TEACHER QUALITY, TEST SCORES AND NON-COGNITIVE SKILLS: EVIDENCE FROM PRIMARY SCHOOL TEACHERS IN THE UK*

SARAH FLECHE
CEP, London School of Economics
s.fleche@lse.ac.uk

ABSTRACT

This paper estimates the importance of teachers in UK primary schools using matched pupil-teacher survey data. The results show that teachers are important inputs in pupil cognitive skills (measured by math test scores) as well as non-cognitive skills (measured by emotional health and social-behaviours). In addition, teacher ability to improve math test scores is weakly correlated with teacher ability to improve non-test score outcomes. I then decompose these measures of teacher effectiveness into different teacher characteristics. Teacher satisfaction and teaching practices (including class streaming, homework, types of incentives used, etc.) contribute more to explaining the variation in estimated teacher quality (up to 25%) than is explained by traditional observable characteristics such as teacher gender or experience.

Keywords: Teacher Quality, Test Scores, Non-Cognitive Skills, Mental Health, Teaching Practices.

JEL Classification Codes: I2; I31.

* I thank Clément Bosquet, Andrew Clark, Richard Layard, Nattavudh Powdthavee, Anna Vignoles and George Ward for useful comments and discussions. Support from the US National Institute on Aging (Grant R01AG040640), the John Templeton Foundation and the What Works Centre for Wellbeing is gratefully acknowledged.

I. INTRODUCTION

There is a general consensus that schooling produces both cognitive skills (reflected by test scores) and non-cognitive skills (e.g. good social behaviours and emotional health), both of which are important determinants of adult outcomes (Heckman and Rubinstein, 2001; Heckman *et al*, 2006; Cunha and Heckman, 2007; Borghans *et al*, 2008; Lindqvist and Vestman, 2011). Yet, most of the education literature has focused on test scores as measure of student skills. Therefore, evaluating schooling effects based on test scores may fail to capture schooling' overall effects and address only one dimension of what matters for child development and adult success.[1]

To address this issue, this paper extends this research by estimating the importance of teachers on both student cognitive and non-cognitive outcomes and tests whether teachers who improve test scores also improve non-test score outcomes. It then decomposes these measures of teacher effectiveness into observable characteristics of the instructors. The data used in this paper provide me with a key advantage: the ability to observe teacher emotional health (e.g. self-esteem, confidence or job satisfaction) as well as teaching practices in addition to teacher traditional characteristics (e.g. gender or experience). Policy makers and researchers agree that teachers are important components of the schooling environment (e.g. Rockoff, 2004; Rivkin *et al*, 2005; Hanushek *et al*, 2005; Aaronson *et al*, 2007; Kane and Staiger, 2008). However, traditional observable teacher characteristics are found to account for at most 10% of the variance in teacher effectiveness (Rivkin *et al*, 2005; Aaronson *et al*, 2007). This provides reason to suspect that other teacher characteristics and teaching practices matter for student cognitive outcomes and most likely for student non-cognitive outcomes. This paper presents one of the first analyses of teacher effectiveness on both student cognitive and non-cognitive outcomes[2] and is the first to investigate the effect of teacher emotional health and teaching practices on both test scores and student non-cognitive ability.[3]

This paper relies on an unusually rich matched pupil-teacher dataset (Avon Longitudinal Study of Parents and Children, ALSPAC) that follows three cohorts of pupils born in the early-1990s in Avon, UK.[4] These data have several advantages. First, they provide precise information on test scores, pupil

[1] Teacher may have important effects on long-run outcomes that are not reflected in their test score value-added (Chetty *et al*, 2014; Chamberlain, 2013).

[2] In existing work, Jackson (2014) finds that teacher have causal effects on test scores and non-test score outcomes measured by absences, suspensions, grades and on-time grade progression. However, he does not test for specific teacher characteristics that could predict teacher effects on cognitive and non-cognitive skills.

[3] In existing work, Hidalgo-Cabrillana and Lopez-Milan (2015) tests to what extent using certain teaching practices and materials in class is related to achievement in maths and reading. See also Schwerdt and Wupperman, 2011; Van Klaveren, 2011; Brewer and Goldhaber, 1997; Lavy, 2011; Bietenbeck, 2014. However, all these studies do not investigate the effects of teaching practices on student non-cognitive outcomes.

[4] Avon was a county in the West of England from 1974 to 1996. The area had a population of 903,870 in 1991 and included the cities of Bath and Bristol.

behavioural problems, pupil wellbeing, in years 3 and 6 of primary school as well as family background.[5] Because this is a birth cohort study and pupils are observed several years from birth, this allows to control for a full range of time-varying pupil characteristics and family background that could drive pupil outcomes. In addition, by controlling for previous pupil test scores and non-cognitive ability, it is possible to focus on changes in pupil outcomes in relation with teacher effects. Second, very precise information on school, classroom and teacher characteristics are available, including school type, school size, classroom composition, class size, teacher experience, teacher emotional health and teaching methods. In contrast, many other studies can only look at school-level characteristics and have access to few observable features. Because I am able to examine the pupil-teacher matches at the classroom level, I have more power to estimate teacher effects than is commonly available. Third, because this is a multi-cohort study, teachers are observed in multiple classrooms over time. Importantly, this allows to separately identify teacher fixed effects from idiosyncratic class effects. Finally, by focusing on variation in pupil outcomes within cohorts and grades (year 3 and year 6), my teacher quality estimates are less likely to be driven by cohort characteristics or grade-level specificities that are also correlated with pupil outcomes.

The results reveal that teachers are important inputs in pupil cognitive and non-cognitive outcomes in primary school. Namely, after controlling for initial ability and other pupil characteristics, teacher effects are statistically important in explaining math test score achievement and in reducing emotional issues or behavioural problems in years 3 and 6 of primary school. In addition, there are large differences in quality among teachers: a one standard deviation improvement in estimated teacher quality translates into an increase in math achievement equal to 0.51 standard deviations, an improvement in pupil emotional state equal to 0.47 standard deviations and an improvement in pupil social skills equal to 0.39 standard deviations. Furthermore, higher teacher quality in math test scores is weakly correlated with higher teacher effectiveness in improving pupil non-cognitive skills.

Consistent with earlier studies, I find that traditional measures of teacher characteristics – such as gender and experience – explain little of the variation in teacher quality on both cognitive and non-cognitive outcomes. In contrast, teacher emotional health (e.g. self-esteem, confidence or job satisfaction) and teaching practices (e.g. homework frequency, homework duration, class streaming, types of incentives used, assessments, etc.) account for between 20 and 25% of the total variation in estimated teacher quality. Hence the inclusion of teacher emotional health and teaching practices almost doubles the share of the variance in estimated teacher quality that is usually explained by traditional teacher characteristics.

[5] Year 3 is an educational year group in schools in UK. It is the third year of compulsory education and incorporates students aged between seven and nine. Year 6 incorporates students aged between ten and twelve.

These results are robust to several checks. First, they are unlikely to be driven by pupil sorting, that is, pupils may be purposely placed in certain schools/classrooms or with certain teachers based on their characteristics. Second, the individual teacher characteristics (e.g. emotional health or teaching practices) are relatively stable over time and do not vary with the characteristics of pupils in the class. Third, the estimates are robust to a range of alternative measures to capture pupil cognitive ability, emotional health and behavioural problems.

While my study focuses on only one area – Avon – in the 1990s, the population of parents and children living in the study area is broadly similar to those of the rest of Great Britain. A bit more than 65% of mothers in the Avon area lived in owner occupied accommodation in 1991, 72% were married and 5% were non-white. Similarly, in the rest of Britain, 63% of mothers lived in owner occupied accommodation in 1991, 72% were married and 8% were non-white (1991 Census). In addition, a comparison of the growth standards (weights and birth lengths) for the ALSPAC children with published national standards provide very similar measures. Therefore, on these dimensions, Avon was quite representative of the Great Britain population.[6]

This paper is organised as follows: Section II presents the data and Section III the empirical strategy. Section IV analyses the results and Section V concludes.

II. ALSPAC DATASET

The unique detail and scope of the ALSPAC data are major strengths of this study. The ALSPAC data are survey data completed by pupils, parents and primary school teachers, across three cohorts, born in the early 1990s. Pupils are linked with the UK national pupil database. These ALSPAC data include a large set of information on pupil characteristics, family background, life events, classroom, teacher characteristics, and school characteristics. I focus on the performance of pupils at the end of year 3 and year 6 of primary school. Appendix Table A1 provides an overview of the ALSPAC data.

2.1 Pupil characteristics

Several pupil characteristics are included: test scores, behaviour and emotional health and family and pupil background measures. In most of previous studies using administrative data, these information are somewhat limited. I here have detailed information on the entire history and family background that allows me to control for past (and present) pupil heterogeneity that could affect pupil school achievement: parental education, number of siblings, parental divorce, parental employment history,

[6] http://www.bristol.ac.uk/alspac/researchers/resources-available/cohort/represent/

parental support, parental emotional health and parental financial problems. Appendix Table A2 includes descriptive statistics for some of the variables available.

In particular, these data include a history of previous test scores that can be used as controls for past performance. In order to measure pupil achievement, I rely on pupil test scores from two math tests, administered by ALSPAC at the end of year 3 and the end of year 6 of primary school,[7] and from two national tests: Key Stage 1 (KS1) and Key Stage 2 (KS2),[8] administered in year 2 and year 6. I limit my analysis to math test scores. Although I have information on English test scores for KS1 and KS2, I choose to focus on math achievement to be able to control for previous test scores in years 3 and 6, respectively. As robustness check, however, Appendix Tables A3 and A4 provide results with English test scores instead of math test scores. Very similar findings are obtained.[9] Another argument is that math test scores seem to have more, or are often to have more predictive power than English scores for future productivity (see, e.g. Murnane *et al*, 1991; Grogger and Eide, 1995; Hanushek and Kimko, 2000).

Multiple math test scores are vital to control for previous cognitive ability of pupils. I rely on a general form of the value-added model of education production in which I regress math test scores in year 6 and end of year 3 on the variables of interest while controlling for initial achievement (hence at the end of year 3 and in year 2, respectively). I observe the two ALSPAC math tests and both KS1 and KS2 test scores for the majority of pupils which provides me with a sample size of roughly 7,000 pupils. Appendix Table A2 includes summary statistics for the math test scores available. Although I have information on two math test scores in year 6, I choose to focus on KS2 math test scores in year 6 as this is a standardised test in the British education system. Results that substitute ALSPAC math test scores in year 6 for KS2 math test scores are shown in the Appendix Tables A3 and A4. Again, similar findings are obtained, with a significant correlation of 0.67 between the two teacher quality estimates.

The key advantage of the ALSPAC data is that it also gives concomitant information on pupil behaviours and emotional states (in addition to math achievement) in years 3 and 6 of primary school. In particular, I rely on the Strength and Difficulties Questionnaire (SDQ) scores that gives a complete

[7] The ALSPAC math test scores are two tests of mathematical reasoning. The items in these tests require very simple arithmetic computations. The mathematical reasoning tasks include three types of items, additive reasoning about quantities, additive reasoning about relations, and multiplicative reasoning items. All items are presented orally with the support of pictures. The children's booklets, where they are asked to write their answers, contain no text, only drawings; the story is read by the teacher to the class. The assessments contains a total of 17 items in year 3 and 35 items in year 6. It is not timed; administration usually takes approximately 25-30 minutes.

[8] The Key Stage Assessments are two standardized tests of mathematical achievement, designed by the UK government and administered and scored by the teachers. One assessment, Key Stage 1 (KS1) is given to the pupils when they are in year 2. The second assessment, Key Stage 2 (KS2) is given to the pupils when they are in year 6. Both KS tests measure a variety of aspects of mathematics and are seen as valid measures of mathematical achievement because of the role that they play in the British education system.

[9] See Section IV (4.4 robustness checks) for a detailed description of the results.

behavioural screening in the following areas: conduct problems, hyperactivity and inattention, emotional symptoms, peer relationship problems and pro-social behaviour (Goodman, 1997). The questionnaires includes 25 items in total. This includes information about "whether the pupil is restless", "overactive", "cannot stay still for long", "considerate of other people's feelings", "would rather be alone than with other youth", "is helpful if someone is hurt", "upset or feeling ill", "has at least one good friend", "often fights with other youth or bullies them", "is often unhappy", "depressed or tearful", "often lies or cheats", "has good attention span" and "saws tasks through to the end". Appendix Table A5 provides a detailed description of the SDQ questionnaire.

Following Goodman *et al.* (2010), I combine the SDQ's emotional and peer subscales into an "internalising" subscale (SDQ INT see after) and the SDQ's behavioural and hyperactivity subscales into an "externalising" subscale (SDQ EXT see after). This provides me with two composite measures on whether pupil has emotional issues and whether pupil has behavioural problems on a 0-20 scale. For robustness checks, it is also possible to run the main analyses using the SDQ scales separately (Appendix Tables A6 and A7). A key advantage of the ALSPAC data is that the SDQ questionnaires were completed by parents and teachers. Hence, instead of using one source of information, it is possible to estimate the relationship between teacher effectiveness and pupil outcomes by measuring pupil outcomes according to the perspective of teachers but also to the perspective of parents. This is of particular interest with such subjective data. Information reported by teachers and parents have different advantages and disadvantages. Teachers' reports about pupil emotional health and behaviours are useful because they provide information on pupil in-class outcomes that might differ from what parents perceive at home. On the other side, teacher responses are also subject to bias. Teachers may answer about pupil SDQ based on their own mental state or influenced by the class context.

Appendix Tables A8 and A9 test whether the main results are robust to teachers' and parents' reports.[10] In addition, Table 1 contains the mean and standard deviation of pupil outcomes and teachers' and parents' answers for SDQs INT and EXT. Average SDQ INT and SDQ EXT reported by parents and teachers are similar for the full sample: 17 out of 20 for SDQ INT and 16 out of 20 for SDQ EXT. However, Table 1 reveals that parents' and teachers' answers are not very much correlated: The coefficient of correlation is 0.28 for SDQ INT and 0.26 for SDQ EXT.[11] This suggests that using both parents' and teachers' responses with different potential reporting bias can provide a broader picture and improve our understanding of the role of teachers on pupil outcomes.

[10] See Section IV (4.4 Robustness checks) for a full description of the results.

[11] To gain further insight about to what extent teachers' answers differ from parents' ones, I calculate the correlation within each class. The standard deviation of the within-class coefficients of correlation is about 0.30 for SDQ INT and 0.25 for SDQ EXT. This means that there are large variations in the extent to which teachers' answers differ from parents' ones across classes. In some classes, teachers' and parents' answers match each other quite well while in other classes this is not the case.

Table 1: Descriptive Statistics and Correlation Matrices of Pupil Outcomes

	Math test scores	SDQ INT – parents assessed	SDQ INT – teacher assessed	SDQ EXT – parents assessed	SDQ EXT – teacher assessed
Math test scores	1.00				
SDQ INT – parents assessed	0.16	1.00			
SDQ INT – teacher assessed	0.22	0.33	1.00		
SDQ EXT – parents assessed	0.27	0.39	0.20	1.00	
SDQ EXT – teacher assessed	0.33	0.15	0.37	0.44	1.00
N	7,651	5,562	10,381	5,554	10,344
Mean	62.3	17.3	17.4	15.6	16.6
SD	20.4	2.74	3.23	3.24	3.94
Min	0	4	1	1	0
Max	100	20	20	20	20

Notes: SDQ INT is a composite measure on pupil emotional health and combines the SDQ's emotional and peer subscales into an "internalising" subscale. SDQ EXT is a composite measure on pupil social behaviour and combines the SDQ's behavioural and hyperactivity subscales into an "externalising" subscale. See Appendix Table A5 for a detailed description of the SDQ INT and SDQ EXT measures. Math test scores are taken from two ALSPAC math test scores given in year 3 and in year 6 and KS1 and KS2 math test scores given in year 2 and in year 6. All math test scores have been rescaled from 0 to 100, for comparison purposes.

2.2 School, classroom and teachers characteristics

Another important feature of the ALSPAC data is the detailed information on school, classroom and teacher characteristics that rarely appear together in other studies. It allows to disentangle the importance of school, classroom and teacher on pupil outcomes. The ALSPAC data include the type of school, school size, school admission policy, frequency of the staff meetings, head teacher gender but also class size, number of exclusions in class, percentage of free school meal pupils in the class, percentage of SEN statemented[12] pupils in the class, percentage of pupils with home concerning problems in class, percentage of pupils for whom English is not the first language, and class age composition.

In addition, this is the first study, to the best of the author's knowledge, that uses such detailed information on primary school teachers including teacher's gender, experience at school, experience everywhere, year of certification, but also teacher's Crown-Crisp Experiential Index (CCEI), teacher Bachman self-esteem, job satisfaction, confidence in teaching and teaching style (homework, incentives, ability groups, activity groups, type of assessments, etc.). Teacher CCEI is a sum of 23 items from the ALSPAC questionnaire which captures whether the "teacher feels upset for no obvious reason", "teacher feels like life is too much effort", "teacher feels uneasy and restless", "teacher has long periods of sadness", "teacher loses ability to feel sympathy", "teacher worries a lot", etc. The

[12] Special education needs (SEN) that affect a child's ability to learn can include their behaviour or ability to socialise, reading and writing (e.g. they have dyslexia), ability to understand things, concentration levels (e.g. they have Attention Deficit Hyperactivity Disorder), physical needs or impairments.

Bachman score of self-esteem consists in a sum of 11 items and measures whether "teacher feels to be a person of worth", "teacher feels to have a number of good qualities", "teacher is a useful person to have around", "teacher does job well", "teacher feels unlucky", "teacher feels their life is not usual". Appendix Table A5 provides a full description of teacher CCEI and teacher Bachman self-esteem. Teacher job satisfaction and teacher confidence in teaching are drawn from the following questions: "Teacher really enjoys teaching (from 1 to 5)" and "teacher's confidence in teaching numeracy (from 1 to 3)". Hence these data allow me to decompose the teacher fixed effect into teacher cognitive and non-cognitive factors and better understand how teacher quality matters for pupil outcomes.

Pupils are assigned to a class and to a teacher at the beginning of the academic year and continue with the same classmates and teacher until the end of the academic year. Note, in addition, that pupils have the same teacher and classmates for the entire school day. In order to identify teacher fixed effects, I construct a teacher identifier based on teacher's gender, experience, year of qualification and school attendance – knowing that a teacher has only one class a year. Appendix Table A10 includes descriptive statistics for the teacher variables available. The teacher file contains 716 teachers in 230 primary schools in year 3 and year 6. There are on average 3 to 4 teachers per schools, which limits the possibility of teacher misidentification. In addition, I assume that if teachers move between schools, they are assigned to different identifiers. 80% of teachers are women, with approximately 15 years of experience. Because this is a multi-cohort dataset, 56% of teachers are observed once, 34% of teachers are observed twice and 10% are observed three times. The average number of pupils observed per teacher is 15.

III. EMPIRICAL STRATEGY

The aim of this paper is to estimate the importance of teacher effects for the production of pupil cognitive and non-cognitive skills. To separate teacher effects from school, classroom and pupil characteristics, I follow a two-step procedure.

3.1 First step

The first step consists in estimating the effect of teachers on pupil outcomes. In the standard education production function, achievement, Y_{ijckt}, of pupil i with teacher j in school k class c at time t is expressed as a cumulative function of own characteristics X and family characteristics F, from age 0 to the current age as well as school, classroom and teacher characteristics. In particular, previous studies that estimate the effect of schooling on pupil cognitive skills rely on value-added models to

control for initial achievement (Hanushek, 1979; Hanushek, 1986; Chetty *et al*, 2014; Krueger, 1999; Rivkin *et al*, 2005; Rockoff, 2004; Aaronson *et al*, 2007). The identifying assumption is that lagged test scores account for the cumulative inputs of prior years.

I adopt the same empirical strategy and estimate the effect of teachers on pupil outcomes using a value-model of the general form:

$$Y_{ijckt} = \rho Y_{ijckt-1} + X_{it}\,\beta + F_i\,\alpha + S_{kt}\,\mu + C_{ct}\,\theta + T_{j\tau}\,\gamma + \lambda_j + \eta_\tau + \varepsilon_{ijck\tau t} \qquad (1)^{13}$$

where Y_{ikcjt} refers to the math test score, SDQ INT or SDQ EXT of pupil *i*, who is enrolled in year *t* (t=3 or t=6) at school *k*, class *c* with teacher *j*. λ_j is a teacher *j* fixed effect. $T_{j\tau}$ includes a non-linear function of teacher experience,[14] length of time teacher *j* taught pupil *i*, and time spent by teacher *j* teaching numeracy. X_{it} is a vector of pupil characteristics, and $Y_{ijckt-1}$ is the lagged dependent variables. F_i records family background. S_{kt} and C_{ct} represent vectors of time-varying school and class characteristics (including school size, frequency of staff meetings, head teacher gender, class size, class age composition, percentage of pupils entitled to free school meals, percentage of SEN statement pupils in class, percentage of pupils for whom English is not the first language and percentage of pupils with home concerning problems). η_τ controls for year started school (cohort τ fixed effects).

Because this is a birth cohort study and pupils are observed several years from birth, this is possible to control for time-varying pupil characteristics, including lagged dependent variables, as well as family background. In addition, because teachers are observed in multiple classrooms over time, this is possible to estimate teacher fixed effects and separately identify teacher effects from idiosyncratic class effects. Finally, the cohort fixed effects allow to control for cohort characteristics that could drive pupil outcomes.

The coefficients reported in subsequent tables are partial correlation coefficients (or β-statistics). They reflect the "power" of each variable to explain the prevalence of cognitive and non-cognitive skills of pupils, holding all other variables in the equation constant. They, therefore, reflect the impact of the variable times its standard deviation.

[13] $i \in \tau, c \in k, j \in k, j \in t$. Each pupil belongs to one cohort, each class to one school, each teacher to one school and teachers are observed either in year 3 or in year 6, but not both.

[14] Teaching experience is measured by a categorical variable equals to 1 if teacher experience (anywhere) is less than 1 year, 2 if teacher experience is between 1 and 2 years, 3 if teacher experience is between 3 and 9 years and 4 if teacher experience is more than 10 years.

This specification has several advantages: first, the value-added model controls for the fact that teachers may be assigned to pupils with different initial ability. Second, including a substantial list of observable pupil and family traits that may be correlated with cognitive, emotional or behavioural changes allows to control for "non-school" factors that may account for differences in teacher effectiveness. Third, school and classroom time-varying characteristics help to separate the teacher effects from school and classroom characteristics that co-vary with individual teacher quality. Fourth, controlling for individual teacher experience, length of time teacher taught pupil and time spent by teacher teaching numeracy allows to separate teacher time-varying characteristics from fixed teacher quality.

This first step also allows to evaluate the respective explanatory power of pupil, school, classroom and characteristics as well as teacher fixed effects and time-varying characteristics in the production of pupil cognitive and non-cognitive skills. Furthermore, the importance of fixed teacher quality can be measured by the variation in the estimated teacher fixed effects. For example, one might measure the expected rise in math test score or SDQ INT/EXT for moving up one standard deviation in the distribution of teacher fixed effects.

One potential concern with this first step, however, is that the teacher fixed effect estimates will be biased when they are based on small populations and hence will suffer from measurement errors (Kane and Staiger, 2002; Aaronson *et al*, 2007). In order to correct for this bias, I analytically adjust the estimates of λ_j by using the standard errors of λ_j and compute the standard deviation of teacher fixed effects with inverse standard error probability weights. I report the unadjusted and adjusted standard deviations in subsequent tables.[15] For statistical reliability, I also preliminarily raise the minimum number of pupils to identify an individual teacher to 5.

3.2 Second step

The second step estimation examines whether teacher quality (measured by the estimated teacher fixed effects from equation (1)) can be explained by demographic and other teacher characteristics. In order to do so, I estimate the following specification:

$$\hat{\lambda}_j = a + T_j\, b + u_j \qquad (2)$$

[15] Aaronson *et al*. (2007) finds that roughly 30% of the standard deviation in estimated teacher quality is due to sampling error.

where $\hat{\lambda}_j$ refers to the estimated teacher j fixed effect. T_j is a vector of teacher invariant characteristics which includes gender, potentially average experience, CCEI, Bachman self-esteem, confidence in teaching, job satisfaction, but also teaching practices such as homework frequency, homework duration, type of assessments, incentives used, class streaming, etc.

Because I cannot observe teachers who switch schools in this dataset, I do not include school fixed effects in equation (1). However, it is possible that some of teacher quality might be attributed to the school. For this reason, I present estimations of the second step procedure with and without school fixed effects.

There are several advantages in using the two-step procedure.[16] First, including teacher dummies in the first step allows for a more general specification than the one that could be made by considering different teacher characteristics. Second, the first step estimates of the coefficients for school and classroom characteristics and for pupil characteristics are independent from the specification chosen for the teacher characteristics effects in the second step. Changing the specification of the second step does not affect estimates from the first step. Third, the two-step procedure allows to consider both individual and aggregate error terms, which deals with heteroscedasticity issues raised by Moulton (1990).

There are still some endogeneity issues. One legitimate concern in estimating teacher effects is the possibility of school/classroom sorting process, that is, pupils may be purposely placed into certain schools/classrooms or with certain instructors based on their learning potential or behavioural characteristics.[17] This issue here is however somewhat limited since I control for an extended set of pupil, classroom, school and teacher characteristics. To assess the extent to which pupils may be sorted, I report in first step regressions the correlation between the effect of pupil characteristics and the teacher fixed effects (Appendix Table A16). I also report the correlation between the effect of school and classroom characteristics and the teacher fixed effects.

Endogeneity biases can also be present in the second step estimation. For instance, teacher emotional health or teaching practice choices are endogenous to teacher quality. High quality teachers are more likely to be satisfied with their job or to choose certain teaching methods than others. One way to deal with this would consist in using natural experiments, with exogenous changes in school policy on teaching methods. I am not aware of any paper in economics that proposes such analysis. Another way to deal with this would be to instrument teacher emotional health using exogenous life events.

[16] See Bosquet and Combes (2015) for a detailed description of the two-step procedure, its advantages and its disadvantages.

[17] Clotfelter, Ladd and Vigdor (2007) examines how the sorting of teachers and students affects estimates of teacher effectiveness. They results suggest that the bias from between-school sorting is large; the bias associated with sorting within schools, by contrast, is more limited. In addition, two characteristics – teacher experience and licensure test scores – emerge as robust determinants of test scores for fifth grade students in the US.

However, the ALSPAC data do not provide such information on teachers. There is probably no way to reject such concerns definitively, but one test is to examine whether individual teacher characteristics (e.g. emotional health or teaching practices) are relatively stable over time and do not vary with the characteristics of pupils in the class (Appendix Table A19). As a variant, I also present estimations for the second step procedure using lagged teacher characteristics (Appendix Tables A17 and A18).

Lastly, as with any large cohort survey, the usual attrition due to dropouts applies. The participated parents did not always answer every single question in every questionnaire, which means that the sample size may vary across different regression equations. In each analysis, I include only observations for which there are replies to the dependent variable, sample sizes being reported in each table.

IV. RESULTS

4.1 The role of teachers

I begin with an overview of the explanatory power of school, classroom and teacher effects in explaining pupil cognitive and non-cognitive skills. More specifically, Table 2 records the R-squared values from a series of regressions of the different dependent variables (Math test scores; SDQ INT; SDQ EXT) on school, classroom characteristics and teacher dummies.[18] The first column for each dependent variable is based on a specification with only pupil characteristics and lagged dependent variable; the second column adds school and classroom characteristics; the third column adds teacher fixed effects; and the final column employs school rather than teacher fixed effects.

The results show quite clearly that teacher fixed effects are significant predictors of pupil math test scores, SDQ INT and SDQ EXT in year 3 and year 6 of primary school. The *p* values for F tests of the joint significance of teacher fixed effects all fall below 0.01. Comparing columns (2) and (3), the inclusion of teacher fixed effects increases the explanatory power by 7 percentage points for math test scores, 18 percentage points for SDQ INT and 12 percentage points for SDQ EXT. Pupil, school and classroom characteristics explain 34% of the variation in math test scores, 10% of the variation in SDQ INT and 24% of the variation in SDQ EXT. Interestingly, the inclusion of school rather than teacher fixed effects reduces the explanatory power by 7 percentage points for math test scores, 9 percentage points for SDQ INT and 7 percentage points for SDQ EXT. This suggests that much of the variation in teacher quality exists within rather than between schools.

[18] Full specifications are shown in Appendix Table A11.

Table 2: Pupil Outcomes and Teacher Quality

	(1)	(2)	(3)	(4)
Included explanatory variables:				
Pupil covariates	Yes	Yes	Yes	Yes
School & classroom characteristics	No	Yes	Yes	No
Teacher fixed effects	No	No	Yes	No
F tests, H0:	--	--	(<0.01)	--
School fixed effects	No	No	No	Yes
F tests, H0:	--	--	--	(<0.01)
	Math Test Scores			
R-squared	0.341	0.353	0.471	0.400
Observations	7,651			
	SDQ INT – Teacher Assessed			
R-squared	0.101	0.117	0.287	0.176
Observations	10,381			
	SDQ EXT – Teacher Assessed			
R-squared	0.237	0.244	0.351	0.286
Observations	10,344			

Notes: All regressions include pupil characteristics (including lagged dependent variable), family background and cohort fixed effects. Column (2) adds school and classroom characteristics. Column (3) adds teacher fixed effects and column (4) substitutes school to teacher fixed effects. Only R-squareds and number of observations are reported in each column. Numbers in parentheses are *p* values from F tests of the joint significance of teacher fixed effects and school fixed effects respectively. All three outcomes (math test scores, SDQ INT and SDQ EXT) are measured in years 3 and 6. SDQ INT measures pupil emotional health and SDQ EXT measures pupil social behaviours.

Table 6 presents details on the distribution of teacher fixed effects, specifically the standard deviation and the 10th, 25th, 50th, 75th, and 90th percentiles.[19] These are expressed in standard deviation on the sample distribution of math test scores, SDQ INT and SDQ EXT in year 3 and year 6. For all dependent variables, the raw standard deviation is quite broad, indicating that there are large variations in teacher quality and that teacher quality has a large impact on pupil outcomes. Moving one standard deviation up the distribution of teacher fixed effects is expected to raise math test scores by about 0.55 standard deviations, SDQ INT by about 0.53 standard deviations and SDQ EXT by about 0.42 standard deviations. Furthermore, the gaps between the 90th percentile and 10th percentile teacher is between 1.05 and 1.34. This means that having a teacher at the 90th percentile of the quality distribution versus the 10th percentile is associated with 27 points higher score in math on a 0-100 scale, and 4 points higher in emotional health and in social skills on a 0-20 scale.[20]

For all three outcomes (math test scores, SDQ INT and SDQ EXT) the adjusted standard deviation is lower than the raw standard deviation. For math test scores, the adjusted measure is reduced by 6

[19] Appendix Figure 1 also plots the sample distribution of the teacher fixed effect estimates for all three outcomes.

[20] This is 90-10 gap*standard deviation. 27=1.34*2.01; 4=1.32*3.23; 4=1.05*3.94.

percent; for SDQ INT by 11 percent; and for SDQ EXT by 7 percent.[21] However, the adjusted measures still imply that teacher quality has a large impact on pupil outcomes. Moving one standard deviation up the distribution of teacher fixed effects is expected to raise math test scores by about 0.51 standard deviations, SDQ INT by about 0.47 standard deviations and SDQ EXT by about 0.39 standard deviations.

These estimates of teacher effectiveness for math test scores are a bit higher than those reported in Rockoff (2004), Rivkin *et al.* (2005) and Aaronson *et al.* (2007). Rockoff (2004) reports a 0.1 standard deviation gain from a one standard deviation increase in teacher quality from two New Jersey surbuban school districts. Rivkin *et al.* (2005) lower bound estimates suggest that a one standard deviation increase in teacher quality increases student achievement by at least 0.11 standard deviations. In Aaronson *et al.* (2007), a one standard deviation increase in teacher quality over a full year implies about a 0.15 standard deviation increase in math test score gains. This difference is probably due to the fact that these studies report within-school estimates, which is not the case here.

Table 3: Distribution of the Teacher Fixed Effects

	Math test scores		SDQ INT – teacher assessed		SDQ EXT – teacher assessed	
	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
10th percentile	-0.677	-0.594	-0.680	-0.453	-0.518	-0.397
25th percentile	-0.335	-0.231	-0.316	-0.139	-0.263	-0.182
50th percentile	0.017	0.092	0..029	0.168	-0.009	0.051
75th percentile	0.348	0.383	0.366	0.460	0.264	0.324
90th percentile	0.661	0.661	0.637	0.743	0.536	0.557
90-10 gap	1.338	1.255	1.317	1.196	1.054	0.954
75-25 gap	0.683	0.614	0.682	0.599	0.527	0.506
SD	0.546	0.511	0.534	0.474	0.417	0.388
Nb of teachers	623		714		714	

Notes: Teacher fixed effects are estimated in regressions that include controls for teacher time-varying experience, length of time teacher taught pupil, and time spent teaching numeracy, class time-varying characteristics, school time-varying characteristics, pupil characteristics, family background, school cohort effects, and lagged pupil dependent variables (lagged math test scores, lagged SDQ INT, and lagged SDQ EXT respectively). Adjusted means that the standard deviations of the estimated teacher fixed effects are computing using inverse standard errors probability weights. All three outcomes (math test scores, SDQ INT and SDQ EXT) are measured in years 3 and 6. SDQ INT measures pupil emotional health and SDQ EXT measures pupil social behaviours.

[21] This is (Unadjusted SD – Adjusted SD)/Unadjusted SD. 6% = (0.546-0.511) / 0.546; 11% = (0.534-0.474) / (0.534); 7% = (0.417-0.388) / 0.417.

4.2 Relationships between teacher fixed effects estimates

These results indicate that teachers have sizeable and economically meaningful effects on math test scores but also on pupil non-cognitive outcomes. Table 4 examines the relationships between these estimated teacher effects, that is, to test whether teachers who improve test scores also improve non-test scores outcomes.

More specifically, I look at the correlation between the teacher fixed effects obtained from the regression on math test scores and the teacher fixed effects obtained from the regressions on SDQ INT and SDQ EXT. Interestingly, the results indicate that the relationships are weak. The correlation coefficients are negative and close to zero (-0.05 and -0.13) between teacher fixed effects on test scores and teacher fixed effects on pupil emotional health and social-behaviours. In contrast, the correlation is positive and statistically significant (0.51) between teacher effects on pupil emotional health (SDQ INT) and teacher effects on pupil social-behaviours (SDQ EXT). Overall, this indicates that teachers who raise test scores are not associated with better non-test score outcomes, and *vice versa*.[22]

Table 4: Correlation between the Teacher Fixed Effects Estimates

	Teacher FE – math test scores	Teacher FE – SDQ INT	Teacher FE – SDQ EXT
Teacher FE – math test Scores	1.00		
Teacher FE – SDQ INT	-0.13*** (0.04)	1.00	
Teacher FE – SDQ EXT	-0.05 (0.05)	0.51*** (0.03)	1.00

Notes: Bootstrapped standard errors are in parentheses. Teacher fixed effects are estimated in regressions that include controls for teacher time-varying experience, length of time teacher taught pupil, and time spent teaching numeracy, class time-varying characteristics, school time-varying characteristics, pupil characteristics, family background, school cohort effects, and lagged pupil dependent variables (lagged math test scores, lagged SDQ INT, and lagged SDQ EXT respectively). All three outcomes (math test scores, SDQ INT and SDQ EXT) are measured in years 3 and 6. SDQ INT measures pupil emotional health and SDQ EXT measures pupil social behaviours. ***, **, * significant at the 1%, 5% and 10% respectively.

If teacher fixed effects on test scores are weak predictors of effects on non-test scores outcomes, this may suggest that teacher test score fixed effects measure certain skills and teacher fixed effects on the non-cognitive skills measure a largely different (but potentially important) set of skills. The next section directly asks to what extent teacher characteristics, such as gender, teacher experience, teacher emotional state or teaching practices, can explain the estimated contribution of teachers to the variance of pupil cognitive and non-cognitive outcomes. Because the ALSPAC data provides several

[22] This is consistent with Jackson (2014) who finds that teacher effects on test scores explain little of the estimated effects on non-test score outcomes.

measures of teacher characteristics, it is possible to relate estimates of teacher fixed effects to these observable characteristics.

4.3 Explaining teacher fixed effects

Table 5 reports the R-squareds of different teacher characteristics in explaining teacher ability to improve math test scores, SDQ INT and SDQ EXT. All the teacher quality estimates are based on the full specification of equation (1).

4.3.1 Teacher gender and experience

First and foremost, traditional observable characteristics – such as gender and experience – explain at most 2% of the total variation in teacher quality (based on all three pupil outcomes). This is consistent with previous studies (Hanushek 1991; Rivkin *et al.* 2005; Aaronson *et al.* 2007) that find a small relationship between teacher characteristics such as gender, experience, educational background and teacher ability to raise student achievement.[23]

In addition, Table 6 provides detailed information on the effects of teacher gender. What is notable is that female teachers are associated with lower math test scores than male teachers while male teachers are associated with lower SDQ INT and SDQ EXT achievement than female teachers. Male teachers increase math test scores by 0.9 standards deviations compared to female teachers.[24] In contrast, female teachers increase SDQ INT by 0.7 standard deviations and SDQ EXT by 0.12 standard deviations compared to male teachers. The coefficient on teacher experience is not statistically significant for all three outcomes.

I also investigate whether teacher gender affects pupil outcomes differently by pupil gender. A brief analysis, including an interaction term in equation (1) between teacher and pupil genders (Appendix Table A12) shows that having a female teacher increases more SDQ EXT outcomes for male pupils compared to female pupils. In contrast, the effect is not significantly different for math test scores and

[23] Hanushek (1971) finds no relationship between teacher quality and experience or master's degree attainment. Rivkin *et al.* (2005) also find no link between education level and teacher quality, although they find a small positive relationship between the first two years of teacher experience and teacher quality. Aaronson *et al.* (2007) find that the vast majority of the total variation in teacher quality is unexplained by observable teacher characteristics, such as gender, ethnicity, experience, advanced degrees and teaching certifications.

[24] See Dee (2005) and Ehrenberg et al. (1995) for a discussion on the influence of teachers' race, gender and ethnicity. However, they mostly focus on how pairings by race, ethnicity and gender influence teachers' perceptions and expectations of students. The evidence is mixed.

SDQ INT. Interestingly, this suggests that how teacher's gender influences pupil behavioural outcomes depends upon the match or the mismatch of teacher and pupil gender.

4.3.2 Teacher emotional health

Given (i) the sizeable effects of teachers on both pupil cognitive and non-cognitive skills and (ii) the limited amount of variation in teacher effectiveness explained by simple characteristics such as gender and experience, a key question is whether other teacher characteristics predict teacher effectiveness and whether these characteristics explain differently teacher effects on test scores and non-test scores outcomes. One potential source of investigation is to examine the influence of teacher emotional state and job satisfaction on pupil test scores, pupil emotional health and pupil behaviours. A rich body of research in economics and psychology shows that individuals' non-cognitive skills and traits – e.g. perseverance, self-control, self-esteem, motivational ability – significantly influence academic and labour market outcomes (Duckworth and Seligman, 2005; Heckman and Rubinstein, 2001). Hence, teacher emotional state and personality traits may exert a significance influence on teacher quality. As of yet, however, rigorous quantitative evidence regarding the effect of teacher non-cognitive characteristics is limited.

I test this hypothesis in Table 5, by estimating the effects of several teacher non-cognitive skills – including CCEI, Bachman self-esteem, job satisfaction and confidence in teaching – on teacher ability to improve pupil outcomes. The results indicate that teacher emotional characteristics are significantly related to estimated teacher quality for the three observable pupil outcomes (math test scores, SDQ INT and SDQ EXT). Strikingly, they explain roughly 3.5 % of the total variation in teacher quality in maths, 3.8 % of the total variation in teacher quality in SDQ INT and 3.1% of the total variation in teacher quality in SDQ EXT. Even if these percentages might seem modest, this is three times the explanatory power of teacher gender and teacher experience for all three outcomes.

In Table 6, I then detail the different effects of these teacher emotional characteristics on teacher quality, running separate regressions.[25, 26] According to column 1 (teacher quality based on math test scores), a one standard deviation increase in teacher emotional health (as measured by CCEI) translates into an increase in teacher quality by 0.11 standard deviations. Similarly a one standard deviation increase in teacher job satisfaction translates into an increase in teacher quality by 0.06 a standard deviations. And a one standard deviation increase in teacher confidence in teaching translates

[25] Note, again, that the coefficients are partial correlation coefficients (or β-statistics). They reflect the "power" of each variable to explain the prevalence of cognitive and non-cognitive skills of pupils, holding all other variables in the equation constant. They therefore reflect the impact of the variable times its standard deviation.
[26] I run separate regressions because teacher emotional characteristics are likely to be correlated (Appendix Table A13). For robustness check, however, I present results in which all teacher emotional characteristics are included at once (Appendix Table A14).

into an increase in teacher quality by 0.13 standard deviations. These relationship are all statistically significant.

Table 5: Share of the Variance in Teacher Fixed Effects Explained by Teacher Characteristics (R^2)

	Teacher FE – math test scores	Teacher FE – SDQ INT	Teacher FE – SDQ EXT
(1) = Teacher gender + experience	1.0 %	1.3 %	2.1 %
(2) = Teacher mental health	3.5 %	3.8 %	3.1 %
(3) = Teaching practices	21 %	14 %	22 %
(4) = (1) + (2)	4.2 %	5.4 %	5.5 %
(5) = (1) + (3)	21 %	15 %	22 %
(6) = (2) + (3)	22 %	18 %	24 %
(7) = (1) + (2) + (3)	22 %	19 %	25 %
(8) = (1) + (2) + (3) + School FE	61 %	53 %	58 %

Notes: Mental Health includes teacher CCEI, Bachman Self-Esteem, job satisfaction, teaching confidence. Teaching practices include homework frequency, homework type, homework duration, the type of assessments, the type of incentives, any class activity groups and any class ability groups. The teacher fixed effects (FE) are estimated in regressions that include controls for length of time teacher taught pupil, and time spent teaching numeracy, class time-varying characteristics, school time-varying characteristics, pupil characteristics, family background, school cohort effects, and lagged pupil dependent variables (lagged math test scores, lagged SDQ INT, and lagged SDQ EXT respectively). Only R-squareds from second step regressions in which teacher fixed effects are decomposed into different teacher characteristics are reported. The second step regressions include teacher gender and experience in row (1). Row (2) includes only teacher mental health characteristics. Row (3) includes only teaching practices. Row (4) includes teacher gender, experience and mental health. Row (5) includes teacher gender, experience and teaching practices. Row (6) includes teacher mental health and teaching practices. Row (7) includes rows (1), (2) and (3) variables and row (8) adds school fixed effects on top of row (7) variables. All three outcomes (math test scores, SDQ INT and SDQ EXT) are measured in years 3 and 6. SDQ INT measures pupil emotional health and SDQ EXT measures pupil social behaviours.

Columns (2) and (3) also indicate that teacher emotional health (as measured by CCEI) is a significant driver of teacher ability to improve pupil SDQ INT and pupil SDQ EXT. As such, a one standard deviation increase in teacher CCEI translates into an increase in teacher quality (as measured by SDQ INT) by 0.17 standard deviations and an increase in teacher quality (as measured by SDQ EXT) by 0.15 standard deviations. I also identify teacher emotional health effects from a specification that includes all these teacher emotional characteristics at once (see Appendix Table A14). The results show that a large and statistically important effect comes from teacher CCEI.

These are striking findings, showing that teacher emotional health is a significant driver of teacher quality. I believe that these results are a first piece of evidence that teacher emotional characteristics matter and go well beyond what has been done in past estimations that have tried to explain the variation in teacher quality. Given the lack of explanatory power of traditional observable characteristics, it is of particular interest that teacher emotional health contributes more to explaining the variation in estimated teacher quality than is explained by teacher gender or experience.

Table 6: Determinants of the Teacher Fixed Effects Estimates – Teacher Mental Health

	Teacher FE – math test scores	Teacher FE – SDQ INT	Teacher FE – SDQ EXT
Teacher gender: female	-0.09**	0.07*	0.12***
	(0.034)	(0.036)	(0.038)
Teacher CCEI	0.11**	0.17***	0.15***
	(0.041)	(0.033)	(0.037)
Teacher self-esteem	0.06***	0.03	-0.02
	(0.019)	(0.061)	(0.038)
Teacher job satisfaction	0.06*	0.00	0.00
	(0.038)	(0.035)	(0.038)
Teaching confidence	0.13***	-0.04	-0.07*
	(0.041)	(0.035)	(0.037)
Observations	623	714	714

Notes: All rows are estimates from separate regressions that include teacher gender + alternatively teacher CCEI or Bachman self-esteem, or job satisfaction or confidence in teaching. The dependent variable is the teacher fixed effects (FE) estimates coming from regressions that include controls for time-varying teacher experience, length of time teacher taught pupil, and time spent teaching numeracy, class time-varying characteristics, school time-varying characteristics, pupil characteristics, family background, school cohort effects, and lagged pupil dependent variables (lagged math test scores, lagged SDQ INT, and lagged SDQ EXT respectively). All three outcomes (math test scores, SDQ INT and SDQ EXT) are measured in years 3 and 6. SDQ INT measures pupil emotional health and SDQ EXT measures pupil social behaviours. Teacher CCEI stands for teacher Crown-Crisp Experiential Index. See Appendix Table A5 for a full description of CCEI and Bachman Self-Esteem. ***, **, * significant at the 1%, 5% and 10% respectively.

4.3.3 Teaching practices

Another line of research to explain teacher effectiveness is to shift the focus to teaching practices, that is, what teachers actually do in the classroom. Again, the evidence on teaching practices is still scare and not conclusive. Even more on pupil non-cognitive skills.

To analyse to what extent using certain teaching practices in class is related to teacher quality and pupil performance in math and non-test score outcomes, I consider different teaching practices – including homework duration (up to 20 minutes: Yes/No), homework frequency (at least once a week: Yes/No), homework type (mostly reading: Yes/No), the type of assessments (always mark written tests: Yes/No), and whether pupils are engaged in any class ability groups (class streaming). A complete set of information is available in year 3, less information are available in year 6.

Table 5 reports the R-squared values from estimating the effect of teacher practices on teacher ability to improve pupil outcomes. The results indicate that including both teacher emotional state and teaching practices explain up to 22 % of the total variation in teacher effects on math test scores; 18% of the variation in teacher effects on SDQ INT scores and 24% of the variation in teacher effects on

SDQ EXT scores. Again this is much larger than the explanatory power attributed to traditional observable teacher characteristics – such as gender or experience.

Some of teacher quality might also be attributed to the school because I did not include school fixed effect in the first step estimations. In order to test for this, I identify teacher characteristics effects from a specification that includes school fixed effects and which allows me to use within-school variation. Including school fixed effects, the R-squared goes up to 61%; 53% and 58% respectively.

Table 7 presents detailed results on the effect of teaching practices. They clearly show that certain teaching practices are correlated with higher teacher quality, but can have different effects on pupil math test scores achievement and pupil emotional health/social behaviours. According to column 1, class ability groups decrease teacher quality in math test scores by 0.09 standard deviations.[27] At the same time, I find suggestive evidence that teachers who place pupils in ability groups perform better on SDQ INT and SDQ EXT achievement. Interestingly, this suggests that pupils might gain in confidence when they are taught in groups with similar ability learners. They may feel less overwhelmed and less overshadowed in such classes.

Table 7: Determinants of the Teacher FE Estimates – Teaching Practices

	Teacher FE – math test scores	Teacher FE – SDQ INT	Teacher FE – SDQ EXT
Class ability groups	-0.09*	0.06*	0.07*
	(0.046)	(0.032)	(0.035)
Homework frequency: at least once a week	0.17***	-0.08**	-0.05
	(0.039)	(0.035)	(0.038)
Homework duration: less than 20 minutes	-0.21***	0.17***	0.23***
	(0.039)	(0.035)	(0.037)
Homework type: mostly reading	-0.11**	0.11***	0.12***
	(0.048)	(0.035)	(0.038)
Always marking written work	0.02	0.03	0.05
	(0.049)	(0.053)	(0.065)
Displaying high quality work	0.03	0.01	-0.02
	(0.057)	(0.050)	(0.052)
Observations	623	714	714

Notes: All rows are estimates from separate regressions that include teacher gender + alternatively a dummy for streaming, a dummy for weekly homework, a dummy whether homework lasts less than 20 minutes, a dummy for always marking written work and a dummy whether teacher displays high quality work. The dependent variable is the teacher fixed effects (FE) estimates coming from regressions that include controls for time-varying teacher experience, length of time teacher taught pupil, and time spent teaching numeracy, class time-varying characteristics, school time-varying characteristics, pupil characteristics, family background, school cohort effects, and lagged pupil dependent variables (lagged math test scores, lagged SDQ INT, and lagged SDQ EXT respectively). ***, ** ,* significant at the 1%, 5% and 10% respectively.

[27] The accumulating research evidence on grouping appears to be contradictory. Streaming students into separate ability groups could disadvantage low-achieving students while benefiting high-achieving students, thereby exacerbating inequality (Epple *et al,* 2002). On the other hand, streaming could potentially allow teachers to more closely match instruction to students' needs, benefiting all students (Duflo *et al*., 2011).

In addition, homework frequency (at least once a week) increases teacher effectiveness. Setting homework at least once a week increases pupil math test scores by 0.17 standard deviations. In contrast, teachers who give homework which do not exceed 20 minutes perform lower by 0.21 standard deviations. And teachers who give mostly readings as homework perform lower by 0.11 standard deviations. At the same time, teachers who give homework which do not exceed 20 minutes perform better on SDQ INT and SDQ EXT achievement. Same for teachers who give mostly readings as homework. This suggests that pupils enjoy having short and easy homework although this is associated with lower math achievement. Displaying high quality work and always marking written work do not seem to have a significant effect on teacher quality (for all three outcomes).

4.4 Robustness checks

In this section, I conduct several robustness checks in order to address potential reservations about the above findings. One usual concern in estimating teacher effects, is the possibility of school/classroom sorting process. Sorting means that pupils may be purposely placed into certain schools/classrooms or with certain teachers based on their characteristics. To assess the extent to which pupils may be sorted based on their characteristics, I calculate the correlation between the effect of pupil characteristics and the teacher fixed effects from equation (1). In other words, I test whether pupils with certain characteristics are more likely to be with certain types of teacher.

Appendix Table A16 displays the results. The correlation between the effect of pupil characteristics and the teacher effects is almost zero (-0.01), which means that individual with certain observed characteristics are not placed with higher quality teachers. This provides some suggestive evidence that no sorting occurs on pupil characteristics. I also examine whether certain types of teachers are more likely to be in certain types of schools or classrooms. To this end, I calculate the correlation between the effect of school and classroom characteristics and teacher fixed effects from equation (1). The results indicate that there is a negative sorting on classroom characteristics (correlation at -0.38) and school characteristics (correlation at -0.09). Hence, if there is no sorting on observable pupil characteristics, I find that teacher fixed effects are related to classroom and school effects (e.g. class size, number of class exclusions and school size effects).

Another concern in estimating the effect of teacher emotional health and teaching practices is the possibility of reverse causality. That is, teacher performing well are more likely to have a better emotional health and to choose certain teaching methods than bad quality teachers. In other words, teachers with lower performing pupils may tend to report lower emotional health. In order to address this bias, I re-do the main results of the paper using lagged teacher characteristics. The results are reported in Appendix Tables A17 and A18. It shows very similar results suggesting that individual

teacher characteristics are relatively stable over time and do not vary accordingly to concomitant pupil characteristics. Another test, if needed, consists in calculating the within-teacher standard deviations for each teacher characteristics (see Appendix Table A19). The results indicate that much of the variation in teacher characteristics exists between teachers rather than within teachers, suggesting that teacher emotional health and teaching practices are not correlated with variations in classroom characteristics.

Assessing teacher effectiveness can also be very sensitive to how math test scores and pupil non-cognitive skills are measured. In the following set of tests, I present sensitivity analysis where I use alternative pupil outcomes. First, one might be concerned by the use of different types of math test scores to assess pupil achievement gains in math. I rerun the main analysis using the ALSPAC math test scores in year 6 as the main outcome variable controlling for the ALSPAC math test scores at the end of year 3 for consistency. The results are reported in Appendix Tables A3 and A4. Overall, the sample distribution of teacher quality (based on ALPAC math test scores) displays very similar pattern with a raw standard deviation of 0.52 and an adjusted standard deviation of 0.51 (compared to 0.54 and 0.51, respectively). In addition, teacher quality estimates in ALSPAC math test scores are highly correlated with the previous estimates based on KS2 math test scores, with a significant correlation of 0.67.

A second test consists in using English test scores instead of math test scores in order to measure pupil test skills. The ALSPAC data provides information on English Test scores at the end of year 2 (KS1) and the end of year 6 (KS2). The results are reported in Appendix Tables A3 and A4. Interestingly, the sample distribution of the teacher fixed effects (based on English test scores) is relatively lower than the one based on math test scores, with a raw standard deviation of 0.43 and an adjusted standard deviation of 0.38. This is consistent with Rockoff (2004) and Rivkin *et al.* (2005), suggesting that there are less variations in teacher ability to improve English test scores than in teacher ability to improve math test scores. Nonetheless, teacher quality estimates in math test scores and teacher quality estimates in English test scores are significantly correlated with a correlation between 0.30 and 0.47.

Third, it is possible to test whether using parents' assessed SDQ INT and parents' assessed SDQ EXT instead of teacher assessed SDQ INT and teacher assessed SDQ EXT biases the results. One might be concerned by the possibility that the results would be driven by how teachers answered the SDQ questionnaire rather than true effects in pupil SDQ outcomes. Appendix Tables A8 and A9 display the main results of the paper when a range of alternative outcomes are used to measure pupil emotional health and pupil social behaviours. It includes parents' assessed SDQ INT, parents' assessed SDQ EXT and parents' assessed SMFQ. The Short Mood and Feelings Questionnaire (SMFQ) is a sum of 13 items which includes questions such as "Child felt miserable or unhappy", "did not enjoy anything

at all", "felt so tired that he/she just sat around and did nothing", "was very restless", "felt he/she was no good any more", etc. in the past two weeks (see Appendix Table A5 for a full description of the SMFQ questionnaire).

The results indicate that there are similarly large variations in teacher quality as measured by SDQ INT parent/teacher assessed, SDQ EXT parent/teacher assessed and SMFQ parents' assessed. A one standard deviation increase in the teacher quality distribution raises SDQ INT parents' assessed by 0.58 standard deviations; raises SDQ EXT parents' assessed by 0.44 standard deviations and raises SMFQ parents' assessed by 0.50 standard deviations. Moreover, the correlation between the teacher quality estimates based on SDQs teacher assessed and SDQs parents' assessed, as reported by Appendix Table A9 is about 0.30 and is statistically significant at the 1% level. This suggests that these teacher quality estimates measure potentially equal teacher abilities. However, one cannot rule out the objection that the difference between those estimates might be due to measurement errors in parents and teacher reports. Reassuringly, however, the effects of teacher characteristics on these teacher quality estimates are quite similar to the ones based on teacher assessed indicators. Female teachers significantly increase SDQ INT and SDQ EXT compared to male teachers. In addition, teachers with better emotional health tend to perform better. Class streaming is positively associated with SDQ INT parents' assessed gains; homework which do not exceed 20 minutes and homework which are mostly readings are positively associated with SDQ INT parents' assessed gains, while weekly homework are negatively associated with SDQ INT parents' assessed outcomes.[28]

Finally, one might want to look at whether teacher fixed effects have an impact on long-run outcomes. Because the ALSPAC dataset is a birth cohort study, I have information on pupil test scores and non-test score outcomes at later ages, up to 18. In particular, it is possible to evaluate the effect of primary school teachers on pupil KS4 achievement and SDQ INT/EXT at the age of 16. To test this, I re-do the first step regressions of the paper using KS4 test scores, SDQ INT (parents' assessed) at age 16 and SDQ EXT (parents' assessed) at age 16 as the main outcomes. Appendix Table A20 reports the results. Columns (3) (Appendix Table A20) shows that primary teacher fixed effects are still statistically significant for all three outcomes. This suggests that teachers have significant long-run outcome effects on students and future adults. It is worthwhile noticing that including teacher fixed effects increases the R-squared by 8, 16 and 14 percentage points for math test scores, SDQ INT and SDQ EXT, respectively. This suggests that teachers have relatively long-run significant effects on cognitive and non-cognitive skills.

[28] The results are available upon request.

V. CONCLUSION

There has been a growing literature in education economics that has tried to assess what makes a good teacher. However, the evidence on teacher effectiveness in improving pupil non-cognitive skills has until now been limited. Here I explore the importance of teachers for both pupil cognitive and non-cognitive achievements. Using a very rich cohort dataset, I demonstrate that teachers are important inputs in pupil math test scores, but also pupil emotional health and pupil social behaviours. I also show that teacher ability to improve math test scores is weakly associated with teacher ability to improve non-test score outcomes.

Furthermore, I find that teacher emotional health and teaching practices contribute more to explaining the estimated variation in teacher quality (as measured by pupil cognitive and non-cognitive skills) than is explained by traditional observable teacher characteristics such as teacher gender or experience. These results go beyond previous studies like Hanushek (1992), Rivkin *et al.* (2005) and Aaaronson *et al.* (2007) who find evidence of significant teacher impacts on student test scores, but find little evidence that any observable teacher characteristics, explain any of the variation in estimated teacher quality.

The implications of these results for social policy are large. First, these findings suggest that test score measures understate the effect of teachers on children and adult outcomes in general and that evaluating teacher effects on non-test score outcomes may greatly improve our ability to predict teachers' overall effects. More generally, this study highlights the need to consider emotional health at school – both children's and teachers' emotional health – in addition to intellectual development. Yet, in many countries most of these objectives are still marginalised.

VI. REFERENCES

Aaronson, D., L. Barrow and W. Sander (2007), 'Teachers and Student Achievement in the Chicago Public High Schools' *Journal of Labor Economics*, 25(1), 95-135.

Bietenbeck, J.C. (2014), 'Teaching Practices and Cognitive skills', *Labour Economics*, 30, 143-153.

Borghans L., B.T. Weel and B.A. Weinberg (2008), 'Interpersonal Styles and Labor Market Outcomes', *Journal of Human Resources,* 43 (4), 815-58.

Bosquet C. and P.P. Combes (2015), 'Do Large Departments Make Academics More Productive? Sorting and Agglomeration Economies in Research', Unpublished.

Brewer D.J. and D.D. Goldhaber (1997), 'Why Don't Schools and Teachers Seem to Matter?', *Journal of Human Resources,* 32, 505-523.

Chamberlain G. (2013), 'Predictive Effects of Teachers and Schools on Test Scores, College Attendance, and Earnings', *PNAS*, no: 1315746110.

Chetty, R., J. Friedman, and J. Rockoff (2014), 'Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood', *American Economic Review*, 104(9), 2633-79.

Clotfelter, C., H. Ladd and J. Vigdor (2007), 'Teacher-student Matching and the Assessment of Teacher Effectiveness' *Journal of Human Resources*, 41(4), 778-820.

Cuhna, F. and J.J. Heckman (2007), 'The Technology of Skill Formation', *American Economic Review,* 97(2), 31-47.

Dee, T.S (2005), 'A teacher like me: Does Race, Ethnicity or Gender Matter?' *American Economic Review Papers and Proceedings*, 95, 158-165.

Duckworth, A. and M. Seligman (2005), 'Self-discipline Outdoes IQ in Predicting Academic Performance of Adolescents' *Psychological Science*, 16(2), 939-944.

Duflo, E., P. Dupas and M. Kremer (2011), 'Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya' *American Economic Review, American Economic Association,* 101(5), 1739-74.

Ehrenberg. R.G, D. D. Goldhaber and D.J. Brewer (1995), 'Do Teachers' Race, Gender and Ethnicity Matter?' *Industrial and Labor Relations Review,* 48(3), 547-61.

Epple D., E. Newlon and R. Romano (2002), 'Ability Tracking, School Competition and the Distribution of Educational Benefits', *Journal of Public Economics,* 83(1), 1-48.

Goodman A, D.L. Lamping and G.B. Ploubidis (2010), 'When to Use Broader Internalising and Externalising Subscales Instead of Hypothesised Five Subscales on the Stengths and Difficulties Questionnaire (SDQ); Data from British Parents, Teachers and Children', *Journal of Abnormal Child Psychology*, 38(8), 1179-91.

Goodman R. (1997), 'The Strengths and Difficulties Questionnaire: A Research Note' *Journal of Child Psychology and Psychiatry*, 38, 581-586.

Grogger J. and E. Eide (1995), 'Changes in College Skills and the Rise in the College Wage Premium', *Journal of Human Resources,* 30(2), 280-310.

Hanushek, E.A. (1971), 'Teacher Characteristics and Gains in Student Achievement', *American Economic Review,* 61(2), 280-88.

Hanushek, E.A. (1979), 'Conceptual and Empirical Issues in the Estimation of Educational Production Functions', *Journal of Human Resources*, 14(3), 351-88

Hanushek, E.A. (1986), 'The Economics of Schooling: Production and Efficiency in Public Schools', *Journal of Economic Literature*, 24(3), 1141-77.

Hanushek, E.A. (1992), 'The Trade-off Between Child Quantity and Quality', *Journal of Political Economy*, 100, no: 1:84-117.

Hanushek, E.A. and D.D. Kimko (2000), 'Schooling, Labor Force Quality and the Growth of Nations', *American Economic Review*, 90, 1184-1208.

Hanushek, E.A., J. Kain, D. O'Brien and S. Rivkin (2005), 'The Market for Teacher Quality' Stanford University. Unpublished.

Heckman, J.J and Y. Rubinstein (2001), 'The Importance of Non-cognitive Skills: Lessons from the GED Testing Program', *American Economic Review*, 145-149.

Heckman, J.J, J. Stixrud and S. Urzua (2006), 'The Effects of Cognitive and Non-cognitive Abilities on Labor Market Outcomes and Social Behavior', *Journal of Labor Economics,* 24(3), 411-482.

Hidalgo-Cabrillana, A. and C. Lopez-Mayan (2015), 'Teaching Styles and Achievement: Student and Teacher Perspectives', Available at http://dx.doi.org/10.2139/ssrn.2569020

Jackson, K. (2014), 'Non-Cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9th Grade Teachers in North Carolina', *NBER working paper*.

Kane, T. and D. Staiger (2008), 'Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation', *NBER working paper.*

Krueger A.B. (1999), 'Experimental Estimates of Education Production Functions' *The Quarterly Journal of Economics*, 114(2), 497-532.

Lavy, V. (2011), 'What Makes an Effective Teacher? Quasi-Experimental Evidence', *NBER working paper ,* no. 16885.

Lindqvist, E. and R. Vestman (2011), 'The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment', *American Economic Journal: Applied Economics,* 3(1), 101-128.

Moulton, B.R. (1990), 'An Illustration of a Pitfall in Estimating The Effects of Aggregate Variables on Micro Units', *Review of Economics and Statistics,* 72(2): 334-338.

Murnane, R.J., J.D. Singer, J.B. Willet, J.J. Kemple and R. Olsen (1991), 'Who Will Teach? Policies that Matter'. Cambridge, MA: Havard University Press.

Rivkin, S.G, E.A. Hanushek and J.F. Kain (2005), 'Teachers, Schools and Academic Achievement', *Econometrica*, 73, 417-458.

Rockoff, J.E. (2004), 'The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data', *American Economic Review Papers and Proceedings*, 94, 247-252.

Schwerdt, G. and A.C. Wuppermann (2011), 'Is Traditional Teaching Really all that Bad? A Within-Student Between-Subject Approach'*, Economics of Education Review*, 30, 365-379.

Van Klaveren C. (2011), 'Lecturing Style Teaching and Student Performance', *Economics of Education Review,* 30, 729-739.

VII. APPENDIX TABLES

Table A1: ALSPAC School Data

	KS1	KS2			
Year of schooling	Y2	Y3	Y4	Y5	Y6
Pupil's age	6-7	7-8	8-9	9-10	10-11
ALSPAC school questionnaires		√			√
SDQ INT reported by carer	√		√	√	√
SDQ INT reported by teacher		√			√
SDQ EXT reported by carer	√		√	√	√
SDQ EXT reported by teacher		√			√
Key Stage test scores	√				√
ALSPAC math test scores			√		√

Notes: √ means that the ALSPAC data contain information on the above characteristics in the corresponding year. KS1 and KS2 means Key Stage 1 and Key Stage 2. Key Stage are stages of the state education system in the UK. KS1 stands for year 1 and year 2 of compulsory education. KS2 covers years 3 to 6 of compulsory education. SDQ measures pupil emotional health and SDQ EXT measures pupil social behaviours.

Table A2: Pupil Characteristics

	N	Mean	Sd	Min	Max
Pupil health	5,563	0.87	0.17	0	1
Pupil age (in months)	10,384	118.70	17.32	92	145
Pupil entitled to free school meal	9,311	0.11	0.31	0	1
Pupil SEN statemented	5,245	0.02	0.14	0	1
Pupil with low birth weight	10,032	0.05	0.21	0	1
Pupil ethnicity: white	8,730	0.96	0.19	0	1
Pupil gender: male	10,390	0.51	0.50	0	1
Pupil's mother education	8,955	2.87	1.24	1	5
Pupil's father education	8,602	2.89	1.43	1	5
Household has major financial difficulties	9,533	0.13	0.34	0	1
Pupil's mother works	9,642	0.71	0.45	0	1
Pupil birth order	9,262	1.86	0.96	1	12
Pupil's mother age at birth	10,036	27.90	4.86	15	44
Pupil's mother is married	9,424	0.76	0.43	0	1
Pupil's birth month	10,385	7.02	3.25	1	12
Pupil's birth year	10,385	1991.61	0.50	1991	1993
Pupil KS1 math test score	11,493	3.21	1.37	0	5
Pupil ALSPAC math test score year 3	4,794	10.51	3.12	0	17
Pupil ALSPAC math test score year 6	7,462	18.91	7.05	0	35
Pupil KS2 math test score	11,488	62.98	21.72	0	100

Table A3: Distribution of the Teacher FE Estimates – English & ALSPAC math test scores

	English test scores		ALSPAC math test scores	
	Unadjusted	Adjusted	Unadjusted	Adjusted
10th percentile	-0.513	-0.438	-0.643	-0.653
25th percentile	-0.255	-0.237	-0.329	-0.326
50th percentile	-0.005	-0.022	0.026	0.027
75th percentile	0.283	0.255	0.335	0.315
90th percentile	0.539	0.493	0.657	0.635
90-10 gap	1.052	0.931	1.300	1.288
75-25 gap	0.538	0.492	0.664	0.641
SD	0.429	0.380	0.516	0.509
Nb of teachers	383		618	

Notes: Teacher fixed effects are estimated in regressions that include controls for time-varying teacher experience, length of time teacher taught pupil, and time spent teaching numeracy, class time-varying characteristics, school time-varying characteristics, pupil characteristics, family background, school cohort effects, and lagged pupil dependent variables (lagged English test scores and lagged ALSPAC math test score, respectively). Adjusted means that the standard deviations of the estimated teacher fixed effects are computing using inverse standard errors probability weights. The two outcomes (English test scores and math test scores) are measured in years 3 and 6. ALSPAC math test scores, means that in year 6, KS2 math test scores have been replaced by ALSPAC math test scores in equation 1.

Table A4: Correlation between the Teacher Fixed Effects Estimates – English & Math test scores

	Teacher FE – Math test scores	Teacher FE – English test scores	Teacher FE – ALSPAC math test scores
Teacher FE – Math test scores	1.00		
Teacher FE – English test scores	0.47*** (0.05)	1.00	
Teacher FE – ALSPAC math test scores	0.67*** (0.03)	0.30*** (0.06)	1.00

Notes: Bootstrapped standard errors are in parentheses. Teacher fixed effects are estimated in regressions that include controls for time-varying teacher experience, length of time teacher taught pupil, and time spent teaching numeracy, class time-varying characteristics, school time-varying characteristics, pupil characteristics, family background, school cohort effects, and lagged pupil dependent variables (lagged math test scores and lagged english test scores). Note the ALSPAC data only contain information on KS1 and KS2 English test scores. ALSPAC math test scores, means that in year 6, KS2 math test scores have been replaced by ALSPAC math test scores in equation 1. ***, **, * significant at the 1%, 5%, 10% respectively.

Table A5: Data Description

Variables	Questions	Answers
Internalising SDQ[29]	Emotional problems scale Often complains of headaches, stomach-aches or sickness Many worries, often seems worried Often unhappy, down-hearted or tearful Nervous or clingy in new situations, easily loses confidence Many fears, easily scared Peer problems scale Rather solitary, tends to play alone Has at least one good friend Generally liked by other children Picked on or bullied by other children Gets on better with adults than with other children	Not True; Somewhat True; Certainly True
Externalising SDQ	Behavioural problems scale Often has temper tantrums or hot tempers Generally obedient, usually does what adults request Often fights with other children or bullies them Often lies or cheats Steals from home, school or elsewhere Hyperactivity scale Restless, overactive, cannot stay still for long Constantly fidgeting or squirming Easily distracted, concentration wanders Thinks things out before acting Sees tasks through to the end, good attention span	Not True; Somewhat True; Certainly True
SMFQ[30]	I felt miserable or unhappy I didn't enjoy anything at all I felt so tired I just sat around and did nothing I was very restless I felt I was no good anymore I cried a lot I found it hard to think properly or concentrate I hated myself I was a bad person I felt lonely I thought nobody really loved me I thought I could never be as good as other kids I did everything wrong	Not True; Somewhat True; Certainly True
CCEI[31]	Feels upset for no obvious reason Troubled by dizziness/shortness of breath Felt like fainting Feels sick Feels life is too much effort Feels uneasy and restless Feels tingling in arms/legs/body	Very often; Often; Not very often; Never

[29] SDQ: Strength and Difficulties Questionnaire
[30] SMFQ: Short Mood and Feelings Questionnaire
[31] CCEI: Crown-Crisp Experiential Index

	Regrets much pas behaviour Sometimes feels panickly Has little or no appetite Wakes unusually early in morning Worries a lot Feels tired/exhausted Has long periods of sadness Feels strung up inside Goes to sleep all right Feels to be going to pieces Often sweats excessively Needs to cry Has had upsetting dreams Loses ability to feel sympathy	
Bachman self-esteem	Feels to be a person of worth Feels to have a number of good qualities Is able to do things as well as others Feels not to have much to be proud of Takes a positive attitude towards self Sometimes thinks to be not good at all Is a useful person to have around Feels cannot do anything right Does job well Feels their life is not useful Feels unlucky	Almost always true; Often true; Sometimes true; Seldom true; Never true

Table A6: Distribution of the Teacher FE Estimates – SDQ separate scales

	SDQ Emo – Teacher assessed		SDQ Peer – Teacher assessed		SDQ Conduct – Teacher assessed		SDQ Hyper – Teacher assessed	
	Unadj	Adj	Unadj	Adj	Unadj	Adj	Unadj	Adj
10th percentile	-0.774	-0.432	-0.666	-0.408	-0.521	-0.270	-0.530	-0.422
25th percentile	-0.303	-0.126	-0.257	-0.076	-0.200	-0.030.	-0.267	-0.192
50th percentile	0.047	0.186	0.065	0.221	0.055	0.163	0.003	0.045
75th percentile	0.393	0.475	0.344	0.447	0.277	0.331	0.259	0.293
90th percentile	0.646	0.766	0.544	0.610	0.410	0.462	0.506	0.532
90-10 gap	1.420	1.198	1.210	1.018	0.931	0.732	1.036	0.954
75-25 gap	0.696	0.892	0.601	0.523	0.477	0.361	0.526	0.485
SD	0.563	0.491	0.488	0.427	0.400	0.310	0.401	0.388
Nb of teachers	714		714		714		714	

Notes: Teacher fixed effects are estimated in regressions that include controls for time-varying teacher experience, length of time teacher taught pupil, and time spent teaching numeracy, class time-varying characteristics, school time-varying characteristics, pupil characteristics, family background, school cohort effects, and lagged pupil dependent variables (lagged SDQ Emo, lagged SDQ Peer, lagged SDQ Conduct and lagged Hyper, respectively). Adjusted means that the standard deviations of the estimated teacher fixed effects are computing using inverse standard erros probability weights. All outcomes are measured in years 3 and 6. See Table A5 for a detailed description of SDQ Emo, SDQ Peer, SDQ Conduct and SDQ Hyper.

Table A7: Correlation between the teacher FE Estimates – SDQ separate scales

	SDQ Emo Parents	SDQ Peer Parents	SDQ Conduct Parents	SDQ Hyper Parents	SDQ Emo Teacher	SDQ Peer Teacher	SDQ Conduct Teacher	SDQ Hyper Teacher
SDQ Emo - Parents	1.00							
SDQ Peer – Parents	0.01	1.00						
SDQ Conduct - Parents	0.51	-0.10	1.00					
SDQ Hyper – Parents	0.02	0.80	-0.03	1.00				
SDQ Emo - Teacher	0.90	-0.05	0.43	-0.00	1.00			
SDQ Peer - Teacher	0.71	0.17	0.29	0.10	0.42	1.00		
SDQ Conduct – Teacher	0.37	-0.02	0.76	0.00	0.29	0.39	1.00	
SDQ Hyper - Teacher	0.45	-0.08	0.92	0.00	0.39	0.29	0.58	1.00

Notes: Teacher fixed effects are estimated in regressions that include controls for time-varying teacher experience, length of time teacher taught pupil, and time spent teaching numeracy, class time-varying characteristics, school time-varying characteristics, pupil characteristics, family background, school cohort effects, and lagged pupil dependent variables (lagged SDQ Emo, lagged SDQ Peer, lagged SDQ Conduct and lagged Hyper, respectively). All outcomes are measured in years 3 and 6. Parents means that SDQ parents assessed are used and teacher means SDQ teacher assessed are used.

Table A8: Distribution of the Teacher FE Estimates – Other Outcomes

	SDQ INT – Carer assessed		SDQ EXT – Carer assessed		SMFQ – Carer assessed	
	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
10th percentile	-0.736	-0.697	-0.550	-0.509	-0.695	-0.389
25th percentile	-0.409	-0.408	-0.274	-0.230	-0.305	-0.151
50th percentile	-0.056	-0.140	-0.010	-0.011	0.035	0.133
75th percentile	0.447	0.311	0.262	0.242	0.354	0.444
90th percentile	0.802	0.746	0.527	0.496	0.669	0.679
90-10 gap	1.538	1.443	1.077	1.005	1.364	1.068
75-25 gap	0.856	0.719	0.536	0.472	0.659	0.595
SD	0.637	0.579	0.451	0.435	0.604	0.504
Nb of teachers	607		607		379	

Notes: Teacher fixed effects are estimated in regressions that include controls for time-varying teacher experience, length of time teacher taught pupil, and time spent teaching numeracy, class time-varying characteristics, school time-varying characteristics, pupil characteristics, family background, school cohort effects, and lagged pupil dependent variables (lagged SDQ INT, lagged SDQ EXT, and lagged SMFQ respectively). Adjusted means that the standard deviations of the estimated teacher fixed effects are computing using inverse standard errors probability weights. All three outcomes are measured in years 3 and 6. SMFQ and SDQ INT measures pupil emotional health. SDQ EXT measures pupil social behaviours. See Appendix Table A5 for a detailed description of SMFQ, SDQ INT and SDQ EXT.

Table A9: Correlation between the Teacher FE Estimates – Other Outcomes

	Teacher FE – SDQ INT Parents assessed	Teacher FE – SDQ EXT Parents assessed	Teacher FE – SMFQ Parents assessed
Teacher FE – SDQ INT Teacher assessed	0.28*** (0.04)	0.13*** (0.04)	-0.02 (0.05)
Teacher FE – SDQ EXT Teacher assessed	0.25*** (0.04)	0.26*** (0.04)	-0.06 (0.06)
Teacher FE – SDQ INT Parents assessed	1.00	0.21*** (0.05)	0.34*** (0.07)

Notes: Bootstrapped standard errors are in parentheses. Teacher fixed effects are estimated in regressions that include controls for time-varying teacher experience, length of time teacher taught pupil, and time spent teaching numeracy, class time-varying characteristics, school time-varying characteristics, pupil characteristics, family background, school cohort effects, and lagged pupil dependent variables. All three outcomes are measured in years 3 and 6. SMFQ and SDQ INT are measures of pupil emotional health. SDQ EXT measures pupil social behaviours. ***, **, * significant at the 1%, 5% and 10% respectively.

Table A10: School, Classroom and Teacher Characteristics

	N	Mean	Sd	Min	Max
School characteristics					
Private school	9,736	0.02	0.14	0	1
School size	7,437	285.8	109.6	43	673
School admission policy: waiting list (no policy)	7,384	1.77	0.42	1	2
School admission policy: priority for special needs	7,384	1.80	0.40	1	2
School admission policy: priority for catchment area	7,384	1.52	0.50	1	2
School admission policy: priority for siblings	7,384	1.56	0.50	1	2
School admission policy: parental request	7,384	1.94	0.24	1	2
School pupil ratio	5,676	19.6	6.43	5	49
Frequency of staff meetings	7,472	1.13	0.42	1	6
Head teacher gender	7,475	0.52	0.50	0	1
Class characteristics					
Class size	10,345	28.9	4.56	9	40
Number of class exclusions	9,678	0.30	0.73	0	7
Number of permanent class exclusions	9,678	0.03	0.33	0	7
Percentage of free school meals in class	9,169	13.1	13.12	0	79
Percentage of Sen statements in class	9,831	3.88	7.64	0	100
Percentage of pupils in class with concerning home pbs	10,128	0.47	0.85	0	5
Percentage of pupils for whom English is not the first language	9,941	2.00	3.83	0	35
Class noise	10,356	2.79	1.17	1	4
Difference in age in class	9,148	1.12	0.51	0	5
Teacher characteristics					
Teacher gender: female	10,390	0.80	0.40	0	1
Teacher CCEI	10,316	27.23	8.75	0	44
Teacher Bachman self-esteem	10,305	31.22	5.71	14	40
Teacher job satisfaction	10,303	4.46	0.83	1	5
Length of time taught pupil	10,157	1.21	0.41	1	2
Teacher confidence score in teaching	10,328	1.79	0.31	1	2
Teacher experience at school	10,390	2.88	1.01	1	4
Teacher overall experience	10,390	3.42	0.75	1	4
Teacher years of qualification	10,378	15.47	11.13	1	40
Homework frequency	10,308	3.57	0.80	1	5
Homework type	10,092	1.97	0.48	1	3
Homework duration	10,103	2.81	1.00	1	5
Assessments: tests	7,632	2.58	0.56	1	3
Assessments: written tests	4,607	2.91	0.29	2	3
Assessments: self-assessments	4,601	2.01	0.55	1	3
Assessments: listening	4,607	2.75	0.45	1	3
Assessments: Q&A	4,607	2.79	0.42	1	3
Assessments: discussion	4,607	2.23	0.54	1	3
Incentives: direct verbal	4,624	1.00	0.00	1	1
Incentives: direct written	4,624	1.00	0.04	0	1
Incentives: direct	4,624	1.00	0.00	1	1
Incentives: naming pupil	4,624	1.00	0.05	0	1
Incentives: free time	4,615	0.58	0.49	0	1
Incentives: competition	4,624	0.89	0.32	0	1
Incentives: displaying high quality work	4,624	0.40	0.49	0	1
Class activity groups	4,600	1.00	0.07	0	1
Class activity groups by attainment	4,600	0.97	0.16	0	1
Class activity groups by gender	4,562	0.15	0.35	0	1
Class activity groups by age	4,579	0.15	0.36	0	1
Class ability groups	10,347	0.94	0.24	0	1
Class ability groups: literacy	10,347	0.89	0.31	0	1
Class ability groups: maths	10,338	0.93	0.26	0	1
Class ability groups: others	10,347	0.19	0.40	0	1

Table A11: Pupil Outcomes and Teacher Quality – Full Specification

	Math Test Scores	SDQ INT	SDQ EXT
Teacher experience	0.03	0.05	0.02
	(0.055)	(0.068)	(0.065)
Length of time teaching pupil	0.08***	0.00	0.01
	(0.021)	(0.025)	(0.024)
Time spent teaching numeracy	0.10	-0.14	-0.12
	(0.082)	(0.088)	(0.089)
Pupil health	0.01	0.15***	-0.05
	(0.032)	(0.037)	(0.031)
Pupil age (in months)	-0.30	0.06	-0.02
	(0.200)	(0.218)	(0.209)
Pupil entitled to free school meal	-0.03***	-0.09***	-0.07***
	(0.011)	(0.014)	(0.013)
Pupil SEN Statemented	-0.08***	-0.12***	-0.06***
	(0.011)	(0.016)	(0.016)
Pupil with low birth weight	-0.03***	-0.02	-0.01
	(0.010)	(0.012)	(0.011)
Pupil ethnicity: white	0.01	-0.11***	-0.03
	(0.023)	(0.022)	(0.024)
Pupil gender: male	0.05***	-0.03***	-0.29***
	(0.009)	(0.010)	(0.010)
Pupil's mother education	0.13***	0.03**	0.04***
	(0.016)	(0.017)	(0.016)
Pupil's father education	0.10***	0.01	0.07***
	(0.015)	(0.016)	(0.016)
Household has major financial difficulties	-0.01	-0.03**	-0.02
	(0.009)	(0.011)	(0.011)
Pupil's mother works	0.01	0.01	0.01
	(0.011)	(0.012)	(0.012)
Pupil birth order	-0.02*	0.02	-0.08***
	(0.013)	(0.015)	(0.014)
Pupil's mother age at birth	0.03*	-0.02	0.08***
	(0.017)	(0.019)	(0.018)
Pupil's mother is married	0.02*	0.04***	0.04***
	(0.012)	(0.014)	(0.013)
Pupil's birth month	-0.12***	-0.00	-0.02
	(0.040)	(0.044)	(0.042)
Pupil's birth year	-19.07***	1.15	-6.67
	(6.118)	(6.675)	(6.380)
Class size	0.05	-0.04	0.01
	(0.036)	(0.045)	(0.040)
Number of class exclusions	0.01	-0.03	-0.02
	(0.024)	(0.027)	(0.026)
Number of permanent class exclusions	0.00	-0.02	0.01
	(0.022)	(0.018)	(0.014)
Percentage of free school meals in class	0.03	-0.04	0.06
	(0.042)	(0.053)	(0.048)
Percentage of SEN statements in class	0.08**	0.04	-0.01
	(0.033)	(0.035)	(0.033)
Percentage of pupils in class with concerning home pbs	0.03	0.02	0.02
	(0.024)	(0.026)	(0.024)
Percentage of pupils for whom English is not the first language.	-0.00	0.10**	0.02

	(0.036)	(0.041)	(0.036)
Class noise	0.01	-0.04	-0.00
	(0.027)	(0.033)	(0.029)
Difference in age in class	0.02	0.07	0.03
	(0.037)	(0.043)	(0.040)
School size	0.04	-0.13	0.09
	(0.075)	(0.086)	(0.078)
Frequency of staff meetings	0.00	0.03	0.07
	(0.051)	(0.050)	(0.053)
Head teacher gender	-0.08	0.08	-0.01
	(0.052)	(0.056)	(0.052)
Lagged math test scores	0.91***		
	(0.022)		
Lagged SDQ INT		0.75***	
		(0.053)	
Lagged SDQ EXT			0.63***
			(0.035)
Observations	7,651	7,648	7,622
R-squared	0.471	0.288	0.356
Pupil covariates	Yes	Yes	Yes
School and Classroom characteristics	Yes	Yes	Yes
Teacher fixed effects	Yes	Yes	Yes

Notes: Robust standard errors are in parentheses. ***, **, * significant at the 1%, 5% and 10% respectively.

Table A12: Interaction effect between teacher gender and pupil gender

	Math test scores	SDQ INT	SDQ EXT
Female teacher * Male pupil	0.04	-0.02	0.09*
	(0.045)	(0.053)	(0.050)
Observations	7,651	7,648	7,622
R-squared	0.471	0.288	0.357
Pupil covariates	Yes	Yes	Yes
School and Classroom characteristics	Yes	Yes	Yes
Teacher fixed effects	Yes	Yes	Yes

Notes: All regressions include pupil characteristics (including lagged dependent variable), family background and cohort fixed effects. Column (2) adds school and classroom characteristics. Column (3) adds teacher fixed effects. An interaction term "Female teacher*Male pupil" is added, which examines the interaction between teacher gender and pupil gender. Teacher gender and pupil gender are also included in the baseline regression. Only the interaction effect is reported. All three outcomes are measured in years 3 and 6. SDQ INT measures pupil emotional health and SDQ EXT measures pupil social behaviours.

Table A13: Correlation between teacher emotional health characteristics

	Female	CCEI	Bachman Self-Esteem	Job Satis-faction	Confi-dence	Exper-ience
Female	1.00					
CCEI	-0.09	1.00				
Bachman self-esteem	-0.11	0.42	1.00			
Job satisfaction	0.08	0.31	0.35	1.00		
Confidence	-0.04	0.07	0.16	0.12	1.00	
Experience	-0.01	-0.12	-0.01	-0.10	0.18	1.00

Notes: These are simple correlation coefficients between teacher emotional characteristics. CCEI means teacher Crown-Crisp Experiential Index. Bachman test scores measures teacher self-esteem. Experience measures teacher overall experience.

Table A14: Determinants of the Teacher FE Estimates – Teacher Mental Health – Full Specification

	Teacher FE – Math Test Scores	Teacher FE – SDQ INT	Teacher FE – SDQ EXT
Teacher gender: female	-0.07**	0.10***	0.14***
	(0.035)	(0.035)	(0.037)
Teacher CCEI	0.09**	0.19***	0.17***
	(0.043)	(0.035)	(0.040)
Teacher Bachman self-Esteem	0.04*	0.01	-0.04
	(0.021)	(0.056)	(0.034)
Teacher job satisfaction	0.01	-0.05	-0.04
	(0.039)	(0.035)	(0.040)
Teaching confidence	0.12***	-0.04	-0.07**
	(0.041)	(0.034)	(0.037)
Observations	623	714	714

Notes: Teacher fixed effects are estimated in regressions that include controls for time-varying experience, length of time teacher taught pupil, and time spent teaching numeracy, class time-varying characteristics, school time-varying characteristics, pupil characteristics, family background, school cohort effects, and lagged pupil dependent variables (lagged math test scores, lagged SDQ INT and lagged SDQ EXT respectively). All three outcomes are measures in years 3 and 6. SDQ INT measures pupil emotional health and SDQ EXT measures pupil social behaviours. Teacher CCEI stands for teacher Crown-Crisp Experiential Index, See Appendix Table A5 for a full description of CCEI and Bachman Self-Esteem. *** p<0.01, ** p<0.05, * p<0.1.

Table A15: Determinants of the Teacher FE Estimates – Teaching Practices – Full Specification

	Teacher FE – Math Test Scores	Teacher FE – SDQ INT	Teacher FE – SDQ EXT
Streaming	-0.04	0.01	0.02
	(0.040)	(0.032)	(0.033)
Weekly homework	0.10***	-0.02	0.02
	(0.035)	(0.034)	(0.035)
Homework duration: less than 20 minutes	-0.07*	0.02	0.06
	(0.039)	(0.039)	(0.042)
Homework type: mostly reading	0.01	0.02	-0.00
	(0.045)	(0.037)	(0.037)
Always marking written work	0.01	0.04	0.05
	(0.051)	(0.053)	(0.051)
Displaying high quality work	0.05	0.01	-0.02
	(0.058)	(0.052)	(0.053)
Observations	623	714	714

Notes: Teacher fixed effects are estimated in regressions that include controls for time-varying experience, length of time teacher taught pupil, and time spent teaching numeracy, class time-varying characteristics, school time-varying characteristics, pupil characteristics, family background, school cohort effects, and lagged pupil dependent variables (lagged math test scores, lagged SDQ INT and lagged SDQ EXT respectively). All three outcomes are measures in years 3 and 6. SDQ INT measures pupil emotional health and SDQ EXT measures pupil social behaviours. *** p<0.01, ** p<0.05, * p<0.1.

Table A16: Test for Pupil Sorting

	Teacher FE Math Test Scores	Teacher FE SDQ INT	Teacher FE SDQ EXT
Individual effects	-0.012	-0.018	0.003
Pupil health	0.028	0.011	-0.012
Pupil age (in months)	-0.392	-0.343	0.433
Pupil entitled to free school meal	0.097	0.024	0.043
Pupil Sen Statemented	0.079	0.033	0.012
Pupil with low birth weight	0.019	0.014	0.011
Pupil ethnicity: white	0.029	-0.036	-0.036
Pupil gender: male	0.005	0.023	-0.008
Pupil's mother education	0.095	-0.006	0.012
Pupil's father education	0.110	-0.012	-0.003
Household has major financial difficulties	0.012	0.001	-0.007
Pupil's mother works	0.024	-0.004	0.006
Pupil birth order	0.023	0.011	-0.011
Pupil's mother age at birth	0.015	-0.022	0.041
Pupil's mother is married	0.063	0.038	0.060
Pupil's birth month	0.029	-0.062	0.006
Pupil's birth year	-0.008	-0.005	0.001
Class effects	-0.384	-0.242	-0.133
Class size	-0.050	-0.102	0.039
Number of class exclusions	-0.091	0.077	0.123
Number of permanent class exclusions	-0.081	-0.063	0.067
Percentage of free school meals in class	-0.263	0.029	-0.057
Percentage of Sen statements in class	-0.287	-0.070	-0.072
Percentage of pupils in class with concerning home pbs	-0.224	-0.099	-0.175
Percentage of pupils for whom English is not the first language.	-0.074	-0.222	-0.165
Class noise	0.061	-0.107	-0.092
Difference in age in class	-0.111	-0.077	-0.005
School effects	-0.087	-0.098	0.120
School size	-0.083	-0.098	0.076
Frequency of staff meetings	-0.055	0.035	0.157
Head teacher gender	-0.020	0.031	-0.068

Notes: This table reports the correlation between the effect of a variable or group of variables (individual, class or school) and the teacher effects (teacher fixed effects) from regressions on math test scores, SDQ INT and SDQ EXT that include teacher experience, length of time teacher taught pupil, and time spent teaching numeracy, teacher dummies, class time-varying characteristics, school time-varying characteristics, pupil characteristics, family background, school cohort effects, and lagged math test scores, lagged SDQ INT, and lagged SDQ EXT. See Appendix Table A11 for the full specification. SDQ INT measures pupil emotional health and SDQ EXT measures pupil social behaviours.

Table A17: Determinants of the Teacher Fixed Effects – Lagged Teacher Mental Health

	Teacher FE – Math Test Scores	Teacher FE – SDQ INT	Teacher FE – SDQ EXT
Teacher gender: female	-0.09**	0.07*	0.12***
	(0.034)	(0.036)	(0.038)
Teacher CCEI	0.09**	0.18***	0.16***
	(0.040)	(0.034)	(0.036)
Teacher Bachman self-Esteem	0.03	0.05	-0.03
	(0.023)	(0.055)	(0.044)
Teacher job satisfaction	0.06	0.02	0.01
	(0.039)	(0.036)	(0.038)
Teaching confidence	0.15***	-0.04	-0.08**
	(0.041)	(0.035)	(0.037)
Observations	623	714	714

Notes: All rows are estimates from separate regressions that include teacher gender + alternatively lagged teacher CCEI or teacher Bachman self-esteem, or teacher job satisfaction or teacher confidence in teaching. The dependent variable is the teacher fixed effects estimates coming from regressions that include controls for time-varying teacher experience, length of time teacher taught pupil, and time spent teaching numeracy, class time-varying characteristics, school time-varying characteristics, pupil characteristics, family background, school cohort effects, and lagged pupil dependent variables. All three outcomes are measured in years 3 and 6. SDQ INT measures pupil emotional health and SDQ EXT measures pupil social behaviours. Teacher CCEI stands for teacher Crown-Crisp Experiential Index. See Appendix Table A5 for a full description of CCEI and Bachman Self-Esteem. *** p<0.01, ** p<0.05, * p<0.1

Table A18: Determinants of the Teacher Fixed Effects – Lagged Teaching Practices

	Teacher FE – Math Test Scores	Teacher FE – SDQ INT	Teacher FE – SDQ EXT
Streaming	-0.09*	0.06*	0.05
	(0.048)	(0.034)	(0.036)
Weekly homework	0.14***	-0.02	0.01
	(0.044)	(0.035)	(0.036)
Homework duration: less than 20 minutes	-0.21***	0.16***	0.22***
	(0.040)	(0.035)	(0.038)
Homework type: mostly reading	-0.11**	0.11***	0.13***
	(0.050)	(0.035)	(0.040)
Always marking written work	0.01	0.02	0.02
	(0.069)	(0.062)	(0.060)
Displaying high quality work	0.02	0.01	-0.05
	(0.058)	(0.049)	(0.052)
Observations	623	714	714

Notes: All rows are estimates from separate regressions that include teacher gender + alternatively lagged teacher CCEI or teacher Bachman self-esteem, or teacher job satisfaction or teacher confidence in teaching. The dependent variable is the teacher fixed effects estimates coming from regressions that include controls for time-varying teacher experience, length of time teacher taught pupil, and time spent teaching numeracy, class time-varying characteristics, school time-varying characteristics, pupil characteristics, family background, school cohort effects, and lagged pupil dependent variables. All three outcomes are measured in years 3 and 6. SDQ INT measures pupil emotional health and SDQ EXT measures pupil social behaviours. *** p<0.01, ** p<0.05, * p<0.1

Table A19: Within Teacher Variation in Teacher Characteristics

	Total SD	Between SD	Within SD
Teacher characteristics			
Teacher gender: female	0.40	0.40	0
Teacher CCEI (inv)	8.22	7.98	1.93
Teacher Bachman self esteem (inv)	5.71	5.55	1.45
Teacher job satisfaction	0.82	0.79	0.28
Length of time taught pupil	0.40	0.38	0.20
Teacher confidence score in teaching	0.31	0.31	0.11
Teacher experience at school	1.01	1.03	0.18
Teacher overall experience	0.74	0.75	0.13
Teacher years of qualification	11.13	11.19	0.36
Homework frequency	0.80	0.79	0.25
Homework type	0.48	0.47	0.18
Homework duration	1.00	0.96	0.29
Assessments: tests	0.56	0.53	0.26
Assessments: written tests	0.29	0.28	0.13
Assessments: self-assessments	0.54	0.54	0.19
Assessments: listening	0.45	0.45	0.19
Assessments: Q&A	0.42	0.40	0.20
Assessments: discussion	0.54	0.51	0.23
Incentives: direct written	0.04	0.03	0.02
Incentives: naming pupil	0.05	0.06	0.02
Incentives: free time	0.49	0.47	0.17
Incentives: competition	0.49	0.47	0.15
Incentives: displaying high quality work	0.48	0.46	0.16
Class activity groups	0.07	0.06	0.03
Class activity groups by attainment	0.16	0.16	0.05
Class activity groups by gender	0.35	0.36	0.11
Class activity groups by age	0.36	0.40	0.09
Class ability groups	0.24	0.24	0.08
Class ability groups: literacy	0.31	0.30	0.11
Class ability groups: maths	0.26	0.26	0.10
Class ability groups: others	0.39	0.39	0.15

Notes: Column (1) reports total sample standard deviations for each teacher characteristics. Column (2) displays between teacher standard deviations for each teacher characteristics and column (3) displays within teacher standard deviations for each teacher characteristics. CCEI stands for Crown-Crisp Experiential Index. See Appendix Table A5 for a full description of CCEI and Bachman Self-Esteem.

Table A20: Pupil Long-Run Outcomes and Teacher Quality

	Math Test Scores- KS4			
R-squared	0.373	0.386	0.453	0.413
Observations	7,195			
	SDQ INT – age 16			
R-squared	0.185	0.196	0.345	0.250
Observations	4,022			
	SDQ EXT – age 16			
R-squared	0.258	0.262	0.395	0.308
Observations	4,031			
Included explanatory variables:				
Pupil covariates	Yes	Yes	Yes	Yes
School & classroom characteristics	No	Yes	Yes	No
Teacher characteristics	No	No	No	No
Teacher fixed effects	No	No	Yes	No
F tests, H0:	--	--	(<0.01)	--
School fixed effects	No	No	No	Yes
F tests, H0:	--	--	--	(<0.01)

Notes: Teacher fixed effects are estimated using regressions that include controls for pupil characteristics (including lagged dependent variable), family background, teacher time-varying experience, length of time teacher taught pupil, and time spent teaching numeracy, school time-varying characteristics, classroom time-varying characteristics, school-cohort effects. Column (2) adds school and classroom characteristics. Column (3) adds teacher fixed effects and column (4) substitutes school to teacher fixed effects. Only R-squareds are reported. Numbers in parentheses are *p* values from F tests of the joint significance of teacher fixed effects and school fixed effects respectively. All three outcomes are measured at age 16. SDQ INT measures pupil emotional health and SDQ EXT measures pupil social behaviours. SDQ INT and SDQ EXT are assessed by parents at age 16.

VIII. FIGURES

Figure 1: Teacher Quality Distribution (Unadjusted)

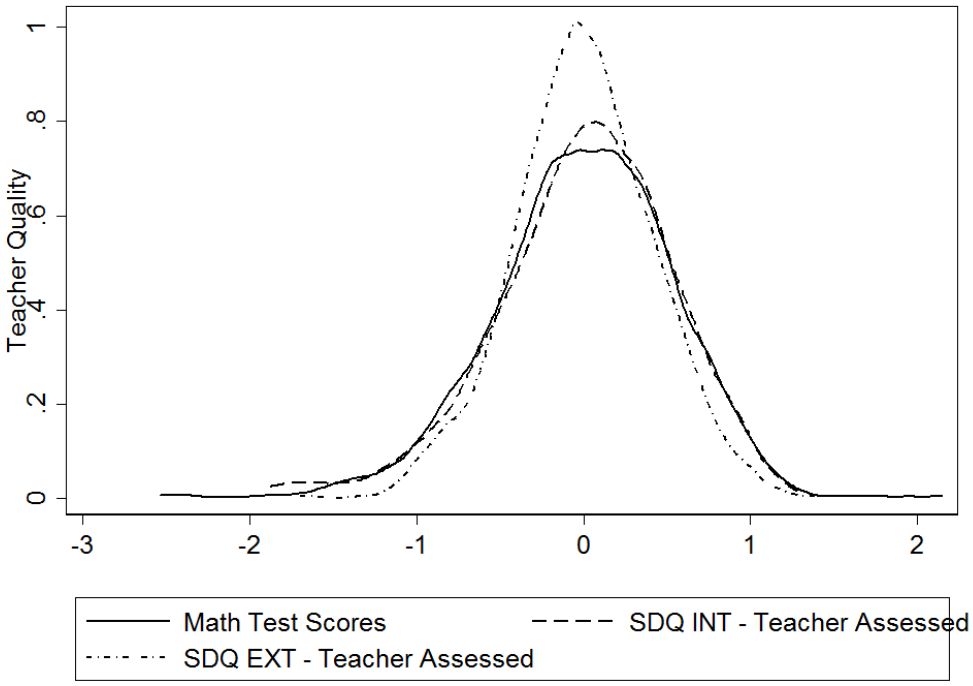
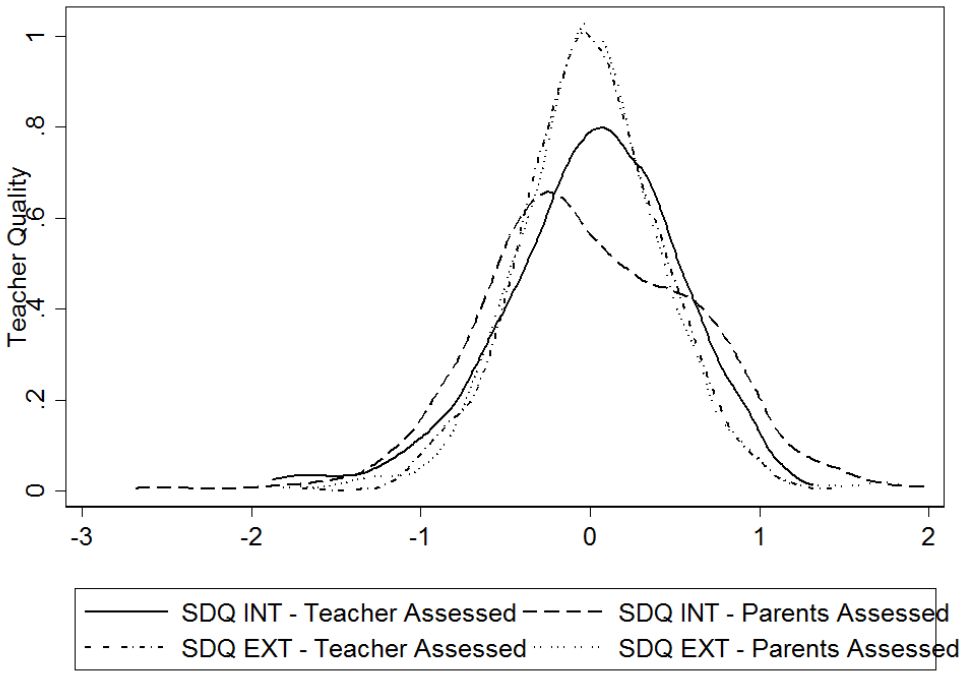


Figure 2: Teacher Quality Distribution (Unadjusted)



Notes: Kernel density estimates of teacher fixed effects. Teacher fixed effects are estimated in regressions that include controls for time-varying teacher experience, length of time teacher taught pupil, and time spent teaching numeracy, class time-varying characteristics, school time-varying characteristics, pupil characteristics, family background, school cohort effects, and lagged pupil dependent variables.

Figure 3: Teacher Quality Distribution (Unadjusted)

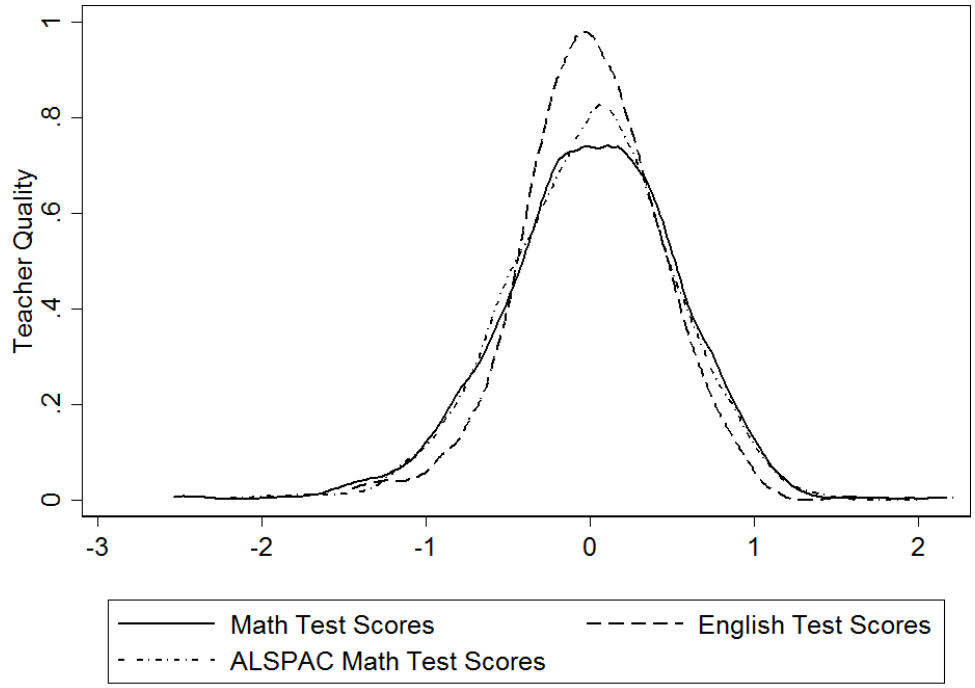
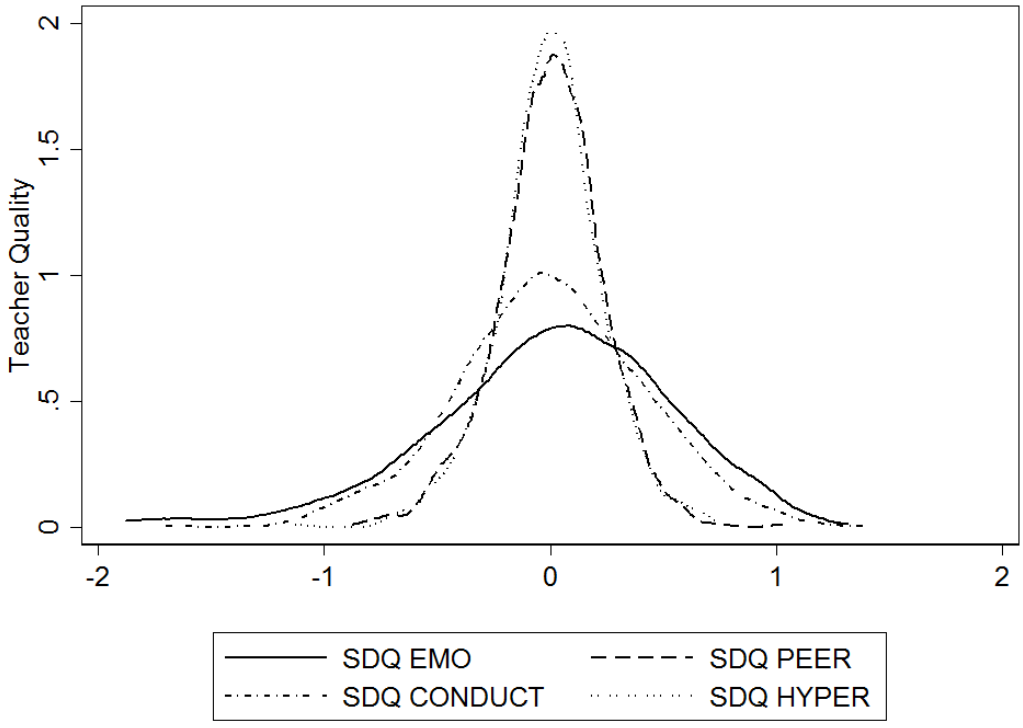


Figure 4: Teacher Quality Distribution (Unadjusted)



Notes: Kernel density estimates of teacher fixed effects. Teacher fixed effects are estimated in regressions that include controls for time-varying teacher experience, length of time teacher taught pupil, and time spent teaching numeracy, class time-varying characteristics, school time-varying characteristics, pupil characteristics, family background, school cohort effects, and lagged pupil dependent variables.