Factors affecting teacher job satisfaction and retention: A causal inference machine learning approach using data from TALIS 2018

Nathan McJames[a,b,*], Andrew Parnell[a,b], Ann O'Shea[b]

[a]Hamilton Institute, Maynooth University, Maynooth, Co. Kildare, Ireland

[b]Department of Mathematics and Statistics, Maynooth University, Maynooth, Co. Kildare, Ireland

[*]Corresponding author. E-mail address: nathan.mcjames.2016@mumail.ie

January 18, 2022

Abstract

Teacher shortages and attrition are problems of international concern. Studies investigating this problem often identify important correlates of these two outcomes, but fail to produce easily implementable recommendations. Accordingly, in this study we have adopted a causal inference machine learning approach to identify practical interventions for improving job satisfaction/retention. We apply our methodology to TALIS 2018 data from England. Our results indicate that participation in continual professional development and induction activities have the most positive effect on both of these outcomes. Out-of-field teaching and part-time contracts are shown to have a negative effect on retention and job satisfaction respectively.

Keywords: Teacher Job Satisfaction, Teacher Retention, Causal Inference, Machine Learning, TALIS

Highlights:

- Specific and implementable measures for improving teacher job satisfaction and retention are identified.
- A causal inference machine learning approach is used for estimating treatment effects.
- Continual professional development and induction schemes are identified as having the most positive impact.
- Out-of-field teaching is shown to increase the rate of attrition of qualified teachers.
- Part-time contracts are shown to reduce levels of job satisfaction.

Funding: This work has emanated from research conducted with the financial support of Science Foundation Ireland under grant number 18/CRT/6049. In addition Andrew Parnell's work was supported by: a Science Foundation Ireland Career Development Award (17/CDA/4695) and SFI Research Centre award (12/RC/2289_P2).

1 Introduction

Teacher supply and demand is an important challenge faced by many countries around the world. Countries currently facing a teacher shortage include England (Hilton, 2017), Ireland (O'Doherty and Harford, 2018), the United States (Wiggan et al., 2021), and many others. Often, teacher shortages are even more pronounced in Science, Technology, Engineering and Mathematics (STEM) subjects (Han and Hur, 2021).

Low numbers of new entrants to the teaching profession and high numbers of qualified teachers leaving their posts contribute to increased teacher shortages. Furthermore, higher levels of teacher turnover have been shown to negatively affect student learning and also incur large economic costs (Levy et al., 2012, Sorensen and Ladd, 2020). Research has shown that job satisfaction is one of the key predictors of a teacher's intention to remain in the profession (Madigan and Kim, 2021). This means that it is of the utmost importance to identify factors that can improve job satisfaction in order to encourage higher retention rates of qualified teachers and to attract new entrants to start their career.

In this paper, we use a causal inference machine learning approach to investigate the effect of a number of factors on teacher job satisfaction. We also investigate the effect of the same factors on a teacher's desire to move to an alternative school. This approach enables us to identify specific and implementable measures that may be introduced to improve job satisfaction or retention rates of qualified teachers. We believe that this is a key advantage of our approach because it allows school principals or other policy makers to determine specific steps that may be taken as part of a school strategy for improving job satisfaction or retention. Our approach uses Bayesian Additive Regression Trees (BART), a cutting-edge modelling tool which enables us to detect non-linear relationships and interactions which would not normally be found in a standard linear model. Additionally, this method allows us to control for a much larger number of background (confounding) variables than would normally be possible when using a linear model. We demonstrate how this approach can be used to identify subgroups of teachers who are most (or least) likely to benefit from the positive effects of a given treatment.

Our study uses data from the third cycle of the Teaching and Learning International Survey (TALIS) which took place in 2018 (OECD, 2019a). TALIS is the world's largest survey of teachers and principals and has taken place every five years since 2008. A fourth cycle is due to take place in 2024. Participating teachers and principals are asked to complete questionnaires on a wide variety of topics such as: personal background; current teaching duties; their perception of the school climate; and job satisfaction. TALIS 2018 is the largest of the surveys to-date with 48 countries participating, and includes data on approximately

260,000 teachers from 15,000 primary, and lower and upper secondary schools. For the purpose of this study, however, we will limit our investigation to the data from England. This subset of the entire dataset contains a representative sample of 4385 primary and lower secondary school teachers. We believe this data set provides us with a good sample size and teachers from both education levels.

With these data, and using a causal inference machine learning approach, we attempt to answer the following research questions: (1) What are the specific and implementable factors that have the most positive (or negative) impact on teacher job satisfaction?, and (2) What are the specific and implementable factors that have the most positive (or negative) impact on retention?. The factors we consider include: participation in induction schemes; high levels of participation in continual professional development; team teaching; observing other teachers; mentorship schemes; teaching in a public school; class size; out-of-field teaching; and having a part-time contract. Our decision to include these factors in our investigation has been informed by previous studies which show they have a strong association with both teacher job satisfaction and retention. We now discuss these findings in more detail in a literature review, focusing on key aspects relevant to our research.

1.1 Induction and Mentoring Programmes

Induction is a broad term used to describe different activities or supports put in place for teachers to assist them in adapting to the ethos or practices of a new school (Allen, 2005, OECD, 2018). Induction programmes are often aimed at newly-qualified teachers, but can also be offered to experienced teachers who have recently begun teaching in a new school.

Mentoring describes the arrangement whereby a newly-qualified teacher is assigned a more experienced member of staff at their school, who will advise and assist them as they begin their career (Allen, 2005, OECD, 2018). Roles of a mentor can vary, as can frequencies of meetings between a mentor and their mentee. An experienced teacher who is new to a school may also be assigned a mentor, who can act as a guide and offer support as they adapt to their new school.

Research shows that participation in induction and mentoring schemes is positively associated with teacher job satisfaction and can also lead to an increase in retention rates. For example, in a non-causal regression analysis of newly-qualified teachers, Renbarger and Davis (2019) found a strong link between the presence of a mentor and considerably higher levels of job satisfaction.

Allen (2005), however, in a report which reviews 91 studies on teacher recruitment and retention, finds only limited evidence that participation in induction and mentoring schemes leads to higher retention rates of qualified teachers. Instead, they emphasise that the effects of induction and mentoring are likely to be context specific. The benefits, they argue, are likely to be greater for certain subgroups of teachers such as those often faced with challenging student behaviours.

Other studies have shown that newly-qualified teachers with a mentor are less likely to leave the profession than teachers who do not have a mentor (Smith and Ingersoll, 2004). These findings are important, as new teachers are known to face many challenges in the classroom after they qualify, and are at a higher risk of attrition than more experienced teachers (Guarino et al., 2006).

The effect of *being* a mentor on job satisfaction and retention has been the focus of relatively little research in comparison to the effect of *having* a mentor. Despite this, there are still studies which show that mentoring arrangements can be mutually beneficial to both the mentor and the mentee. Lunsford et al. (2018) for example, show that teachers with either a mentor or a mentee are on average more satisfied than teachers who do not.

It is important to note that all of the above findings are based on observational data, and therefore the positive correlation between induction and job satisfaction cannot be claimed to be causal in nature. In an important contribution, the above findings were corroborated by a longitudinal study by Gray and Taie (2015), which tracked a sample of newly-qualified teachers over the first 5 years of their career. At each follow-up visit, teachers who were assigned a mentor during their first year in the classroom were more likely to still be teaching than those who did not receive this extra support, thus showing a temporal association between induction and retention.

1.2 Continual Professional Development

Continual professional development (CPD) for teachers can refer to a wide range of activities designed to assist teachers as they build upon and improve their professional skills (OECD, 2018). Higher levels of participation in CPD have frequently been linked to improved teacher job satisfaction in a number of different studies (Wang et al., 2020, Yoon and Kim, 2022). Sims (2017) using data from TALIS 2013 was able to show that this relationship holds in both the national and international context. In two separate analyses, they demonstrated that there is a non-causal positive correlation between CPD and job satisfaction, firstly using data for England only, then again for a combined dataset of more than 50,000 teachers from 38 different countries.

CPD has been shown to be related to higher levels of teacher retention. Coldwell (2017) surveyed more than 500 teachers following their completion of a professional development course. Results showed that teachers who were more engaged with the CPD course were more likely to respond that the course had a positive effect on their intention to remain teaching. This link was less strong for teachers who only engaged moderately or weakly with the course. Similar results were reported by Allen and Sims (2017), who showed that participation in science subject-specific CPD courses was associated with a two percentage point reduction in department turnover rates two years on. This finding is especially important given that STEM subject teachers are known to be at higher risk of attrition (Han and Hur, 2021).

Despite these benefits, one challenge often faced by teachers is that there may be barriers to their attendance at different CPD activities due to factors such as timetabling issues, cost of travelling to CPD events, or a lack of suitable events being organised (Zhang et al., 2020). It is unsurprising then, that the presence of barriers to attending quality CPD activities has also been linked to lower levels of job satisfaction (Renbarger and Davis, 2019).

1.3 Teacher Cooperation

Higher levels of cooperation between teachers and staff within schools has been identified as a strong correlate of job satisfaction in previous research (Lopes and Oliveira, 2020). Examples of factors contributing to high levels of cooperation within a school could include team teaching, observation of other teachers' classes, or sharing of teaching materials and resources (OECD, 2018). Sims (2017), using international data from TALIS 2013, found teacher cooperation to be the most significant predictor of job satisfaction when accounting for other working conditions and teacher characteristics.

Similar results were reported by Toropova et al. (2021) using Swedish data from TIMSS 2015. They found cooperation to be one of the strongest predictors of job satisfaction. They also noted a strong interaction with gender, suggesting higher levels of cooperation are more important for male teachers.

Teacher cooperation was identified as a key predictor of teacher turnover by Nguyen (2021). They found that teachers reporting higher levels of cooperation were less likely to want to leave their current school. They also found, however, that higher levels of cooperation were not associated with lower probabilities of teachers wanting to leave the teaching profession entirely.

Findings from qualitative studies tend to agree with the above. Skaalvik and Skaalvik (2015), using the results of 30 interviews with Norwegian teachers, reported that teamwork and cooperation was often cited as a source of enjoyment by teachers. Although often seen as positive, teamwork can also have negative effects. Indeed, as was noted that in the same study, 16 teachers also referred to teamwork as a source of stress resulting from disagreements with colleagues when they could not choose who they collaborated with.

1.4 Characteristics of the Student Body

Challenging behaviour and discipline problems are some of the most widely discussed student-related factors associated with teacher job satisfaction in the literature (e.g. Wang et al., 2020, Kengatharan, 2020). Malinen and Savolainen (2016), in a longitudinal study of 642 teachers from Finland, found that teachers who initially reported better student behaviour in their classes had higher levels of job satisfaction at the end of the school year.

Teacher self-efficacy has been found to moderate the relationship between student behaviour and teacher job satisfaction. Toropova et al. (2021) found that teachers who reported higher levels of self-efficacy were less negatively affected by bad student behaviour. This suggests that teachers with more confidence in their ability cope better with the challenges posed by student behaviour and find these challenges more manageable.

Schools with a catchment area that includes students of a predominantly lower socioeconomic status have been linked to lower levels of teacher job satisfaction and retention (Ingersoll, 2001, Borman and Dowling, 2008). Other studies, however, have conflicting results that show socioeconomic status is either not a strong predictor of job satisfaction when controlling for other factors, or that teachers in schools with students from a predominantly lower socioeconomic status have higher rates of retention (Hughes, 2012). In fact, Matsuoka (2015) showed, using a structural equation modelling approach, that socioeconomic status is a predictor of student discipline which in turn predicts job satisfaction, thus showing that there is a link between socioeconomic status and job satisfaction albeit not a direct one.

1.5 Other Factors

Class size is a variable which is commonly used in studies investigating student achievement. Larger class sizes and larger student teacher ratios have often been shown not to have a large effect on student achievement (Woessmann and West, 2006, Li and Konstantopoulos, 2017), and a clear connection between class size and job satisfaction has not been established. Perrachione et al. (2008), in an interview study of

200 teachers, found that class size was one of the top 3 reasons reported by teachers giving explanations for their current job satisfaction. Reeves et al. (2017), however, found that class size was not a major driver of American or Japanese teacher job satisfaction when controlling for other working conditions.

A second factor which is less commonly examined in relation to teacher job satisfaction and retention, and is normally confined to studies focused on student achievement, is the practice of out-of-field teaching. Out-of-field teaching as a practice has been linked to lower student achievement in a number of studies (Dee and Cohodes, 2008, Hill and Dalton, 2013), but the literature available on the effects that out-of-field teaching has on job satisfaction or retention is quite limited. Olmos (2010) and Provasnik and Dorfman (2005) found that out-of-field teachers were more prone to attrition, though other studies have not found as substantial an effect (e.g. Shen, 1997).

Finally, one additional factor which has not been the subject of much research in relation to teacher job satisfaction or retention is contract-type. Our search for studies relating factors associated with the terms of a teacher's employment and their job satisfaction or intention to continue teaching returned few results. Furthermore, those studies which we did find were not focused primarily on terms of employment, but instead used it as one of a variety of control variables. Gil-Flores (2017), in an investigation of the effect of personal characteristics on teacher job satisfaction, included contract-type (permanent vs. fixed-term) as a variable, and identified teachers with permanent contracts as being less satisfied on average. Conversely, Capone and Petrillo (2020) found teachers with permanent contracts to have higher levels of job satisfaction and well-being. Other studies which have investigated the effects of part-time or full-time contracts have not found clear differences in job satisfaction (e.g. Ferguson et al., 2012).

2 Data and Pre-Processing

This study uses data from England from TALIS 2018 (OECD, 2019a) which provides us with a representative sample of 4385 primary and lower secondary school teachers. Each observation includes more than 30 scales describing various teacher and school characteristics such as self-efficacy, participation in CPD, and perceived cooperation among staff. The individual survey responses upon which these scales are based are also provided, as well as personal and background details for each of the teachers such as gender; school level; qualification; and years' experience. A full list of all variables used can be found in Appendix C. A description of our handling of missing data in these variables can be found at the end of this section.

The main variables of interest in this study are teacher job satisfaction and intention/desire to move to another school. Teacher job satisfaction (Variable Code: T3JOBSA) in the TALIS data is actually a composite of two subscales (Variable Codes: T3JSENV and T3JSPRO), both consisting of four items which gauge a teacher's overall contentment and happiness with their current working environment and profession. All eight questions share a common stem which reads "We would like to know how you generally feel about your job. How strongly do you agree or disagree with the following statements?". An example item for measuring satisfaction with the working environment is "I enjoy working at this school", and an example item for satisfaction with the profession is "The advantages of being a teacher clearly outweigh the disadvantages". Possible responses to these items lie on a 4 point Likert scale, with options ranging from strongly disagree (1) to strongly agree (4). The responses to these eight questions have been combined by the survey organisers using confirmatory factor analysis to give a measure of overall levels of job satisfaction. The resulting job satisfaction scale (after combining primary and lower secondary school teachers) has a mean of 12.42, and a standard deviation of 2.28. Also note the negative skew observed in Figures 1, 8, and 9.

To measure the impact of different factors on retention or attrition we have used the following item from the survey (Variable Code: TT3G53C): "I would like to change to another school if that were possible". In our analysis we have labelled teachers who responded with agree or strongly agree as "movers", and all other teachers who responded with disagree or strongly disagree as "stayers". This is the closest link to retention we could find in the survey, and there is no analogous question that directly asks if a teacher is considering leaving the profession. Therefore it is important to bear in mind that our measure of retention/attrition is based on movement between different schools, not exiting from teaching entirely, and we are measuring it indirectly as opposed to directly.

As can be seen from Figure 1, there is a very clear relationship between our two outcome variables. Teachers who do want to move school score lower on the job satisfaction scale on average. It should be noted, however, that there are some very satisfied teachers who do want to move school, and also some very unsatisfied teachers who do not.

To ensure a representative sample is collected during the data collection stages of TALIS, a stratified two-stage probability sampling design is used within each country. Each teacher within the TALIS dataset is therefore assigned a number of weights for the purposes of rigorously calculating population parameters of interest and their associated standard errors. The sampling weights resulting from this design were

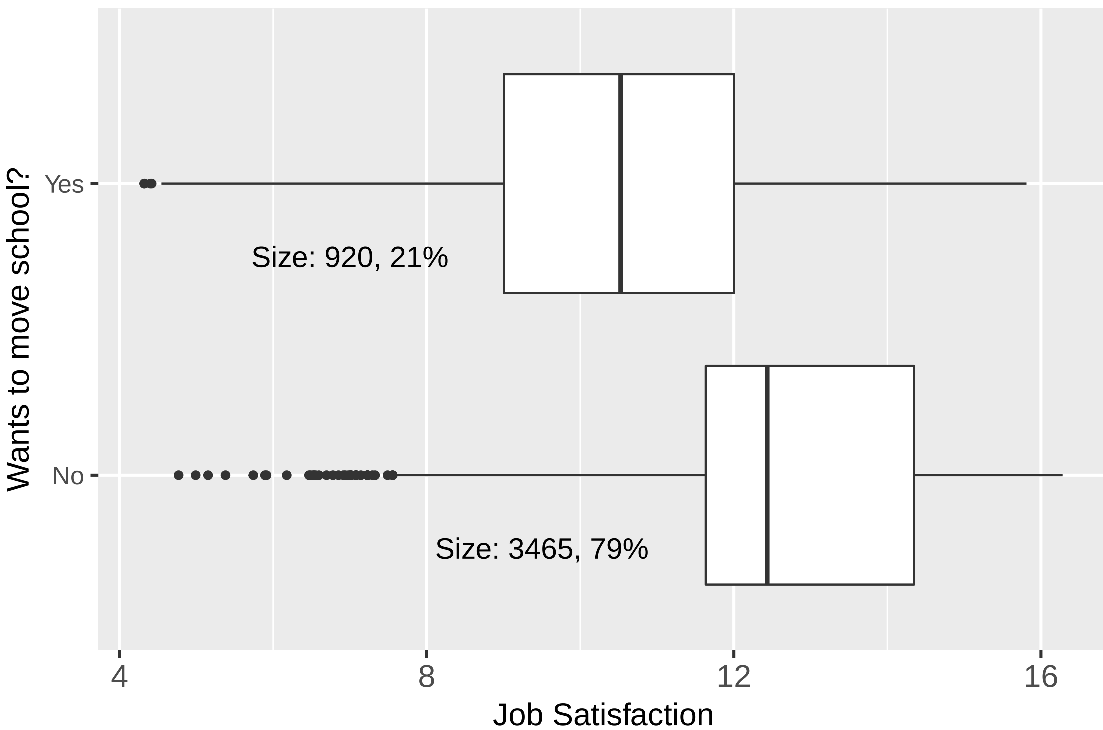

Figure 1: Relationship between job satisfaction and retention. Teachers who wish to move school tend to have lower values of job satisfaction.

fully accounted for in our analysis by estimating standard errors using the balanced repeated replication procedure (OECD, 2019b).

Data from the survey can be missing for a number of reasons. Some teachers did not reach every question, and others did not answer personal questions such as those concerning their age. Of the variables we have used, 52 contained missing values, with on average 8% of the data missing. In order to maximise the data available for use, we have imputed these missing responses with the R package `missRanger` (Mayer, 2019). This procedure involves substituting missing values with responses based on an individual's answers to all of the other questions in the survey. This enables us to retain information that would otherwise be lost if missing cases were deleted, and is more accurate than other approaches which simply use the mean value for imputation (Stekhoven and Bühlmann, 2012).

3 Methods

3.1 Traditional Approaches

Commonly used approaches in international large scale assessments to investigate the relationship between a set of independent variables, X, and a dependent variable, Y, include ordinary least squares regression and other more sophisticated modelling approaches such as hierarchical linear models. These approaches are very useful but have a number of limitations. Firstly, they assume a linear relationship between each independent variable and the outcome of interest (unless higher order terms or interactions are manually included by the analyst). This can lead to biased parameter estimates in some cases and can lead one to believe that there is no relationship between two variables when in fact there is. For example, the relationship between teacher attrition and age has been found to be $\cup$ shaped in a number of different studies (Guarino et al., 2006, Boe et al., 1997).

A second limitation is that due to the cross sectional and observational nature of the survey data, it is not possible to make any causal claims. In a linear model, relationships are expressed as $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ where β_0 is an intercept, β_1 is a slope, and ϵ_i is a residual error term. In this model β_1 should be interpreted as "the expected increase in y *associated* with a unit increase in x", and not as "the expected increase in y *caused by* a unit increase in x", since for example another variable may be causing both x and y to change. Furthermore, the direction of the relationship is not always possible to determine. Teacher self-efficacy, for example, has usually been assumed to be an antecedent of job satisfaction in much of the literature, but a recent study by Burić and Kim (2021) finds the causal direction may actually be the opposite.

Thirdly, linear models can become difficult to interpret when a large number of covariates have been included as explanatory variables. This means that it can be difficult to control for a large number of factors simultaneously when investigating the association of one variable of interest with another while still maintaining the required interpretability. Consequently, researchers often limit their analysis to a smaller subset of the available data. However, not controlling for some variables may bias parameter estimates.

In Section 4 we have concentrated on factors which relate to measures that school principals or other policy makers could introduce immediately with the view to improving job satisfaction and retention levels. By contrast, much of the existing literature on teacher job satisfaction uses scale scores of different psychological constructs which have been validated using approaches such as confirmatory factor analysis (e.g. McInerney et al., 2018, Saloviita and Pakarinen, 2021). Such approaches are certainly useful because, for example, they have demonstrated a link between higher levels of teacher self-efficacy and job satisfaction. Teachers' levels of self-efficacy are not easily changed however, and so these results do not provide a directly implementable process that can be used to improve job satisfaction or the outcome of interest. Therefore instead of considering these variables as treatments, we have instead included them as potential confounders in our models.

3.2 Bayesian Additive Regression Trees for Causal Analysis

With the above considerations in mind, this study aims to investigate the effect of a number of binary factors, which we call treatments, on teacher job satisfaction. For each of these factors, we will also conduct a secondary analysis in which we investigate its effect on a teacher's likelihood of wanting to move to a different school.

Our approach will be to use the R package `bartCause` which is a causal inference machine learning package for the R programming language (Dorie and Hill, 2020, R Core Team, 2021). The `bartCause` package allows us to estimate causal effects, and has been demonstrated to be highly competitive in causal inference machine learning competitions (Dorie et al., 2019). The package owes its success to the impressive prediction capabilities of Bayesian Additive Regression Trees (BART), a Bayesian non-parametric machine learning algorithm which is well suited to a wide variety of regression and classification problems (Chipman et al., 2010). Below we describe both the BART model, and the steps by which it is used to perform causal inference.

BART is known as a sum of trees model which can flexibly and accurately predict an outcome of interest Y using a set of covariates X. It can be seen as an extension of regression modelling that automatically identifies interactions and non-linear relationships between the variables. BART proceeds by assuming that $Y = f(X) + \epsilon$, where ϵ is a normally distributed error term. In the case of a single tree model, BART makes predictions by establishing a set of decision rules which allocate a covariate value X to a terminal node in the tree T; see Figure 2 for an example of a decision tree. The model learns the decision rules from the data as well as the terminal node parameter which provides the prediction.

More often than not, a single decision tree is insufficient to capture the complicated relationship that can exist between X and Y. For this reason, a larger number of n decision trees is often used. When this happens, each terminal node in a decision tree contributes only a small amount to the final prediction, and the resulting model is defined instead by summing the contributions from the terminal nodes across all of the trees.

The ability of BART to identify non-linear relationships and interactions between variables makes it well suited to a variety of different regression and classification problems. For a recent example of a study which uses BART to estimate the causal effects of private tuition on student achievement, see Suk et al. (2021). BART has also been used extensively in other fields outside of education, and is a popular choice for many quantitative researchers (e.g. Prado et al., 2021).

3.3 Treatment Effect Estimation

To evaluate whether a particular covariate is a causal predictor of the outcome variable we adopt the Neyman-Rubin causal model (Splawa-Neyman et al., 1990, Rubin, 1974, Sekhon, 2008). Central to the Neyman-Rubin causal model is the concept of potential outcomes. Consider a binary treatment variable Z. $Z = 1$ indicates an individual has been assigned to a treatment group, for example has been assigned to a school induction programme. $Z = 0$ indicates that an individual has been assigned to the control group (i.e. has not been assigned to an induction programme). We want to know what effect being assigned to the treatment group will have on an outcome variable of interest, Y, such as teacher job satisfaction. To do this, we would need to know the job satisfaction of an individual i both under treatment, denoted $Y_i(Z = 1)$, and under control, $Y_i(Z = 0)$. The individual treatment effect would then be given by $\tau_i = Y_i(Z = 1) - Y_i(Z = 0)$. Observing individual i simultaneously under both treatment and control is impossible, however, and this is known as the fundamental problem of causal inference.

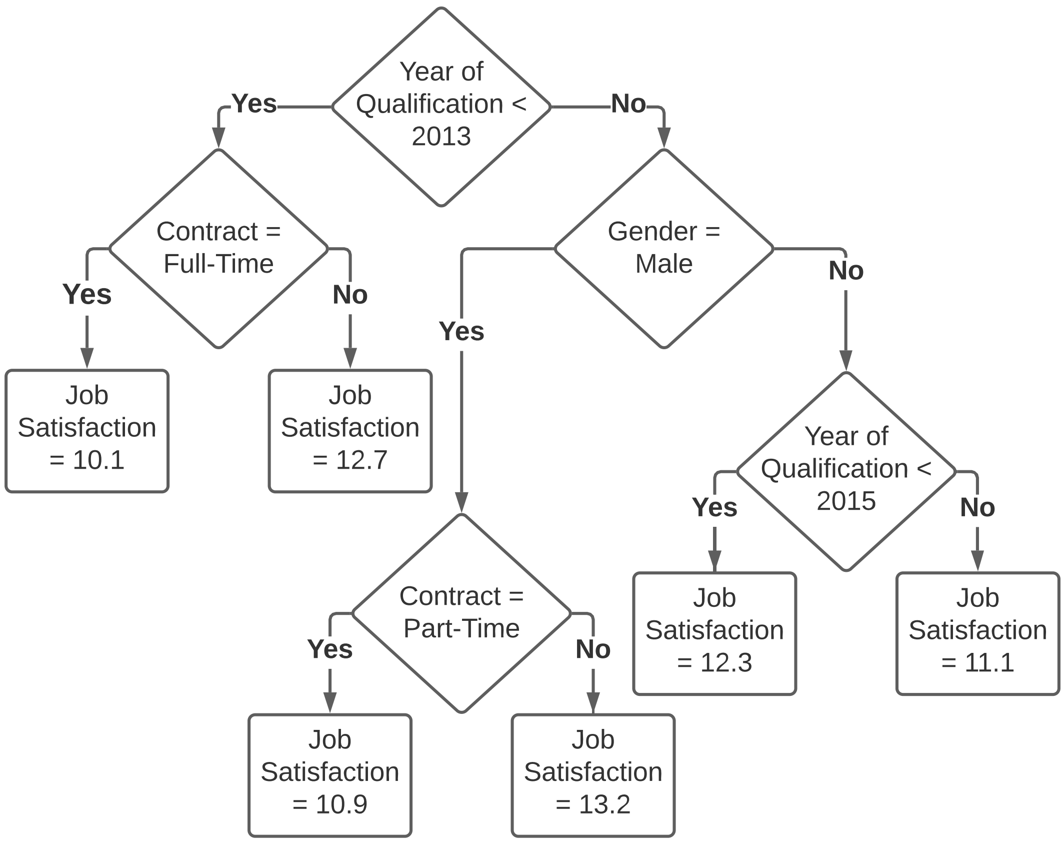

Figure 2: Example of single decision tree for the TALIS data. Each teacher's information can be fed into the tree by following the decision rules. The terminal nodes provide the predictions for the job satisfaction of each teacher. In practice the BART model works by creating many different decision trees and summing the predictions together.

Estimation of τ is a difficult task, especially in the case of observational data. Challenges posed by observational data to estimating causal effects include the fact that individuals are not randomly assigned to the treatment and control groups, and that our observation of the data may not include all variables which have an influence on the outcome of interest or the non-random assignment mechanism. It is, however, possible to identify causal effects with a number of key assumptions (Kurz, 2021). These assumptions include:

1. The stable unit treatment value assumption (SUTVA). It requires that the treatment status Z of an individual i does not affect the potential outcome of any other individual j.
2. The ignorability assumption. This requires that the potential outcomes of individual i must be independent of their treatment status conditioned on their observed covariates. In other words, we require there to be no confounding variables we have not observed.
3. The overlap assumption. This assumes that for all individuals i, $0 < P(Z_i = 1) < 1$. This means that the propensity scores for each individual (the probability of being assigned to the treatment group) must not be equal to zero or one. Essentially, every individual must have some probability of being assigned to both treatment conditions.

Assuming the above assumptions hold, the `bartCause` package estimates treatment effects by using a BART model $Y = f(X, Z) + \epsilon$ to flexibly model the response surfaces of the control and treatment groups. Hill (2011) showed that the average treatment effect (ATE), can then be estimated as:

$$\frac{1}{N}\sum_{i=1}^{N} E[Y_i(1)|X_i] - E[Y_i(0)|X_i] = \frac{1}{N}\sum_{i=1}^{N} \tau_i$$

$$= \frac{1}{N}\sum_{i=1}^{N} f(X_i, Z = 1) - f(X_i, Z = 0).$$

The performance of this approach relies on the ability to accurately estimate the potential outcomes for all individuals. This is why the high predictive performance of BART and its ability to capture interactions and non-linear relationships makes it well suited to this task. In the case of a continuous variable Y, such as a scale score for teacher job satisfaction, the ATE should be interpreted as the mean increase (or decrease) in Y that would be observed if the entire population was moved from $Z = 0$ to $Z = 1$. In the case of a binary variable Y, such as a yes/no indication that a teacher would like to change schools, the ATE should be interpreted as the average change in probability of teachers responding yes to the question that would be observed if the entire population was moved from $Z = 0$ to $Z = 1$.

3.4 Including Propensity Scores in Causal Models

Following the advice of Hahn et al. (2020), we will include an additional independent variable in the design matrix X for our model. This additional variable is known as the propensity score, and is defined as an individual's probability of being assigned to the treatment group. This probability can be estimated from an individual's characteristics such as their gender, year of qualification, degree type etc. Logistic regression is a common choice for this task, but we have chosen to use BART instead to keep our approach as consistent as possible and retain the superior predictive approach.

Due to the non-random treatment assignment mechanism in observational studies, certain cohorts may be more likely to be assigned to the treatment group than others. For example, newly-qualified teachers are more likely to be assigned a mentor than experienced members of staff. The propensity score is often used in an approach called propensity score matching, in which people with similar propensity scores but different treatment assignments are compared in an attempt to estimate causal effects, see Caliendo and Kopeinig (2008) for a summary. This ensures that only individuals with similar baseline characteristics are compared with each other. We avoid this restriction by simply including the propensity score as an extra covariate in each BART model.

The inclusion of the propensity score is designed to help avoid a phenomenon called regularisation induced confounding (Hahn et al., 2020). Besides this practical advantage, it can also be interesting to look at different trends in the propensity scores for individual teachers. Analysing such trends allows us to identify, for example, which subgroups of teachers are particularly likely to belong to positive or negative treatment groups. This process can identify specific subgroups of teachers who need to be given extra support, or who would benefit from being assigned to a particular treatment group. We highlight some examples of this in our next section.

3.5 Choice of Treatment Variables

We calculate average treatment effects for each of the following assignment (Z) options (short names or abbreviations used in Figures are shown in brackets):

1. Did the teacher take part in at least 4 CPD activities in the past year (CPD)?

2. Did they take part in a formal/informal induction programme when they started teaching at their current school (Induction)?
3. Do they take part in observing other teachers (Observing)?
4. Do they take part in team teaching (Team Teaching)?
5. Do they have a mentor (Has Mentor)?
6. Are they a mentor to another teacher (Is Mentor)?
7. Do they teach in a publicly managed school (Public School)? (Full definition in Appendix A)
8. Do they have ≥ 30 students in their class (30+ Students)?
9. Are they an out-of-field teacher (Out-of-field)?
10. Do they have a part-time contract (Part-Time)?

In each of the cases above, the ATE is calculated independently of the other treatments. This means we proceed through the list one by one, changing the definition of Z each time depending on the treatment investigated. The set of control variables included in X remains unchanged, as we control for the same covariates in every assignment option (with a few exceptions). For an exact definition of each of these treatments see Appendix A. Appendix C identifies any variables which were removed from X for a specific assignment option. For example, it would not make sense to control for the number of students in a class when investigating the effect of teaching a class with ≥ 30 students (Variable Code: TT3G38).

Figure 3 shows the control and treatment group sizes for the different factors that we have created and are investigating. The control group size for CPD is 1618, meaning that 37% (unweighted) of teachers in the sample did not take part in 4 or more CPD events over the course of the past year. The treatment group size for this assignment option is 2767, corresponding to a 63% participation rate in at least 4 CPD events. The other segments of the plot have similar interpretations.

As can be seen from Figure 3, 30% of teachers met our criteria for teaching out-of-field. A more in depth analysis of these numbers reveals that 24% of secondary school teachers meet this criteria, and 37% of primary school teachers do. Further investigations also show that the subjects being taught out-of-field by teachers are different across the two school levels. We bring this point to the reader's attention to make clear that these teachers are all treated identically, and we do not make careful distinctions between reasons for teaching out-of-field. Furthermore, we do not distinguish between primary vs. secondary school teachers for this treatment effect (or indeed any of the other treatment effects).

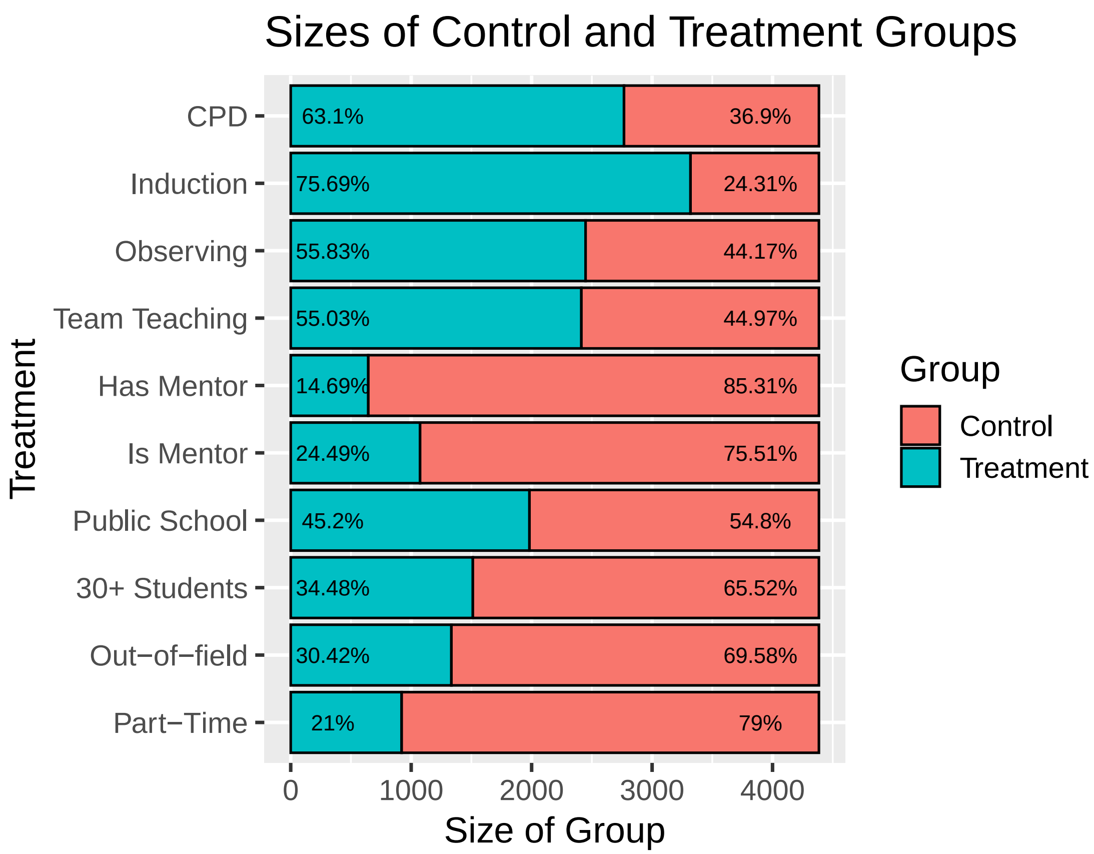

Figure 3: Percentage of teachers belonging to the control and treatment groups under investigation. There are different levels of balance across the groups.

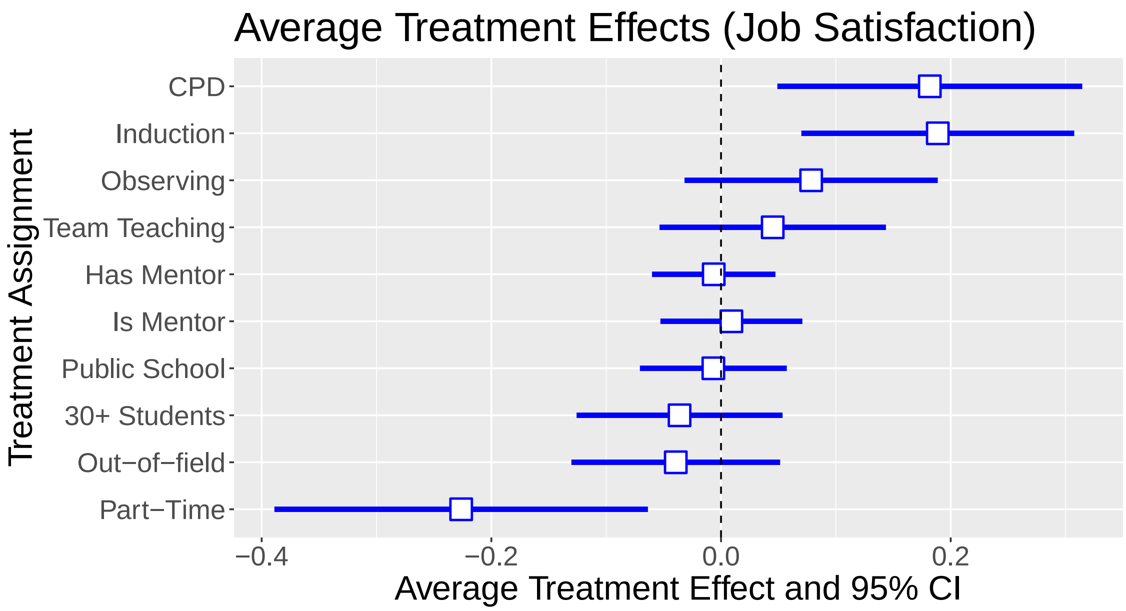

Figure 4: Plot of Average Treatment Effects (Effect on Job Satisfaction)

The assignment options listed have varying levels of balance across the two groups. Team teaching and working in a public school have the highest levels of balance across the two groups, with close to 50% in each. Having a mentor is the least balanced, as there is fewer than 15% of the population in the minority group.

4 Results

This section describes the results from:

1. Choosing a treatment assignment option to consider from the list in Section 3.5.
2. Calculating the average treatment effect of this assignment option on both job satisfaction and desire/intention to move to a different school.

For a visual representation of these results, see Figures 4 and 5 which indicate the final estimate and 95% Bayesian credible interval for each of the treatment effects. For a summary of goodness of fit statistics for each of the models used to calculate these results, the reader is referred to Appendix D.

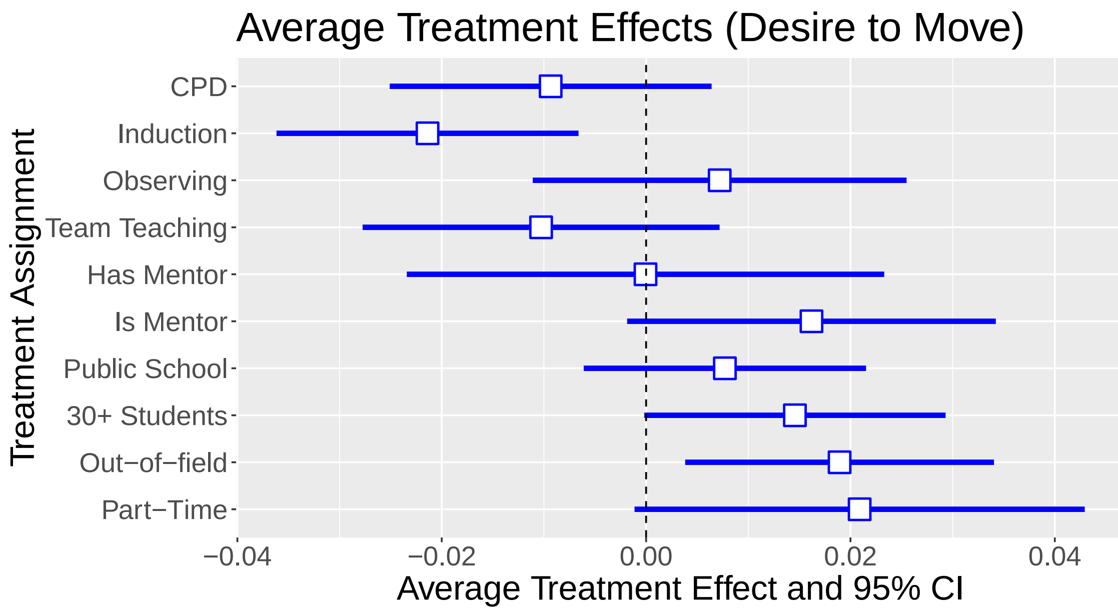

Figure 5: Plot of Average Treatment Effects (Effect on Likelihood of Wanting to Move School)

4.1 Continual Professional Development

Our results identify participation in at least 4 CPD events over the course of a year as having a positive effect on teacher job satisfaction. The 95% credible interval for this average treatment effect is [0.049, 0.315]. To give an idea of the magnitude of this treatment effect, consider that the teacher job satisfaction scale has a mean of 12.42, and a standard deviation of 2.28. Therefore, the centre-point of this credible interval which is at 0.182 would correspond to an increase in job satisfaction of 0.08 standard deviations, which is a small but positive improvement.

When we consider the effect of CPD on desire/intention to move to another school we find that there is also a positive effect. Our results show that on average, higher levels of participation in CPD leads to a 1% reduction in desire to move school. This is based on a 1% lower probability of teachers agreeing or strongly agreeing with the statement “I would like to change to another school if that were possible”. The 95% credible interval for this percentage is [-2.509, 0.641]. Therefore, the associated uncertainty allows for the possibility that CPD could in fact lead to an even greater reduction in movement between schools.

4.2 Induction and Mentoring Programmes

Our results show that taking part in induction when starting at a new school has a positive effect on both job satisfaction and retention. The 95% credible intervals for the average treatment effect on job satisfaction and retention are [0.070, 0.308] and [-3.617, -0.661] respectively. Therefore, taking part in an induction scheme is associated with a mean increase in job satisfaction of 0.189, and an average reduction in desire to change school of 2.139%. This result means that induction schemes are the most beneficial of all of the treatment assignment options we have considered, both in relation to job satisfaction and retention.

Mentoring, however, is not identified as having a strongly positive effect on teacher job satisfaction or retention. As can be seen from the 95% credible intervals in Figures 4 and 5, this is true for both mentors and mentees.

4.3 Observation and Team Teaching

Team teaching and observing the lessons of other teachers are both identified as having a positive effect on job satisfaction. The uncertainty in these estimates is quite large however, and this is reflected in the wide 95% credible intervals shown in Figure 4 which both include 0 within their range. The effect of these factors on teacher retention is also unclear, with the treatment effect estimates being close to 0, and with considerable uncertainty. Given the large credible intervals it may be that there are large effects of these variables, but the data here do not provide us with enough information to estimate them precisely. Alternatively there may be sub-groups for whom the causal effect is particularly high or low. This, however, would also be difficult to ascertain with a high degree of statistical confidence.

4.4 Other Factors

Of the remaining factors we considered, the treatment assignment option with the largest effect in relation to job satisfaction is the possession of a part-time contract of less than 90% of a typical full-time contract's hours. This factor has an effect of reducing job satisfaction on average by 0.226, 95% credible interval [-0.389, -0.064]. We also see that this factor has a similar negative effect on retention, increasing the likelihood of teachers wanting to change school by on average 2.090%. The results from analysing the propensity scores for this factor show an interesting trend. Figure 6 shows that experienced female teachers generally have much higher propensity scores (probability of being assigned to the treatment group) than their male colleagues.

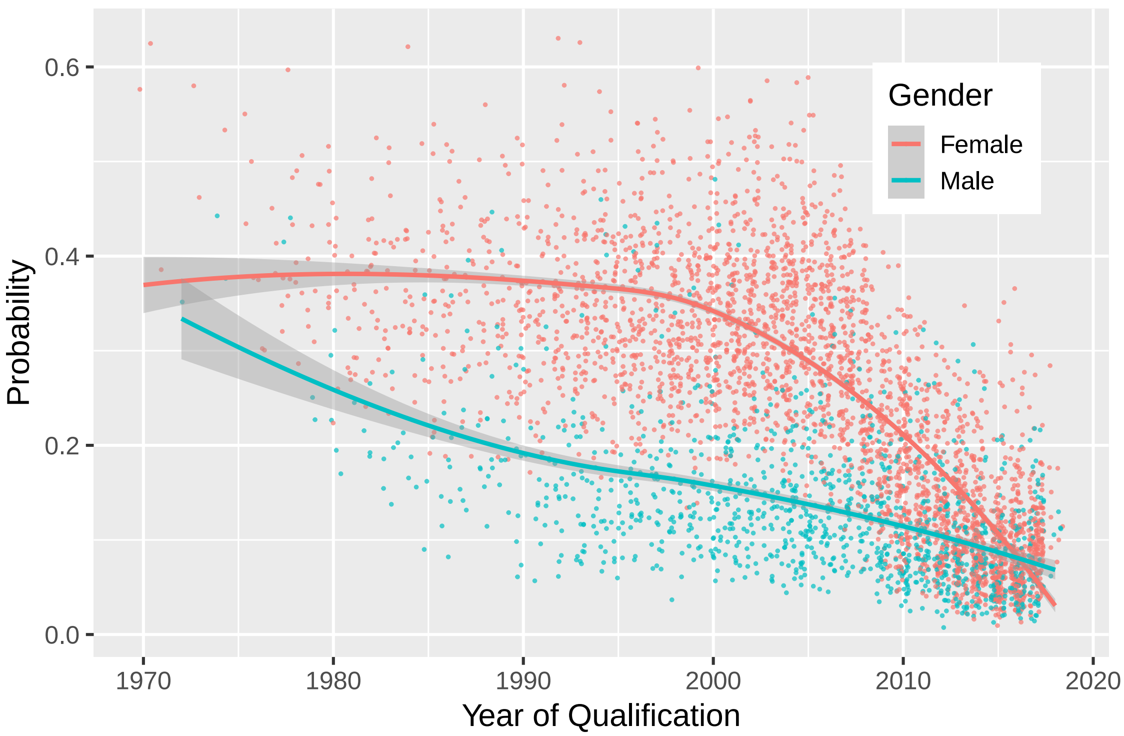

Figure 6: Probability of having a part-time contract. Female teachers have higher probabilities than male teachers, especially more experienced female teachers.

The only other assignment option having a considerable effect on either job satisfaction or retention is out-of-field teaching, which increases the probability of wanting to move to another school by an average of 1.894%. The effect of out-of-field teaching on job satisfaction is also negative, but the effect is quite small and is associated with a large amount of uncertainty.

The other factors we have considered are working in a public vs. a private school, and teaching a class with 30 or more students. According to our results, neither of these factors are associated with a strong effect on job satisfaction or retention. We emphasise again that these factors may indeed be very important, but the precision with which the data allows us to estimate these effects is insufficient to make such claims with a high degree of statistical confidence in this case.

5 Discussion

We begin by discussing our main findings in more detail, and go on to highlight some key aspects of this study which make a new and important contribution to the literature on teacher job satisfaction and retention. We finish this section by drawing the reader's attention to some limitations of this study, and by suggesting areas for future research.

5.1 Main Findings

5.1.1 Continual Professional Development

Our results identify high levels of participation in CPD as having a positive effect on teacher job satisfaction. This is in line with previous studies which investigate the relationship between CPD and job satisfaction (Yoon and Kim, 2022, Wang et al., 2020). Crucially, our result supports these previous findings by verifying the strong positive effects of CPD using a causal machine learning approach, and thus we are able to infer results about causation and not just correlation. We also find a strong positive effect of high levels of CPD on decreasing the likelihood of a teacher intending to move to a different school. Therefore, our results support previous findings which have linked CPD to higher levels of retention (e.g. Coldwell, 2017, Allen and Sims, 2017). Furthermore we have ensured that our treatment effect estimates are as unbiased as possible, by removing the effect of possible confounding variables on our two outcomes of interest.

We noted that only 63% of teachers in the English dataset have reached this high level of CPD. Barriers to participation in CPD are known to be a key predictor of both job satisfaction and retention (Zhang et al., 2020). Our results therefore also provide strong support for this body of work by demonstrating the positive gains that can be made by removing such barriers and encouraging and enabling more teachers to engage in CPD events.

We highlight the fact that our investigation has only considered a binary version of CPD. In reality, however, levels of attendance at CPD belong on a spectrum, not just high/low. Furthermore, the benefits from CPD are likely to depend on many factors such as the quality and relevance of the training to a teacher's needs. These factors warrant further investigation but were beyond the scope of this study. Despite this, we do find clear evidence in favour of recommending CPD as a measure for improving both job satisfaction and retention.

5.1.2 Induction and Mentoring Programmes

Our finding that induction schemes have a very positive effect, both in relation to job satisfaction and retention, agrees with prior work from Ronfeldt and McQueen (2017). Contrary to the review by Allen (2005), we did not detect high levels of heterogeneity in the treatment effect estimates of this assignment option. Therefore we argue that induction schemes should be made available for all new teachers at a school, regardless of number of years qualified or levels of difficulties faced within the classroom.

An analysis of the propensity scores from the English data shows a greater tendency for more recently qualified teachers to participate in induction schemes (Appendix E). This could reflect the fact that teachers are more likely to receive support when beginning their career, or could be a sign that induction schemes have become more common. However only 80% of the most recently qualified teachers reported that they availed of an induction scheme. Our results indicate that there are significant benefits to be gained from encouraging and enabling the remaining 20% to participate in an induction scheme.

Contrary to previous studies (e.g. Ingersoll and Strong, 2011), our results do not identify the presence of a mentor as being beneficial for job satisfaction or retention. There are a number of plausible reasons for this. Firstly, there may be some unobserved or unaccounted for confounding variables common to schools with mentorship schemes which bias the estimates of these analyses. A second plausible explanation is that mentoring usually involves new teachers. This means that the treatment and control groups consist mostly of two distinct subgroups of the population - new teachers and experienced teachers. Therefore, a comparison of these two groups is likely to be less reliable, leading to an unexpected result. Thirdly, we did not consider other aspects related to mentoring, such as the subject area of the mentor. Research has shown that a mentee is more likely to benefit from a mentoring arrangement if their mentor is a teacher from the same grade level (Parker et al., 2009). Other factors such as the mentoring quality and the frequency of meetings can also be important (Richter et al., 2013). Accounting for this extra influence may have lead to different results.

An indicator of whether or not a teacher is currently a mentor to another member of staff was also included as a component in our analysis. Similarly, we did not find that this treatment was associated with an appreciable increase or decrease in job satisfaction or retention. Again, this could be a result of our binary view of mentoring relationships, in which we only consider the presence or absence of a mentee, and fail to account for other aspects such as the quality of the mentoring relationship, which has been demonstrated to be an important predictor of job satisfaction (Lunsford et al., 2018).

5.1.3 Observation and Team Teaching

The fact that we have not found a clear link between team teaching or observation with job satisfaction or retention may initially appear to be strange. The literature reviewed consistently pointed towards higher levels of teamwork and cooperation as having a positive effect on teacher job satisfaction and retention. Therefore, we might have expected to see this reflected in our results also.

One plausible explanation for this is that higher levels of teamwork and cooperation within a school are difficult to attribute to a small number of specific practices such as team teaching and observation. Higher levels of teamwork and cooperation within a school are characterised by many different aspects such as sharing resources with colleagues and collaborating together on different projects etc. As a result, it is difficult to capture the true impact of higher levels of teamwork and cooperation as a whole by only considering two of a much larger number of indicators. Therefore, the absence of a large effect size here does not necessarily mean that team teaching and observation are not useful practices. Rather, the results indicate that only implementing one or two of these factors is unlikely to yield significant improvements in job satisfaction or retention, and efforts should instead be focused on improving teamwork and cooperation as a whole. This is made clear by the very small treatment effect sizes that result from us considering two of these such practices in isolation.

5.1.4 Other Factors

We investigated whether working at a publicly owned and managed school affects job satisfaction and retention. We consider this treatment assignment option to be an excellent test for our approach, because if all other aspects of a school's working environment (such as workload, cooperation, and characteristics of the student body) are kept constant, we would not expect to see a change in job satisfaction of the staff in a school based solely on a change of ownership. In other words, we expect there to be no direct causal influence. The results from our approach, as expected, do not identify a significant causal effect for this treatment assignment. This result is in line with work by Dahler-Larsen and Foged (2018) who come to the same conclusion, and attribute the difference in job satisfaction between public and private schools to differences in organisational characteristics.

In line with research by Reeves et al. (2017), our results show that teaching a class with 30 or more students does not have a large effect on job satisfaction or desire to move school. We should note, however, that our finding is based on a cut-off point of 30 students. This value was chosen to ensure an approximately even split of teachers in the treatment and control groups. It is possible, however, that a different value would yield different results, and teachers at the more extreme end of the distribution with greater than 35 students may experience a more negative effect from this treatment.

Given the lack of research linking out-of-field teaching to job satisfaction or retention, we thought it was important to include this as a factor in our study. Out-of-field teaching stands out as an interesting example of an assignment option where we have found a relatively large increase in desire to move school, but only a very small reduction in teacher job satisfaction. This might suggest that even though job satisfaction and retention are strongly correlated, they are in fact distinct entities, and the size of an effect that a given intervention has may be quite different for these separate outcomes. Therefore, it may be important to interpret the results of studies which focus solely on job satisfaction with caution if attempting to identify factors which will boost retention.

As in the study by Ferguson et al. (2012), the contract-type used in our study refers to full-time or part-time contracts. We have chosen this as it will allow us to have more evenly balanced control and treatment groups. Our results show that teachers on a part-time contract are less satisfied with their career than their full-time colleagues. Also, an analysis of the propensity scores for this treatment effect shows that experienced female teachers have much higher probabilities of being on part-time contracts than their male counterparts. Future research should investigate the reasons for this, and supports that might be put in place for teachers with childcare responsibilities.

To summarise this section, and to answer our two research questions, CPD and participation in induction schemes are the two factors we have identified which have the strongest positive influence on teacher job satisfaction and retention. We have also found that possessing a part-time contract can have a negative effect on job satisfaction, but the effect of this factor on retention is less certain. Conversely, we have also found that out-of-field teaching can lead to a significant increase in desire to move school, but we did not find a substantial effect on the job satisfaction of these teachers. These two observations highlight the need to consider the effect of factors on job satisfaction and retention separately, as we have done in this study.

5.2 Contribution of This Paper

We believe this study makes three main contributions to the current literature on teacher job satisfaction and retention. Firstly, we have employed a causal inference machine learning approach. The advantages of this method include the ability to flexibly model job satisfaction without assuming a linear relationship which is a common feature of most conventional statistical models. The approach is also well suited to detecting interactions between variables and allows us to include a much wider variety of covariates than would normally be possible when using linear models. This is absolutely crucial, because it enables us to model the response surface using the propensity score along with a large number of other variables. This allows us to account for many potential sources of confounding which could bias treatment effect estimates.

Secondly, instead of identifying important characteristics related to a teacher's working environment such as cooperation, quality of school leadership, or personal traits such as self-efficacy, we have instead established several specific and implementable measures that may be introduced in an attempt to improve job satisfaction and retention. This is important because although it may be known that certain factors such as stress are negatively correlated with job satisfaction (Klassen and Chiu, 2010), it is not always obvious how best to reduce stress levels among teachers, or if a set of proposed changes will have the desired effect. This study therefore bridges that gap by identifying factors such as induction schemes which can be beneficial for job satisfaction and retention, while also identifying the negative effects of factors such as out-of-field teaching. We suggest that future research should investigate more binary features like this which could impact on job satisfaction and retention.

The propensity scores described in Section 3.4, although not the primary focus of this study, provide us with an interesting insight into the types of teachers more likely to belong to the treatment and control groups we have investigated. This can help us to identify certain subgroups of teachers who have not availed of positive treatments, and we can then ensure that these activities are made available to them. This can also help us to identify subgroups of teachers who are more likely to be exposed to the negative effects of a treatment, such as experienced female teachers who we found were significantly more likely to have a part-time contract.

5.3 Limitations and Areas for Future Research

As discussed in the methodology, the causal inference approach that we have employed makes a number of important assumptions. Among these is the ignorability assumption, which requires that we have accounted for all potential sources of confounding when investigating a given treatment. Despite including a wide variety of control variables in our design matrix, X, it is certainly still possible that there may be some confounding variables not collected as part of the survey. Teachers with young children for example, may be more likely to work part-time, but there is no indication in the TALIS data whether teachers have young children. Future research could include a detailed assessment of the reasonableness of these assumptions in relation to TALIS by incorporating data from external sources, and using different diagnostic methods designed to assess these assumptions.

A second limitation of our approach is that some aspects of the working environment such as teamwork and cooperation are very difficult to capture with binary variables. Therefore, it may be less meaningful to investigate binary factors in relation to aspects such as this, because levels of teamwork and cooperation can not be fully characterised by a simple dummy variable. Also, hours of CPD attended and the number of students in a class are both continuous variables. Therefore, their impact on job satisfaction and retention can not be fully appreciated by artificially converting them into a binary factor. Additional studies which use causal inference machine learning algorithms designed to handle continuous treatment variables may be better suited to this task.

Finally, our investigation of factors affecting teacher retention uses a measure of a teacher's intention/desire to move to a different school, which is not the same as actual numbers of teachers leaving to teach in different schools. A second important distinction to make is that we have only considered retention in the context of moving between different schools. Therefore, our results may not provide a good indication of the effect of different assignment options on a teacher's intention to stay teaching or leave the profession entirely. Future research could include studies using a similar approach to ours, but applied to other observational data sets in which there is a direct measure of attrition.

6 Conclusion

Many studies which investigate factors associated with job satisfaction and retention use continuous scales such as cooperation, workload, or self-efficacy to predict these two outcomes. In order to improve teacher satisfaction or retention, however, it is vital to identify specific and implementable measures that may be put in place. It is also important to be aware of the limitations of using observational data for these purposes. This study focuses on these two gaps in the research by employing a causal inference machine learning approach to estimate the causal effects of a number of treatment options. Future research could build on these two areas by investigating other specific steps which can be taken to improve job satisfaction or retention, or by employing other causal inference approaches, or both.

A Questions Used to Define Treatment Groups

Treatment	Question	Condition
CPD	During the last 12 months, did you participate in any of the following professional development activities?	Teachers who responded “yes” to any 4 of the 10 available options.
Induction	Did you take part in any induction activities?	Teachers who responded “yes” to either taking part in a formal or informal induction programme at their current school.
Observing	On average, how often do you do the following in this school?	Teachers who did not respond “never” to the option “Observe other teachers’ classes and provide feedback.”
Team Teaching	On average, how often do you do the following in this school?	Teachers who did not respond “never” to the option “Teach jointly as a team in the same class.”
Has Mentor	Are you currently involved in any mentoring activities as part of a formal arrangement at this school?	Teachers who responded “yes” to having a mentor.
Is Mentor	Are you currently involved in any mentoring activities as part of a formal arrangement at this school?	Teachers who responded “yes” to being a mentor.
Public	Is this school publicly or privately-managed?	Teachers with a principal who indicated their school is publicly-managed.
30+ Students	How many students are currently enrolled in this class?	Teachers who answered 30 or more students.
Out-of-field	Were the following subject categories included in your formal education or training, and do you teach them during the current school year to any students in this school?	Teachers who indicated that at least one option given was not included in their education, but that they do currently teach it.
Part-Time	What is your current employment status as a teacher, in terms of working hours?	Teachers who indicated they do not have a full time contract at their current school.

B Definitions of Key Terms Given in TALIS Questionnaire

Key Term	Definition
CPD	In this section, ‘professional development’ is defined as activities that aim to develop an individual’s skills, knowledge, expertise and other characteristics as a teacher.
Induction	‘Induction activities’ are designed to support new teachers’ introduction into the teaching profession and to support experienced teachers who are new to a school, and they are either organised in formal, structured programmes or informally arranged as separate activities.
Mentoring	‘Mentoring’ is defined as a support structure in schools where more experienced teachers support less experienced teachers. This structure might involve all teachers in the school or only new teachers. It does not include mentoring of student teachers doing teaching practice at this school.
Public School	This is a school managed by a public education authority, government agency, municipality, or governing board appointed by government or elected by public franchise.
Private School	This is a school managed by a non-government organisation; e.g. a church, trade union, business or other private institution.

C List of Potential Confounders Used

TALIS Variable Code	Description	Removed from X for treatment
IDCNTPOP	Primary/Secondary School.	
TT3G01	Gender.	
TT3G03	Highest level of formal education completed.	
TT3G04	How did you receive your first teaching qualification?	
TT3G05	Year of Qualification.	
TT3G08	Was teaching your first choice as a career?	
TT3G09	Permanent/Fixed-Term Contract.	
TT3G10A	Working hours at this school.	Part-Time Contract.
TT3G10B	Working hours altogether.	Part-Time Contract.
TT3G11A	Year(s) working as a teacher at this school.	
TT3G11B	Year(s) working as a teacher in total.	
TT3G11C	Year(s) working in other education roles.	
TT3G11D	Year(s) working in non education roles.	
TT3G12	Do you currently work as a teacher at another school?	
TT3G14	Number of students in class with special needs.	
TT3G37	Subject taught.	
TT3G38	Number of students in class.	30+ Students.
TT3G39A	% of time spent on administrative tasks.	
TT3G39B	% of time spent keeping order in classroom.	
TT3G39C	% of time actually spent teaching.	
T3STBEH	Student behaviour stress.	
T3CLAIN	Clarity of instruction.	
T3CLASM	Classroom management.	
T3COGAC	Cognitive activation.	
T3COLES	Professional collaboration in lessons among teachers.	
T3EFFPD	Effective professional development.	
T3EXCH	Exchange and co-ordination among teachers.	
T3PDBAR	Professional development barriers.	
T3DISC	Teachers' perceived disciplinary climate.	
T3PERUT	Personal utility motivation to teach.	
T3PDIV	Needs for professional development for teaching for diversity.	
T3PDPED	Needs for professional development in subject matter and pedagogy.	
T3VALP	Perceptions of value and policy influence.	
T3SATAT	Satisfaction with target class autonomy.	
T3SECLS	Self-efficacy in classroom management.	
T3SEINS	Self-efficacy in instruction.	
T3SEENG	Self-efficacy in student engagement.	
T3SEFE	Self-related efficacy in multicultural classrooms.	
T3SOCUT	Social utility value.	
T3STAKE	Participation among stakeholders, teachers.	
T3TEAM	Team innovativeness.	
T3STUD	Teacher-student relations.	
T3WELS	Workplace well-being and stress.	
T3WLOAD	Workload stress.	
T3TPRA	Teaching practices, overall.	
T3COOP	Teacher co-operation.	
T3SELF	Teacher self-efficacy.	
T3DIVP	Diversity practices.	
T3JOBSA	Overall job satisfaction.	All.
TT3G53C	Would like to move school.	All.

D Goodness of Fit Statistics

To calculate the goodness of fit statistics in the table below we have split the data completely at random into a 70% training set, and 30% test set. For each treatment under investigation we have then trained the required propensity score, job satisfaction, and desire to move school models before testing them on the unseen data. For the propensity score and desire to move school models we have calculated the Area Under Receiver Operating Characteristic Curve (AUC) metric, and for the job satisfaction model we have calculated the Root Mean Squared Error (RMSE).

Treatment	Propensity Score Model AUC	Job Satisfaction Model RMSE	Wants to Move Model AUC
CPD	0.756	1.739	0.600
Induction	0.642	1.657	0.617
Observing	0.850	1.741	0.606
Team Teaching	0.850	1.697	0.622
Has Mentor	0.833	1.678	0.646
Is Mentor	0.667	1.659	0.600
Public	0.653	1.732	0.615
30+ Students	0.571	1.750	0.600
Out-of-field	0.609	1.736	0.613
Part-Time	0.748	1.737	0.643

E Supplementary Figures

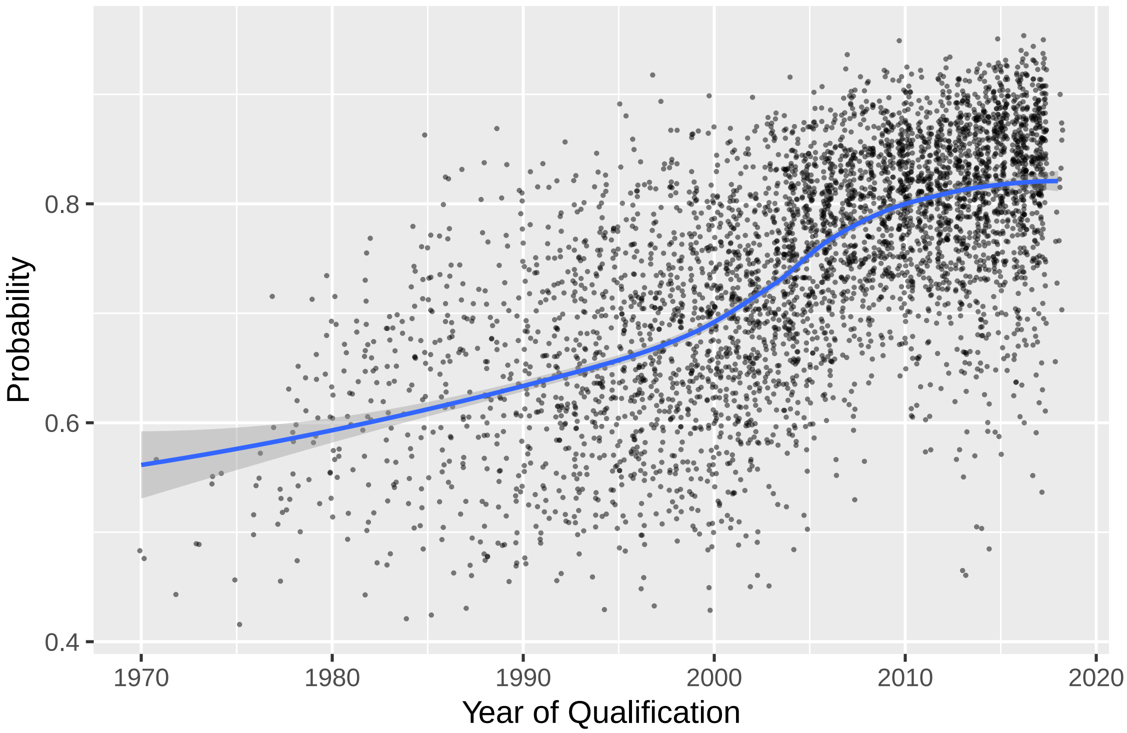

Figure 7: Probability of taking part in induction. More recently qualified teachers have higher probabilities.

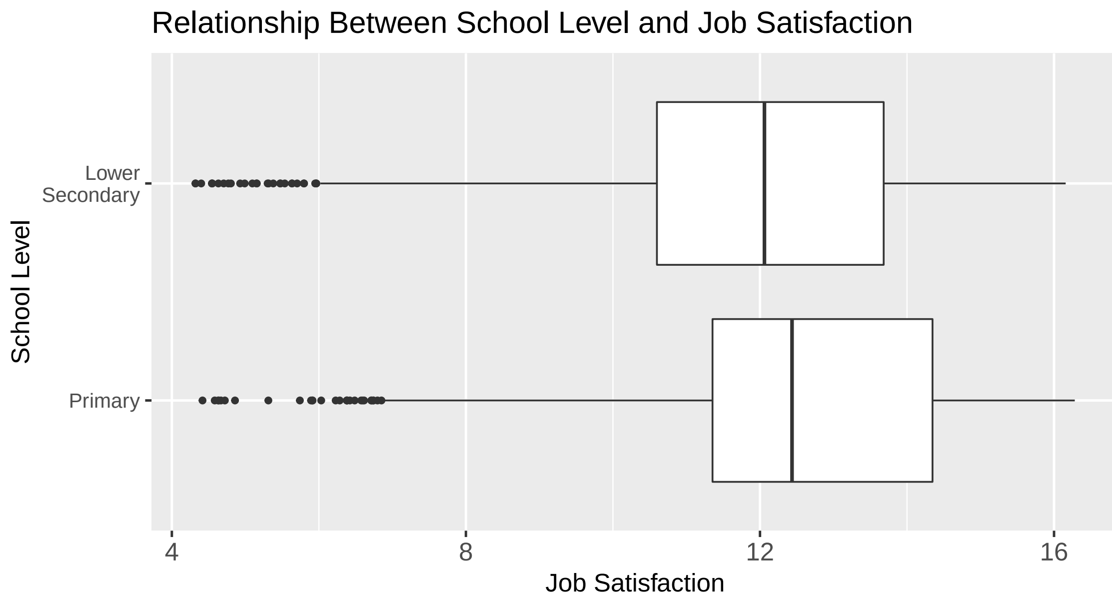

Figure 8: Job Satisfaction Grouped by School Level

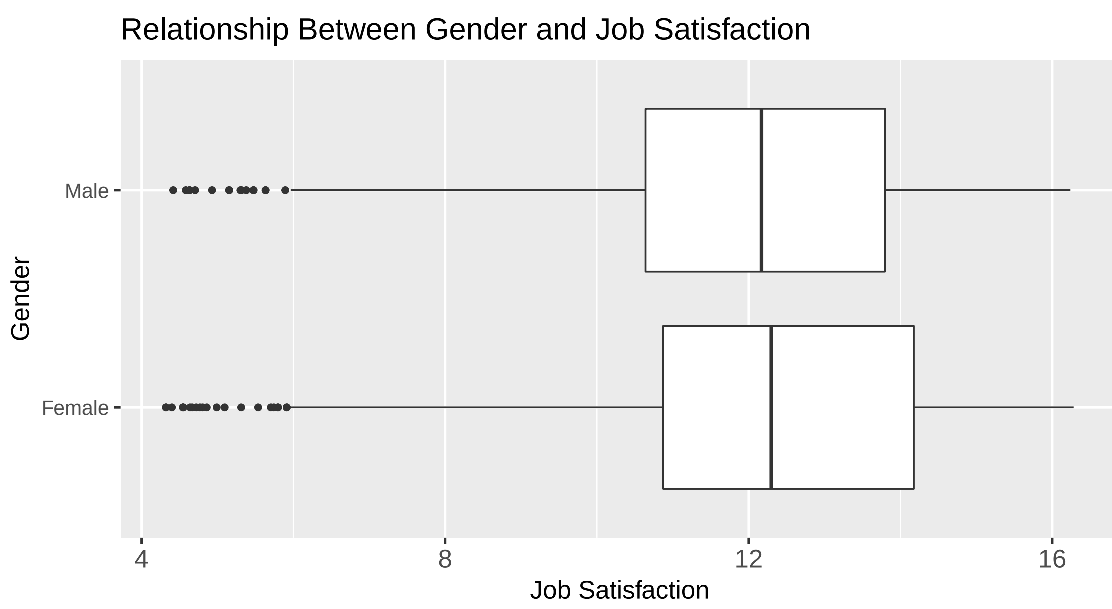

Figure 9: Job Satisfaction Grouped by Gender

References

Allen, M. B. (2005). Eight questions on teacher recruitment and retention: What does the research say?. *Education Commission of the States (NJ3)*.

Allen, R. and Sims, S. (2017). Improving science teacher retention: Do national STEM learning network professional development courses keep science teachers in the classroom. *Education Research in STEM Subjects: Main Collection*.

Boe, E. E., Bobbitt, S. A., Cook, L. H., Whitener, S. D., and Weber, A. L. (1997). Why didst thou go? Predictors of retention, transfer, and attrition of special and general education teachers from a national perspective. *The journal of special education*, 30(4):390–411.

Borman, G. D. and Dowling, N. M. (2008). Teacher attrition and retention: A meta-analytic and narrative review of the research. *Review of educational research*, 78(3):367–409.

Burić, I. and Kim, L. E. (2021). Job satisfaction predicts teacher self-efficacy and the association is invariant: Examinations using TALIS 2018 data and longitudinal Croatian data. *Teaching and Teacher Education*, 105.

Caliendo, M. and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22(1):31–72.

Capone, V. and Petrillo, G. (2020). Mental health in teachers: Relationships with job satisfaction, efficacy beliefs, burnout and depression. *Current Psychology*, 39(5):1757–1766.

Chipman, H. A., George, E. I., and McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1):266–298.

Coldwell, M. (2017). Exploring the influence of professional development on teacher careers: A path model approach. *Teaching and Teacher Education*, 61:189–198.

Dahler-Larsen, P. and Foged, S. K. (2018). Job satisfaction in public and private schools: Competition is key. *Social Policy & Administration*, 52(5):1084–1105.

Dee, T. S. and Cohodes, S. R. (2008). Out-of-field teachers and student achievement: Evidence from matched-pairs comparisons. *Public Finance Review*, 36(1):7–32.

Dorie, V. and Hill, J. (2020). Package 'bartCause'. `https://cran.r-project.org/web/packages/bartCause/bartCause.pdf`.

Dorie, V., Hill, J., Shalit, U., Scott, M., and Cervone, D. (2019). Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition. *Statistical Science*, 34(1):43–68.

Ferguson, K., Frost, L., and Hall, D. (2012). Predicting teacher anxiety, depression, and job satisfaction. *Journal of teaching and learning*, 8(1).

Gil-Flores, J. (2017). The role of personal characteristics and school characteristics in explaining teacher job satisfaction. *Revista de Psicodidáctica (English ed.)*, 22(1):16–22.

Gray, L. and Taie, S. (2015). Public school teacher attrition and mobility in the first five years: Results from the first through fifth waves of the 2007-08 beginning teacher longitudinal study. First look. *National center for education statistics*.

Guarino, C. M., Santibanez, L., and Daley, G. A. (2006). Teacher recruitment and retention: A review of the recent empirical literature. *Review of educational research*, 76(2):173–208.

Hahn, P. R., Murray, J. S., and Carvalho, C. M. (2020). Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects (with discussion). *Bayesian Analysis*, 15(3):965–1056.

Han, D. and Hur, H. (2021). Managing turnover of STEM teacher workforce. *Education and Urban Society*.

Hill, J. G. and Dalton, B. (2013). Student math achievement and out-of-field teaching. *Educational Researcher*, 42(7):403–405.

Hill, J. L. (2011). Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1):217–240.

Hilton, G. L. (2017). Disappearing teachers: An exploration of a variety of views as to the causes of the problems affecting teacher recruitment and retention in England. *Bulgarian Comparative Education Society*.

Hughes, G. D. (2012). Teacher retention: Teacher characteristics, school characteristics, organizational characteristics, and teacher efficacy. *The Journal of Educational Research*, 105(4):245–255.

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American educational research journal*, 38(3):499–534.

Ingersoll, R. M. and Strong, M. (2011). The impact of induction and mentoring programs for beginning teachers: A critical review of the research. *Review of educational research*, 81(2):201–233.

Kengatharan, N. (2020). The effects of teacher autonomy, student behavior and student engagement on teacher job satisfaction. *Educational Sciences: Theory & Practice*, 20(4).

Klassen, R. M. and Chiu, M. M. (2010). Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress. *Journal of educational Psychology*, 102(3):741.

Kurz, C. F. (2021). Augmented inverse probability weighting and the double robustness property. *Medical Decision Making*.

Levy, A. J., Joy, L., Ellis, P., Jablonski, E., and Karelitz, T. M. (2012). Estimating teacher turnover costs: A case study. *Journal of Education Finance*, pages 102–129.

Li, W. and Konstantopoulos, S. (2017). Does class-size reduction close the achievement gap? Evidence from TIMSS 2011. *School Effectiveness and School Improvement*, 28(2):292–313.

Lopes, J. and Oliveira, C. (2020). Teacher and school determinants of teacher job satisfaction: A multilevel analysis. *School Effectiveness and School Improvement*, 31(4):641–659.

Lunsford, L., Baker, V., and Pifer, M. (2018). Faculty mentoring faculty: Career stages, relationship quality, and job satisfaction. *International Journal of Mentoring and Coaching in Education*.

Madigan, D. J. and Kim, L. E. (2021). Towards an understanding of teacher attrition: A meta-analysis of burnout, job satisfaction, and teachers' intentions to quit. *Teaching and Teacher Education*, 105.

Malinen, O.-P. and Savolainen, H. (2016). The effect of perceived school climate and teacher efficacy in behavior management on job satisfaction and burnout: A longitudinal study. *Teaching and Teacher Education*, 60:144–152.

Matsuoka, R. (2015). School socioeconomic context and teacher job satisfaction in Japanese compulsory education. *Educational Studies in Japan*, 9:41–54.

Mayer, M. (2019). Package 'missRanger'. `https://cran.r-project.org/web/packages/missRanger/missRanger.pdf`.

McInerney, D. M., Korpershoek, H., Wang, H., and Morin, A. J. (2018). Teachers' occupational attributes and their psychological wellbeing, job satisfaction, occupational self-concept and quitting intentions. *Teaching and Teacher Education*, 71:145–158.

Nguyen, T. D. (2021). Linking school organizational characteristics and teacher retention: Evidence from repeated cross-sectional national data. *Teaching and Teacher Education*, 97.

OECD (2018). TALIS 2018 teacher questionnaire. `https://www.oecd.org/education/school/TALIS-2018-MS-Teacher-Questionnaire-ENG.pdf`.

OECD (2019a). TALIS 2018 and TALIS starting strong 2018 user guide. `https://www.oecd.org/education/talis/TALIS_2018-TALIS_Starting_Strong_2018_User_Guide.pdf`.

OECD (2019b). TALIS 2018 technical report. Chapter 9. Estimation weights, participation rates and sampling error. `https://www.oecd.org/education/talis/TALIS_2018_Technical_Report.pdf`.

Olmos, F. (2010). *Square peg in a round hole: Out-of-field teaching and its impact on teacher attrition*. University of California, Irvine and California State University, Los Angeles.

O'Doherty, T. and Harford, J. (2018). Teacher recruitment: Reflections from Ireland on the current crisis in teacher supply. *European Journal of Teacher Education*, 41(5):654–669.

Parker, M. A., Ndoye, A., and Imig, S. R. (2009). Keeping our teachers! Investigating mentoring practices to support and retain novice educators. *Mentoring & Tutoring: Partnership in Learning*, 17(4):329–341.

Perrachione, B. A., Rosser, V. J., and Petersen, G. J. (2008). Why do they stay? Elementary teachers' perceptions of job satisfaction and retention. *Professional Educator*, 32(2).

Prado, E. B., Parnell, A. C., McJames, N., O'Shea, A., and Moral, R. A. (2021). Semi-parametric Bayesian additive regression trees. *arXiv preprint arXiv:2108.07636*.

Provasnik, S. and Dorfman, S. (2005). Mobility in the teacher workforce findings from the condition of education 2005. *National Center for Education Statistics*.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reeves, P. M., Pun, W. H., and Chung, K. S. (2017). Influence of teacher collaboration on job satisfaction and student achievement. *Teaching and Teacher Education*, 67:227–236.

Renbarger, R. and Davis, B. (2019). Mentors, self-efficacy, or professional development: Which mediate job satisfaction for new teachers? A regression examination. *Journal of Teacher Education and Educators*, 8(1):21–34.

Richter, D., Kunter, M., Lüdtke, O., Klusmann, U., Anders, Y., and Baumert, J. (2013). How different mentoring approaches affect beginning teachers' development in the first years of practice. *Teaching and Teacher Education*, 36:166–177.

Ronfeldt, M. and McQueen, K. (2017). Does new teacher induction really improve retention? *Journal of teacher education*, 68(4):394–410.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.

Saloviita, T. and Pakarinen, E. (2021). Teacher burnout explained: Teacher-, student-, and organisation-level variables. *Teaching and Teacher Education*, 97.

Sekhon, J. S. (2008). The Neyman-Rubin model of causal inference and estimation via matching methods. *The Oxford handbook of political methodology*, 2:1–32.

Shen, J. (1997). Teacher retention and attrition in public schools: Evidence from SASS91. *The Journal of Educational Research*, 91(2):81–88.

Sims, S. (2017). *TALIS 2013: Working Conditions, Teacher Job Satisfaction and Retention. Statistical Working Paper.* ERIC.

Skaalvik, E. M. and Skaalvik, S. (2015). Job satisfaction, stress and coping strategies in the teaching profession-what do teachers say? *International education studies*, 8(3):181–192.

Smith, T. M. and Ingersoll, R. M. (2004). What are the effects of induction and mentoring on beginning teacher turnover? *American educational research journal*, 41(3):681–714.

Sorensen, L. C. and Ladd, H. F. (2020). The hidden costs of teacher turnover. *AERA Open*, 6(1).

Splawa-Neyman, J., Dabrowska, D. M., and Speed, T. (1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science*, pages 465–472.

Stekhoven, D. J. and Bühlmann, P. (2012). missForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1):112–118.

Suk, Y., Kim, J.-S., and Kang, H. (2021). Hybridizing machine learning methods and finite mixture models for estimating heterogeneous treatment effects in latent classes. *Journal of Educational and Behavioral Statistics*, 46(3):323–347.

Toropova, A., Myrberg, E., and Johansson, S. (2021). Teacher job satisfaction: The importance of school working conditions and teacher characteristics. *Educational review*, 73(1):71–97.

Wang, K., Li, Y., Luo, W., and Zhang, S. (2020). Selected factors contributing to teacher job satisfaction: A quantitative investigation using 2013 TALIS data. *Leadership and Policy in Schools*, 19(3):512–532.

Wiggan, G., Smith, D., and Watson-Vandiver, M. J. (2021). The national teacher shortage, urban education and the cognitive sociology of labor. *The Urban Review*, 53(1):43–75.

Woessmann, L. and West, M. (2006). Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS. *European Economic Review*, 50(3):695–736.

Yoon, I. and Kim, M. (2022). Dynamic patterns of teachers' professional development participation and their relations with socio-demographic characteristics, teacher self-efficacy, and job satisfaction. *Teaching and Teacher Education*, 109.

Zhang, S., Shi, Q., and Lin, E. (2020). Professional development needs, support, and barriers: TALIS US new and veteran teachers' perspectives. *Professional Development in Education*, 46(3):440–453.