Delays in project milestones could impact the overall schedule and reporting deadlines.  | Medium     | Medium | Implement a detailed project management plan with regular progress reviews and contingency buffers for critical tasks.  |
| <b>Ill health of key personnel in team</b>   | Operational            | All   | Ill health or similar causes key personnel members to be absent for part of project. The risk is highest for Becky and Burak since John Jerrim and Sam Sims have some overlapping expertise and there is greater availability of qualitative researcher skills for Loic. | Low        | High   | Ensure everything is fully documented so that others can stay up to speed. In her absence, Iain Ford takes over project management to ensure success of data collection. May need to delay analysis stage. Burak - creates analysis plan early on in the project so that Sam can implement it in his absence. Review whether there are paired conjoint experiment experts to provide consultancy. |

|   |                        |     |  |        |        |  |
|---|------------------------|-----|--|--------|--------|--|
| <b>Participant shares personal data during a focus group or open question</b> | Ethical and compliance | All | We do not plan to collect personal data during this project but it could be created if a participant shares their name or personal, private information during a focus group or in an open-ended question. | Medium | Low    | Write script for sign-up and start of focus group to discourage participants from using full name or disclosing personal information. Post-transcription, manually check the scripts to remove personal data before it is analysed. In open text responses, scan responses and remove personal data before analysis. |
| <b>Inadequate data protection and privacy measures</b>                        | Ethical and compliance | All | Breaches of participant confidentiality could lead to ethical violations and loss of trust.  | Low    | High   | Adhere strictly to GDPR guidelines, use secure data storage solutions, and obtain necessary ethical approvals before commencing the study.   |
| <b>Inadequate specification of attributes in RQ1</b>                          | Methodology            | RQ1 | Failure to correctly specify job attributes in RQ1 could lead to inaccurate results. Risk of underestimating value of attribute is higher than overestimating it.  | Medium | High   | Conduct thorough literature review, pilot testing, and focus groups to refine the attribute specifications before the main study.  |
| <b>Complexity in designing paired conjoint experiments for mobile screens</b> | Methodology            | RQ1 | Difficulty in presenting complex job advertisements on small mobile screens might lead to response errors.   | Medium | Medium | Collaborate with the app development team to optimise the visualisation of paired conjoint experiment questions and test designs for usability.  |
| <b>Satisficing and masking in responses</b>                                   | Analysis               | RQ1 | Respondents might over- or under-interpret the attribute text we provide leading to biased results.  | Medium | Medium | Limit the number of attributes displayed and conduct preliminary testing within the pilot stage (i.e. soft attributes) to balance comprehensiveness and respondent fatigue.  |
| <b>Insufficient power in detecting interaction effects</b>                    | Analysis               | RQ1 | Small sample sizes for certain subgroups might reduce the power of detecting interaction effects in paired conjoint experiments.   | Low    | Medium | Ensure all demographic groups are specified such that they are sufficiently large (i.e. never less than a fifth of the overall sample). Ensure large overall sample size and conduct power calculations in advance to confirm adequate subgroup representation.  |
| <b>Low participation into teacher focus groups</b>                            | Stakeholder engagement | RQ1 | Poor attendance could limit insights into the RQ1 job attributes (the topic of the focus groups isn't exactly compelling).   | Medium | Low    | Use Teacher Tapp team to help work out how we 'sell' the topic of these focus groups to teachers. Over-recruit for no-shows. Explain incentives in sign-up.  |
| <b>Low response rate for headteacher surveys</b>                              | Data collection        | RQ2 | Insufficient headteacher responses may affect the reliability of RQ2 results.  | Medium | High   | If we fail to collect sufficient sample of day one, then open questions to headteachers who haven't yet answered on additional days. (We could consider opening to deputies but would rather not if possible).   |
| <b>Low participation in headteacher focus groups</b>                          | Stakeholder engagement | RQ2 | Poor attendance in focus groups could limit insights into barriers to adopting new working practices. The pool of headteachers eligible for these focus groups is much smaller.                            | Medium | Medium | 'Sell' the online focus groups as workshops and get well-known innovators to come along to talk about flexible working initiatives.  |

|   |                 |     |   |        |        |  |
|---|-----------------|-----|---|--------|--------|--|
| <b>Teacher Tapp holds inaccurate current school information</b> | Data collection | RQ3 | Whilst we encourage panellists to regularly update their school information, some do not and this accuracy is essential for RQ3.                  | Medium | Medium | (1) Messages to encourage them to check this information is correct in weeks before questions; (2) discard from sample all teachers who name a "nearby school" that clearly isn't near the named school we hold. |
| <b>Social desirability bias in teacher responses</b>            | Analysis        | RQ3 | Teachers may underreport reluctance to work in disadvantaged schools, skewing results, even though our design specifically aims to overcome this. | Low    | Medium | John, Becky and Sam to spend more time discussing how the size of this bias could be better estimated in non-experimental settings.  |

## Timeline

| Date                         | Activity  | Staff responsible/ leading                    |
|------------------------------|---|---|
| July, August 2024            | Project set-up, planning, ethics                            | Becky Allen, Karen Wespeiser, Burak Sonmez    |
| July, August, September 2024 | Teacher Tapp app adjustments for paired conjoint experiment | Iain Ford, Darren Royle                       |
| August, September 2024       | Policy literature review                                    | Becky Allen, Sam Sims, John Jerrim            |
| September 2024               | Technical literature review                                 | Burak Sonmez                                  |
| August 2024                  | RQ1 - focus groups - design and data collection             | Loic Menzies, Karen Wespeiser                 |
| September 2024               | RQ1 - focus groups - analysis                               | Loic Menzies                                  |
| September 2024               | RQ1 - Pilot paired conjoint experiment                      | Becky Allen, Iain Ford, Sam Sims, John Jerrim |
| October 2024                 | RQ1 - data collection                                       | Becky Allen, Iain Ford, Burak Sonmez          |
| November 2024                | RQ1 - analysis quant  | Burak Sonmez                                  |
| November 2024                | RQ1 - analysis free text qual data                          | Loic Menzies                                  |
| November 2024                | RQ3 - data collection                                       | Becky Allen, Iain Ford                        |
| December 2024                | RQ3 - analysis quant  | Becky Allen, Iain Ford, Sam Sims, John Jerrim |
| December 2024                | RQ3 - analysis free text qual data                          | Loic Menzies                                  |
| January 2025                 | RQ2 - focus groups - design, data collection, analysis      | Loic Menzies                                  |
| February 2025                | RQ2 - data collection                                       | Becky Allen, Iain Ford, Burak Sonmez          |
| March 2025                   | RQ2 - analysis quant  | Burak Sonmez                                  |
| March 2025                   | RQ2 - analysis free text qual                               | Loic Menzies                                  |
| March 2025                   | RQ2 - headteacher policy focus groups                       | Loic Menzies                                  |
| March 2025                   | Draft final report  | Becky Allen and co-Is                         |
| April, May 2025              | Revisions to final report                                   | Becky Allen and co-Is                         |

## References

- Abramson, S. F., Koçak, K., & Magazinnik, A. (2022). *What do we learn about voter preferences from conjoint experiments?* *American Journal of Political Science*, 66(4), 1008-1020.
- Allen, R., Ford, I., Hallahan, G., and Hannay, T. (2024). *Teacher Recruitment and Retention in 2024: An exploration of recruitment challenges in disadvantaged schools*, Teacher Tapp Research Report, <https://teachertapp.co.uk/publications/recruitment-and-retention-in-2024/>.
- Allen, R., Ford, I., and Hannay, T. (2023). *Teacher Recruitment and Retention in 2023: Teacher views on coping with shortages, job attachment and flexible working*, Teacher Tapp Research Report, [https://teachertapp.co.uk/app/uploads/2023/06/Teacher-Recruitment-and-Retention-in-2023-TT\\_SchoolDash-Final.pdf](https://teachertapp.co.uk/app/uploads/2023/06/Teacher-Recruitment-and-Retention-in-2023-TT_SchoolDash-Final.pdf).
- Allen, R. and Sims, S. (2018). *Do pupils from low-income families get low-quality teachers? Indirect evidence from English schools*. *Oxford Review of Education*, 44 (4), 441-458.
- Allen, R. and McInerney, L. (2019). *The Recruitment Gap: Attracting teachers to schools serving disadvantaged communities*, Sutton Trust Report, <https://www.suttontrust.com/wp-content/uploads/2019/12/The-Recruitment-Gap.pdf>
- Bansak, K., Hainmueller, J., Hopkins, D. J., and Yamamoto, T. (2019). *Conjoint Survey Experiments in Druckman, James N., and Donald P. Green (eds)*. *Cambridge Handbook of Advances in Experimental Political Science*, New York: Cambridge University Press.
- Bansak K., Hainmueller J., Hopkins D.J., and Yamamoto T. (2021). *Beyond the breaking point? Survey satisficing in conjoint experiments*. *Political Science Research and Methods*, 9(1):53-71. doi:10.1017/psrm.2019.13
- Bansak, K., Hainmueller, J., Hopkins, D. J., Yamamoto, T., Druckman, J. N., & Green, D. P. (2021b). *Conjoint survey experiments*. *Advances in experimental political science*, 19, 19-41
- Brutger, R., Kertzer, J.D., Renshon, J., Tingley, D., and Weiss, C.M. (2023). *Abstraction and Detail in Experimental Design*. *American Journal of Political Science*, 67: 979-995. <https://doi.org/10.1111/ajps.12710>
- Burge, P., Lu, H., and Phillips, W. D. (2021). *Understanding Teaching Retention: Using a Discrete Choice Experiment to Measure Teacher Retention in England*, RAND Research Report, [https://www.rand.org/pubs/research\\_reports/RRA181-1.html](https://www.rand.org/pubs/research_reports/RRA181-1.html).
- Campoamor, N. B., Guerrini, C. J., Brooks, W. B., Bridges, J. F., and Crossnohere, N. L. (2024). *Pretesting discrete-choice experiments: a guide for researchers*. *The Patient-Patient-Centered Outcomes Research*, 17(2), 109-120.
- Carson, R.T., Louviere, J.J., Anderson, D.A., Arabe, P., Bunch, D.S., Hensher, D.A., Johnson, R.M., Kuhfeld, W.F., Steinberg, D., Swait, J. and Timmermans, H. (1994). *Experimental analysis of choice*. *Marketing letters*, 5, pp.351-367.
- Chong, S.W., Oxley, E., Negrea, V., Bond, M., Liu Q., and Kong, M.S. (2024). *Teacher recruitment and retention in schools in socio-economically disadvantaged areas in England - review of practice*, Education Endowment Foundation Research Report.
- Egami, N., & Imai, K. (2019). *Causal interaction in factorial experiments: Application to conjoint analysis*. *Journal of the American Statistical Association*.

Hainmueller, J., Hopkins, D.J., and Yamamoto, T. (2014). *Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments*. *Political Analysis*, 22(1): 1–30.

Hainmueller, J., Hangartner, D., & Yamamoto, T. (2015). *Validating vignette and conjoint survey experiments against real-world behavior*. *Proceedings of the National Academy of Sciences*, 112(8), 2395-2400.

Hensher, D. A. (2010). Hypothetical bias, choice experiments and willingness to pay. *Transportation Research Part B: Methodological*, 44(6), 735-752.

Horiuchi, Y., Markovich, Z., & Yamamoto, T. (2022). *Does conjoint analysis mitigate social desirability bias?* *Political Analysis*, 30(4), 535-549

Horng, E. L. (2009). *Teacher tradeoffs: Disentangling teachers' preferences for working conditions and student demographics*. *American Educational Research Journal*, 46(3), 690-717.

Jerrim, J. (2023). *Teacher surveys: The pros and cons of random probability surveys versus teacher panels*, *Review of Education*, 11 e3428, 1-19. <https://doi.org/10.1002/rev3.3428>

Jerrim, J. and Sims, S. (2019). *The Teaching and Learning International Survey (TALIS) 2018*. Department for Education.

Laricchia, F. (2024). *Smartphone usage in the United Kingdom (UK) 2012-2023, by age*, <https://www.statista.com/statistics/300402/smartphone-usage-in-the-uk-by-age/>.

Leeper T.J., Hobolt S.B., and Tilley, J. (2020). *Measuring subgroup preferences in conjoint experiments*. *Political Analysis* 28(2): 207–221.

Lovison, V. S., and Hyunjung Mo, C. (2024). *Investing in the Teacher Workforce: Experimental Evidence on Teachers' Preferences*. *American Educational Research Journal*, 61(1), 108-144. <https://doi.org/10.3102/00028312231208956>.

Lumley, T., Gao, P. and Schneider, B. (2024). *Package 'survey' for Analysis of Complex Survey Samples*, CRAN Repository documentation, <https://cran.r-project.org/web/packages/survey/survey.pdf>.

Perryman, J. and Calvert, G. (2020). *What Motivates People to Teach, and Why Do They Leave? Accountability, Performativity and Teacher Retention*, *British Journal of Educational Studies*, 68 (1), pp. 3–23. <https://doi.org/10.1080/00071005.2019.1589417>.

Schuessler, J., and Freitag, M. (2020). Power analysis for conjoint experiments. *SocArXiv*, 2020. 42.

See, B. H., Morris, R., Gorard, S., Kokotsaki, D., and Abdi, S. (2020) *Teacher Recruitment and Retention: A Critical Review of International Evidence of Most Promising Interventions*, *Education Sciences*, 10 (10), p. 262. <https://doi.org/10.3390/educsci10100262>.

Sharp, C., Smith, R., Worth, J. and Van den Brande, J. (2019). *Part-Time Teaching and Flexible Working in Secondary Schools*. National Foundation for Educational Research: <https://www.nfer.ac.uk/publications/part-time-teaching-and-flexible-working-in-secondary-schools/>.

Stefanelli, A., and Lukac, M. (2020). *Subjects, trials, and levels: Statistical power in conjoint experiments*, *SocArXiv*, 2020. 42.

Worth, J., Lynch, S., Hillary, J., Rennie, C. and Andrade, J. (2018). *Teacher Workforce Dynamics in England*, National Foundation for Educational Research, <https://www.nfer.ac.uk/publications/teacher-workforce-dynamics-in-england/>

Worth, J. (2023). *Short Supply: Addressing the Post-Pandemic Teacher Supply Challenge in England*, National Foundation for Educational Research: <https://www.nfer.ac.uk/publications/short-supply-addressing-the-post-pandemic-teacher-supply-challenge-in-england/>

Zhirkov K. (2022). Estimating and Using Individual Marginal Component Effects from Conjoint Experiments, *Political Analysis*, 30(2), pp.236-249. doi:10.1017/pan.2021.4.