Teacher Quality, Test Scores and Non-Cognitive Skills: Evidence from Primary School Teachers in the UK

Sarah Fleche*

February 14, 2017

JOB MARKET PAPER

Abstract

Schooling can produce both cognitive and non-cognitive skills, both of which are important determinants of adult outcomes. Using very rich data from a UK birth cohort study, I estimate teacher value added (VA) models for both pupils' test scores and non-cognitive skills. I show that teachers have large effects on pupils' non-cognitive skills - above and beyond their effects on test scores. This finding extends the economics literature on teacher effects, which has primarily focused on pupils' test scores and may fail to capture teachers' overall effects. In addition, the large estimates reveal an interesting trade-off: teacher VA on pupils' test scores are weak predictors of teacher VA on non-cognitive skills, which suggests that teachers recourse to different techniques to improve pupils' cognitive and non-cognitive skills. Finally, I find that teachers' effects on pupils' non-cognitive skills have long-run impacts on adult outcomes such as higher education attendance, employment and earnings, conditional on their effects on test scores. This result indicates that long-run outcomes are improved by a *combination* of teachers increasing pupils' test scores and non-cognitive skills and has large policy implications.

JEL Codes: I21, J00
Keywords: Teacher quality, test scores, non-cognitive skills, long-run impacts, teaching practices.

*Centre for Economic Performance. London School of Economics. Email: s.fleche@lse.ac.uk. I am extremely grateful to Esteban Aucejo, Clement Bosquet, Raj Chetty, Andrew Clark, Julien Grenet, Richard Layard, Alan Manning, Stephen Machin, Sandra McNally, Barbara Petrongolo, Jorn-Steffen Pischke, Jesse Rothstein and Claudia Senik as well as seminar and conference participants at the AEA meetings, LSE, PSE and AMSE economics departments, the Royal Economic Society, Louis-Andre Gerard Varet (LAGV), HEIRS, and ECCP. This study uses the ALSPAC data. The UK Medical Research Council and Wellcome (Grant: 102212/2/13/2) and the University of Bristol provide core support for ALSPAC. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. I am very grateful to Richard Layard, Andrew Clark, Nattavudh Powdthavee, Nele Warrinnier and George Ward for making this project possible. Support from the US National Institute on Aging (Grant R01AG0400640), the John Templeton Foundation and the What Works Centre for Wellbeing is gratefully aknowledged.

I. Introduction

Recent studies have shown that cognitive and non-cognitive skills accumulated during childhood have important impacts on adult outcomes (e.g. Heckman and Rubinstein, 2001; Heckman et al., 2006; Borghans et al., 2008).[1] Also, schooling can produce both cognitive skills and non-cognitive skills. However, most of the education literature in economics has focused on test scores as measures of students' skills. Much less is known about the effect of schooling on non-cognitive skills (e.g. self-esteem, perseverance, adaptability, social skills, etc.). Accordingly, evaluating schooling effects based on test scores may fail to capture schooling overall effects and addresses only one dimension of what matters for child development and adult success.

This paper speaks to this issue by estimating the importance of teachers on both pupils' cognitive and non-cognitive outcomes. Policy makers and researchers agree that teachers are one of the most important school-related factors. Previous work has shown that during one year with a teacher in the 85th percentile according to value added scores (VA), pupils gain 40% more in their learning than they would with a teacher in the 15th percentile (e.g. Rockoff, 2004; Rivkin et al., 2005; Aaronson et al., 2007; Kane and Staiger, 2008; Chetty et al., 2014a). US school districts have begun to produce estimates of teachers' VA on pupils' test scores to evaluate teachers. However, it is surprising that most of the discussions on teachers' VA almost exclusively focuses on measures of cognitive ability. Accordingly, it is critical for policy that these measures reflect teachers' overall effects.

Much of the neglect of non-cognitive skills in analysis of schooling effects is certainly due to the lack of any reliable measure of them. However, in recent research, economists and psychologists have constructed measures of non-cognitive skills and have provided evidence that they predict meaningful outcomes (e.g. Borghans et al., 2008; Almlund et al., 2011; Heckman et al., 2015). In a very short period of time, non-cognitive skill measures have started to be included in large scale surveys and administrative registers. These measures include, among others, teacher assessments of social skills, parental reports of behaviours, self-reported beliefs about personal control, and administrative records of

[1] An increasing body of empirical literature sheds light on the importance of non-cognitive skills and finds that non-cognitive skills are good predictors of adult outcomes, such as labour market success, crime behaviours and health (see Heckman et al. (2015) for a review of the literature). In particular, the more recent economics literature on non-cognitive skills comes into prominence with two studies by James Heckman and co-authors. Heckman and Rubinstein (2001) find that GED recipients are more likely to engage in drug use and to commit minor crimes than either conventional high school graduates or high school dropouts, and infer that the absence of a positive economic return to GED recipiency is due to a shortfall in non-cognitive skills among those who receive this credential. Heckman et al. (2006), using adolescent measures of self-efficacy and self-esteem in the National Longitudinal Survey of Youth 1979 as indicators of non-cognitive abilities, find that non-cognitive and cognitive skills are equally important in the determination of a variety of economic and social outcomes.

school suspensions. In particular, behavioural problem indices that include measures of internalising and externalising behaviours, as well as reports of persistence, ability to focus, and social skills, have been extensively used by psychologists and education specialists, and are available in large-scale datasets.

In this paper, I rely on a very rich UK birth cohort study, the Avon Longitudinal Study of Parents and Children (ALSPAC), which provides a behavioural screening test, known as the Strength and Difficulties Questionnaire, as indicator of non-cognitive abilities. It is a well-accepted questionnaire in developmental, genetic, social, clinical and educational studies. It includes 25 items on non-cognitive attributes, which are divided between 5 scales: emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and pro-social behaviour. The ALSPAC data also provide teacher assignments in years 3 and 6 of primary school when the pupils were aged 8 and 11. The data are merged with the National Pupil Database which contains detailed information on pupils' test scores and exam results spanning 1991-2009.

The strength of these data allows me to make four important contributions to the literature on teacher effectiveness. First, I construct VA estimates for the teachers in my dataset, based on pupils' math test scores and non-cognitive abilities. My approach to estimate VA parallels closely that used by previous work estimating teacher VA on pupils' test scores (e.g. Kane and Staiger, 2008; Chetty et al., 2014a), except that I also provide one of the first VA estimates on non-cognitive skills.[2] Second, I use the results to test whether teachers who raise test scores also improve non-cognitive skills. If there is a weak correlation between a teacher's ability to increase cognitive and non-cognitive skills, this has important implications for how teachers are evaluated: a teacher who is good at developing pupils' non-cognitive skills, but not efficient at raising their test scores, might be rated as ineffective, thus undervaluing her contribution to pupils' learning.

Third, I leverage my research design to provide the first estimates of teachers' non-cognitive effects on long-run outcomes such as higher education attendance, earnings, unemployment, and full-time job.[3] Previous work has shown that teachers' impacts on

[2]In existing work, Jackson (2012) finds that teachers have causal effects on test scores and proxies for non-cognitive skills (e.g. absences, suspensions, grades and on-time grade progression). Araujo et al. (2016) find that teachers have substantial effects on students' executive function. Mihaly et al. (2013) estimate teachers' effects on non-test score outcomes to better predict teachers' effects on test scores. Ruzek et al. (2014) find that teachers influence their students' motivation, as measured by mastery and performance achievement goals. Finally, Blazar and Kraft (2015) find that upper-elementary teachers have large effects on self-reported measures of students' self-efficacy in math, and happiness and behaviour in class. Gershenson (2016) finds that teachers have important effects on students' absences. Yet, all these studies rely on proxies for non-cognitive skills.

[3]Only one study, Jackson (2012), has attempted to investigate this issue. He finds that teacher effects on absences, suspensions, course grades and on-time grade progression predict high school completion. However, his work relies on proxies for non-cognitive skills and he does not look at long-run effects on other adult outcomes such as labour market outcomes.

test scores fade out very rapidly (Rothstein, 2010; Jacob et al., 2010; Chetty et al., 2014b). Despite this fade-out, there is evidence that teachers' impacts on test scores do create persistent improvements in successful lifetime outcomes (Chetty et al., 2014b). This would suggest that teachers may have important effects on long-run outcomes that are not reflected in their test score VA and that might be related to their non-cognitive skills VA.[4] This paper addresses this issue by investigating (i) whether teachers have influence on pupils' non-cognitive skills and (ii) whether teachers' effects on non-cognitive skills have long-run impacts that are not measured by teachers' effects on cognitive skills. This may help reconcile the apparent paradox of the long-term impacts of teachers despite the rapid fade-out on test scores.

Fourth, I turn my attention to the mechanisms through which teachers affect pupils' cognitive and non-cognitive skills. If long-run outcomes are improved by a combination of teachers skilled at increasing pupils' test scores and those able to raise pupils' non-cognitive skills, how can a school system reinforce the importance of both? Rivkin et al. (2005) find that teachers have powerful effects on reading and math achievement, but less than 10% of the variation in teacher quality is explained by observable teacher characteristics such as education or experience. The disjuncture between estimates of teacher quality and the explanatory power of observed teacher characteristics creates a clear dilemma for policy makers. This paper complements previous studies on the determinants of teacher effectiveness by (i) analysing to what extent teachers' VA on both pupils' cognitive and non-cognitive skills are associated with a number of teacher characteristics (including teacher non-cognitive skills) and teaching practices, and (ii) by testing whether different teacher characteristics and teaching practices are associated with teachers' ability to improve pupils' cognitive and non-cognitive skills.

The results are as follows. I show that teachers have large influences on pupils' non-cognitive skills. The VA estimates indicate that a one standard deviation (SD) improvement in teacher VA raises normalised internalising behaviour by 0.22 SD and externalising behaviour by 0.12 SD. For comparison, a one SD improvement in teacher math VA raises normalised math test scores by 0.13 SD, consistently with estimates in prior studies.[5] One concern in interpreting these results, however, is that the assignment of pupils to teachers is not random. Such sorting can lead to biased estimates of teachers' VA for both pupils' test scores and non-cognitive skills. To address this issue, I implement standard tests in the recent literature and estimate the degree of bias in my VA estimates from omitting parent characteristics and lagged measures of cognitive and non-cognitive

[4]Similarly, Chamberlain (2013) finds that predictions based on test score effects have small predictive power for college attendance.

[5]For instance, Chetty et al. (2014a) find that a one SD improvement in teacher VA raises normalised test scores by approximately 0.14 SD in math. See Section 3.2 for a full description of the results.

skills (e.g. Chetty et al., 2014a). I find that the selection of pupils to teachers based on cognitive and non-cognitive VA is only limited in my database: the selection bias from omitting parent characteristics such as parents' education, financial difficulties, marital status, mother's age at birth, and employment history is at most 1% for both outcomes. Similarly, the selection bias from omitting additional lagged measures of pupils' cognitive and non-cognitive ability is at most 2%.

The validity of my empirical results as evidence regarding the effects of teachers on pupils' non-cognitive skills might also depend on how pupils' non-cognitive skills are measured. Non-cognitive skills in the ALSPAC data are reported by parents and teachers. Results based on teacher-assessed non-cognitive skills could be driven by how teachers answered the questionnaire rather than "true" effects on pupils' non-cognitive skills. To obtain robust estimates of teachers' VA which account for teachers' reporting bias, I replicate the main analysis using parent-assessed non-cognitive skills. I also develop two additional approaches: a principal component analysis and instrumental regressions, predicting teachers' reports using parents' reports of pupils' non-cognitive skills. The results indicate that the VA estimates are fairly robust to the use of these different methods. There are similarly large variations in teachers' VA estimates on non-cognitive skills.

I find that teachers' effects on test scores and non-cognitive skills are not strongly correlated, so that many teachers who increase non-cognitive skills do not raise test scores, and *vice versa*. The correlations range from 0.01 between teachers' VA on math and internalising behaviour and 0.02 between teachers' VA on math and externalising behaviour. In contrast, the correlation is positive and statistically significant (approximately 0.5) between teachers' VA estimates on non-cognitive skills.

Teachers' effects on both test scores and non-cognitive skills predict substantial effects on the probability of higher education attendance, future earnings and employment. I find that a one SD improvement in teachers' VA on non-cognitive skills raises the probability of higher education attendance at age 20 by approximately 0.7 percentage points, relative to a sample mean of 45%. Improvements in teachers' VA on non-cognitive skills also raise pupils' earnings. At age 20, the oldest age at which I currently have information on pupils' earnings, a 1 SD increase in teachers' VA on pupils' non-cognitive skills raises annual earnings by roughly 2%.[6] I also find that improvements in teachers' VA significantly reduce the probability of ever having been unemployed and increase the probability of being in full time job at age 20. Overall, including teachers' effects on non-cognitive skills significantly increase the predictive power of teachers' VA on long-run outcomes. This in-

[6]For comparison, a one SD increase in teacher VA on math test scores raise annual earnings by 2% as well.

dicates substantial long-run teacher non-cognitive effects on pupil outcomes - conditional on their effects on test scores.

Since the evidence indicates persistent teacher effects on long-run outcomes, my findings are hard to reconcile with the fact that teachers' impacts on test scores fade out very rapidly (Rothstein, 2010; Jacob et al., 2010; Chetty et al., 2014b). Turning the focus to non-cognitive skills, I find suggestive evidence that improvements in teachers' VA on non-cognitive skills not only raise long-run outcomes but also subsequent math test scores. A one SD increase in non-cognitive skill teachers' VA significantly raises math test scores 3 to 5 years after. These findings (i) suggest that teachers' VA on non-cognitive skills have more persistent effects over time and (ii) constitute a first piece of evidence that teachers' VA on non-cognitive skills reinforce teachers' VA on cognitive skills in subsequent years. In other words, having a teacher who increases pupils' non-cognitive skills in primary school is likely to increase academic achievement throughout the schooling process.

The ALSPAC data provide very rich information on teaching practices, including homework, assessments, incentives used, classroom organisation, and the teachers' sense of responsibility. I combine these variables into five categories of teaching practices, following a common and accepted terminology in the education literature: (i) instilment of knowledge and enhancement of comprehension; (ii) instilment of analytical and critical skills; (iii) instilment of capacity for individual study; (iv) instilment of social and moral behaviours and (v) individual treatments of pupils.[7] The results suggest that including teaching practices explains about 15% of the variation in teachers' ability to enhance both pupils' cognitive and non-cognitive skills. In addition, I show that teaching that emphasises the instilment of knowledge and comprehension, often termed "traditional"-style teaching, is negatively correlated with teachers' ability to increase pupils' non-cognitive skills. By way of contrast, the use of classroom techniques that endow pupils with analytical and critical skills and a capacity for individual study, ("modern" teaching), has some positive payoffs. In addition, the individual treatment of pupils, such as class activities by attainment groups, giving homework to pupils according to their ability, and providing individual reviews, has negative effects on teachers' ability to increase pupils' cognitive skills but positive effects on pupils' non-cognitive skills.

This paper contributes to several strands of the literature. First, the importance of teachers' effects on non-cognitive skills helps to explain the previous findings (Chamberlin, 2013; Chetty et al., 2014b) that the effects of test scores VA on long-run outcomes do not reflect the total effect of teachers. The importance of non-cognitive skills also offers a potential explanation for school interventions with test score effects that "fade-out" over time but have lasting effects on adult outcomes (Cascio and Staiger, 2012; Heckman et al.,

[7]See for instance Bloom (1956) and Lavy (2011).

2013). Second, the importance of teaching practices in explaining variations in teachers' ability to improve pupils' cognitive and non-cognitive skills complements the previous findings that traditional teacher characteristics (such as education and experience) are only little correlated with teacher VA estimates (Rivkin et al., 2005; Aaronson et al, 2007). More generally, this paper is one of the first to demonstrate that non-cognitive skills can identify teachers who have large influence on pupils' short-run and long-run outcomes - but are no more effective than average in improving math test scores. In doing so, my research complements the extensive literature that assesses teachers' effects on students' test scores (e.g. Hanushek, 1971; Rockoff, 2004; Rivkin et al., 2005; Aaronson et al., 2007; Kane and Staiger, 2008; Chetty et al, 2014a; Rothstein, 2014; Bacher-Hicks et al., 2014) as it provides evidence for teachers' effects that are not reflected in their test score VA. These findings have direct policy implications.

The rest of this paper is organised as follows. Section II presents the data. Section III describes a simple conceptual framework to illustrate how teachers' effects can be decomposed into teachers' effects on test scores and non-cognitive skills and presents the empirical strategy. Section IV analyses the short-run teacher effects. Section V investigates teacher VA long-run impacts. Section VI tests the potential mechanisms through which teachers can influence pupils' cognitive skills and non-cognitive skills. Section VII discusses further implications of this study and concludes.

II. Data

The unique detail and scope of the ALSPAC data are major strengths of this study. This section describes the ALSPAC data and then provides descriptive statistics.

2.1. ALSPAC

The ALSPAC survey is a UK birth cohort study that recruited over 14,000 pregnant women who were due to give birth between April 1991 and December 1992 in Bristol and its surrounding areas, including some of Somerset and Gloucestershire. These women and their families have been followed ever since.[8] The bulk of my analysis focuses on when the participants entered primary school. Because they were born between April 1991 and December 1992, they were assigned to three different school-year cohorts. School questionnaires in years 3 and 6 of primary school have been completed by parents, teachers and the children themselves. In addition, the ALSPAC children have been linked with the UK National Pupil Database which contains information on math and English national

[8] Please note that the study website contains details of all the data that is available through a fully searchable data dictionary. http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/

test scores. These ALSPAC data thus include a large set of information on pupil characteristics, family background, life events, classroom, teacher characteristics and school characteristics for about 10,000 pupils in primary school.

Pupil characteristics - A number of pupil characteristics are included: test scores, non-cognitive skills as well as family and pupil background measures. In most of previous studies using administrative data, these types of information are somewhat limited. I here have detailed information on the respondent's entire history and family background that allows me to control for past (and present) pupil heterogeneity that could affect pupils' school achievement: parental education, number of siblings, parental marital status, parental employment history, parental financial problems, and mother's age at birth.

In particular, these data include a history of previous test scores that can be used as controls for past performance. In order to measure pupils' achievement, I rely on pupils' test scores from two math tests,[9] administered by ALSPAC at the end of year 3 and the end of year 6 of primary school, when the pupil was aged 8 and 11, and from two national tests: Key Stage 1 (KS1) and Key Stage 2 (KS2),[10] administered in year 2 and year 6, when the pupil was aged 7 and 11. I limit my main 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 Table A1 provide results with English test scores instead of math test scores. Very similar findings are obtained.[11] Another argument is that math test scores seem to have more, or are often perceived to have more, predictive power than English scores for future productivity (e.g. Murnane et al, 1991; Grogger and Eide, 1995; Hanushek and Kimko, 2000).

Multiple math test scores are vital to control for the previous cognitive ability of pupils.

[9]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 contain 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.

[10]The Key State Assessments are two standardised 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 (aged 7). The second assessment, Key Stage 2 (KS2) is given to the pupils when they are in year 6 (aged 11). 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.

[11]A one SD improvement in teacher VA raises normalised English test scores by approx. 0.26 SD. This is a bit higher than previous estimates in the literature (see Chetty et al., 2014a). This is probability due to the fact that English test scores are measured only twice in my dataset, which makes the identification less robust.

I rely on a general form of the VA 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).[12]. 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 10,000 pupils. 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 Appendix Table A1. Again, similar findings are obtained, with a significant correlation of 0.8 between the two teacher quality estimates.

The key advantage of the ALSPAC data is that it also gives contemporaneous information on pupils' non-cognitive skills (in addition to academic achievement) in years 3 and 6 of primary school (when the pupil was aged 8 and 11). In particular, I rely on the Strength and Difficulties Questionnaire (SDQ) which is commonly used in developmental, genetic and clinical studies and gives a complete behavioural screening in the following five areas: conduct problems, hyperactivity and inattention, emotional symptoms, peer relationship problems and pro-social behaviour (Goodman, 1997). The questionnaire 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", "hast at least one good friend", "often lies or cheats", "has good attention span" and "saws tasks through to the end". Appendix B1 provides a detailed description of the SDQ questionnaire.

Following Goodman et al. (2010), I use two broader sub-scales, as in low-risk samples such as the ALSPAC respondents the five finer sub-scales may not be able to detect distinct aspect of pupil non-cognitive skills. The SDQ's emotional and peer subscales are combined into an "internalising" subscale (Internalising behaviour see below) and the SDQ's behavioural and hyperactivity subscales into an "externalising" subscale (Externalising behaviour see below). This provides me with two composite measures on whether the pupil has emotional issues or behavioural problems on 0-20 scales. I reverse these two scales so that higher values indicate better outcomes. For robustness checks, it is also possible to run the main analyses using the five SDQ scales separately (Appendix Table A1). 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 pupils' outcomes by measuring pupil outcomes from the perspective of both teachers and parents.

[12]See the timeline graphic in Figure A.1

This is of particular interest with subjective data. The information reported by teachers and parents has different advantages and disadvantages. Teachers' reports about pupils' internalising and externalising behaviours are useful because they provide information on pupil in-class outcomes that might differ from what parents perceive at home. On the other hand, teacher responses are also subject to bias. Teachers may answer about pupils' behaviours based on their own mental state or as a function of the class context.

School, classroom and teacher 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. This allows us 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 staff meetings, head-teacher's gender as well as the class size, the number of exclusions in class, the percentage of free school meal pupils in the class, the percentage of SEN statemented[13] pupils in the class, the percentage of pupils with home concerning problems in class, the 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 very detailed information on primary school teachers, including teacher's gender, experience at school, experience everywhere, year of certification, but also the teacher's Crown-Crisp Experiential Index (CCEI), Bachman self-esteem, job satisfaction, confidence in teaching and teaching style.[14] In particular, information on teaching practices is very detailed in year 3 and 6 of primary school and includes information on: homework (type, frequency, duration), assessments (written, individual discussions, etc.), incentives used (naming pupils, competition, etc.) and classroom organisation (class ability groups, class activity groups, etc.): these are listed in Appendix B2. I group the items under five categories that describe the teacher's pedagogical practices in the classroom: (i) instilment of knowledge and enhancement of comprehension; (ii) instilment of analytical and critical skills; (iii) instilment of capacity for individual study; (iv) instilment of social and moral behaviours;

[13]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.

[14]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 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", etc. Appendix B3 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)".

and (v) individual treatment of pupils. These categories of teacher pedagogical practices correspond to a common and accepted terminology in the educational-psychology literature (see for instance Bloom, 1956; Lavy, 2011). By relying on this categorisation, I avoid any arbitrariness in grouping the items in different categories, even though some may disagree with the appropriate placement of certain items. Based on these teaching categories, the data allow me to decompose teachers' VA estimates into different teaching practices and better understand the mechanisms through which teachers influence pupils' cognitive and non-cognitive outcomes.

Pupils are assigned to a class and 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 estimate teachers' VA estimates, 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 A2 includes descriptive statistics for the teacher variables available. The teacher file contains 1061 teachers in 217 primary schools in year 3 and year 6. There are on average 3 to 4 teachers per school in the database, 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, 32% of teachers are observed twice and 22% are observed three times. The average number of pupils observed per teacher per year is 14.

Long-term outcomes - Because the ALSPAC data is a birth cohort study, pupils are observed almost every years from birth to age 20. 37.5% of the pupils in the primary school sample are still observed at age 20. Hence this information can be used to analyse the long-term impacts of teachers.[15] I define pupils' outcomes in adulthood as follows.

<u>Higher Education Attendance</u>. Higher education attendance is as an indicator for being full-time or part-time in higher education at age 20. All colleges and universities as well as vocational schools and other post-secondary institutions are taken into account. Comparisons to administrative data records suggest that I capture the higher education enrollment rate accurately.[16]

<u>Earnings</u>. Information on earnings is available in the ALSPAC data at age 20. I measure earnings as the annual total take home pay (after tax and any national insurance).

[15] As with any large cohort survey, the usual attrition bias due to dropout applies. The participated parents did not always answer every single question in every questionnaire, which means that the sample size vary across years. In section 4, I perform analyses controlling for potential bias due to attrition.

[16] The UK department of education reports that the highest educational initial participation rate is about 48% in 2014/2015 in the UK. In the data, the average higher-education attendance rate is about 50%.

42% of individuals in the sample report having earnings at age 20 and 50% of individuals in the sample report earnings above £12,280 per year.

Ever been unemployed. Information on labour force status is available in the ALSPAC data at age 20. Ever been unemployed is an indicator for ever having been unemployed at age 20. 67% of individuals in the sample report never having been unemployed at age 20.

Full-time job. Similarly 33% of individuals in the sample report being in a full-time job at age 20. This does not include people in full-time education.

Math test scores. The ALSPAC data also provide information on subsequent math test scores and non-cognitive skills, following the primary school. In order to measure pupil achievement in later years, I rely on pupils' test scores from three national exams: Key Stage 3 (KS3), Key Stage 4 (KS4), and Key Stage 5 (KS5) administered at ages 14, 16 and 18. I also rely on parent-assessed internalising and externalising behaviours based on the SDQ questionnaire, measured at ages 14 and 16.

2.2. Summary Statistics

Table 1 shows the summary statistics for the main sample used to estimate teacher VA models in primary school. The mean age at which pupils are observed in primary school is 9.8 years. 12% of pupils are eligible for free school meal and 3% are pupils in special education. Regarding parent characteristics, 14% have had major financial difficulties since child birth and 70% of the mothers are currently working.

While my study focuses on only one Area - Avon - in the 2000s, the population of parents and children in ALSPAC is broadly similar to those of the rest of Great Britain. 14% of pupils were eligible for free school meals in the 2000s in Britain, and 3% were pupils in special education. 65% of the mothers were in the labour force.

If we examine the sample characteristics at child birth, 79% of mothers in ALSPAC lived in owner occupied accommodation in 1991, 79% were married and 2% were non-white. In 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 ALSPAC children and published national figures shows that they are very similar measures. Overall the sample is broadly representative of the national population of mothers with children born in the 1990's, although higher socio-economic status groups as well as people of white ethnicity are over-represented compared to the national population.[17]

Table 2 presents the means and standard deviations of pupil test scores, and teachers'

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

and parents' answers for internalising and externalising behaviours. The mean math test score in the sample is 62.5 with a standard deviation of 20.8 on a 0-100 scale. Average internalising and externalising behaviours reported by parents and teachers are similar for the full sample: 17 out of 20 for internalising behaviour and 16 out of 20 for externalising behaviour. However, Table 2 reveals that parents' and teachers' answers are not strongly correlated: the coefficient of correlation is 0.3 for internalising behaviour and 0.5 for externalising behaviour. This suggests that using both parents' and teachers' responses with different potential reporting bias can paint a broader picture and improve our understanding of the role of teachers on pupil outcomes.

Appendix Table A4 further investigates the differences in teachers' and parents' reports. For over half of pupils, the correlation between teachers' reports and parents' reports is above 0.4. Overall, it seems that there are teachers who are better at assessing pupils' non-cognitive skills (i.e. who closely match with parents' reports) while others are not. In Appendix Table A4, I estimate the correlations for each pupil between teachers' and parents' reports and then regress these on teachers' characteristics and teaching practices. I find that teachers who have taught the pupil for longer, teachers who use class activity groups and teachers who report having the responsibility to help pupils develop in their own way, are those who report non-cognitive skills that more closely match the parents' reports.

Table 2 also reports unconditional correlations of test scores and internalising and externalising behaviours and reveals some interesting patterns. The first is that test scores in math and English are relatively strongly correlated with each other (correlation = 0.7) but are weakly correlated with internalising and externalising behaviours. Specifically, the correlations between internalising behaviour is 0.2 with math test scores and a bit under 0.2 with English test scores.[18] The analogous figures for externalising behaviour are slightly higher at 0.3 and 0.4. This suggests that while pupils who tend to have better math and English test scores also tend to have better non-cognitive skills, the ability to predict non-cognitive skills based on math test scores is relatively limited. In other words, pupils who score well on standardised tests are not necessarily those who have better emotional health, and many pupils who are not well-behaved have good standardised tests.

The second notable pattern is that internalising and externalising behaviours are slightly more correlated with each other. For example, the correlations between teacher-assessed internalising and externalising behaviours is 0.4 (slightly higher than the correlations between internalising on the one hand, and externalising behaviours with test scores on the other). Similarly, the correlations between parent-assessed internalising and

[18]Within-teacher correlations reveal similar patterns.

externalising behaviours is 0.4. This suggests that pupils who have good emotional health tend to be better behaved.

Finally, Table 3 looks at the correlations between short-run outcomes (math test scores, internalising and externalising behaviours in primary school) and long-run outcomes (higher education attendance, earnings, ever been unemployed and being in a full-time job). Most of the previous literature on schooling effects is based on higher test scores predicting better adult outcomes. To demonstrate that non-cognitive outcomes also matter, I show that both test scores and non-cognitive measures are correlated with long-run outcomes. Note that Table 3 only reports correlations and may not represent causal relationships. In particular, these correlations do not account for socioeconomic status, demographics and school characteristics. Table 3 shows that the correlation of higher education attendance with math test scores is about 0.4 and that with internalising and externalising behaviours between 0.1 and 0.2. Similarly, the correlations of earnings and labour market outcomes with math test scores is about 0.1 and 0.1 with internalising and externalising behaviours. Overall, this suggests that interventions improving both types of skills - cognitive and non-cognitive skills - can have positive effects on long-run outcomes.

III. Teacher Impacts on Pupil Non-Cognitive Skills

This section outlines the strategy used to estimate and predict teacher effects on pupil cognitive and non-cognitive skills in primary school. I then show the empirical results and suggest that these effects are robust to a number of tests.

3.1. A Model of Pupil Ability and Teacher Ability

I first define a simple model following Heckman et al. (2006) and Jackson (2012) that formalises the use of both cognitive and non-cognitive outcomes to measure overall teacher effects. The main insight of this model comes from moving from a single to a multidimensional model of pupils' ability. I assume that pupil ability is two-dimensional:

$$a_i = (a_{c,i}, a_{n,i}) \tag{1}$$

Here a_i is a pupil i's ability vector, where $a_{c,i}$, denotes cognitive ability and $a_{n,i}$ denotes non-cognitive ability.

Pupil i's cognitive and non-cognitive ability are potentially affected by teacher j. Each teacher j has a two-dimensional ability vector $t_j = (t_{c,j}, t_{n,j})$, where $t_{c,j}$ denotes how much teacher j affects her pupils' cognitive ability and $t_{n,j}$ denotes how much teacher j affects her pupils' non-cognitive ability.

The total ability of pupil i with teacher j can then be modeled as a function of both pupil i's ability and teacher j's ability vectors:

$$b_{ij} = a_i + t_j \tag{2}$$

The objective of this paper is to identify the difference in pupils' outcomes between teacher j with $t_j = (t_{c,j}, t_{n,j})$ and an average teacher with $t_j = (0, 0)$. Note that the teachers' estimates are normalised to be mean zero. I try to answer the following simple question: If a given classroom of pupils were to have teacher j with $t_j = (t_{c,j}, t_{n,j})$ rather than an average teacher with $t_j = (0, 0)$, how different would their average ability $\overline{b_{ij}}$ be?

3.2. Estimating Teacher VA

To estimate teacher effects on pupil cognitive skills and non-cognitive skills, I follow the standard practice in the literature and estimate teacher VA models following previous work such as Kane and Staiger (2008) and Chetty et al. (2014a), except that I also predict VA estimates on non-cognitive skills. The main idea is that teachers' VA are estimated using the average test scores and average non-cognitive skills of pupils that she taught in other years.

I assume that school principals assign each pupil i in school year t to a classroom c. Principals then assign a teacher j to each classroom c. For simplicity, assume that each teacher j teaches one class per year, as in primary schools.

Pupils' residuals. Within each grade-level (year 3 and year 6), I construct the test score and internalising/externalising behaviour residuals Y^*_{it}, by regressing the raw standardised pupils' outcomes Y_{it}, on a vector of covariates and teacher fixed effects. I control for the lagged dependent variable. I also control for pupil age, ethnicity, gender, health, indicators for special education needs, eligibility for free school meals, low birth weight, and number of siblings. I also include the following family controls: mother's education, father's education, family major financial difficulties, mother's age at child birth, parental marital status and mother's employment history. Finally, I also include the following class- and school-level controls: (i) class size, class-year means of the percentage of pupils eligible to free school meal, of pupils SEN statemented and of class exclusions; (ii) school size, school admission policy, frequency of staff meeting, gender of head-teacher; and (iii) school-cohort dummies.

The residual of pupil outcomes after removing the effect of observable characteristics is:

$$Y^*_{it} = Y_{it} - X_{it}\beta \tag{3}$$

where β is estimated using within-teacher variation from an OLS regression of the form:

$$Y_{it} = X_{it}\beta + \mu_j \tag{4}$$

Y_{it} refers to the math test scores, internalising or externalising behaviours of pupil i who is enrolled in year t (t=3 or t=6) in class c with teacher j, and μ_j is a teacher j fixed effect. As ALSPAC is a birth-cohort study and pupils are observed on numerous occasions after birth, it is possible to control for time-varying pupil characteristics, including lagged dependent variables, as well as family background. In addition, classroom characteristics, school characteristics, school-cohort and grade fixed effects allow us to control for classroom, school, cohort and grade characteristics that could drive pupil outcomes.

Selection issues. A key issue that VA estimates have to address is the potential non-random assignment of teachers to classrooms, i.e. how to identify "similar" classrooms for the counterfactual of what pupils' outcomes would have been with the assignment of a different teacher. In other words, that there is no selection of pupils to teachers within class. The specification (in equation (4)) addresses this issue in a number of ways: first, the VA 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 characteristics that may be correlated with cognitive and non-cognitive outcomes allows us to control for "non-school" factors that may account for differences in teacher VA. Third, including school, classroom characteristics, and grade dummies in the VA model allows us to compare outcomes within groups of pupils in the same type of school, and classroom, and in the same grade. This removes some of the influence of the selection to school and classroom on the estimated teacher effects. Note that, because I cannot observe teachers who switch schools in this dataset, I do not include school fixed effects in equation (4). Therefore, I cannot reject the possibility that some of teachers' VA might be attributed to the school.

Predicted VA estimates. I would like to compare the residual outcomes Y^*_{it} of pupils who are exposed to teachers with different ability. The simplest way to do this is to compare the class-level means of pupil residual outcomes in year t between teachers. Let $\overline{Y^*_{j,t}}$ denote the mean residual outcome in the class that teacher j teaches in year t:

$$\overline{Y^*_{j,t}} = \sum_{i=1}^{n} Y^*_{i,t} \tag{5}$$

Under random assignment of teachers to classrooms, these average residuals are consistent estimates of teacher j's effects on pupil outcomes in year t.

However, because there might be common shocks (for instance, sampling variation or classroom shocks) that affect Y^*_{it} that are unrelated to the teacher quality in year t, it is important that the estimated teacher effects in year t not be based on the pupils who are observed in year t. Doing so produces endogeneity mechanically. To address this issue, I follow a strategy very similar to that in Chetty et al. (2014a) to form a prediction of how much each teacher will improve her pupils' test scores or non-cognitive outcomes in a given year t, based on her performance in all other years (i.e., based on the test scores and non-cognitive skills of a different set of pupils). This method produces an estimate of the variability in a teacher's predicted effect that is persistent over time.[19]

I obtain the predicted effect of teacher j for the current year t based on the estimate of her effect in all other years in two steps:

Step 1: I regress the mean class-level outcome residuals in year t on class-level outcome residuals in other years:

$$\overline{Y^*_{j,t}} = \psi_1 \overline{Y^*_{j,1}} + ... + \psi_{t-1} \overline{Y^*_{j,t-1}} \tag{6}$$

Step 2: I use the estimated coefficients ψ_1, ..., ψ_{t-1}, to predict the VA in year t based on the mean pupil outcome residuals in other years for each teacher j.

$$\hat{\mu}_{j,t} = \sum_{s=1}^{t-1} \hat{\psi}_s \overline{Y^*_{j,s}} \tag{7}$$

Note that were performances in the past to be perfect predictors of current performance, then $\hat{\psi}$ would equal 1. However, because the mean residual outcomes are estimated with error, $\hat{\psi}$ is less than 1, so that the prediction "shrinks" the VA estimates toward zero. In other words, ψ parallels the shrinkage factor typical in empirical Bayesian analysis. The underlying idea of the empirical Bayesian approach is to multiply a noisy estimate of teachers' VA (e.g. $\overline{Y*_{jt}}$, the mean residuals of a teacher's pupils from a VA regression) by an estimate of its reliability. Less reliable estimates are shrunk back toward the mean (zero, since the teachers' estimates are normalised to be mean zero) to reduce the mean-squared error. Nearly all recent applications have used a similar approach to

[19]Table 2 provides VA estimates without implementing this adjustment. The results are broadly similar, except that the SD of the teachers' VA estimates are higher: 0.23 in math; and approx. 0.2-0.3 in non-cognitive skills.

estimate VA (Kane and Staiger, 2008; Jackson, 2012; Chetty et al, 2014a; Rothstein, 2014; Bacher-Hicks et al., 2014).

It is important to note that $\hat{\mu}_{j,t}$ simply represents the best linear predictor of the future outcomes of pupils assigned to teacher j in my data. This prediction does not necessarily capture the causal effect of teacher j on pupils' outcomes in year t, because the prediction could be driven by the sorting of pupils to teachers based on unobservable factors.

3.3. Results: Teacher Effects on Pupil Non-Cognitive Skills

3.3.1. The Role of Teachers

Before presenting my empirical estimates of teachers' VA, I begin with an overview of the explanatory power of school, classroom and teacher effects in explaining pupils' cognitive and non-cognitive skills. More specifically, Table 4 shows the R-squared and adjusted R-squared values from a series of regressions of the different dependent variables (math test scores, internalising behaviour and externalising behaviour) on school, classroom characteristics and teacher dummies (as in equation (4)). The first column for each dependent variable is based on a specification with only pupil characteristics, family background, lagged dependent variable, school-cohort and grade dummies. 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 reveal a number of interesting features. First, teacher fixed effects are significant predictors of pupils' math test scores, and internalising and externalising behaviours in years 3 and 6 of primary school, when the pupil was aged 8 and 11. The p values for F tests of the joint significance of the teacher fixed effects all fall below 0.01. Comparing columns (2) and (3), the inclusion of teacher fixed effects increases the explanatory power by 18 percentage points for math test scores, 20 percentage points for internalising behaviour and 13 percentage points for externalising behaviour. Second, the inclusion of school rather than teacher fixed effects reduces the explanatory power by 7 percentage points for math test scores, 8 percentage points for internalising behaviour and 6 percentage points for externalising behaviour. This indicates that much of the variation in teacher quality exists within rather than between schools. Last, pupil characteristics, family background, school-cohort, grade dummies and lagged dependent variables explain 27% of the variation in math test scores, 10% of the variation in internalising behaviour and 24% of the variation in externalising behaviour, indicating a considerable influence of "non-school factors" on pupil outcomes.

3.3.2. VA Estimates on Pupils' Non-Cognitive Skills

Table 5 presents details on the distribution of the teachers' VA estimates, specifically the SD and the 10th, 25th, 50th, 75th and 90th percentiles. These are expressed in standard deviation on the sample distribution of math test scores, internalising and externalising behaviours in year 3 and year 6. The empirical distributions of the teachers' VA estimates are also plotted in Figures 1-6.

For all the dependent variables, the SD of the teachers' VA estimates is quite high, so that variations in teacher quality can potential have a large impact on pupil outcomes. The SD of the teachers' VA estimates is 0.13 of a SD in pupil performance in math, 0.22 in internalising behaviour and 0.12 in externalising behaviour in primary school. The results suggest that moving one SD up the distribution of teachers' VA estimates is expected to raise math test scores by about 3 points on a 0-100 scale, internalising behaviour by 1 point on a 0-20 scale and externalising behaviour by 0.5 point on a 0-20 scale.

Furthermore, the gap between the 75th percentile and 25th percentile teacher is between 0.14 SD and 0.24 SD. This means that having a teacher at the 75th percentile of the quality distribution versus the 25th percentile is again associated with 3 points higher score in math on a 0-100 scale, 1 point higher in internalising behaviour scores and 0.5 point higher in externalising behaviour scores on a 0-20 scale.

These estimates of teacher effectiveness for math test scores are in line with those reported in Rockoff (2004), Rivkin et al. (2005), Aaronson et al. (2007), Kane et al. (2008) and Chetty et al. (2014a). Rockoff (2004) reports a 0.10 SD gain from a one SD increase in teacher quality from two New Jersey suburban school districts. Rivkin et al. (2005) lower bound estimates suggest that a one SD increase in teacher quality increases student achievement by at least 0.11 SD. In Aaronson et al. (2007), a one SD increase in teacher quality over a full year implies about a 0.15 SD increase in math test score gains. In Chetty et al. (2014), the SD of teachers' VA estimates is 0.14 in math in elementary school.

These results also provide the first estimates of teachers' VA on pupils' non-cognitive skills. Although there have been previous studies attempting to evaluate teachers' effects on non-test score outcomes, they all rely on proxies for non-cognitive skills. For instance, Jackson (2012) reports that a one SD increase in teacher quality decreases suspensions by 0.15 SD. In addition, as it is the case here, he finds that teachers have on average larger effects on non-cognitive outcomes (measured by a combined measure of absences, suspensions and on-time grade progression) than on math or English test scores. Similarly, Araujo et al. (2016) find that a one SD increase in teacher effectiveness within one classroom is associated with a 0.07 SD growth in students' executive function scores, which measure a child's ability to regulate her thoughts, actions and emotions. However, they also show that a one SD increase in teacher quality led to a higher increase in language and math test scores (0.11 SD) and conclude that the effects' sizes are larger for cognitive than for non-cognitive skills. Such results hint a possible cause of such disparate findings: the dependent variables are different. My results offer the first estimates of teachers' VA, which rely on a number of pupils' non-cognitive skills, including conduct, hyperactivity and inattention, emotional health, peer relationship and pro-social behaviours, and which use a well-accepted behavioural screening test (the SDQ).

3.3.3. Comparing VA Estimates Across Models

In this subsection, I conduct several robustness checks in order to address potential reservations about the above estimates. Each row of Table 6 considers a different VA specification. Table 7 also reports correlations between the VA estimates obtained from each model and the baseline estimates.

The first row of the table replicates the baseline VA model as a reference. In row 2, I replicate the main specification controlling in each model by prior test scores and prior internalising/externalising behaviour scores simultaneously. This specification controls more extensively for prior achievement and thus can address the suspicion that pupils are purposely placed into certain schools/classrooms or with certain teachers based not only on their previous math test scores but also their previous non-cognitive skills. In practice, I obtain very similar results: a one SD improvement in teacher VA raises math test scores by 0.13 SD; a one SD improvement in teacher VA raises internalising behaviour by 0.22 SD and a one SD improvement in teacher VA raises externalising behaviour by 0.12 SD.

Row 3 tests whether the previous estimates were sensitive to the use of teachers' reports for pupils' non-cognitive skills. I replicate the main specification using parents' reports instead. I find that a one SD improvement in teacher VA raises internalising and externalising behaviours by 0.03 SD and 0.06 SD, respectively. Note that the estimates are significantly lower than the ones based on teachers' reports of internalising and externalising behaviours. One might argue that both teachers' and parents' reports of pupils' non-cognitive skills suffer from measurement errors. An alternative strategy would then be to perform a principal component analysis, using the latent component of these two variables to measure pupils' internalising and externalising behaviours. Row 4 reports the results. I find that a one SD improvement in teacher VA raises internalising and externalising behaviours by 0.12 SD and 0.29 SD, respectively. Another strategy is to instrument teachers' reports using parents' reports assuming that teachers' reporting biases and parents' reporting biases are not correlated. Row 5 reports the results. I find that a one SD improvement in teacher VA raises internalising and externalising behaviours by

0.07 SD and 0.07 SD, respectively.[20]

Another potential concern with these teachers' VA estimates, is that they might be biased when they are based on small populations and hence might again suffer from measurement errors (Kane and Staiger, 2002; Aaronson et al., 2007). For instance, Aaronson et al. (2007) find that roughly 30% of the SD in estimated teacher quality is due to sampling error. In order to test this, I successively raise the minimum number of pupils to identify an individual teacher to 5, 6, 7, 8, 9 and 10. Rows 6-11 reports the results. Overall, the results are robust to this statistical reliability test. However, another bias emerges - when restricting the sample to teachers with at least 10 pupils observed per year - we are left with very few teachers (33).

Overall my VA estimates are robust to these different checks. However, they also highlight that teachers' VA on pupils' non-cognitive skills can vary depending on the measure used. The correlations reported in Table 7 show that teachers' VA regarding internalising behaviour using teachers' reports and parents' reports are not strongly correlated (0.14). Similarly, teachers' VA on externalising behaviour show a correlation of 0.28. Teachers' VA using principal component analysis are more highly correlated with teachers' VA on teachers- and parents-assessed internalising and externalising behaviours. The correlation is between 0.58 and 0.71. Finally, the correlations with teachers' VA based on instrumental variables range from 0.19 to 0.99 (see Table 7).

3.3.4. Estimating Pupil Sorting using Parent Characteristics and Lagged Scores

Another concern is that these teachers' VA estimates will be biased due to the selection of teachers to pupils, that is pupils may be purposely placed into certain schools/classrooms or with certain instructors based on their learning potential or behavioural characteristics. This problem has been dealt with to a certain extent as I control for an extended set of pupil, family, classroom, and school characteristics. In this subsection, however, I assess the extent to which pupils may be sorted in the estimation sample, according to observable characteristics such as parent characteristics and additional lagged outcomes.

Parent characteristics - In order to test this, I generate predicted test scores and non-cognitive scores for each pupil based on parent characteristics (e.g. father's education, mother's education, family major financial difficulties, mother's age at child birth, parental marital status and mother's employment history) and regress the predicted scores

[20] First stage regressions indicates coefficients of 0.33 and 0.42, strongly significant at the one percent level between teachers' reports and parents' reports.

on teachers' VA for all three pupil outcomes. If there is no selection of pupils to teachers, there would be no systematic relationship between predicted outcomes and predicted teachers' effects.

Table 8 reports the results. There is little evidence of positive selection for test-score VA and non-cognitive factors VA. The coefficients are respectively 0.015, 0.008 and 0.010. This implies that the degree of bias due to selection on these parent characteristics is 1.5%, 0.8% and 1%, respectively.[21]

Another way to assess the degree of selection on these parent characteristics is to control for the same parent characteristics when estimating the impact of teacher VA on pupil outcomes. Column 3 of Table 8 shows the results. The coefficients on VA are respectively 0.944, 0.914 and 1.018. The difference between the point estimates in columns 1 and 3 is roughly 0.010. These differences coincide exactly with the estimates reported in column 2.

Intuitively, the degree of bias due to selection on parent characteristics is very small for several reasons. First, variations in test scores, internalising and externalising behaviours, that correlate with parent characteristics are captured by lagged ability and other controls such as school and classroom characteristics. In other words, pupils from "better observable characteristics" families have higher test scores and better non-cognitive skills not just in the current year but also in the previous school year. Thus previous scores and previous non-cognitive skills capture a large portion of the variation in family characteristics. Second, the correlation between teachers' VA estimates and parent characteristics is small and between 0.02-0.05 (Appendix Table A3). This means that the variation in teachers' VA after controlling for X_{it} is essentially unrelated to parent characteristics.[22]

Prior Ability - Another potential source of bias relates to prior test scores and prior non-cognitive skills (as in Rothstein, 2010 and Chetty et al. 2014a). One might wonder whether controlling for additional lags substantially affects VA estimates once I control for $Y_{i,t-1}$. I assess the bias due to sorting on lagged outcomes using the same approach as with parent characteristics. Panel B replicates Panel A of Table 8 using predicted score residuals based on pupils' outcomes at the entry of primary school. The coefficients on teachers' VA are 0.011, 0.014 and 0.020, respectively. I conclude that the bias due to omitting additional lagged pupils' cognitive and non-cognitive skills is small. The same two explanations as above apply.

Overall, this suggests that selection on two important predictors of test scores and

[21] Excluding pupils in private school provides similar results.

[22] See Chetty et al. (2014a), Rothstein (2014) and Bacher-Hicks et al. (2014) for further discussion on forecasting bias.

non-cognitive skills that are usually excluded from the baseline VA models - parent characteristics and additional lagged outcomes - assure negligible bias in the baseline VA estimates on both pupils' outcomes. There might be other sources of selection based on unobservable characteristics. However, using a quasi-experiment, Chetty et al. (2014a) have found that bias due to sorting on unobservables is minimal in models which control for lagged test scores. One can expect similar conclusions in models which control for lagged non-cognitive skills.

3.3.5. The relationship between Teacher Effects on Cognitive Skills and Teacher Effects on Non-Cognitive Skills

Having established that teachers have significant effects on test scores and non-cognitive skills, this section documents the relationships between these estimated effects. Do teachers who improve math test scores also improve pupils' internalising and externalising behaviours? This question has considerable implications for how teachers are evaluated. A teacher who is good at developing pupils' non-cognitive skills, but not efficient at increasing their test scores, might be rated as ineffective, thus undervaluing her contribution to pupils' learning.

To get a sense of whether teachers who improve test scores also improve other outcomes, I calculate the correlations between the predicted teacher effects for math test scores, and internalising and externalising behaviours. The results are reported in Tables 9 and 10. The bootstrapped standard errors appear in parentheses.

I find that teachers with higher math test score effects are associated with better internalising and externalising behaviours, but that the correlation is only small. The correlation between teachers' VA's on math test scores and internalising behaviour is 0.01. Similarly, the correlation between teachers' VA's on math test scores and externalising behaviour is 0.2. This indicates that while teachers who raise test scores may also be associated with better non-cognitive outcomes, most of the effects on non-cognitive outcomes are unrelated to the effects on test scores.[23]

By contrast, the effects on internalising behaviour are more highly correlated with the effects on externalising behaviour, with a correlation of 0.5, consistent with rather high correlations between internalising and externalising behaviours (see Table 2). The results from parent-assessed behaviours, principal component analysis and instrumental variables are similar (see Table 10). Teacher effects on math test scores are not strongly correlated with teacher effects on internalising and externalising behaviours and the relationship between teachers' effects is robust to the measure used to capture pupils' non-cognitive

[23]Taking into account measurement errors and multiplying the correlation estimates by the inverse of the signal-to-noise ratio provide similar results. I obtain correlations of 0.04 and 0.04 respectively.

skills. Overall, teacher effects on test scores are weak predictors of teacher effects on non-cognitive skills. This might suggest that teachers who raise test scores are not the "same" as teachers who increase non-cognitive skills. In other words, teacher test score effects might measure certain skills, and teacher effects on non-cognitive skills might measure a largely different but potentially important set of skills.

The validity of my interpretation of these results, however, depends on whether confouding mechanisms can produce the same findings. I consider two alternative explanations: (i) small correlations between teacher effects simply mirror the small correlations between pupils' cognitive skills and non-cognitive skills and (ii) weak correlations between teachers' effects are due to systematic bias in how teachers/parents report pupils' cognitive and non-cognitive skills. The former builds on the hypothesis that teachers who are good at improving math test scores may have little effects on pupils' non-cognitive skills, as a result of an increase in math test scores being little correlated with an increase in non-cognitive skills. The reverse holds for teachers who are good at improving pupils' non-cognitive skills. While this explanation is consistent with the small correlation between teachers' VA estimates, it is also possible that teachers have independent effects on both pupils' outcomes. The latter posits that low correlations between teachers' VA reflect low correlations between pupils' cognitive and non-cognitive skill reports. However, this explanation cannot account for the finding that weak correlations between teachers' VA are robust to the measure of pupils' non-cognitive skills used (e.g. teachers' and parents' reports).

IV. The Long-Run Impacts of Teachers' Ability to Improve Non-Cognitive Skills

This section examines the long-term impacts of teachers' VA and compares the outcomes of pupils who were assigned to high math teacher VA versus high non-cognitive skills teacher VA. If there is a small correlation between a teacher's ability to increase cognitive skills and to increase non-cognitive skills, we would like to know which type of teacher is best at improving pupils' lifetime outcomes, and so in which type of teachers school systems should invest.

I estimate the long-term impacts of teachers' VA based on cross-section comparisons across classrooms. I thus compare the outcomes of pupils who were assigned to teachers with different VA, controlling for a rich set of observable characteristics. I implement this approach by regressing long-term outcomes on the test score VA estimates and the non-cognitive VA estimates described in the sections above. The identification assumption underlying this approach is selection on observables: the unobserved determinants of outcomes in adulthood have to be unrelated to teachers' VA conditional on observable characteristics. Although this is a very strong assumption, Chetty et al. (2014b) have shown that this approach closely matches the quasi-experimental estimates, supporting its validity.[24]

4.1. Predicting Long-Run Effects

I present a simple empirical model of pupils' long-term outcomes as a function of their teachers' VA. The model is described with the higher-education attendance rate as the main dependent variable. Throughout the paper, I also replicate the analysis for other adult outcomes such as earnings, the probability of never having been unemployed and being in full-time job at age 20.

Let HE denote pupil i's higher education attendance in the future. Throughout the analysis, I focus on the probability of higher-education attendance residuals after removing the effect of observable characteristics. I estimate the higher education attendance residuals, HE^*_{it}, using the following equation:

$$HE_{it} = X_{it}\beta + \mu_j + \epsilon_{it} \tag{8}$$

where μ_j is a teacher fixed effect and X_{it} is a vector of baseline controls including pupil, family, school and classroom characteristics.

I then model the relationship between the higher-education attendance residuals and teachers' VA in year t using the following specification:

$$HE^*_{it} = a + \kappa \frac{\mu_{jt}}{\sigma(\mu_{jt})} + u_{it} \tag{9}$$

Here $m_{jt} = \mu_{jt}/\sigma_\mu$ denotes teacher j's "normalised VA" (i.e. teacher quality scaled in SD units of the teachers' VA distribution). The parameter κ represents the impact of one SD increase in teacher's VA on higher education attendance.

There are several important aspects when interpreting this parameter κ. For example, teachers' VA on both cognitive and non-cognitive skills can affect other educational inputs, that will in turn affect higher-education attendance. For example, parents might behave differently depending on changes in teacher quality, or higher VA teachers might be assigned to higher-achieving pupils. Such selection bias would lead us to overestimate the

[24] In their paper, Chetty et al. (2014b) exploit teacher turnover as a quasi-experimental source of variation in teacher quality. Building on this idea, they estimate teachers' impacts by regressing changes in mean adult outcomes across consecutive cohorts of children within a school on changes in the mean VA of the teaching staff. Seel also Rothstein (2014) and Hacher-Hicks et al. (2014).

impacts of higher teacher VA in year t holding fixed the quality of teachers in subsequent grades.[25][26].

There is also the concern that higher teacher quality would increase both pupils' math test scores and pupils' non-cognitive skills, and would have beneficial effects on pupils' long-term payoffs through these two channels (Chetty et al., 2014b). Without controlling for teachers' VA on non-cognitive skills, this would lead us to attribute the impacts of improving teacher quality in math on higher education attendance, while the positive effect might instead be driven by combined effects on pupils non-cognitive skills. It is hence important to identify the impact of having a higher VA teacher in math holding fixed teacher quality in internalising and externalising behaviours and *vice versa*. This will tell us whether higher-education attendance is more likely to be correlated with having a higher VA teacher in math or a higher VA teacher in non-cognitive skills.

Note that it is also important to check that attrition does not affect the long-term estimates of teachers' VA. Only 37.5% of the pupils in the primary school sample are observed at age 20. Attrition can be problematic if pupils who remain in my sample are those who had a high (or low) teacher VA in test scores and non-cognitive skills. To control for this, I estimate logit regressions for the probability of dropping out of the sample by age 20. The independent variables in this attrition equation are a vector of pupil and family characteristics in primary school, including pupil's gender, ethnicity, free school meal eligibility, special education needs, low birth weight, number of siblings, father's education, mother's education, family financial difficulties, mother's age at child birth and mother's employment history. I use the outcomes from the logit attrition regression to calculate inverse-probability weights and include them in my main specifications: these give more weight to observed individuals who have similar characteristics to those who are likely to attrit from the study.

4.2. Results: Long-run Impact of Teachers' Ability to Increase Non-Cognitive Skills

I find that both teachers' influence on pupils' math test scores and teachers' influence on pupils' non-cognitive skills have substantial impacts on adult outcomes. In addition, these effects are to some extent independent. A one SD improvement in teachers' VA for internalising and externalising behaviours raises the probability of higher education attendance at age 20 by 0.70 percentage points, relative to a sample mean of 45%. Simi-

[25]For instance, the within-pupil correlation between having a high math teacher VA in year 3 and in year 6 is 0.08. Similarly this is 0.05 for internalising and externalising behaviour.

[26]This is possible to replicate the analysis including secondary school fixed effects. The results do not change qualitatively

larly, a one SD improvement in math teacher VA in primary school raises the probability of higher education attendance at age 20 by 0.91 percentage points. In addition, pupils who were assigned higher VA teachers in internalising and externalising behaviours have higher earnings. The same holds for pupils who were assigned higher VA teachers for math test scores. I also find that improvements in teacher quality on both pupils' test scores and pupils' non-cognitive skills reduce the probability of ever been unemployed and increase the probability of being in a full-time job at age 20, with an effect that is slightly larger for teachers' VA on pupils' non-cognitive skills.

4.2.1. Higher Education Attendance

I begin by analysing the impact of teachers' test-score VA and teachers' non-cognitive skills VA on higher-education attendance at age 20, the oldest age at which I have information on educational achievement in my sample. Figures 7-8 plot residual higher-education attendance rates for pupils at age 20 vs $\hat{m_{jt}}$, the primary school estimates of their teacher's VA. To obtain this graph, I first residualise higher-education attendance rates with respect to the control vectors, X_{it}, using within-teacher variation to estimate the coefficients on the controls as described above. I then divide the VA estimates $\hat{m_{jt}}$ into 20 equal-sized groups (vingtiles) and plot the mean of the higher education attendance residuals in each bin against the mean of $\hat{m_{jt}}$ in each bin. The regression coefficients and standard errors reported in these figures are estimated using equation (9), with standard errors clustered by school-cohort.

Figure 7 provides evidence that being assigned to a higher VA teacher in math in primary school raises a pupil's probability of attending higher-education significantly. The null hypothesis that teachers' VA has no effect on higher education attendance is rejected with a t-statistics above 10 ($p < 0.001$). A one SD increase in a teacher's test score VA in primary school increases the probability of higher-education attendance at age 20 by 0.91 percentage points, relative to a mean higher education attendance rate of 45%. These results are in line with Chetty et al. (2014b)'s findings. In their paper, they find that a one SD increase in teacher's test score VA in a single grade, increases the probability of college attendance at age 20 by 0.82 percentage points, relative to a mean college attendance rate of 37%.

Figure 8 also shows that being assigned a higher VA teacher in non-cognitive skills in primary school raises a pupil's probability of attending higher-education significantly. A one SD increase in a teacher's non-cognitive skills VA increases the probability of higher-education attendance at age 20 by approx. 0.70 percentage points.

I test the validity of the estimates to the use of alternative measures of non-cognitive skills in Table 11. Panel A of Table 11 replicates the specification with teacher VA estimated on teachers' reports in Figures 7-8 as the reference. Panel B replicates Panel A, using parents' reports. The estimates in Panel B are quite similar in magnitude to those in Panel A, although the coefficients are not statistically significant for internalising and externalising behaviours. Panels C and D replicates Panel A using principal component analysis and instrumental variables. Again, the coefficients do not change qualitatively. These alternative measures of non-cognitive skills then provide similar estimates of κ suggesting that the long-run estimates are not biased by how teachers and parents assess pupils' non-cognitive skills.

If both teachers' VA in math test scores and in non-cognitive skills affect pupil long-run outcomes, is it better to be assigned a high teacher VA in math or non-cognitive skills? Column 4 replicates Column 1, adding teacher VA in math test scores and teacher VA in non-cognitive skills in the same regression. The teacher VA in non-cognitive skills consists of the teacher-year level means of teacher VA in internalising behaviour and teacher VA in externalising behaviour (as in Figure 8). The estimates in Column 4 indicate that teachers' influence in both math test scores and non-cognitive skills significantly predict pupils' higher-education attendance and that their effects are not substituable: the coefficients on κ barely change when teacher VA in non-cognitive skills is added. The two coefficients - on teacher VA in math and non-cognitive skills - are significant in each panel. These results suggest that the two competences are complements and policy makers should aim to improve teachers' ability to increase both test scores and non-cognitive skills in order to increase higher-education attendance rates.

4.2.2. Earnings

I now turn to earnings at age 20, the only age at which I have information on pupils' future earnings. Although pupils with earnings in full-time paid jobs are a rather selected sample, this can provide a first piece of information on teachers' VA impacts on pupils' future earnings. In addition, early earnings are found to be good predictors of earnings at later ages (Haider and Solon, 2006).

Table 12 reports the results. Higher VA teachers in both math and non-cognitive skills have significant impacts on earnings, with the null hypothesis of $\kappa = 0$ rejected with a p-value below 0.01 in each case. A one SD increase in teachers' VA increases earnings at age 20 by 2% for math and 2% for non-cognitive skills of mean earnings in the regression sample. This is in line with Chetty et al. (2014b)'s findings who show that a one SD increase in teacher VA in a single grade increases earnings at age 28 by 1.7% of mean earnings in the regression sample.

Panels B-D evaluate the robustness of these estimates to the use of alternative measures of non-cognitive skills. These specifications mirror Panel A of Table 12, but use teacher VA based on parent-assessed non-cognitive skills, principal component analysis and instrumental variables as the independent variables. As with higher-education attendance, using these alternative measures has small qualitative effects on the point estimates, supporting the idea that long-term estimates are not biased by how teachers and parents assess pupils' non-cognitive skills, except for the coefficients on internalising behaviour which are no longer statistically significant. The largest of the three estimates implies that a one SD increase in teachers' VA in non-cognitive skills raises total take-home pay by 3%.

Column 4 adds teachers' VA in math test scores and teachers' VA in non-cognitive skills in the same regressions. Here teachers' influence on non-cognitive skills does not remain statistically significant when holding constant teachers' influence on math test scores to predict pupils' future earnings. This suggests that, at least for earnings, teachers' VA's in math are better predictors of future success than teachers' VA's in non-cognitive skills.

Extensive Margin - When analysing earning effects, one might want to distinguish between the extensive and intensive margins. In Table 13, I regress an indicator for having never been unemployed at age 20 on teachers' VA in math test scores, teachers' VA in internalising behaviour and teachers' VA in externalising behaviour, successively. A one SD increase in teachers' VA in math raises the probability of having never been unemployed at age 20 by 0.83 percentage points. A one SD increase in teachers' VA internalising behaviour raises the probability of having never been unemployed at age 20 by 0.96 percentage points. And a one SD increase in teachers' VA externalising behaviour raises the probability of having never been unemployed at age 20 by 1.97 percentage points.

Column 4 of Table 13 replicates the baseline specification in Columns 1-3, adding simultaneously teachers' VA in math test scores and teachers' VA in non-cognitive skills, where teachers' VA in non-cognitive skills is a mean indicator of teachers' VA in internalising behaviour and teachers' VA in externalising behaviour, as previously. Again, the effects do not change qualitatively from specifications where they are entered one by one. In addition, it appears in Panels, A, C and D, that teachers' VA in non-cognitive skills have *larger* effects than teachers' VA in math test scores. This indicates that while teachers' VA in math might have larger effects on the intensive margins, teachers' VA in non-cognitive skills have actually larger effects on the extensive margins. Including teachers' effects on non-cognitive skills increases substantially the predictive power of teachers' effects on the probability of never having never been unemployed by 7%.

Finally, Table 14 replicates the analysis using the probability of being in a full-time job at age 20 as the dependent variable. Similar results are obtained. Having a higher

VA teacher in primary school who improve pupils' cognitive and non-cognitive skills, increases the probability of being in full-time jobs at age 20 among those who are on the labour market by approx. 0.5 percentage point, with an effect that is slightly larger for teachers' VA in non-cognitive skills. This is independent of the effect of having a higher VA teacher in primary school on the probability of being in higher education at age 20.

Hence, teacher's influence on both math test scores and non-cognitive skills are important for pupils' long-term outcomes. These have significant effects on education and labour market outcomes at age 20. In addition, their effects appear to be complementary, and in some cases larger for teachers' VA non-cognitive skills. These results complement the previous literature finding that cognitive skills and non-cognitive skills predict a variety of adult outcomes, including academic achievement, employment and financial stability (see Heckman et al. (2015) for a review). Researchers have also found that non-cognitive skills are more predictive of long-term outcomes than are test scores (Chetty et al., 2011; Heckman and Rubinstein, 2001; Lindquist and Vestman, 2011; Mueller and Plug, 2006).

4.2.3. Fade-out or Persistent Effects?

The final set of outcomes I consider are teachers' impacts on test scores and non-cognitive skills in subsequent years. Figure 9 plots the impacts of teachers' VA (in math test scores and non-cognitive skills) on subsequent math test scores at $t+3$, $t+5$ and $t+7$. See Table 15 for the underlying coefficients. To construct this figure, I residualise raw test scores $Y_{i,t+s}$ with respect to the baseline controls using within-teacher variation and then regress the residuals $Y^{*}_{i,t+s}$ on $\hat{\mu_{jt}}$. I scale teachers' VA in units of pupils' outcome SDs in these estimates - by using μ_{jt} as the independent variable instead of m_{jt} - to facilitate the interpretation of the regression coefficients, which are plotted in Figure 9. The coefficient for math teacher VA at $s = 0$ is not statistically different from 1. The results then suggest that teachers' impacts on math test score fade out rapidly in subsequent years. Again these results align with existing evidence that improvements in education raises contemporaneous scores, then fade out in later years, only to reemerge in adulthood (Deming, 2009; Heckman et al,. 2010; Chetty et al., 2014b).

What about teachers' non-cognitive VA and subsequent math test scores? The story is different here. There is suggestive evidence that teachers' impacts on non-cognitive skills persist in subsequent years. The coefficients are positive and statistically significant at $s = 3$ and $s = 5$, with a positive trend. This finding is very interesting as it directly addresses the question of a reemergence effect in adulthood. In addition, it adds to the idea that teachers' non-cognitive skills effects are more predictive of long-term outcomes than are test score effects because (1) they are more persistent and (2) they reinforce the influence of teachers on math test scores. To confirm these results, it would be of great interest to replicate the analysis with more frequent math test scores in subsequent years.

V. Explaining Teacher VA Estimates

The estimates in the previous sections show that (i) teachers influence both pupils' academic skills and pupils' non-cognitive skills; (ii) teachers vary in their ability to enhance pupils' cognitive and non-cognitive skills and there seems to be a weak correlation between a teacher's ability to increase cognitive and non-cognitive skills; and (iii) long-run outcomes are improved by a *combination* of teacher's ability to increase cognitive skills and non-cognitive skills.

One can then ask which teacher traits are associated with improvements in pupils' cognitive and non-cognitive skills. In this section, I decompose the teachers' VA estimates based on cognitive skills and non-cognitive skills into a range of teacher characteristics and teaching practices to better understand the relationship between teachers' total estimated impacts on pupils' outcomes and teaching capacities. If teachers' VA on test scores are weak predictors of teachers' VA on non-cognitive skills, this may suggest that teachers hence recourse to different techniques to improve pupils' test scores and non-cognitive skills (Jackson, 2012; Blazar and Kraft, 2015; Gershenson, 2016).

5.1. Decomposition of Teacher VA Estimates

I model the relationship between teachers' VA, teacher characteristics and teaching practices using the following linear specification:

$$\mu_{jt} = a + T_{jt}\lambda + TP_{jt}\gamma + u_{jt} \tag{10}$$

where μ_{jt} refers to the estimated teacher j effect in year t. T_{jt} is a vector of teacher characteristics which includes gender, experience, self-esteem, confidence in teaching, job satisfaction and TP_{jt} is a vector of teaching practices. There are several advantages in using a two-step procedure (e.g. first estimating teachers' VA and then decomposing the teachers' VA into different teacher components). First, estimating teachers' VA in a first step allows for a more general specification than the one that could be made by considering teacher characteristics. Second, the first step estimates in equation (4) 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 (equation (10)). Changing the specification in the second step does not affect the estimates from the first step. Third, the two-step procedure allows us to consider both individual and aggregate error terms, which deals with the heteroscedasticity issues raised by Moulton (1990).

There are still some endogeneity issues. One legitimate concern is that teacher characteristics or teaching practices 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 be a quasi-natural experiment, with exogenous changes in school policy or teaching methods.[27] Another would be to instrument teacher satisfaction using exogenous life events. There is probably no way to reject such concerns definitively, but one test is to examine whether individual teacher characteristics (for instance, self-esteem, confidence in teaching, job satisfaction) and teaching practices are relatively stable over time and do not vary with the characteristics of pupils in the class. Appendix Table A5 reports the results. There is no evidence of significant changes in teacher practices over time. Note in addition, that because the teachers' VA estimates are based on the pupil that the teacher taught in other years, this bias could be fairly limited here.

5.2. Results: The Decomposition of Teacher VA Estimates

Table 16 reports the R-squared of different teacher characteristics in explaining teachers' VA to improve math test scores, internalising and externalising behaviours. All the teachers' VA estimates are based on the full specification described in previous sections.

5.2.1. Teacher's Gender and Experience

First and foremost, traditional observable characteristics - such as gender and experience - explain at most 2% of the total variation in teacher VA (based on all three pupil outcomes). This is consistent with previous work (Hanushek, 1971; Rivkin et al., 2005; Aaronson et al., 2007) finding a small relationship between teacher characteristics such as gender, experience, educational background and teacher ability to raise student achievement.[28]

In addition, Table 17 provides detailed information on the effects of teacher gender. It is notable that female teachers are associated with better teachers' VA for all three outcomes, with significant effects for internalising behaviour. Female teachers' VA in

[27] The ALSPAC data provides information on head teacher and school policy that can be exploited in future research.

[28] Hanushek (1971) finds no relationship between teacher quality and experience or master's degree attainment. Rivkin et al. (2005) also find no link between educational 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.

internalising behaviour is 0.02 SD higher than male teachers' VA in math test scores.[29] However the results are not robust to replicating the analysis using teachers' VA on parents' reports (Panel B of Table 17). Similarly, the coefficients on teacher experience are not statistically significant for all three outcomes.

5.2.2. Teacher's Non-Cognitive Skills

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 relate differently to teachers' VA's on math test scores and non-test score outcomes.

One potential question is the influence of teacher's non-cognitive skills on pupils' test scores, internalising and externalising behaviours. While individuals' non-cognitive skills and traits significantly influence academic and labour-market outcomes, teacher's non-cognitive skills may exert a significance influence on teacher quality. There is not much rigorous quantitative evidence regarding the effect of teacher's non-cognitive characteristics.

I test this hypothesis in Table 16, by estimating the effects of a number of teacher non-cognitive skills - including CCEI, Bachman self-esteem, job satisfaction and confidence in teaching - on teacher ability to improve pupils' outcomes. The results indicate that teacher emotional characteristics are significantly related to estimated teacher quality in pupils' cognitive and non-cognitive skills. Strikingly, they explain roughly 1.5% of the total variation in teachers' VA in maths, 2.7% of the total variation in teachers' VA in internalising behaviours and 4.3% of the total variation in teachers' VA in externalising behaviour. Even if these percentages look modest, this is twice the explanatory power of teacher gender and teacher experience for pupils' non-cognitive skills.

In Table 17, I then detail the effect of teacher's CCEI, Bachman self-esteem, job satisfaction and confidence in teaching on teacher quality.[30] A one SD increase in teacher emotional health (as measured by CCEI) translates into an increase in teacher quality of 0.04 SD for internalising behaviour and 0.02 SD for externalising behaviour. These relationships are all statistically significant. In addition, teacher self-esteem and teacher confidence increase teacher non-cognitive quality by approx. 0.1 SD. Finally, teacher job

[29]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.

[30]Note 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.

satisfaction has positive effects on teacher quality in math. Note that these results are to some extent robust to the use of alternative teachers' VA based on parents' reports. Panel B of Table 17 reports the results. While teacher CCEI is no longer significant, there is still a positive and significant effect of teacher confidence in teaching and teacher job satisfaction on teacher quality.

These are interesting findings, showing that teachers' non-cognitive skills are significant drivers of teacher quality. I believe that these results are a first piece of evidence that teacher's non-cognitive skills matter and go well beyond what has been shown 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's non-cognitive skills contribute more to explaining the variation in estimated teacher quality on pupils' non-cognitive skills than teacher gender or experience.

5.2.3. 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. Previous evidence on teaching practices is not conclusive and especially so far on pupil non-cognitive skills.

To analyse to what extent certain teaching practices in class are related to teacher quality and pupil performance in math and non-test score outcomes, I group the described teaching practices under five categories: (i) instilment of knowledge and enhancement of comprehension; (ii) instilment of analytical and critical skills; (iii) instilment of capacity for individual study; (iv) instilment of social and moral behaviours and (v) individual treatments of pupils. A complete set of information is available in year 3, less information is available in year 6 (see Appendix B2 for a full description).

Table 16 reports the R-squared values from estimating the effect of teacher practices on teacher's ability to improve pupils' outcomes. The results indicate that including both teacher characteristics and teaching practices explains up to 14% of the total variation in teachers' effects on math test scores, 13% of that on internalising behaviour and 16% of that on externalising behaviour. Again this is larger than the explanatory power attributed to traditional observable teacher characteristics, such as gender or experience.

Table 18 presents detailed results on the effect of teaching practices. They clearly show that certain teaching practices are correlated with higher teacher VA, but can have different effects on pupils' math test scores and non-cognitive skills. Table 18 reports the estimates for each category of teaching practice, controlling for teacher characteristics such as teacher's gender, emotional health and experience. In columns (1), (3) and (5), the estimates are from separate regressions in which each teaching practice enters as a single treatment variable. In columns (2), (4) and (6), the estimates come from one regression that includes all the teaching practice measures as multiple treatments. In Panel B, I replicate the analysis using teachers' VA estimates based on parents' reports.

Focusing first on the teachers' reports estimates of the effects of teachers' pedagogical methods in Panel A, most of the coefficients are not statistically significant at the 10 percent level. However, some interesting patterns emerge. First, the estimates of measures that capture elements of "traditional teaching practices", e.g. instilment of knowledge and enhancement of comprehension, have negative and significant effects on teachers' ability to enhance pupils' internalising behaviours. Similarly the estimated effect is negative on teachers' ability to increase pupils' math test scores. Another noteworthy feature of the estimates is the negative effects of more "modern teaching practices", e.g. instilment of analytical and critical skills and instilment of capacity for individual study on teachers' ability to increase pupils' math test scores. However, the coefficients are no longer significant when all the teaching practices are introduced at the same time. By way of contrast, there is suggestive evidence that these "modern teaching practices" are associated with better teachers' VA in pupils' non-cognitive skills. Panel B of Table 18 presents estimates based on parents' reports. The estimated effects are again strongly significant for modern practices. These specifications allow us to estimate the effects of teaching practices controlling for any teachers' reporting bias that might affect the correlation between teacher quality and teaching practices.

Row 5 presents estimates based on the individual treatment of pupils. This category includes whether teachers group children by attainment groups for classroom activities, whether teachers use competition in relation to academic work, whether teachers display high quality work as incentives, etc. The effects are negative for teachers' VA in math, although not significant, while they are positive on teachers' non-cognitive VA.[31] This seems to suggest 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. By way of contrast, grouping pupils by ability has detrimental effects on their academic achievement.

Overall the evidence in Table 18 suggests that three of the five teaching styles and methods tested have positive effects on pupils' non-cognitive skills, while traditional teaching practices have detrimental effects. The most important of these in terms of effect size is the indicator of the extent to which teachers make sure that their pupils have the capacity to study individually. When this measure increases by one SD, teachers' ability to

[31]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 a recent paper, Algan et al. (2013), have shown horizontal teaching practices (i.e. students work in groups) are positively correlated with student self-confidence and positive attitude.

increase pupils' internalising behaviour by 0.02 SD, which results in an increase of 0.01 SD in internalising behaviour. Alternatively, a one SD increase in traditional teaching methods, would decrease teachers' ability to increase pupils' internalising behaviour by 0.09 SD, which results in a decrease of 0.02 SD in internalising behaviour.

VI. Discussion and Concluding Remarks

Analysing teachers' effects on pupils' cognitive and non-cognitive skills, I have shown that teachers have a large influence on pupils' non-cognitive skills - above and beyond their effects on test scores. I also shed light on long-term impacts of teachers who increase pupils' non-cognitive skills and found that long-run outcomes are improved by a *combination* of teachers increasing pupils' cognitive and non-cognitive skills. I argued that my findings can provide potential explanation for school interventions with test score effects that "fade-out" over time but have lasting effects on adult outcomes. My research design also allowed me to decompose the teachers' effects on pupils' cognitive and non-cognitive skills into different teaching practices. The analysis revealed that teachers who increase pupils' cognitive skills and teachers who increase pupils' non-cognitive skills use different teaching practices, thus supporting the idea that higher teacher effects in maths are weak predictors of teacher effects in non-cognitive skills. A fruitful avenue for future research would be to see how different types of school interventions amplify or weaken the effects of teachers on pupils' cognitive and non-cognitive skills.

More generally, my findings contribute to the extensive literature that assesses teacher effects on student test scores (e.g. Hanushek, 1971; Rockoff, 2004; Rivkin et al., 2005; Aaronson et al., 2007; Kane and Staiger, 2008; Chetty et al., 2014a,b; Rothstein, 2014; Hacher-Hicks et al., 2014) as it provides evidence for teachers' effects that are not reflected in their test score VA. In doing so, my research could be ultimately used to estimate the optimal weighting of teachers' effects on pupils' cognitive and non-cognitive skills to evaluate teacher quality in school districts.

Although these findings are encouraging, several questions might need to be resolved before we can use these types of measures for policy. For example, in this paper I use different types of measures for pupils' non-cognitive skills. One might ask which of these measures best captures teacher effects on non-cognitive skills. How much do the results depend on the measure used? Moreover, using VA measures on pupils' non-cognitive skills to evaluate teachers could induce teachers to answer the questionnaires differently. If behavioural responses substantially alter the quality of teachers' VA measures, policy makers may need to develop metrics that are more robust to such responses.

In addition, there are many aspects of teachers' long-run impacts that remain to be explored, and which would have considerable policy implications. For example, in this paper, I identify the impact of primary school teachers on long-run outcomes at age 20. Are teacher impacts different over time? Does having a good teacher who improves cognitive skills matter more in primary school while having a good teacher who improves non-cognitive skills is more effective later? Are teacher impacts additive over time? Similarly, it would be interesting to develop analyses which go beyond the mean treatment effects that I have estimated here. We could ask whether different types of teacher quality are more effective in helping different types of students. For instance, are higher non-cognitive VA teachers better with boys or girls or lower achievers rather than high achievers?

Whether or not teachers' effects on non-cognitive skills can be used in teachers' evaluations, my results underline the value of teachers who increase pupils' non-cognitive skills for future outcomes. Hence, this study highlights that considering non-cognitive skills in addition to intellectual development in school objectives is likely to have substantial economic and social benefits. At the time of writing, in many countries this objective remains only marginal.

VII. 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.

Algan, Y, P. Cahuc and A. Shleifer (2013), 'Teaching Practices and Social Capital' *American Economic Journal: Applied Economics*, 5(3), 189-210.

Almlund, M., A.L. Duckworth, J.J. Heckman and T.D. Kautz (2011) 'Personality Psychology and Economics' in *Handbook of the Economics of Education*, volume 4, e.d. Eric A. Hanushek, Stephen Machin and Ludger Woessmann, 1-182. Amsterdam, North Holland: Elsevier Science.

Araujo, M.C., Carneiro, P.M., Cruz-Aguayo, Y. and N. Schady (2016),'Teacher Quality and Learning Outcomes in Kindergarten', *Quarterly Journal of Economics*, Forthcoming.

Bacher-Hicks, A., T.J. Kane, and D.O. Staiger (2014),'Validating Teacher Effects Estimates Using Changes in Teacher Assignments in Los Angeles', *Unpublished Manuscript.*

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

Blazar, D. and M.A. Kraft (2015),'Teacher and Teaching Effects on Students' Academic Behaviors and Mindsets', *Mathematic Policy Research.* Working Paper 41.

Bloom, B.S. (1956),'Taxonomy of Educational Objectives', *Handbook 1: Cognitive Domain*, New York, N.Y. Longmans, Green and co.

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

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

Cascio E. and D. Staiger (2012),'Knowledge, Tests and Fade out in Educational Interventions', *NBER working paper No.18038.*

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 (2014a), 'Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates', *American Economic Review*, 104(9), 2593-2632.

Chetty, R., J. Friedman, and J. Rockoff (2014b), '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.

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*, 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.

Gershenson, S. (2016),'Linking Teacher Quality, Student Attendance and Student Achievement', *Education Finance and Policy*, 125-149.

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 Strengths 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. (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.

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.

Heckman J.J., R. Pinto and P. Savelyev (2013),'Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes', *American Economic Review*, 103(6), 2052-2086.

Heckman J.J., T. Kautz, R. Diris, B. Ter Weel and L. Borghans (2015),'Fostering and Measuing Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success", *NBER working paper*.

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. (2012), 'Non-Cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9th Grade Teachers in North Carolina', *NBER working paper.*

Jacob B.A., L. Lefgren and D. Sims (2010),'The Persistence of Teacher-Induced Learning Gains', *Journal of Human Resources*, 45(4), 915-943.

Kane, T. and D. Staiger (2008), 'Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation', *NBER working paper. No14607*

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.

Mihaly K., D.F. McCaffrey, D. Staiger and J.R Lockwood (2013),'A Composite Estimator of Effective Teaching', *Gate Foundation Research Paper.*

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.

Mueller, G. and E. Plug (2006),'Estimating the Effect of Personality on Male and Female Earnings', *Industrial and Labor Relations Review*, 60(1), 3-22.

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

Rothstein, J. (2010),'Teacher Quality in Educational Production: Tracking, Decay and Student Achievement', *Quarterly Journal of Economics*, 125(1), 175-214.

Rothstein, J. (2014),'Revisiting the Impacts of Teachers', *Mimeo.*

Ruzek, E.A., T. Domina, A.M. Conley, G.J. Duncan and S.A. Karabenick (2014), 'Using Value-Added Models to Measure Teacher Effects on Students' Motivation and Achievement', *The Journal of Early Adolescence*, 35(5-6), 852-882.

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.

VIII. Figures

Figure 1: Teacher Quality Distribution (Teacher Reports)

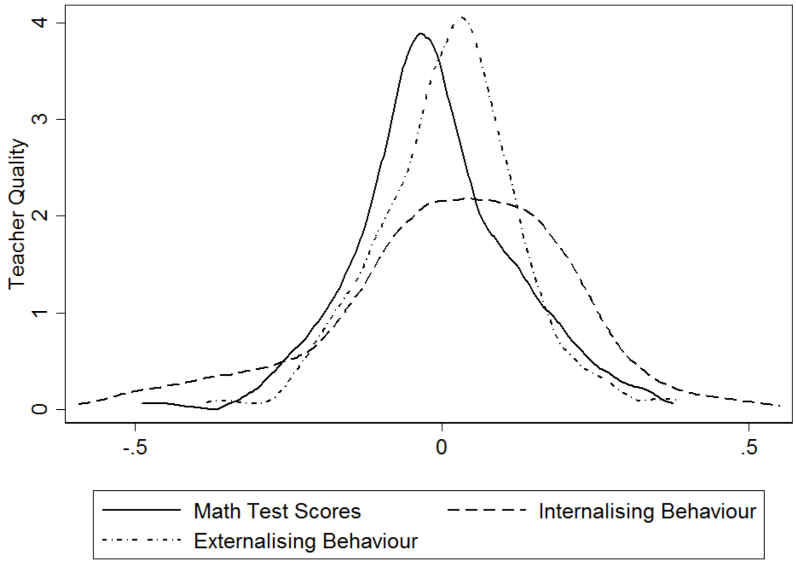


Figure 2: Teacher Quality Distribution (Teacher Reports, Controlling for Prior Cognitive and Non-Cognitive Ability)

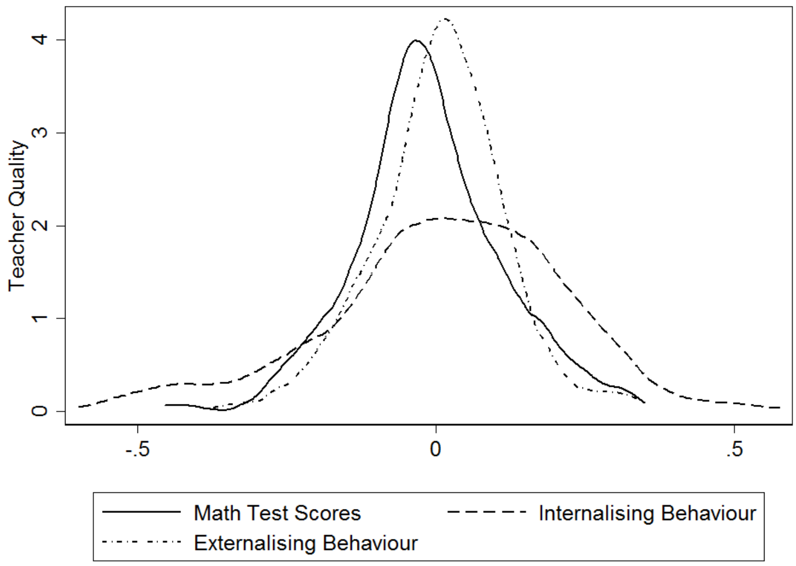


Figure 3: Teacher Quality Distribution (Parents reports)

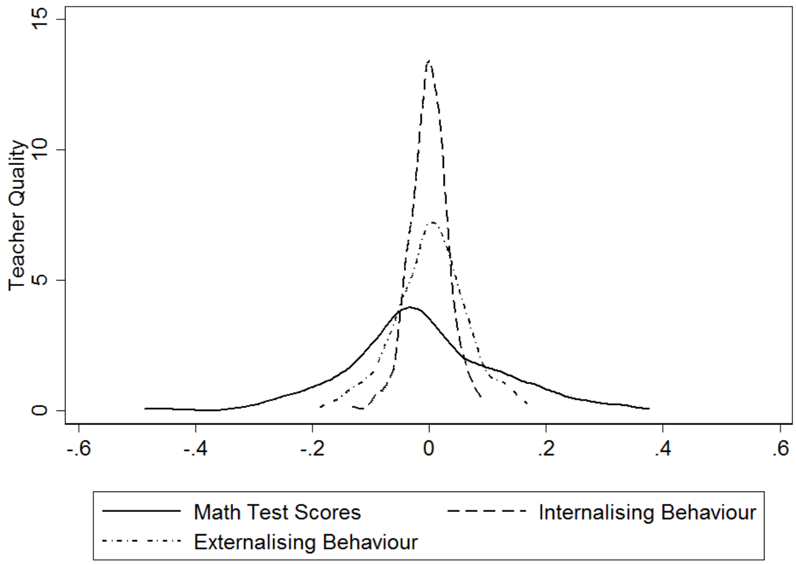


Figure 4: Teacher Quality Distribution (Principal Component Analysis)

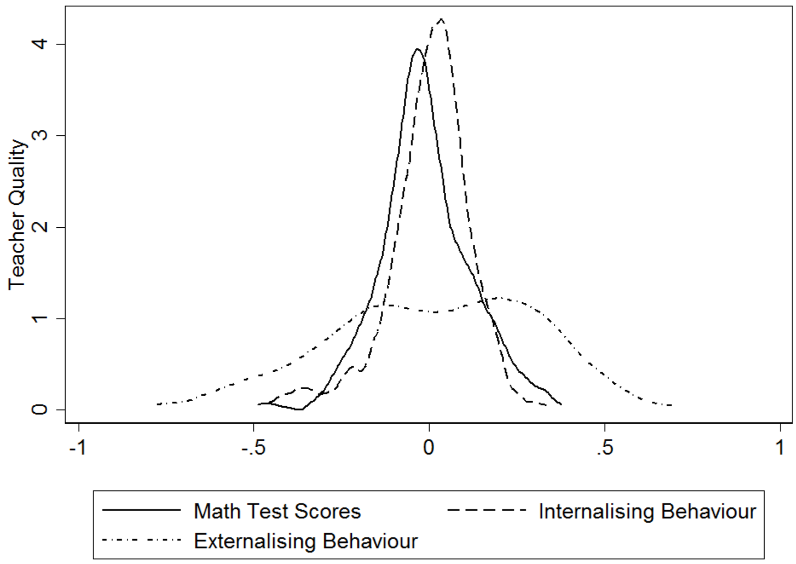


Figure 5: Teacher Quality Distribution (Instrumental Variables)

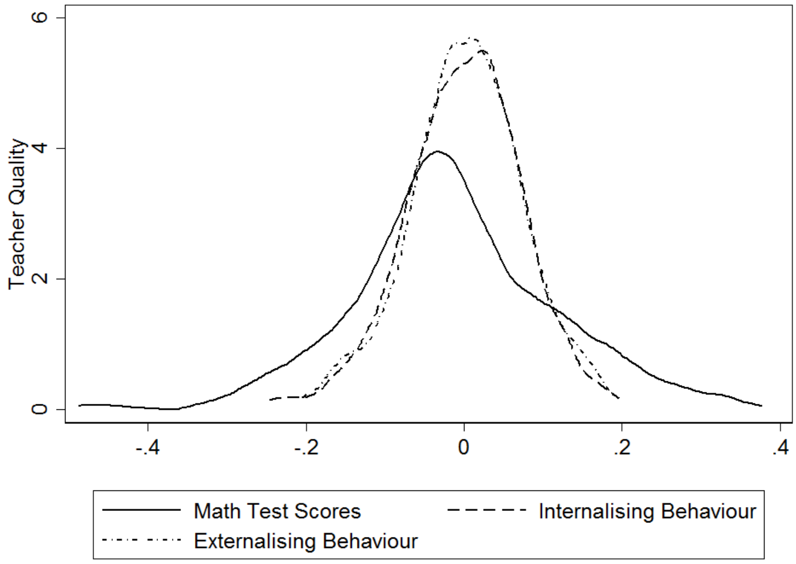
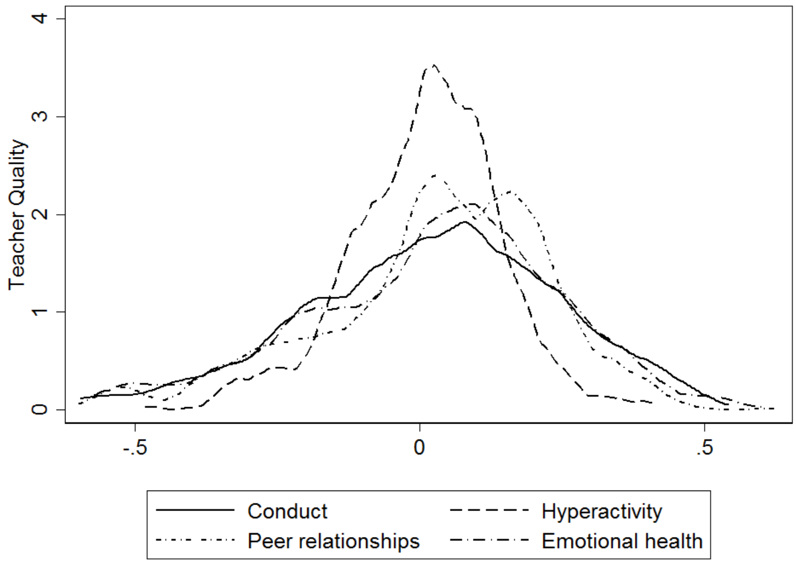


Figure 6: Teacher Quality Distribution (SDQ Subscales)



Notes: Figures 1-6 report kernel distribution of teacher VA estimates. Teacher VA are estimated in regressions that include controls for class characteristics, school characteristics, pupil characteristics, family background, school cohort effects, grade dummies and lagged pupil dependent variables.

Figure 7: Effects of Math Teacher Quality on Higher Education Attendance

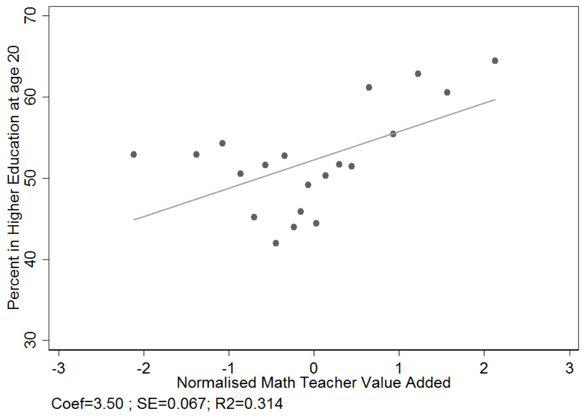
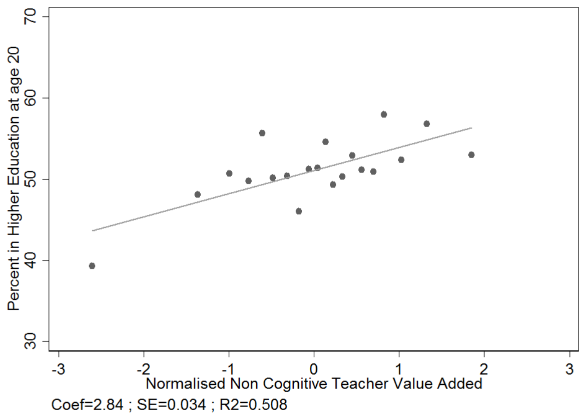
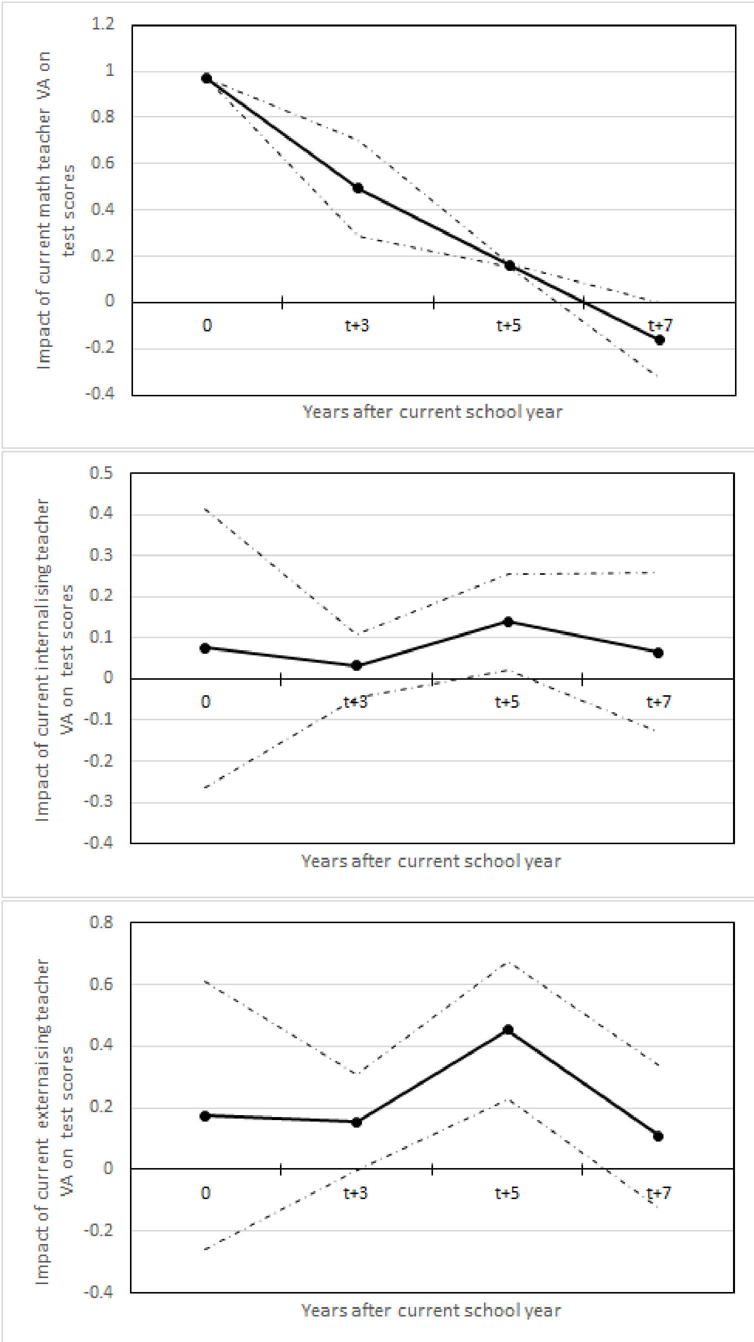


Figure 8: Effects of Non-Cognitive Teacher Quality on Higher Education Attendance



Notes: These figures are drawn using one observation per pupil per year. Figure 7 is a binned scatter plot of higher education attendance rates versus normalised teacher VA on math test scores. Figure 8 is a binned scatter plot of higher education attendance rates versus normalised teacher VA on non-cognitive skills. These plots corresponds to the regressions in Table 11. To construct these binned scatter plots, I first residualise the y-axis variable with respect to the baseline class-level control vector (e.g. pupil, family, school and classroom characteristics defined in the text) using within teacher variation to estimate the coefficients on the controls. I then divide the VA estimates into twenty equal-sized groups (vingtiles) and plot the means of the y-variable residuals within each bin against the mean value of teacher VA within each bin. The solid line shows the best linear fit estimated using OLS. The coefficient show the estimated slope of the best-fit line, with standard errors clustered at the school-cohort level reported in parentheses.

Figure 9: Effects of Teacher VA on Future Outcomes



Notes: These figures show the effect of current teacher VA on test scores at the end of current and subsequent school years. To construct these figures, I regress outcomes in year t=s on teacher VA, in year t varying s from 3 to 7. I control for the baseline control vector (defined in section 3), using within teacher variation to identify the coefficients on controls. The dashed lines depict 90% confidence intervals on each regression coefficient, with standard errors clustered by school-cohort. The coefficients and standard errors from the underling regressions are reported in Table 15.

IX. Tables

Table 1: Summary Statistics for Sample Used To Estimate VA Model

Variable	Mean (1)	SD (2)
Student characteristics		
Male	51.50%	
Age (years)	9.83	1.41
Free school meal eligible	12.00%	
Special education needs	2.80%	
White	94.90%	
Low birth weight	5.50%	
Good health	86.90%	
Class size	28.5	4.96
Class exclusion	2.60%	
Parent characteristics		
Age at child birth	27.8	4.99
Mother's education (1-5)	2.92	1.27
Father's education (1-5)	2.94	1.45
Major financial difficulties	14.20%	
Mother in labour force	69.50%	
Married	74%	

Notes: All statistics are from the ALSPAC data and are measured during primary school. Free school meal eligible is an indicator for receiving free school meals. Mother's and father's education are measured on a 1-5 scale. Marital status is measured by whether the natural parents are married at child birth. Age at child birth is the age of the mother at child birth. Major financial difficulties is an indicator for whether the household has been in major financial difficulties since child birth. Mother in labour force is an indicator about whether mother is currently working at child birth.

Table 2: Correlations between Pupils' Cognitive and Non-Cognitive Skills

Variable	Math test scores	English test scores	Intern. behaviour (parents report)	Extern. behaviour (parents report)	Intern. behaviour (teacher report)	Extern. behaviour (teacher report)
Correlation matrix						
Math test scores	1					
English test scores	0.71	1				
Internalising behaviour index (parents report)	0.17	0.14	1			
Externalising behaviour index (parents report)	0.28	0.34	0.40	1		
Internalising behaviour index (teacher report)	0.23	0.24	0.33	0.2	1	
Externalising behaviour index (teacher report)	0.35	0.45	0.16	0.46	0.37	1
Pupil average	62.5	57.0	17.19	15.5	17.39	16.56
Individual level SD	20.8	16.3	2.79	3.34	3.25	3.97
Within-year variance: (on standardised outcomes)						
Individual level SD	0.72	0.63	0.91	0.92	0.79	0.73
Class + Teacher SD	0.23	0.28	0.19	0.14	0.21	0.38

Notes: Table 2 reports the correlation, the mean and the standard deviations of math test scores, English test scores, internalising behaviour and externalising behaviour in years 3 and 6 of primary school for the sample used in estimating the baseline VA model. Internalising behaviour and externalising behaviour are reported by the parents and by the teacher in both school years for each pupil. Math test scores are measured by KS2 math test scores in year 6 and ALSPAC math test scores at the end of year 3. English test scores are measured by KS2 and KS1 English test scores in year 6 and year 2 of primary schools. Internalising behaviours and externalising behaviours are two composite indicators computed from the Strength and Difficulties Questionnaire (see Goldman et al., 1997). "Pupil average" reports the mean of unstandardised math test scores, English test scores, internalising behaviour and externalising behaviour at the pupil level. Individual level SD reports the standard deviation of unstandardised math test scores, English test scores, internalising behaviour and externalising behaviour at the pupil level. The last two rows are based on standardised math test scores, English test scores, internalising behaviour, and externalising behaviour residuals, used in the teacher VA estimates computation. Individual level SD reports the within-year standard deviation of pupils' outcomes residuals based on pupils' variation. Class-teacher level SD reports the within-year standard deviation of pupils' outcomes residuals based on classroom and teacher-level variation.

Table 3: Correlations between Pupils' Cognitive, Non-Cognitive Skills and Long-Run Outcomes

Variable	Higher Education Attendance at age 20	Total Take Home Pay at age 20	Never been unemployed at age 20	Full time job at age 20
Correlation matrix				
Math test scores	0.36	0.09	0.08	-0.25
English test scores	0.35	0.08	0.09	-0.28
Internalising behaviour index (parents report)	0.07	0.03	0.10	-0.02
Externalising behaviour index (parents report)	0.18	0.00	0.10	-0.14
Internalising behaviour index (teacher report)	0.08	0.04	0.11	-0.03
Externalising behaviour index (teacher report)	0.20	-0.05	0.14	-0.14
Pupil Average	50%	£12,444	67%	33%
Individual level SD		5196		

Notes: Tables 3 reports the correlation of higher education attendance, total take home pay, the probability of never been unemployed and the probability of being in a full time job at age 20 with primary school outcomes such as math test scores, English test scores, internalising behaviour and externalising behaviour in year 3 and 6. Internalising behaviour and externalising behaviour are reported by the parents and by the teacher in both school years for each pupil. Math test scores are measured by KS2 math test scores in year 6 and ALSPAC math test scores at the end of year 3. English test scores are measured by KS2 and KS1 English test scores in year 6 and year 2 of primary schools. Internalising behaviours and externalising behaviours are two composite indicators computed from the Strength and Difficulties Questionnaire (see Goldman et al., 1997). Higher education attendance is an indicator for being in higher education at age 20 (all types of higher education institutions). Total take home pay is measured in pounds and is the annual total take home pay (after NI and tax) at age 20 if in full time job. Never been unemployed is an indicator for never having been unemployed at age 20. Full time job at age 20 is an indicator for being in a full time job at age 20 (it excludes being in full time education and being in a part time job). Pupil average reports the mean of higher education attendance, total take home pay, never been unemployed and full time job at age 20, at the pupil level. Individual level SD reports the standard deviations of these long-run outcomes.

Table 4: Pupils' Cognitive and Non-Cognitive Skills - Explanatory Powers

	(1)	(2)	(3)	(4)
Included explanatory variables				
Pupil and family covariates	Yes	Yes	Yes	Yes
School & classroom characteristics	No	Yes	Yes	Yes
Teacher fixed effects	No	No	Yes	No
F tests, HO:			(<0.01)	
School fixed effects	No	No	No	Yes
F tests, HO:				(<0.01)
	Math Test Scores			
R-squared	0.271	0.279	0.451	0.387
Adjusted R-squared	0.269	0.275	0.384	0.353
Observations	10377	10377	10377	10377
	Internalising behaviour			
R-squared	0.102	0.118	0.305	0.218
Adjusted R-squared	0.099	0.114	0.231	0.182
Observations	12533	12533	12533	12533
	Externalising behaviour			
R-squared	0.242	0.250	0.374	0.317
Adjusted R-squared	0.240	0.247	0.308	0.286
Observations	12479	12479	12479	12479

Notes: All regressions include pupil characteristics (including lagged dependent variable), family background, grade fixed effects and school-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, adjusted 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 separately. All three outcomes (math test scores, internalising and externalising behaviours) are measured in years 3 and 6 of primary schools. Internalising and externalising behaviours are reported by the teacher.

Table 5: Teacher VA Model Estimates

	Math Test Scores	Internalising behaviour	Externalising behaviour
Teacher VA (SD)	0.129	0.219	0.119
10th percentile	-0.165	-0.273	-0.148
25th percentile	-0.084	-0.091	-0.062
50th percentile	0.019	0.027	0.017
75th percentile	0.062	0.148	0.075
90th percentile	0.164	0.228	0.140
90-10 gap	0.329	0.501	0.288
75-25 gap	0.146	0.239	0.137
Number of teachers	294	310	310
Number of school-years per teacher	2.32	2.4	2.4
Avg number of pupils per teacher per year	14	14	14

Notes: Teacher VA are estimated in regressions that include controls for school and classroom characteristics, pupil characteristics, family background, school-cohort effects, grade dummies and lagged pupil dependent variables. All three outcomes are measured in years 3 and 6 of primary school. Internalising and externalising behaviour scores are reported by the teacher.

Table 6: Sensitivity Analysis

	Math Test Scores	Internalising behaviour	Externalising behaviour
Teacher VA (SD):			
Baseline estimates	0.129	0.219	0.119
Controlling for prior cog. and non cog. skills	0.123	0.229	0.113
Using parents reports	–	0.033	0.062
Using principal component analysis	–	0.119	0.290
Using instrumental variables	–	0.073	0.071
Minimum number of obs per teacher per year			
More than 5	0.165	0.239	0.113
More than 6	0.177	0.202	0.108
More than 7	0.206	0.212	0.14
More than 8	0.251	0.238	0.192
More than 9	0.282	0.157	0.106
More than 10	0.460	0.363	0.132

Notes: Table 6 reports teacher VA estimates using alternative specifications. Row 1 reports the baseline estimates where teacher VA controls for school and classroom characteristics, pupil characteristics, family background, school-cohort effects, grade dummies and lagged dependent variables, separately. Row 2 presents results from a specification which includes lagged dependent variables simultaneously. Row 3 presents results where parents' reports for internalising and externalising behaviours are used instead of teacher reports. Row 4 presents teacher VA estimates where internalising and externalising behaviours are obtained from a principal component analysis on teachers' and parents' reports. Row 5 presents teacher VA estimates where internaliting and externalising behaviours are obtained from a regression of teachers' reports on parents' reports. The next 6 rows present teacher VA estimates, where I restrict the sample to teachers who have at least 5 observations (pupils) per year, 6 observations per year, 7 observations per year, etc.

Table 7: Comparisons of Estimates Across VA Models

	Internalising behaviour (parents report)	Externalising behaviour (parents report)
Internalising behaviour (teacher report)	0.14*** (0.01)	
Externalising behaviour (teacher report)		0.28*** (0.01)
	Internalising behaviour (PCA)	**Externalising behaviour (PCA)**
Internalising behaviour (teacher report)	0.71*** (0.01)	
Externalising behaviour (teacher report)		0.61*** (0.01)
Internalising behaviour (parents report)	0.58*** (0.01)	
Externalising behaviour (parents report)		0.71*** (0.01)
	Internalising behaviour (IV)	**Externalising behaviour (IV)**
Internalising behaviour (teacher report)	0.19*** (0.01)	
Externalising behaviour (teacher report)		0.33*** (0.01)
Internalising behaviour (parents report)	0.30*** (0.01)	
Externalising behaviour (parents report)		0.99*** (0.00)

Notes: Bootstrapped standard errors are in parentheses. Teacher VA estimates are based on specifications that include school and classroom characteristics, pupil characteristics, family background, school-cohort effects, grade dummies and lagged dependent variables, separately. All outcomes are measured in years 3 and 6 of primary school. Internalising and externalising behaviour scores are reported by the teacher, the parents, computed from a principal component analysis (PCA) and from instrumental variables (IV). See Section 3. for a full description.

Table 8: Estimation of Pupil Sorting Using Parent Characteristics and Lagged Dependent Variables

	Math test scores	Predicted math test scores using parent char.	Math test scores	Predicted math test scores using lags
Panel A				
Math teacher VA	0.949***	0.015**	0.944***	0.011**
	(0.044)	(0.002)	(0.054)	(0.002)
Observations	4564	5907	4564	5907
	Intern. behaviour	Predicted intern. behaviour using parent char.	Intern. behaviour	Predicted intern. behaviour using lags
Panel B				
Intern. teacher VA	0.914***	0.008**	0.914***	0.014**
	(0.031)	(0.001)	(0.039)	(0.001)
Observations	6501	6507	6501	6507
	Extern. behaviour	Predicted extern. behaviour using parent char.	Extern. behaviour	Predicted extern. behaviour using lags
Panel C				
Extern. teacher VA	1.028***	0.010**	1.018***	0.020***
	(0.008)	(0.008)	(0.003)	(0.005)
Observations	6479	6507	6479	6507

Notes: Each column reports coefficients from an OLS regression, with bootstrapped standard errors clustered by school-cohort in parentheses. The regressions are run on the sample used to estimate the baseline VA models. There is one observation for each pupil school year in all regressions. Teacher VA are scaled in units of pupils' test scores, and pupils' internalising and externalising behaviours and are estimated using data from classes taught by the same teacher in other years, following the procedure described in section 3. Teacher VA is estimated using the baseline control vector which includes: prior math test scores and prior internalising and externalising behaviours; pupil's gender, ethnicity, free school meal eligibility, age, SEN statement, health; family background such as parents' education, parental marital status, mother employment history, mother age at birth, financial difficulties; school and classroom characteristics; grade and school-cohort dummies. In columns (1) and (3) the dependent variable is the pupil's math test scores (internalising or externalising behaviour) in a given year. In column 2 and 4, the dependent variable is the predicted value generates from a regression of test scores (internalising or externalising behaviour) on in (2): parents' education, parental marital status, mother employment history, mother age at birth, financial difficulties and in (4): lagged test scores (at the entry of primary school) and lagged internalising or externalising behaviours measured at age 6.

Table 9: Correlation Between the Teacher VA Model Estimates

	Math Test Scores	Internalising behaviour (teacher report)	Externalising behaviour (teacher report)
Math Test Scores	1		
Internalising behaviour index (teacher report)	0.01 (0.01)	1	
Externalising behaviour index (teacher report)	0.19*** (0.01)	0.52*** (0.01)	1

Notes: Bootstrapped standard errors are in parentheses. Teacher VA estimates are based on specifications that include school and classroom characteristics, pupil characteristics, family background, school-cohort effects, grade dummies and lagged dependent variables, separately. All outcomes are measured in years 3 and 6 of primary school. Internalising and externalising behaviour scores are reported by the teacher.

Table 10: Comparaisons of Estimates Across VA Models

	Math Test Scores
Math Test Scores	1
Internalising behaviour (parents report)	0.01 (0.01)
Externalising behaviour (parents report)	0.14*** (0.01)
	Math Test Scores
Math Test Scores	1
Internalising behaviour (PCA)	0.01 (0.01)
Externalising behaviour (PCA)	0.28*** (0.01)
	Math Test Scores
Math Test Scores	1
Internalising behaviour (IV)	0.23*** (0.01)
Externalising behaviour (IV)	0.17*** (0.01)

Notes: Bootstrapped standard errors are in parentheses. Teacher VA estimates are based on specifications that include school and classroom characteristics, pupil characteristics, family background, school-cohort effects, grade dummies and lagged dependent variables, separately. All outcomes are measured in years 3 and 6 of primary school. Internalising and externalising behaviour scores are reported by the parents, computed by principal component factor analysis and from instrumental variables (IV). See Section 3 for a full description.

Table 11: Impacts of Teacher VA on Higher Education Attendance

	Higher Education Attendance at Age 20 (%)			
Panel A: Teachers' reports				
Math Teacher VA	0.908**			0.887**
	(0.357)			(0.370)
Internalising Teacher VA		0.583***		
		(0.145)		
Externalising Teacher VA			0.902***	
			(0.127)	
Int. + Ext. Teacher VA				0.405***
				(0.122)
Panel B: Parents' reports				
Math Teacher VA	0.908**			1.173***
	(0.357)			(0.248)
Internalising Teacher VA		0.461		
		(0.302)		
Externalising Teacher VA			0.651	
			(1.285)	
Int. + Ext. Teacher VA				0.827
				(0.805)
Panel C: Using principal component analysis				
Math Teacher VA	0.908**			1.137**
	(0.357)			(0.237)
Internalising Teacher VA		0.453		
		(0.353)		
Externalising Teacher VA			0.442	
			(2.333)	
Int. + Ext. Teacher VA				0.429
				(1.255)
Panel D: Using instrumental variables				
Math Teacher VA	0.908**			1.219***
	(0.357)			(0.273)
Internalising Teacher VA		1.281		
		(0.790)		
Externalising Teacher VA			0.529	
			(1.401)	
Int. + Ext. Teacher VA				1.484
				(1.306)
Mean of Dep. Var.	44.90%	45.30%	45.30%	44.90%
Observations	1120	1181	1181	1081

Notes: Each column reports coefficients from an OLS regressions with standard errors clusteed by school-cohort in parentheses. Teacher VA are estimated as described before. The dependent variable is higher education attendance at age 20. See Section 4 for more details on the construction of the variable. Each column reports weighted estimates, using inversed-probability weights for being in the sample at age 20.

Table 12: Impacts of Teacher VA on Earnings

	Total Take Home Pay (£)			
Panel A: Teachers' reports				
Math Teacher VA	257.0***			259.5***
	(34.73)			(38.120)
Internalising Teacher VA		267.3***		
		(66.44)		
Externalising Teacher VA			322.1***	
			(83.86)	
Int. + Ext. Teacher VA				33.64
				(100.19)
Panel B: Parents' reports				
Math Teacher VA	257.0***			263.6***
	(34.73)			(34.25)
Internalising Teacher VA		71.17***		
		(16.73)		
Externalising Teacher VA			329.9***	
			(57.53)	
Int. + Ext. Teacher VA				5.38
				(10.65)
Panel C: Using principal component analysis				
Math Teacher VA	257.0**			252.6***
	(34.73)			(26.26)
Internalising Teacher VA		8.182		
		(48.38)		
Externalising Teacher VA			573.3***	
			(168.49)	
Int. + Ext. Teacher VA				119.3
				(108.99)
Panel D: Using instrumental variables				
Math Teacher VA	257.0***			256.5***
	(34.73)			(26.58)
Internalising Teacher VA		92.92		
		(112.84)		
Externalising Teacher VA			346.4***	
			(55.59)	
Int. + Ext. Teacher VA				62.54
				(67.92)
Mean of Dep. Var.	£12,256	£12,227	£12,227	£12,256
Observations	4862	5105	5100	4649

Notes: Each column reports coefficients from an OLS regressions with standard errors clusteed by school-cohort in parentheses. Teacher VA are estimated as described before. The dependent variable is total take home pay at age 20. See Section 4 for more details on the construction of the variable. Each column reports weighted estimates, using inversed-probability weights for being in the sample at age 20.

Table 13: Impacts of Teacher VA on Being Never Unemployed

	Never been unemployed (%)			
Panel A: Teachers' reports				
Math Teacher VA	0.827**			0.670**
	(0.327)			(0.355)
Internalising Teacher VA		0.964*		
		(0.591)		
Externalising Teacher VA			1.969***	
			(0.184)	
Int. + Ext. Teacher VA				1.795***
				(0.197)
Panel B: Parents' reports				
Math Teacher VA	0.827**			0.761*
	(0.371)			(0.467)
Internalising Teacher VA		0.199		
		(0.639)		
Externalising Teacher VA			1.025*	
			(0.547)	
Int. + Ext. Teacher VA				0.822
				(0.548)
Panel C: Using principal component analysis				
Math Teacher VA	0.827**			0.716*
	(0.327)			(0.417)
Internalising Teacher VA		0.192		
		(1.504)		
Externalising Teacher VA			1.132**	
			(0.461)	
Int. + Ext. Teacher VA				0.158
				(0.795)
Panel D: Using instrumental variables				
Math Teacher VA	0.827**			0.731*
	(0.327)			(0.440)
Internalising Teacher VA		1.076*		
		(0.771)		
Externalising Teacher VA			0.946	
			(0.686)	
Int. + Ext. Teacher VA				0.114
				(1.012)
Mean of Dep. Var.	64.20%	64.10%	64.10%	64.20%
Observations	1092	1152	1152	1053

Notes: Each column reports coefficients from an OLS regressions with standard errors clusteed by school-cohort in parentheses. Teacher VA are estimated as described before. The dependent variable is the probability of never been unemployed at age 20. See Section 4 for more details on the construction of the variable. Each column reports weighted estimates, using inversed-probability weights for being in the sample at age 20.

Table 14: Impacts of Teacher VA on Being in Full Time Job

	Being in Full Time Job (%)			
Panel A: Teachers' reports				
Math Teacher VA	0.443***			0.390**
	(0.064)			(0.101)
Internalising Teacher VA		0.207		
		(0.287)		
Externalising Teacher VA			0.474**	
			(0.139)	
Int. + Ext. Teacher VA				0.430**
				(0.196)
Panel B: Parents' reports				
Math Teacher VA	0.443***			0.380**
	(0.064)			(0.091)
Internalising Teacher VA		0.294		
		(0.191)		
Externalising Teacher VA			0.420***	
			(0.070)	
Int. + Ext. Teacher VA				0.161
				(0.157)
Panel C: Using principal component analysis				
Math Teacher VA	0.443***			0.239**
	(0.064)			(0.119)
Internalising Teacher VA		0.616***		
		(0.245)		
Externalising Teacher VA			0.877***	
			(0.209)	
Int. + Ext. Teacher VA				0.764***
				(0.217)
Panel D: Using instrumental variables				
Math Teacher VA	0.443***			0.341***
	(0.064)			(0.068)
Internalising Teacher VA		0.622***		
		(0.116)		
Externalising Teacher VA			0.617***	
			(0.057)	
Int. + Ext. Teacher VA				0.369***
				(0.038)
Mean of Dep. Var.	33%	33%	33%	33%
Observations	5906	6204	6199	5653

Notes: Each column reports coefficients from an OLS regressions with standard errors clusteed by school-cohort in parentheses. Teacher VA are estimated as described before. The dependent variable is the probability of being in a full time job at age 20. See Section 4 for more details on the construction of the variable. Each column reports weighted estimates, using inversed-probability weights for being in the sample at age 20.

Table 15: Impacts of Teacher VA on Current and Future Outcomes

	Subsequent math test scores (Standardised)			
	0	T+3	T+5	T+7
Math Teacher VA	0.967***	0.494***	0.161***	-0.164*
	(0.049)	(0.107)	(0.004)	(0.084)
Observations	3353	4099	5187	3161
Internalising Teacher VA	0.075	0032	0.140**	0.064
	(0.173)	(0.039)	(0.060)	(0.099)
Observations	3384	4149	5246	3189
Externalising Teacher VA	0.175	0.153*	0.452***	0.108
	(0.222)	(0.080)	(0.115)	(0.118)
Observations	3384	4149	5246	3189

Notes: The table shows the effect of current teacher VA on test scores and non-cognitive skills at the end of the current and subsequent school years. I regress end of grade test scores and non-cognitive skills in year t+s on teacher VA in year t, varying s from 3 to 7. I scale teacher VA in units of pupils test score SD's and pupils non-cognitive skills SD's. I control for the baseline control vector (pupil, family, classroom and school characteristics, defined in the text) using within-teacher variation to identify the coefficients on controls. The table reports the coefficients and bootstrapped standard errors clustered at the school-cohort level in parentheses.

Table 16: Share of the Variance in Teacher VA Explained by Teacher Characteristics and Teaching Practices (R-squared)

	Math Teacher VA	Internalising Teacher VA	Externalising Teacher VA
(1) = Teacher gender + experience	0.020	0.010	0.005
(2) = Teacher non cog. skills	0.015	0.027	0.043
(3) = Teaching practices	0.103	0.084	0.126
(4) = (1) + (2)	0.033	0.040	0.051
(5) = (1) + (3)	0.126	0.098	0.135
(6) = (2) + (3)	0.114	0.110	0.149
(7) = (1) + (2) + (3)	**0.136**	**0.125**	**0.164**

Notes: Teacher non-cog skills includes teacher CCEI, teacher Bachman self-esteem, job satisfaction, and teaching confidence. Teaching practices includes all the teaching practices listed in Appendix B2. The teacher VA are estimated in regressions that include the baseline control vector described in Section 3. Only R-squared from second step regressions in which teacher VA are decomposed into different teacher characteristics and teaching practices are reported. The second step regressions include teacher gender and experience in row (1). Row (2) includes only teacher non-cognitive skills. Row (3) includes only teaching practices. Row (4) includes teacher gender, experience and non-cognitive skills. Row (5) includes teacher gender, experience and teaching practices. Row (6) includes teacher non-cognitive skills and teachng practices. Row (7) includes rows (1), (2) and (3) variables.

Table 17: Estimates of the Effect of Teacher Characteristics on Teacher VA

	Math Teacher VA	Internalising Teacher VA	Externalising Teacher VA
Panel A: Teachers' reports			
Teacher-gender	0.008	0.018*	0.007
	(0.007)	(0.010)	(0.006)
Teacher CCEI	-0.003	0.041***	0.018***
	(0.008)	(0.012)	(0.007)
Teacher self-esteem	0.002	0.004	0.012*
	(0.008)	(0.013)	(0.007)
Teacher job satisfaction	0.016*	-0.021*	-0.000
	(0.008)	(0.013)	(0.007)
Teacher confidence in teaching	0.012*	0.009	0.013**
	(0.007)	(0.011)	(0.006)
Teacher experience	-0.004	-0.009	-0.010
	(0.007)	(0.012)	(0.006)
Panel B: Parents' reports			
Teacher-gender	0.008	0.000	0.002
	(0.007)	(0.002)	(0.003)
Teacher CCEI	-0.003	0.001	0.005
	(0.008)	(0.003)	(0.004)
Teacher self-esteem	-0.002	-0.003	0.003
	(0.008)	(0.003)	(0.004)
Teacher job satisfaction	0.016*	-0.001	0.001
	(0.008)	(0.003)	(0.004)
Teacher confidence in teaching	0.012*	-0.001	0.017***
	(0.007)	(0.002)	(0.004)
Teacher experience	-0.004	0.002	-0.006
	(0.007)	(0.003)	(0.004)
Teaching practices	Yes	Yes	Yes

Notes: This tables reports OLS regressions of teacher characteristics on teacher VA. The dependent variables are the teacher VA, estimated in regressions that include the baseline control vector described in Section 3. Teacher characteristics include teacher gender, teacher emotional health (CCEI), teacher Bachman self-esteem, teacher job satisfaction, teacher confidence in teaching math and teacher experience and all the teaching practices described in Appendix B2. See Appendix B3 for a full description of teacher CCEI and teacher Bachman self-esteem.

Table 18: Estimates of the Effect of Teaching Practices on Teacher VA

	Math Teacher VA		Internalising Teacher VA		Externalising Teacher VA	
	Each measure included separately (1)	All measures included jointly (2)	Each measure included separately (3)	All measures included jointly (4)	Each measure included separately (5)	All measures included jointly (6)
Panel A: Teachers' reports						
Knowledge	-0.037*	-0.030	-0.087*	-0.091	0.020	0.016
	(0.021)	(0.039)	(0.050)	(0.059)	(0.031)	(0.032)
Analytical and critical skills	-0.018*	-0.011	-0.003	-0.034	0.006	-0.006
	(0.011)	(0.025)	(0.024)	(0.037)	(0.015)	(0.020)
Capacity for individual study	-0.017*	-0.007	0.038	0.062*	0.009	0.001
	(0.009)	(0.023)	(0.024)	(0.033)	(0.012)	(0.018)
Social and moral behaviours	-0.017	-0.009	0.009	-0.004	0.016	0.017
	(0.010)	(0.024)	(0.024)	(0.035)	(0.013)	(0.019)
Individual treatment of pupils	-0.022	-0.022	0.057	0.040	0.028	0.024
	(0.015)	(0.031)	(0.043)	(0.013)	(0.022)	(0.026)
Panel B: Parents' reports						
Knowledge	-0.037*	-0.030	-0.012	-0.017	0.030	0.030
	(0.021)	(0.039)	(0.015)	(0.013)	(0.020)	(0.019)
Analytical and critical skills	-0.018*	-0.011	0.018***	0.006	0.004	0.001
	(0.011)	(0.025)	(0.006)	(0.008)	(0.009)	(0.012)
Capacity for individual study	-0.017*	-0.007	0.023***	0.018**	0.007	0.008
	(0.009)	(0.023)	(0.006)	(0.007)	(0.008)	(0.011)
Social and moral behaviours	-0.017	-0.009	0.018***	0.005	-0.002	-0.008
	(0.010)	(0.024)	(0.006)	(0.008)	(0.008)	(0.011)
Individual treatment of pupils	-0.022	-0.022	0.018	0.015	-0.011	-0.014
	(0.015)	(0.031)	(0.012)	(0.010)	(0.013)	(0.015)

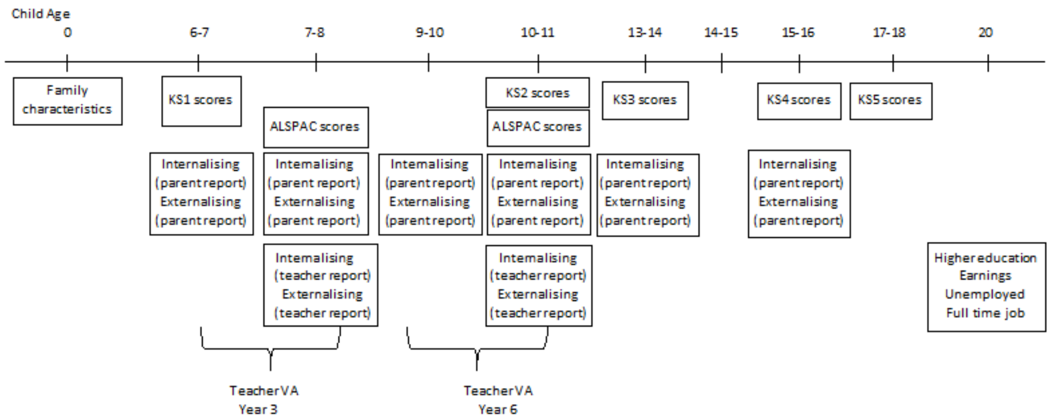
Notes: This table reports OLS estimates of the effect of five teaching practices measures on teacher value-added estimates measured on pupils' math test scores, pupils' internalising behaviours and pupils' externalising behaviours. The unit of analysis is one observation per teacher per year. Robust standard errors are reported in parentheses. The regressions include teacher gender, teacher CCEI and teacher experience. The estimates presented in the odd columns are from regressions when each of teaching practices is used as the only treatment variable in the regression. The estimates presented in the even columns are from regressions where all five teaching practices measures are used simultaneously as treatment variables in the regressions.

X. Appendix

Appendix A: Figures

.

Figure A.1: Timeline



Appendix A: Tables

.

Table A.1: Sensitivity Analysis

	Teacher VA - SD
English Test Scores	0.255
SDQ - conduct	0.240
SDQ - hyperactivity	0.136
SDQ - peer relationships	0.231
SDQ - emotional problems	0.255
SDQ - pro-social	0.207
Math Test Scores (ALSPAC)	0.256

Notes: This table reports teacher VA estimates using alternative specifications. Row 1 reports the teacher VA estimates on English test scores. Rows 2-6 report the teacher VA based on SDQ subscales (teacher-assessed). Row 7 reports the teacher VA estimates on ALSPAC math test scores.

Table A.2: Summary Statistics for Teachers Used to Estimated VA Model

	Mean	SD	Min	Max
Female	79.5 %			
Teacher CCEI	13.61	7.97	1	40
Teacher Bachman self-esteem	30.98	5.55	13	40
Teacher job satisfaction	4.39	0.86	1	5
Length of time teaching pupil	1.17	0.38	1	2
Teacher confidence in teaching	1.54	0.51	0	2
Teacher experience	14.85	11.12	0	42
Homework frequency	3.470	0.878	1	5
Type of homework	1.908	0.507	1	3
Standardised tests	2.429	0.591	1	3
Written tests	2.835	0.391	1	3
Self-assessed tests	1.960	0.545	1	3
Listen to pupils	2.669	0.491	1	3
Individual discussions and review	2.186	0.533	1	3
Written incentives	0.994	0.075	0	1
Naming pupils in the classroom	0.989	0.106	0	1
Free time as incentive	0.560	0.497	0	1
Competition as incentive	0.562	0.497	0	1
Displaying work	0.319	0.467	0	1
Class groups: by attainment	0.966	0.181	0	1
Class ability groups	0.921	0.270	0	1
Class math ability groups	0.905	0.293	0	1
Teacher responsibility: develop skills	1.043	0.213	1	3
Teacher responsibility: moral and behaviours	1.177	0.446	1	3
Teacher responsibility: equip skills for society	1.126	0.365	1	3
Teacher responsibility: develop individual	1.408	0.628	1	4
Teacher responsibility: being obedient	1.406	0.621	1	4
Teacher responsibility: capacity to think	1.138	0.401	1	3
Teacher responsibility: prepare for occupation	2.298	1.032	1	5
Teacher responsibility: respect	1.244	0.518	1	4
Teacher responsibility: work cooperatively	1.287	0.540	1	4
Teacher responsibility: interest in learning	1.128	0.382	1	3
Teacher responsibility: able to organise	1.562	0.675	1	4
Teacher responsibility: self confidence	1.102	0.327	1	3
Teacher responsibility: considerate to others	1.140	0.398	1	3
Teacher responsibility: show respect	1.439	0.519	1	4
Sanction if homework not done	1.312	0.464	1	2
Homework: to the most able or least able	2.998	0.209	1	4

Table A.3: Correlations between Teacher VA and Parent Characteristics

	Math Teacher VA	Internalising Teacher VA	Externalising Teacher VA
Parent characteristics effects	0.054***	0.021***	-0.017
	(0.013)	(0.013)	(0.014)

Notes: This table reports the correlation between the effect of parental variables (mother age at birth, parental education, parental marital status and parent employment history, major financial difficulties) and the teacher VA from estimates described before. Bootstrapped standard errors in parentheses.

Table A.4: Correlations between Teachers' Reports, Parents' Reports and Teaching Practices

	Correlation Teacher / Parents Internalising	Correlation Teacher / Parents Externalising
Teacher gender	0.044***	-0.022*
	(0.014)	(0.013)
Teacher CCEI	0.002***	0.002**
	(0.001)	(0.001)
Length of time taught pupil	0.052***	0.029***
	(0.014)	(0.012)
Teacher experience	0.007	-0.002
	(0.005)	(0.005)
Standardised tests	-0.045***	-0.008
	(0.011)	(0.010)
Individual discussions and review	0.042***	-0.032**
	(0.015)	(0.013)
Naming pupils in the classroom	0.274**	0.062
	(0.116)	(0.101)
Class groups: by attainment	0.095**	0.086**
	(0.047)	(0.044)
Teacher responsibility: develop individual	0.042***	0.054***
	(0.014)	(0.012)
Teacher responsibility: capacity to think	0.049*	-0.080***
	(0.026)	(0.023)
Teacher responsibility: considerate to others	0.023	0.061***
	(0.022)	(0.020)
Observations	5884	5913

Notes: This table reports OLS estimates of different teacher characteristics and teaching practices on a variable describing the correlation between teachers' reports and parents' reports. The correlation between teacher's reports and parents' reports is calculated at the pupil-year level.

Table A.5: Summary Statistics - Within and Between Teacher Variation - for the Sample Used to Estimate VA Model

	Between SD	Within SD
Teacher CCEI	7.86	2.25
Teacher Bachman self-esteem	5.50	1.64
Teacher job satisfaction	0.82	0.27
Length of time teaching pupil	0.38	0.18
Teacher confidence in teaching	0.49	0.18
Teacher experience	11.14	0.51
Homework frequency	0.86	0.25
Type of homework	0.48	0.20
Standardised tests	0.56	0.25
Written tests	0.36	0.12
Self-assessed tests	0.53	0.23
Listen to pupils	0.45	0.20
Individual discussions and review	0.51	0.24
Written incentives	0.06	0.04
Naming pupils in the classroom	0.08	0.05
Free time as incentive	0.47	0.18
Competition as incentive	0.48	0.16
Displaying work	0.46	0.17
Class groups: by attainment	0.16	0.07
Class ability groups	0.24	0.10
Class math ability groups	0.26	0.11
Teacher responsibility: develop skills	0.22	0.10
Teacher responsibility: moral and behaviours	0.45	0.15
Teacher responsibility: equip skills for society	0.38	0.16
Teacher responsibility: develop individual	0.63	0.23
Teacher responsibility: being obedient	0.63	0.23
Teacher responsibility: capacity to think	0.41	0.14
Teacher responsibility: prepare for occupation	1.02	0.35
Teacher responsibility: respect	0.53	0.21
Teacher responsibility: work cooperatively	0.54	0.19
Teacher responsibility: interest in learning	0.39	0.13
Teacher responsibility: able to organise	0.66	0.27
Teacher responsibility: self confidence	0.33	0.17
Teacher responsibility: considerate to others	0.41	0.14
Teacher responsibility: show respect	0.50	0.20
Sanction if homework not done	0.45	0.19
Homework: to the most able or least able	0.21	0.04

Appendix B: Description of the Variables

Internalising SDQ - Strength and Difficulties Questionnaire

- *Emotional problems scale*
  - Often complains of headaches, stomach-aches or sickness (0-2)
  - Many worries, often seems worried (0-2)
  - Often unhappy, down-hearted or tearful (0-2)
  - Nervous or clingy in new situations, easily loses confidence (0-2)
  - Many fears, easily scared (0-2)
- *Peer problems scale*
  - Rather solitary, tends to play alone (0-2)
  - Has at least one good friend (0-2)
  - Generally liked by other children (0-2)
  - Picked on or bullied by other children (0-2)
  - Gets on better with adults than with other children (0-2)

Externalising SDQ - Strength and Difficulties Questionnaire

- *Behavioural problems scale*
  - Often has temper tantrums or hot tempers (0-2)
  - Generally obedient, usually does what adults request (0-2)
  - Often fights with other children or bullies them (0-2)
  - Often lies or cheats (0-2)
  - Steals from home, school or elsewhere (0-2)
- *Hyperactivity scale*
  - Restless, overactive, cannot stay still for long (0-2)
  - Constantly fidgeting or squirming (0-2)
  - Easily distracted, concentration wanders (0-2)
  - Thinks things out before acting (0-2)
  - Sees tasks through to the end, good attention span (0-2)
  - I did everything wrong (0-2)

Teaching Practices: Five Categories Based on the ALSPAC Questionnaire

- *Instilment of knowledge and enhancement of comprehension*
  - The teachers give homework in term time (0-1)
  - The homework includes assignments due for teachers' checking (0-1)
  - The teachers use standardised tests and marked written work (0-1)
  - The teachers use questions and answers in the class (0-1)
  - The teachers feel the responsibility to develop basic skills and build up knowledge (0-1)

- *Instilment of analytical and critical skills*
  - The teachers feel the responsibility to develop the child's capacity to think (0-1)
  - The teachers feel the responsibility that an interest in learning is aroused (0-1)
  - The teachers feel the responsibility to equip child with skills and attitudes which will enable her/him to take a place effectively in society (0-1)
  - The teachers feel the responsibility to fit the child for an occupational role in society (0-1)

- *Instilment of capacity for individual study*
  - The teachers use pupils' self-assessment (0-1)
  - The teachers feel the responsibility that the child should be an individual/developing in his/her own way (0-1)
  - The teachers feel the responsibility that children should be able to organise their work (0-1)
  - The teachers feel the responsibility to develop child's self-confidence (0-1)

- *Instilment of social and moral behaviours*
  - The teachers feel the responsibility that the child should be obedient to parents and teachers (0-1)
  - The teachers feel the responsibility that the child acquired respects for her own property and others (0-1)
  - The teachers feel the responsibility that children learn how to work cooperatively (0-1)
  - The teachers feel the responsibility that children should be kind and considerate to others (0-1)

- *Individual treatmnents of pupils*
  - Teachers group children by attainment groups for classroom activities (0-1)
  - In this class, there are ability groups (0-1)
  - The teachers give homework to the most able or the least able (0-1)
  - The teachers use individual reviews or discussions (0-1)
  - The teachers use the following incentives in relation to academic work: naming of children (0-1)
  - The teachers use the following incentives in relation to academic work: competition (0-1)
  - The teachers use the following incentives in relation to academic work: displaying work (0-1)

Teacher Non-Cognitive Skills Measures from the ALSPAC Questionnaire

- *Teacher Crown-Crisp Experiential Index (CCEI)*

  The Crown-Crisp Experiential Index (CCEI), has been widely used for measuring neurotic traits and symptoms. It has been validated in different studies. It is composed of 21 items:
  - Feels upset for no obvious reason (1-4)
  - Troubled by dizziness/shortness of breath (1-4)
  - Felt like fainting (1-4)
  - Feels sick (1-4)
  - Feels life is too much effort (1-4)
  - Feels uneasy and restless (1-4)
  - Feels tingling in arms/legs/body (1-4)
  - Regrets much pas behaviour (1-4)
  - Sometimes feels panincking (1-4)
  - Has little or no appetite (1-4)
  - Wakes unusually early in morning (1-4)
  - Worries a lot (1-4)
  - Feels tired/exhausted (1-4)
  - Has long periods of sadness (1-4)
  - Feels strung up inside (1-4)
  - Goes to sleep all right (1-4)
  - Feels to be going to pieces (1-4)
  - Often sweats excessively (1-4)
  - Needs to cry (1-4)
  - Has had upsetting dreams (1-4)
  - Loses ability to feel sympathy (1-4)

- *Teacher Bachman Self-Esteem*

  The Bachman Self-Esteem score is composed of 11 items:

  - Feels to be a person of worth (1-5)
  - Feels to have number of good qualities (1-5)
  - Is able to do things as well as others (1-5)
  - Feels not to have much to be proud of (1-5)
  - Takes a positive attitude towards self (1-5)
  - Sometimes thinks to be not good at all (1-5)
  - Is a useful person to have round (1-5)
  - Feels cannot do anything right (1-5)
  - Does job well (1-5)
  - Feels their life is not useful (1-5)
  - Feels unlucky (1-5)