DISCUSSION PAPER SERIES

IZA

IZA DP No. 9955

"Teaching to Teach" Literacy

Stephen Machin
Sandra McNally
Martina Viarengo

May 2016

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

“Teaching to Teach” Literacy

Stephen Machin
University College London,
CEP, LSE and IZA

Sandra McNally
University of Surrey,
CEP, LSE and IZA

Martina Viarengo
The Graduate Institute Geneva,
CID, Harvard University and IZA

Discussion Paper No. 9955
May 2016

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 9955
May 2016

ABSTRACT

“Teaching to Teach” Literacy*

Significant numbers of people have very low levels of literacy in many OECD countries and, because of this, face significant labour market penalties. Despite this, it remains unclear what teaching strategies are most useful for actually rectifying literacy deficiencies. The subject remains hugely controversial amongst educationalists and has seldom been studied by economists. Research evidence from part of Scotland prompted a national change in the policy guidance given to schools in England in the mid-2000s about how children are taught to read. We conceptualise this as a shock to the education production function that affects the technology of teaching. In particular, there was phasing in of intensive support to some schools across Local Authorities: teachers were trained to use a new phonics approach. We use this staggered introduction of intensive support to estimate the effect of the new ‘teaching technology’ on children’s educational attainment. We find there to be effects of the teaching technology (‘synthetic phonics’) at age 5 and 7. However, by the age of 11, other children have caught up and there are no average effects. There are long-term effects only for those children with a higher initial propensity to struggle with reading.

JEL Classification: I21, I28

Keywords: literacy, phonics

Corresponding author:

Stephen Machin
Department of Economics
University College London
30 Gordon Street
London WC1H 0AX
United Kingdom
E-mail: s.machin@ucl.ac.uk

* We would like to thank Simon Brown, Marilyn Joyce, Michele Mann, Winter Rogers, Helen Walker and Edward Wagstaff of the Department for Education for data and detailed information about the policy evaluated in this paper. We thank the NPD team at the Department for Education and Jon Johnson and Rachel Rosenberg of the Institute of Education for provision of data. We thank participants at conferences hosted by CESifo Economics of Education, the European Association for Labour Economics, the Association of Education, Finance and Policy; and seminars at the Centre for Economic Performance LSE, the University of Sheffield, the Institute of Education, Lancaster University and the IFAU in Uppsala. In particular, we would like to thank Sandra Black, David Figlio and John Van Reenen for helpful comments. We thank Andy Eyles for helpful research assistance. Viarengo gratefully acknowledges the support received from the British Academy and the Royal Society in the framework of the Newton International Fellowship.

1. Introduction

Learning to read and write is an essential skill for modern life, yet about 15% of the adult population in OECD countries have not mastered the basics,[1] being unable, for example, to fully understand instructions on a bottle of aspirin. These literacy problems are especially serious in England where younger adults perform no better than older ones (Kuczera et al., 2016). In this context, it is unsurprising to see that not having basic literacy skills generates significant and sizable wage and employment penalties in the labour market (Vignoles, 2016).

How can the situation be improved? It is well understood that good teaching is important for pupil learning and their educational trajectories through school. There is a solid evidence base that teachers, and teaching methods, can matter both for literacy (e.g. Jacob, 2016; Machin and McNally, 2007; Slavin *et al*., 2009) and more generally (e.g. Aaronson et al., 2007; Chetty et al. 2014a, 2014b; Hanushek et al., 2005). But this still leaves open the question as to how we obtain better teaching. One approach is to attract and retain people with higher quality teaching skills. Another approach is to upgrade the skills of any given stock of teachers. A key question is can good teaching be taught?

When it comes to learning to read, many argue that there are pedagogies which are transformative in their effects. If this were true, it would provide a simple policy solution for getting the whole population literate – policy makers could just insist that all teachers adopt a particular pedagogy for teaching children how to read. In fact, this centralised policy approach to education is something done by English policy makers in this area. Although they encourage schools to be autonomous is some respects (e.g. the new academy schools as described in Eyles and Machin, 2015), successive governments have been happy to advocate

[1] The results of PIAAC (OECD 2013), show that 15.5% of adults have a proficiency of 'level 1' or below. See Table 2.2. http://skills.oecd.org/documents/SkillsOutlook_2013_Chapter2.pdf

and recommend how reading should be taught to primary school children, with little sound evidence to back them up. This continues to be highly controversial.[2]

How reading should be taught in schools has been and remains hotly debated amongst educationalists.[3] Learning to read and write in English is difficult relative to other languages because of phonological complexity of syllable structures and an inconsistent spelling system (Wyse and Gosmani, 2008). In countries like Greece, Finland, Italy and Spain, syllable structure is simple and there are 1:1 mappings between letters and sounds. This is far from the case for English where many words look alike but sound different (and vice versa). Despite various unsuccessful attempts at reforming the alphabet over time (most famously by George Bernard Shaw), the language of 26 letters and 45 phonemes that can be spelled in at least 350 ways (Pollack and Pickarz, 1963) is more objectively challenging than learning to read in other languages.

Perhaps because of its complexity, there has been much disagreement about how to teach English. The historic division has been between proponents of 'whole language' versus 'phonics' approaches. The approaches each encompass different methods. In essence 'whole language' is about being introduced to language through context (e.g. through stories, picture books etc.) whereas 'phonics' is about a more systematic method of teaching how spelling patterns correspond to sounds. The building blocks of the language are assembled before stories are introduced. The 'phonics' method was the norm until the mid-19th Century, but in the 1930s and 1940s, the 'whole word' model became popular (Hempenstall, 1997) – whereby words were introduced through their meaning and should be recognised by sight, using the cue of their shape and length.

[2] For example, the UK literacy association has criticised the government for excessive concentration on phonics in its instructions to schools (UKLA, 2010). Also controversial is the Phonics Screening Check, which now has to be taken by all 6 year olds. This was undertaken for the first time in 2012 and only 58% of children passed the test.

[3] See Mike Baker's synopsis around the time of the 2005 controversy. http://news.bbc.co.uk/1/hi/education/4493260.stm

Only relatively recently has ‘systematic phonics’ instruction been advocated in English-speaking countries: in 2000, by the US National Reading Panel (NICHD, 2000), in 2005 by the Australian government (Australian Government, Department of Education Science and Training 2005), and in 2006 by a review commissioned by the English government (Rose, 2006) that was subsequently implemented in all schools. In England, the policy adopted was narrower than in other English-speaking countries (Wyse and Gosmani, 2008) because it advocated a more extreme view of how exactly phonics should be taught (known as ‘synthetic phonics’) and then obliged all schools to implement the approach. In the research we have undertaken we are able to evaluate it because a pilot was established to inform the review itself and because subsequently training in how to implement the new approach was rolled out in an iterative manner to Local Authorities before it became properly embedded in the system as a whole.

In this paper, we compare pupils in schools who were exposed to the original pilot (that ran concurrently with the Rose review) and pupils in schools in the first wave of the programme (post Rose review) with pupils in schools that were subsequently targeted for training in the use of the programme as it was rolled out to different Local Authorities (LAs). We view the intensive training provided as part of the roll-out as a shock to schools that changes the productivity of teachers. We observe an instant effect of the programme at age 5 that is as large as the initial effect of lower class size revealed by Project STAR (Krueger, 1999; Krueger and Whitmore, 2001). However, the policy is of much lower cost, as it involves employing a literacy consultant working with 10 schools per year to deliver intensive support as well as arranging for dissemination and training opportunities throughout the Local Authority. We are able to view whether the programme effect lasts after the intensive training is complete and whether it is stronger for those exposed to it at a younger age (and for longer) as they progress through school. We find that effects are evident up to age 7 and stronger for those with greater exposure to the programme.

We are also able to follow cohorts as they go through primary school to see if any initial effects lasted until the end of primary school (age 11). Most children learn to read eventually and we do not find evidence of average effects at this age for reading, a broader measure of English attainment or maths. However, we explore whether there is heterogeneity in the estimated effect of the treatment for those with a high probability of being struggling readers on school entry (i.e. those from disadvantaged backgrounds and/or those who are non-native speakers of English). Effects persist at age 11 for young people in this category (even though the treatment stopped 4 years earlier). The effect sizes for the most disadvantaged group seem high enough to justify the costs of the policy. This study therefore shows that good teaching can indeed be taught and this is an example of a 'technology' which his helpful in closing the gap between students who start out with disadvantages (whether economically or in terms of language proficiency) compared to others.

The rest of the paper is structured as follows. In Section 2, we explain the English education system, our data, and how phonics has been used in schools before and after the policy change in the mid-2000s. In Section 3, we outline our conceptual framework and empirical strategy. In Section 4, we discuss our results, firstly in the context of an 'events study' for 5 year olds, then based on an analysis of programme effects as relevant cohorts progress through the school system (at age 5, 7, and 11) and then we evaluate whether the policy has a heterogeneous effect depending on whether the student is classified as disadvantaged or a non-native English speaker. We also conduct various placebo tests and robustness checks, such as whether the policy effects subjects other than reading. We conclude in Section 5.

2. The English Education System

2.1. Assessment and Data

The national curriculum in England is organised around 'Key Stages'. In each 'Key Stage' there are various goals made out for children's learning and development and it ends with a formal assessment: the Foundation Stage at age 5, and Key Stages 1 through 4 at ages 7, 11, 14 and 16. The assessments at age 11 and 16 are set and marked externally. These Key Stage 2 and 4 tests are at the end of primary and secondary school respectively and are 'high stakes' for the school in that they are the basis of the School Performance Tables, which are publicly available. At the other ages pupils are assessed by their own teachers. However, there is extensive guidance on how the assessment should be made and it is moderated.

Children must start school the September after they turn 4 years old and there is no grade repetition. For most children, their first assessment takes place at the end of reception year (i.e. the first year) of primary school[4], when the child is at age 5. This Foundation Stage of education is made against 13 assessment scales comprising 6 areas of learning: personal social and emotional development (3 scales), communication, language and literacy (4 scales), mathematical development (3 scales), knowledge and understanding of the world (1 scale), physical development (1 scale) and creative development (1 scale). Points are allocated within each scale. We can sum points over all scales to get a total score or sum points within each sub-category. In this paper, we focus on the score for 'communication, language and literacy'. The first year for which this information is produced is 2003. Between 2003 and 2006, the assessment was only done for a 10% child-level sample.[5] From 2007 onwards, all children in England have been assessed in this way.

The Key Stage 1 assessments take place when the pupil is at age 7. Head teachers have a statutory duty to ensure that their teachers comply with all aspects of the Key Stage 1

[4] Some children may be assessed in settings such as nursery schools and playgroups which receive Government funding.

[5] In our data, all schools are represented in roughly the same proportion from 2003-2006.

assessment and reporting arrangements. The assessments are in reading, writing, speaking and listening, mathematics and science. We will focus on the teacher assessments for reading, although we do examine whether there are effects on other subjects (described in Section 4.4 below). Local Authorities (and other recognised bodies) are responsible for moderation of schools. Thus, although teachers make their own assessments of students (and therefore are susceptible to potential bias), there is a process in place to ensure that there is a meaningful assessment that is standardised over all of England. At age 7, students are given a 'level' (i.e. there is no test score as such). However, following standard practice, we transform National Curriculum levels achieved in reading, writing and mathematics into point scores using Department for Education point scales.

In Key Stage 2, at the end of primary school, pupils take national tests in English, maths and science. These are externally set and marked. There is a continuous measure of achievement in all subjects. An important target for schools is the percentage of pupils that achieve level 4 or above – because this is what matters for the performance tables, which are publicly available.

The National Pupil Database (NPD) is a census of all pupils in the state system in England. During the primary phase of education, this accounts for the vast majority of children. We exclude a small number of independent and special schools from the analysis. We mainly use data between 2003 and 2012, because the age 5 assessment was introduced in 2003. It was originally a 10 per cent child-level sample, but the information was reported for all children from 2007 onwards.

The NPD gives information on all the assessments described above and basic demographic details of pupils – such as ethnicity, deprivation (measured by whether they are eligible to receive free school meals), gender, and whether or not English is their first language. As we know the school attended, we can control for school fixed effects in our analysis – and we can track students if they change schools. For a small minority of areas, there is a structure where pupils attend one type of school from about age 5-10 and then transfer to middle school before going to secondary school. However, in most places, there is no middle school and pupils make the transition to secondary school at the age of 11 (in the autumn after the Key Stage 2 assessment).

For the period covered by our study schooling was organised at the local level into Local Education Authorities (of which there are 152). Schools are largely self-governing and the main functions of the Local Authority are in building and maintaining schools, allocating funding, providing support services, and acting in an advisory role to the head teacher regarding school performance and implementation of government initiatives. The Department for Education have provided us with details of the Local Authorities and schools involved in initial phonics pilot (EDRp) and how support was phased-in across Local Authorities and schools in subsequent years (through the CLLD programme). We describe this below in detail, after first discussing the use of phonics in schools.

2.2. The Use of Phonics in Schools

There are two main approaches to learning the alphabetic principle: synthetic phonics and analytic phonics. The former is used in Germany and Austria and is generally taught before children are introduced to books or reading. It involves learning to pronounce the sounds (phenomes) associated with letters ‘in isolation’. These individual sounds, once learnt, are then blended together (synthesised) to form words. By contrast, analytic phonics does not involve learning the sounds of letters in isolation. Instead children are taught to recognise the beginning and ending sounds of words, without breaking these down into the smallest constituent sounds. It is generally taught in parallel with, or sometime after, graded reading books, which are introduced using a ‘look and say’ approach.[6] One of the reasons the debate between educationalists is so divisive is because those advocating ‘synthetic phonics’ argue this should be taught before any other method. The other side argue that one size does not fit all and it is possible to teach other aspects of reading at the same time.

Up to 2006, the English literacy strategy recommended analytic phonics as one of four ‘searchlights’ for learning to read in the National Literacy Strategy (in place since 1998) – the others were knowledge of context, grammatical knowledge, word recognition and graphic knowledge. However, a review of this approach was prompted by a study in a small area of Scotland (Clackmannanshire), which claimed very strong effects for children taught to read using synthetic phonics (Johnston and Watson, 2005). The outcome of the review was the ‘Rose Report’ (DfES, 2006), after which government guidelines were updated to require the teaching of synthetic phonics as the first and main strategy for reading. According to Wyse and Goswani (2008), one of main differences with the previous ‘searchlights’ model is that the new ‘simple view of reading’ separates out word recognition processes and language comprehension processes. There was a detailed programme called ‘Letters and Sounds: principles and practice of high quality phonics’ which teachers were expected to follow (Primary National Strategy, 2007). This is summarised (as in Wyse and Goswani, 2008) in Table 1.

At the same time as the review was taking place (before it was published), there was a pilot in 172 schools and nurseries that was principally to give intensive training to teachers on the use of synthetic phonics in early years. After the Rose report, training was rolled out to different Local Authorities (LA). The LAs were given funding for a literacy coordinator who would work intensively in about 10 schools per year but also disseminate best practice

[6] Children are typically taught one letter sound per week and are shown a series of alliterative pictures and words which start with that sound, e.g. car, cat, candle, caste, caterpillar. When the 26 initial letter sounds have been taught, children are introduced to final sounds and to middle sounds. At this point, some teachers may show children how to sound and blend the consecutive letters in unfamiliar words.

throughout the LA by offering courses. The programme was rolled out iteratively to different Local Authorities – only reaching all Local Authorities by the school year 2009/10. Thus, it was not anticipated that all schools would update their early years' teaching overnight, even though the government guidelines had changed.[7]

More specifically, the "*The Early Reading Development Pilot"* (ERDp) was introduced in 2005 to test out the pace of phonics teaching and, in terms of timing, ran alongside the Rose review.[8] This involved 18 Local Authorities (LAs) and 172 schools and settings in the school year 2005-06.[9] "*The Communication, Language and Literacy Development Programme"* (CLLD) was launched in September 2006 to implement the recommendations of the Rose Review, replacing the EDRp. A further 32 LAs were invited to join the original 18 LAs, each receiving funding for a dedicated learning consultant. The next wave of the CLLD was introduced from April 2008. This involved another 50 LAs. Then the last third of LAs (i.e. another 50) joined the CLLD programme in April 2009.

The essential model of support was similar across the EDRp and the CLLD (in successive waves). In the EDRp, LAs received funding to engage leadership teams and Foundation Stage practitioners in pilot schools, run an initial cluster meeting for pilot schools and ensure schools complete an audit of their provision. The intention was to disseminate information and build capacity across the Local Authorities and not just those identified as part of the Pilot. For the CLLD, all LAs received £50,000 to support the appointment of a specialist consultant to work across early years and Key Stage 1 (i.e. the stages of the

[7] In 2010, a government spokesman implied that the 'Communication, Language and Literacy programme' was necessary to enable schools to make the necessary changes.
http://www.theguardian.com/education/2010/jan/19/phonics-child-literacy

[8] It was requested by Andrew Adonis, the then Minister of State for education, in response to the findings of the Select Committee on the teaching of early reading.

[9] As some pre-school settings were involved (i.e. nurseries), we have fewer primary schools that this in our data – roughly 160 schools. However, it has been confirmed that the Reception year in these primary schools was the main initial focus for this policy.

curriculum supporting children from age 4-7), with a further £15,000 to allocate to schools and settings.

LAs were asked to employ their funded CLLD consultant to providing coaching support to at least ten schools per year. The consultant works mainly in the Reception year (first year of school) and Year 1, but also in Year 2 and nursery. This includes termly collection of pupil progress data. Developing the role of a lead within the school for early literacy was a key part of the programme in order to build capacity and enable schools to sustain improvements. Schools were expected to exit from intensive support in a year if possible. The consultant also provided support to other schools and settings in the Local Authority, usually through the provision of courses. In most cases, such 'Continuing Professional Development' courses were offered to all schools.

The consultant support involved an initial audit and assessment visit to help schools get started on the programme. This included drawing up a 'CLLD action plan', making observations and detailed assessments of children. In a second visit, the consultant would model or co-teach the adult-led activity or the discrete teaching session and help teachers and practitioners to plan further learning and teaching opportunities over the following few weeks. At this and subsequent visits, the consultant would work with teachers, practitioners and leadership teams to review children's learning and identify the next steps for teaching.

2.3. Selection of Schools and Local Authorities

The selection of Local Authorities and schools into the initial EDRp pilot and subsequent iteration of the CLLD programme to LAs/schools in successive waves was not done in a systematic way according to specific criteria. In relation to the 18 LAs selected for the EDRp pilot in 2005/06, communication with officials in the Department of Education reveals the following: selection of Local Authorities was based on current involvement with the ‘Intensifying Support Programme’[10]; capacity to deliver at short notice; existing expertise around early years learning, reading and phonics teaching; effective working relationships across Early Years and Literacy/School Improvement teams; mix of LA type and representation across regions; commitment to advocacy for early reading pilot approach; willingness to support dissemination. The decision regarding the selection of schools into the pilot was made by the Local Authority. As described by officials in the Department of Education, the criteria were as follows: willingness and capacity to engage with the pilot at all levels (i.e. headteacher, early years coordinator, relevant teachers…); commitment by the school/setting to improve the quality of teaching of early reading in the Foundation Stage; need to improve children’s outcomes in communication, language and literacy; quality of teaching in the Foundation Stage must be at least satisfactory; at least two of the ten schools/settings identified in a single authority would have the potential to become leading practice schools in terms of early reading – building long-term capacity in the authority area.

In September 2006, the Communication, Language and Literacy Development Programme (CLLD) was launched to implement the recommendations of the Rose Review, replacing the EDRp. A further 32 LAs were invited to join the original 18 LAs, each receiving funding for a dedicated learning consultant. Details are similarly vague on how the additional 32 LAs were selected. We are told that they were selected after consultation with the National Strategy regional teams on the basis of several factors including data, LA capacity and the need to encompass a range of different sorts of LAs.

A second group of 50 LAs were invited to join the CLLD programme from April 2008, making 100 LAs in total. The selection was based on the number of young children in the LA who were in the 30% most deprived ‘super output areas’ so that the programme could support work in ‘closing the gap’ in attainment at Foundation Stage. LAs were advised to

[10] This was a programme introduced in 2002. 13 Local Authorities with a number of local attaining schools were invited to join this two-year pilot to work with their schools in challenging circumstances. The programme was further extended to 76 LAs in 2004-05.

select their target schools on the basis of their data for attainment at ages 5 and 7 (i.e. Foundation Stage Profile and Key Stage 1 – as described in Section 3.1), taking into account local knowledge about capacity. However, the consultant's remit was to work beyond the targeted schools to disseminate effective practice as widely as possible in the LA. The CLLD programme was extended to all authorities from April 2009 with the same guidance offered on the selection of targeted schools.

Thus, we do not have clear, transparent criteria for selection of schools for 'intensive support' or how the programme was iterated through Local Authorities. This means looking at the data to define treatment and control groups is an important task. We are interested to establish whether pupils attending schools in the first round of EDRp and CLLD (i.e. two separate 'treatment groups') perform differently to those in schools that subsequently enrolled in the CLLD as this was spread across different Local Authorities between 2008 and 2010. The groups are summarised in Table 2. Our approach will involve a 'difference-in-differences' analysis, comparing outcomes before and after the policy was introduced (conditional on other attributes of schools and pupils). The credibility of the methodology rests on whether these groups show parallel trends in outcome variables pre-policy (below we show that they do) rather than whether they match closely based on observable characteristics at a point in time. However, the advantage of this approach is that all schools in the treatment and control groups were deliberately selected for 'intensive support' – and thus have more in common (for the purposes of evaluating this policy) than all those schools that were not selected.[11]

In Table 3, we show key characteristics of different groups of schools in the pre-EDRp year (2004/05). This is designed to understand the selection process of Local

[11] Other reasons for not using non-selected schools in treated Local Authorities as a control group is that the literacy consultant was supposed to disseminate best practice throughout the Local Authority, as discussed in Section 2.2. When we do use these schools as a control group, estimated effects are smaller but for the most part, qualitatively similar to the current analysis. Results available on request.

Authorities and schools. Columns (1)-(6) show the following groups: (1) all schools; (2) schools in the original EDRp pilot; (3) non-selected schools in the 18 EDRp pilot Local Authorities; (4) schools in the first wave of the CLLD programme (within 50 Local Authorities); (5) schools that were not selected as part of the first Wave of the CLLD programme within the same 50 LAs; (6) schools in the first Wave of the CLLD for the other 100 Local Authorities that entered the programme between 2008 and 2010. Thus, columns (2) and (4) show statistics for the two treatment groups of interest (EDRp and first wave of CLLD respectively) and column (6) shows statistics for the control group.

We show summary statistics for our main outcome variables at age 5 and 7.[12] They are the communication, language and literacy score (standardised to have mean zero and a unit standard deviation) from the age 5 Foundation Stage and the age 7 Key Stage 1 score (similarly standardised) in reading. We also show three important demographic variables[13]: the proportion of children eligible to receive free school meals (an indicator of socio-economic disadvantage); the proportion of native English speakers; and the proportion of children who are classified as 'White British or Irish'.

We learn from the Table that within the two treatment groups (i.e. columns (2) and (4)), schools selected for the treatment are (on average) lower performing than other schools within the Local Authorities of interest (i.e. as shown in columns (3) and (5)). They also tend to include a higher proportion of disadvantaged children, a lower proportion of native English speakers and a lower proportion of children classified as 'White British/Irish'. If we consider the Local Authorities selected for the treatment based on their schools not selected for intensive support in the first year (i.e. columns (3) and (5)), they do not look too different from the national average (column (1)) on most of the reported indicators, although they are a

[12] In the analysis, we link age 7 outcomes to age 11 outcomes for students in the treatment and control group respectively. The policy only applies to children during Key Stage 1 – and some children move school between Key Stages 1 and 2 (i.e. between age 7 and 11).

[13] Apart from outcome variables measured at age 5 and 11, all summary statistics relate to children of age 7 in 2005 (the pre-pilot year).

little more disadvantaged (particularly the EDRp Local Authorities). The control group (column (6)) is a lot more similar to schools in the treatment groups (columns (2) and (4)) compared to schools that were not selected for intensive support in treatment Local Authorities (columns (3) and (5)) and to the overall sample. However, there are still significant differences at baseline between treatment and control groups and it will be important to establish that there is no differential pre-trend in outcome variables. We show this in the context of an 'event study' in Section 4 (see Figure 1) and in a regression context. These approaches very clearly show that that the parallel trends assumption is reasonable and there is no pre-policy differential effect of being in a treated school before the policy was introduced. Before we show these findings, we next turn to explain the conceptual framework and empirical strategy.

3. Conceptual Framework and Empirical Strategy

One way of conceptualising the introduction of intensive support to schools in the teaching of phonics is as a shock to the education production function (where teachers are one of the inputs). Teachers are effectively being trained in the use of a 'new technology', which should lead to an increase in their effectiveness as teachers (if the 'new technology' is actually an improvement).

Consider the following general form of the education production function:

$$A_{ist} = f(T_{st,}X_{st}, Z_{ist}) \tag{1}$$

In (1), student *i*'s attainment (*A*) in school *s* at time *t* is influenced by teachers (*T*) in the school they attend, a vector of other school inputs (*X*) and a vector of personal/family inputs (*Z*). The teaching input T_{st} (and for that matter the other inputs into the production function) can be thought of as reflecting time varying and non-time varying components, say a fixed teaching skill component and one that may change in different teaching years. One way to parameterise this in terms of teacher skills (or efficiency) as $T_{st} = f(S_{st,}\overline{S_s})$ with a bar denoting a time mean. Suppose in time period *t+1*, new information comes to light that we view as a change in 'teaching technology' that teachers need instruction in. This potentially changes the effectiveness of the time varying part of the teaching input (S_{st}) whilst leaving other inputs and the fixed teacher skill component unchanged. In this way an effective introduction of the new teaching technology can be thought of as generating a positive shock to the education production function.

In our empirical analysis, we make use of the differential timing of the phasing-in of intensive support to schools as a 'natural experiment' to identify the causal effect of teacher training in the 'new technology' or pedagogy. As discussed above, we use two treatment groups of schools whose teachers were trained to deliver phonics teaching: (1) the initial schools in the pilot that was set up to inform the Rose review (i.e. EDRp); (2) the schools in the first Wave of Local Authorities that were exposed to intensive support to implement the findings of the Rose Review (i.e. CLLD). The control group consists of schools that were selected for intensive support as soon as their Local Authorities were enrolled in CLLD programme (three years after the 'EDRp treatment group'; two years after the 'CLLD treatment group'). Details of the groups and timing of entry to intensive support are provided in Table 1.

Denoting schools treated by phonics exposure and control schools by a binary indicator variable P (equal to 1 for treatment EDRp or CLLD phonics programme schools and 0 for control schools) we can model the shock to teaching skills by recasting the education production function as the following difference-in-differences equation:

$$A_{ist} = \beta_0 + \beta_1(P_s * I(t \geq p)) + \beta_2(Z_{ist}) + \beta_3(X_{st}) + \gamma_t + u_s + \varepsilon_{ist} \qquad (2)$$

where $I(t \geq p)$ is an indicator function representing time periods after time p when the phonics programmes were introduced. This research design enables us to estimate the effect of a 'phonics shock' (P) in a school s affected by the treatment at a given time t under the (plausible) assumption that this is the only relevant time-varying shock that affects the treated schools relative to the control schools. In fact the phased introduction makes it highly unlikely that another shock to teaching skills occurred at the same time, and thus we have a coherent research design for studying what is a relatively unusual policy in that it is inexpensive but has significant potential to reduce literacy inequalities in the early years of school.

In equation (2) β_1 is the coefficient of interest. The specification in equation (2) controls for school fixed effects (u_s), which includes the baseline effect of being a 'treated school' as well as any other school-level characteristics that do not change over time (including the time invariant teacher skills component). We control for a set of time dummies (γ_t). Variables included in the vector of personal/family characteristics (Z) include gender, ethnicity; whether he/she is a native speaker of English; whether he/she is eligible to receive free school meals (an indicator of poverty) and whether he/she receives a statement of Special Educational Needs. Variables included in the vector of time-varying school characteristics (X) include the percentage of students in the year group according to each of the above-named personal characteristics.

Since we are interested in estimating effects as the affected cohorts age (through their schooling), we set most regressions up as interactions with birth cohorts rather than year. Thus, we estimate β_1 when the treatment cohort is at age 5, 7 and 11 relative to control cohorts. For the EDPp treatment, this is the cohort of children born in 2001 whereas for the CLLD treatment, this is the cohort of children born in 2002. The treatment was initially focussed on the youngest age group but could have an effect on multiple age groups within the same year (i.e. children aged between 5 and 7). The cohort of children born in 1998 is completely unaffected at any stage. However, we show a full set of treatment x cohort interactions for those born between 1998 and 2001 (and 2002 when analysing the effect of CLLD).

Finally, we look at heterogeneity by selecting the 1998 birth cohort and the two main 'treatment' cohorts of interest (2001 for EDRp; 2002 for CLLD). We estimate

$$A_{ist} = \alpha_0 + \alpha_1(D_{ist} * P * T_{st}) + \alpha_2(D_{ist} * T_{st}) + \alpha_3(D_{ist}) + \alpha_4(Z_{ist}) + \alpha_5(X_{st}) + \gamma_t + u_s + \omega_{ist} \quad (3)$$

More precisely, we estimate whether there is a differential treatment effect according to whether the student is classified as: (a) being eligible to receive free school meals; and (b) a native English speaker. In equation (3), the characteristic of interest is represented as (*D*). Again, we estimate regression as the student ages through the school system (at ages 5, 7 and 11). We set the regressions up such that the treatment effect is separately identified for each group (i.e. 'free school meal' and 'non-free school meal' children; native and non-native speakers of English). In a final specification, we estimate the two-way interactions.

4. Results

4.1 Event Study

We can see at first glance whether the policy had an effect by an 'event study' based on 5 year-olds. They were the initial target of the intensive support in schools and there is no ambiguity about the year in which we should start to see an effect. It should be the year in which the policy was introduced in both the EDRp and CLLD schools respectively. Furthermore, we should expect the effects to decline once the control group schools receive the treatment.

Having estimated equation (2), the estimated coefficient for the treatment effect (β_1) and the associated 95% confidence interval are plotted in Figure 1 for the EDRp treatment v control and the CLLD treatment v control. The regression estimates are shown in Appendix

Table A1. The dependent variable is the standardised score for ‘communication, language and literacy’ at age 5. The Figure shows zero effect for the two available pre-policy years for EDRp v control and the three available years for CLLD v control. However, as soon as the treatment is introduced, the effect jumps to over 0.2 standard deviations in both cases. Note that the year ‘*t*’ is different for the EDRp and the CLLD groups, yet the effect sizes are similar (and the control group is the same). Furthermore, the EDRp treatment stays high (at least 0.2 standard deviations) for each year until the control group receive the treatment (at *t+3*), where the effect size falls and drops to no longer being statistically different from zero. The pattern is similar for the CLLD treatment, except that the effect size does not fall as quickly when the control group enters the programme at *t+2* (and also remains statistically different from zero).[14]

The fact that the treatment effect stays high up until the control schools enter the programme (and for some time after than in CLLD) shows that any effect of the programme is not simply down to the presence of the literacy consultant in the school. The intensive support was only on offer for one year (except in cases where schools had difficulties). Thus the effect sizes reflect the effect of the training and not the presence of the trainer.

4.2 Main Results by Cohort

Tables 4a and 4b show estimated effects of the policy for the EDPp treatment (Table 4a) and the CLLD treatment (Table 4b) relative to the control group for different birth cohorts as they progress through the school system. The omitted category is the 1998 birth cohort. In each case, the cohorts fully exposed to the treatment throughout their entire early phase of primary education (i.e. age 5-7) and observable at age 11 are the 2001 cohort (for EDRp) and the 2002 cohort (for CLLD). However, other birth cohorts are partially treated.

[14] We identify the effect of the policy through the staggered nature of the intervention. Inclusion and exclusion for time-varying school and pupil characteristics makes little or no difference to estimated effects of the treatment. When we include a measure of the number of teachers (as an attempt to proxy potential teacher turnover), this makes no difference to the results.

For example, the cohort born in 2000 is potentially affected from the age of 6 if receiving the EDRp treatment and at the age of 7 if receiving the CLLD treatment. The cohort born in 1999 might be affected by the EDRp treatment at the age of 7.

We look at effects at the ages of 5, 7 and 11. In each case, the dependent variable is the standardised test score and so the reported estimates can be viewed in units of a standard deviation σ. The data for those undertaking Key Stage 1 assessments at age 7 is linked to the same individuals' assessments at age 11. Thus, we follow the student exposed to the 'treatment' whether or not he/she changes school between the age of 7 and 11.[15] In any school, the 'treatment' is only defined by what happens between the age of 5 and 7. Thereafter the student is in the 'Key Stage 2' phase of primary education (culminating in a test at age 11) and should not be directly affected by the phonics programme.

Focusing on the results for the cohort that receives the treatment throughout their early schooling and observable at age 11 (i.e. the 2001 cohort for EDRp and the 2002 cohort for CLLD), Table 4a and 4b shows that the initial effect on age 5 results is very high (as also shown in Figure 2). It is close to 0.3σ for the EDRp and 0.22σ for the CLLD. By the age of 7, the effect of the policy has reduced by at least two-thirds (although the test score is more coarsely defined at age 7 and therefore not exactly comparable to that at age 5). However, it is still of a reasonable size of about 0.07σ for both the EDRp and the CLLD and is statistically significant. However, at age 11, the results suggest an effect that is close to zero.

For partially treated cohorts, there is an effect which seems to increase over time. We see this when we look at results for age 7 (i.e. column 2). For the EDRp, the effect goes from 0.037σ to 0.04σ to 0.075σ from first exposure to the programme at age 7, 6 and 5 respectively. For the CLLD, the effect goes from 0.031σ to 0.046σ to 0.073σ at these same ages. Hence, earlier exposure and/or length of exposure has an increasing effect on

[15] We do not do this between the age of 5 and 7 because the age 5 test score is only available for a 10% sample of schools between 2003 and 2006. Instead, treatment and control schools are separately merged to the age 5 and 7 data.

educational attainment. Furthermore, it suggests an impact of the programme on children when the intensive support actually stops (as it was only supposed to last one year in treatment schools). Thus, we can also infer that the effect is coming from training in the use of the programme – not from the fact of having a consultant come to the school. However, the effect never persists to age 11.

A final insight from Table 4 is that it is possible to run various placebo tests: did the policy appear to have an effect for cohorts to which it was not exposed? Of course, this might indicate differential trends in treatment and control schools. Coefficients in italics are those estimated for cohorts that could not have been affected by the policy because of the stage they were at in school when the policy was introduced. In all cases, the coefficients are close to zero and statistically insignificant, suggesting no evidence of differential pre-policy trends.

4.3. **Heterogeneous Effects**

We next consider whether the policy has a heterogeneous effect. We might expect any effects of the programme to be stronger for pupils with characteristics that are likely to make them lower achieving on average in reading when they first go to school (like being from a low income background, or not speaking English as a first language). We can look at this at age of school entry using the Millennium Cohort Study (MCS). This longitudinal study began in the years 2000 and 2001 and follows around 20,000 children from birth.[16] We look at the age 5 wave to study test score differences at about the time of school entry.

Table 5 shows regressions of age 5 cognitive test scores (measuring 'naming vocabulary', 'pattern construction' and 'pattern similarity') on indicators of whether MCS cohort members are eligible for free school meals and whether there home language is not

[16] See Hansen, Joshi and Dex (2010) for more detail on the MCS data and a range of studies of cohort members up to age 5.

English.[17] As the estimates show, both of these groups enter school at age 5 with significantly lower test scores, especially in vocabulary skills. The difference in the vocabulary score for native and non-native speakers of English is close to 1 standard deviation whereas it is about 0.6 standard deviations for those from poor and non-poor family backgrounds (as measured by eligibility to receive free school meals). This vocabulary deficit at time of school entry clearly places children with these characteristics at a significant literacy disadvantage then and, if such deficits hold them back, as they get older. Other measures of cognitive ability (pattern construction and pattern similarity) also show large and significant differences between these groups – but the gap is much smaller than that for vocabulary skills. So it is interesting to ask whether intensive training in the use of phonics has a differential impact across these groups, both in terms of when they first faced the programme and at later ages.

In Table 6, we examine the impact of the treatment for the group most strongly impacted by the policy (i.e. receiving the treatment from age 5 onwards) relative to the control group. Thus, the first three columns show impacts for the 2001 cohort relative to the 1998 cohort for the EDRp treatment and the next three columns show impacts for the 2002 cohort relative to the 1998 cohort for the CLLD treatment. In each case, we show heterogeneous effects of the two treatments at ages 5, 7 and 11 by estimating equation 3.

The upper panel (A) compares the effect of the treatment for native and non-native English speakers. For non-native English speakers, the effect size is stronger at age 5 for the EDRp treatment (though not statistically different from the effect for native English speakers) whereas it is similar for these two groups for the CLLD treatment. However, at age 7, a difference has emerged in both cases – the estimated effect is at least twice as large for non-native speakers (p-values of the difference in the estimated treatment effects for native and

[17] Precise definitions of the three tests are given in the descriptive review of the age 5 (third wave) of the MCS in Jones and Schoon (2005). They are aimed to capture cognitive skills at age in verbal, pictorial reasoning and spatial abilities (as in Elliott, 1996, or Hill, 2005).

non-native speakers are 0.115σ and 0.055σ for the EDRp and CLLD respectively). By age 11, the coefficient is positive for non-native English speakers – but only statistically significant for the CLLD cohort. The effect size is 0.068σ and this is statistically different from that estimated for native English speakers (for whom we see no effect).

The middle panel (B) shows effects of the treatment for disadvantaged students and other students (based on their eligibility for free school meals). The effect sizes are similar at age 5. However, we see differences at age 7 for both the EDRp and the CLLD treatment groups. Disadvantaged students benefit more from the programme than other students in each case. The differences are statistically significant and similar for both the EDRp and CLLD treatments. Whereas the effect for more advantaged students (i.e. non free school meals) is 0.042σ and 0.045σ for the EDRp and CLLD treatments respectively, it is 0.135σ and 0.136σ for students eligible to receive free school meals. By the time students get to age 11, the effect size for disadvantaged students is 0.06σ in both cases. However, this is only statistically significant for the CLLD treatment. For non-disadvantaged students, the EDRp cohort is shown to have a negative effect (of 0.06σ, which is significant at the 10% level) whereas for CLLD students, there is zero effect. It is difficult to know what to make of the former (especially in view of the fact that they appeared to benefit at age 7). In a robustness test (below) we look at whether effect sizes are similar if we consider the following cohort (2002 rather than 2001).

Finally, in panel (C), we show effects where we estimate interactions between disadvantaged status and whether the student is a native speaker of English. We show estimates of the treatment on four groups: native English speakers and eligible to receive free school meals; native English speakers and not eligible to receive free school meals; non-native English speakers and eligible to receive free school meals (i.e. the most 'disadvantaged group') and non-native English speakers who are not eligible to receive free school meals.

These regressions show that for both the EDRp and the CLLD treatments, the effect sizes are strongest for the most disadvantaged group (i.e. non-native English speakers AND eligible to receive free school meals) at both the age of 7 and 11. In both cases, the treatment increases test scores by around 0.2σ at age 7. With regard to effects estimated at age 11, the treatment increases scores by 0.18σ for the EDRp treatment and by 0.10σ for the CLLD. For the CLLD treatment, the effect persists to age 11 for only one other group: non-native speakers who are not eligible to receive free school meals (raising scores by 0.07σ). However, for the EDRp there remains a negative coefficient estimated for one group (i.e. native students who are not eligible to receive free school meals). It is difficult to know what to make of this estimate. In our robustness checks, we will examine whether effects are similar or different for the EDRp treatment if we look at the 2002 birth cohort (who were also fully exposed to the phonics treatment before the control group was entered). It will also be of interest to check whether the very high effect on the most disadvantaged group persists to the next cohort.

4.4. Further Empirical Findings

To consider the robustness of our results, we first check estimates of heterogeneous effects (Table 6) in various ways. Then we investigate whether or main effects (Table 4) vary by subject area and by whether we reclassify outcome variables at age 7 and 11 according to a binary variable indicating whether the student passes a threshold deemed to be the 'expected level' for their age according to the National Curriculum.

Heterogeneous Effects by Student Characteristics

We firstly check whether heterogeneous effects for the EDRp persist for the 2002 birth cohort, Secondly, we check how sensitive our results are to imputation of an exam score (at age 11) for those students who were not entered to the exam because they were deemed to be 'below level' of the test by the teacher. Thirdly, we estimate four-way interactions (between language and free school meal status) where we substitute the variable 'native

English speaker' with whether or not the students' first language is based on the Latin script. The Latin script is the basis for the largest number of alphabets of any writing system and is the most widely adopted in the world. However, one might hypothesise that a more structured approach to learn the English language is particularly important for those who have even more reliance on schools for learning the essential building blocks of the language. As this information is only derivable from 2009 onwards, we use the information when estimating effects for pupils of age 11. Finally, we estimate four-way interactions for girls and boys separately.

Table A3 shows estimated effects for the EDRp treatment for the 2002 cohort (rather than the 2001 cohort, to which our main effects pertain). This enables us to look at the effects for a group who entered treatment schools the year after they had received intensive support (as a result of the EDRp pilot). Comparing the original estimates (columns (1)-(3)) with the estimates using the 2002 cohort (columns (4)-(6)) shows that many of the estimated coefficients are similar. Interestingly, the negative effect for native English speakers at age 7 (and English speakers who are not eligible to receive free school meals in panel C) that we found for the 2001 cohort goes away for the 2002 cohort. Furthermore, the high effect estimated for non-native English speakers who are eligible to receive free school meals is exactly the same for this cohort relative to the control group. The treatment raises the age 11 score by 0.18σ whether we consider the 2001 or 2002 birth cohort.

As a second robustness test, we check whether estimated results at age 11 are sensitive to imputation of missing values on test scores where we know that the reason the children have not been entered for the test is because they are working 'below the required level'. This applies to about 4% (for the pre-policy cohort) – and is no different between the treatment and control group. In this case, we assign missing values to the lowest score given at the school that the student attended at this age. In Table A4, we show average results for the whole cohort (replicating the analysis reported for Table 4) and when we interact the treatment for native/non-native speakers of English and eligible/non-eligible for free school meals (replicating the analysis in panel C of Table 5). Columns (1) and (2) show results for the EDRp and columns (3) and (4) show results for the CLLD. Columns (1) and (3) replicate the results of our main analysis for comparison. We learn that the imputation has no implications for average results – they all suggest an effect which is close to zero and not statistically significant. In the bottom panel, we show that results are very comparable when we examine whether the treatment has a heterogeneous effect. The only result that changes is that the impact of the treatment on the group classified as 'non-native and free school meals' declines from 0.18σ to 0.13σ – making it closer to that estimated for the CLLD treatment (of about 0.10σ).

In Table A5, we substitute the 'non-native speaking' indicator for whether the students' first language uses the Latin script. We estimate this for students of age 11 only as we can derive the measure only for later years (from 2009). This shows effects that are similar to when we used the 'non-native speaking' indicator, although they are a little higher. For students whose language does not use the Latin script AND who are disadvantaged, the treatment effect at age 11 is 0.21σ and 0.13σ for the EDRp and CLLD respectively. For the CLLD treatment, there is an estimated effect even for these students if they are not classified as disadvantaged (0.089σ) but this is not the case for the EDRp treatment where there is no effect.

In Table A6, we show the four-way interactions from our main specification for boys and girls respectively at age 5, 7 and 11. The standard errors are larger (as we are splitting the sample) but produces results that are qualitatively similar and not systematically different for boys and girls. Results for the EDRp suggests that effects are stronger for girls at age 11, but the opposite is true for the CLLD.

Other Outcome Variables

We investigate whether the phonics treatment has any impact on other subjects at age 7 and age 11. We show results for reading, writing and maths at age 7 and for reading, English and maths at age 11.[18] This is shown in Table A7 (A.7.1 and A.7.2). The results at age 7 show that effect sizes are larger for writing than for reading, and also show the pattern of increasing effects for cohorts exposed younger (and for longer) to the new way of teaching reading. The results are also positive for maths. Results at age 11 show no overall effect of the treatment on reading, English or maths.

Finally in Table A8 (A.8.1 and A.8.2), we show results when we redefine outcomes at age 7 and 11 by whether the students achieved the 'expected level' at the age of 7 and 11 respectively. In the pre-policy year (2005), the percentage of students achieving the 'expected level' in the control group was 80% in reading at age 7 and 77% in English at age 11. The results show that for the group longest exposed to the policy, the treatment increased the probability of achieving the 'expected level' at age 7 in reading and writing by about 3 percentage points and by about 2 percentage points in maths. There is no average effect on any subject at age 11 (apart from a small effect for maths for the CLLD treatment).

5. Conclusion

The economics of education literature has well established that good teachers matter. But a critical, yet much less studied question, is whether 'good teaching' can be taught? Our empirical analysis shows that intensive training in the use of a 'new pedagogy' or technology produced strong effects for early literacy acquisition amongst young students. We are able to provide convincing evidence of causal effects because of the way in which training was

[18] We only have an overall English mark up to 2012 (and not a separate writing test). The writing test was changed about this time and we have no separate writing or English test that can be used in 2013. Thus, we can estimate the effect of the EDRp on English but not the CLLD (i.e. the relevant cohort did their Key Stage 2 tests in 2013).

staggered across different Local Authorities (and hence different schools). The initial effects are large and comparable to the early effects of project STAR in reducing class size. Furthermore, the costs were very modest because they only involved employing a literacy consultant to work with a school for a year. If effects only reflected the active involvement of the literacy consultant, one would not expect effects to persist for young students. The fact that effects are observed for younger students in years *after* the literacy consultant had been at the school (at least up until the control group enter the programme) suggests that the training and not the presence of the trainer explains the treatment effect. Effects are stronger for those exposed to the programme earlier (and for longer). It appears that the training really benefits measures of reading attainment (as well as writing) for young people.

However, most students learn to read eventually. This is the simplest explanation for why we do not see any overall effect of the intervention by age 11. There may of course be (unmeasured) benefits of learning to read well at an earlier age. However, these are not reflected in tests that we can observe at age 11 (in English and maths). Most interestingly, there are long-term effects at age 11 for those with a high probability of starting their school education as struggling readers. The results for our study suggests that there is a persistent effect for those classified as non-native English speakers and economically disadvantaged (as measured by free school meal status). The effect persists for these children who enter school with significant literacy deficits and is at least 0.10 of a standard deviation on the reading test at age 11. This is impressive given that the phonics approach is only actively taught up to the age of 7. Without a doubt it is high enough to justify the fixed cost of a year's intensive training support to teachers. Furthermore, it contributes to closing gaps based on disadvantage and (initial) language proficiency by family background.

Finally, and to conclude, that a relatively inexpensive policy introduced to primary schools administered by local authorities reduced literacy inequalities in such a way takes on an added significance given the radical and far-reaching schools policies underway in England. All schools are set to become academy schools which operate entirely outside of local authority control by the end of 2022. It is still unclear what roles local authorities may play in schooling, but it will certainly be massively diminished, and perhaps non-existent, once full academisation has happened. Thus the kind of policy we have studied in this paper will not be feasible once this has taken place. Of course, this has wider ramifications and relevance for other countries that are currently, or planning to, decentralise education in similar ways.

References

Australian Government. Department of Education, Science and Training (2005). Teaching reading. Report and recommendations. National enquiry into the teaching of literacy. Barton, Australia. Department of Education, Science and Training.

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

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, 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, 2633-79.

Elliott, C. (1996) The British Ability Scales II, Windsor, Berkshire: Nfer Publishing Company.

Eyles A. and S. Machin (2015) The Introduction of Academy Schools to England's Education, Centre for Economic Performance Discussion Paper 1368.

Hansen, K., H. Joshi and S. Dex (2010) Children of the 21st century: The First Five Years, Policy Press.

Hanushek, E., S. Rivkin and J. Kain (2005) Teachers, Schools and Academic Achievement, Econometrica, 73, 415–458.

Hempenstall, K. (1997) The Whole Language-Phonics Controversy: An Historical Perspective, Educational Psychology, 17, 399-418.

Hill, V. (2005) Through the Past Darkly: A Review of the British Ability Scales Second Edition, Child and Adolescent Mental Health, 10, 87-98.

Jacob, B. (2016) When Evidence is Not Enough: Findings From a Randomized Evaluation of Evidence-Based Literary Instruction (EBLI), National Bureau of Economic Research Working Paper 21643.

Johnston, R. and J. Watson (2005) The Effects of Synthetic Phonics Teaching on Reading and Spelling Attainment, Report for the Scottish Executive.

Jones, E. and I. Schoon (2008) Child Cognition and Behaviour, in Hansen, K. and Joshi, H. (eds.) Millennium Cohort Study Third Survey: A User's Guide to Initial Findings, London: Centre for Longitudinal Studies.

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

Krueger, A. and D. Whitmore (2001) The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project Star, Economic Journal, 111, 1-28.

Kuczera, M., S. Field and H. Windisch (2016). Building Skills for All: A Review of England. OECD.

Machin, S. and S. McNally (2008) The Literacy Hour, Journal of Public Economics, 92, 1441-62.

National Institute of Child Health and Human Development (NICHD) (2000. Report of the National Reading Panel. Teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups. NIH publication no. 00-4754. Washington CD. US Government Printing Office.

Pollack, M., and J. Pickarz (1963). Reading Programs and Problem Readers, New York. David McKay.

Primary National Strategy (2007) Letters and Sounds: Principles and Practice of High Quality Phonics, Department for Education and Skills, UK. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/190599/Letters_and_Sounds_-_DFES-00281-2007.pdf

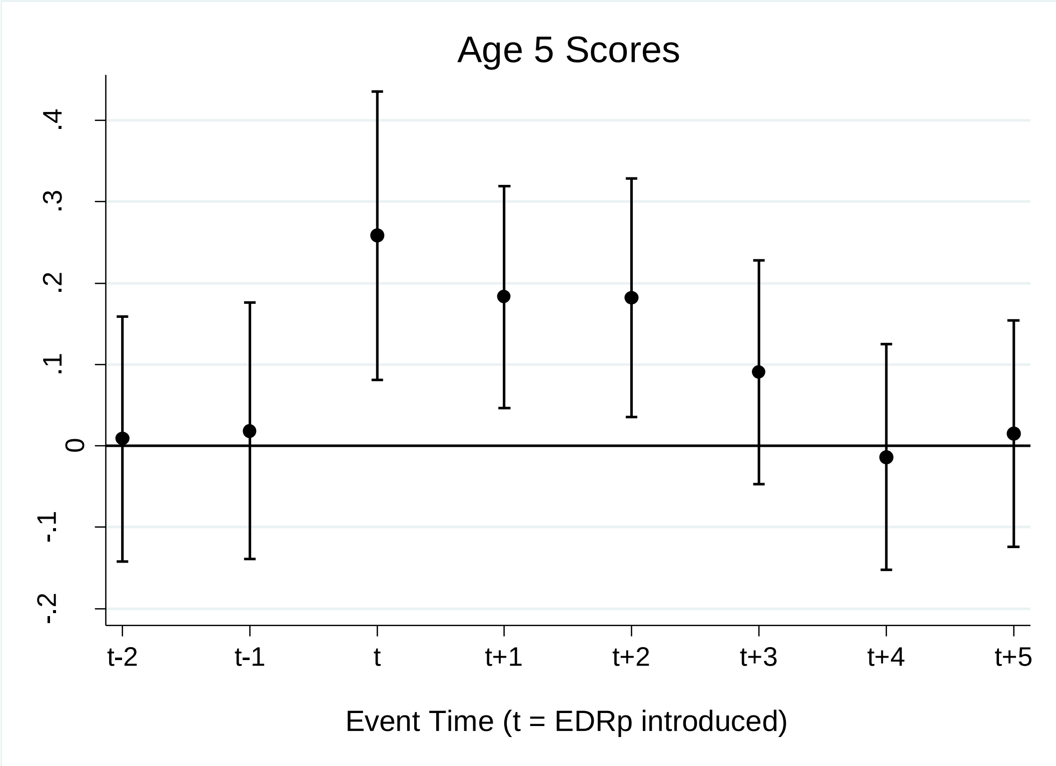
Rose, J. (2006) Independent Review of the Teaching of Early Reading, Department of Education and Skills.

Slavin, R.E., C. Lake, S. Davis, and N. Madden, (2009), Effective Reading Programs for the Elementary Grades: A Best Evidence Synthesis. Review of Educational Research 79(4): 1391-1466.

Vignoles, A. (2016) What is the Economic Value of Literacy and Numeracy? Basic Skills in Literacy and Numeracy are Essential for Success in the Labor Market, IZA World of Labor,2016: 229

Wyse, D. and U. Goswani. (2008) Synthetic Phonics and the Teaching of Reading, British Educational Research Journal, 34, 691-710.

Figure 1: Age 5 Reading Scores - Treatment x Year Coefficients
(Controlling for all observable variables)



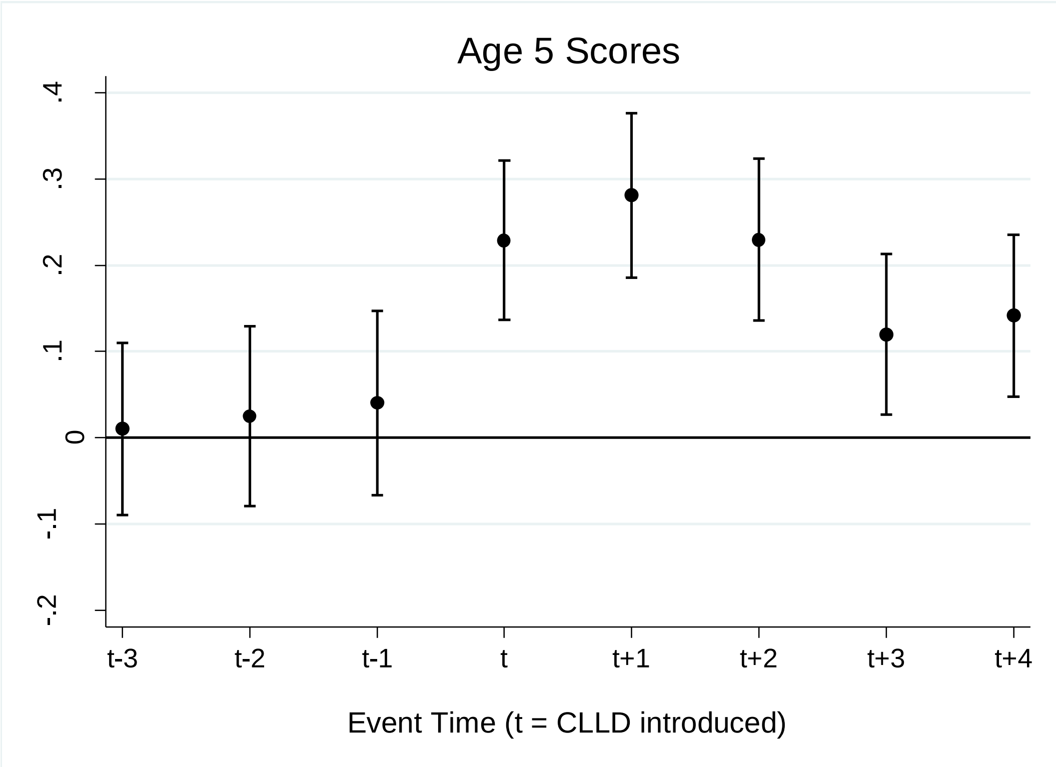


Table 1: How to Teach Reading Post Rose Review.

'Letters and Sounds: principles and practice of high quality phonics' (Primary National Strategy, 2007) As summarised by Wyse and Gosmani (2008) Following the teaching of general orientation to sound discrimination in nursery years, daily lessons for a six week period to feature 'discrete phonics teaching'. Teachers must 'teach at least 19 letters, and move children on from oral blending and segmentation to blending and segmenting with letters (p.48). Application of this knowledge during the Letters and Sounds lessons is limited to 'read or write a caption (with the teacher) using one or more high-frequency words and words containing the new letter (week 3 onwards)' (p.49). This is following by further discrete teaching, lasting up to 12 weeks. The purpose of this phase is to 'teaching another 25 graphemes, most of them comprising two letters (e.g. oa) so the children can represent each of about 42 phonemes by a grapheme' (p.74). Application at this stage is to 'read or write a caption or sentence using one or more tricky words and words containing the graphemes' (p.75). This pattern of a limited context for application of grapheme-phoneme correspondences continues through year one (age 5 to 6) until year two (age 6 to 7) at which point phonics instruction moves to an emphasis on spelling.

Table 2: Description of Groups

Groups	Phonics Programme	LA	Entry	Birth Cohort of Students First Exposed to Programme	Year of Age 5 Assessment	Year of Age 7 Assessment	Year of Age 11 Assessment
Treatment Group 1	EDRp	Schools in 18 LAs	2005/06	2001	2006	2008	2012
Treatment Group 2	CLLD	Schools in same 18 LAs + 32 new LAs	2006/07	2002	2007	2009	2013
Control group		Schools in next 50 LAs	2008/09 and 2009/10	2004	2009	2011	2015
		Schools in next 50 LAs	2009/10	2005	2010	2012	2016

Note: schools in the first 50 LAs (i.e. treatment groups) did come into the scheme in subsequent years. These schools are not included in the analysis.

Table 3: Summary Statistics for Groups of Schools Pre-Policy (2005)

	(1)	**(2)**	(3)	**(4)**	(5)	**(6)**	(7)	(8)
	All Primary Schools	**Treatment Group 1: Original EDRp Pilot (2006)**	Non-Selected Schools in 18 Local Authorities of EDRp Pilot (Not in First Year)	**Treatment Group 2: Schools in Post-Rose Report Programme: CLLD (First Wave, 2007)**	Non-Selected Schools in 50 Local Authorities of CLLD (Not in First Wave of EDRp)	**Control Group: Schools in Other 100 LAs That Entered Later (2009 and 2010)**	P-value: (2) – (6)	P-value: (4) – (6)
Age 5 Score, Communication, Language and Literacy	0	**-0.126**	-0.014	**-0.364**	-0.006	**-0.250**	0.049	0.006
Age 7 Reading Score	0	**-0.091**	-0.059	**-0.286**	-0.023	**-0.196**	0.002	0.000
Proportion Entitled to Free School Meals	0.181	**0.263**	0.230	**0.340**	0.210	**0.273**	0.563	0.000
Proportion Native English Speakers	0.880	**0.817**	0.860	**0.756**	0.884	**0.823**	0.814	0.000
Proportion White British/Irish	0.791	**0.694**	0.763	**0.641**	0.776	**0.722**	0.348	0.000
Number of Schools	16,429	**164**	2,264	**523**	5,500	**1,007**	1,171	1,530

Note: Treatment (columns (2) and (4)) and control groups (column (6)) in bold. The age 5 and age 7 scores are standardised to have a mean of 0 and a standard deviation of 1.

Table 4a: EDRp Treatment on Reading Assessments at Ages 5, 7 and 11

	(1)	(2)	(3)
	Age 5 Score	Age 7 Score	Age 11 Score
Treatment*1999 Birth Cohort	0.005	0.037***	0.003
(Treatment: Only Age 7)	(0.077)	(0.021)	(0.028)
Treatment*2000 Birth Cohort	0.072	0.040	-0.001
(Treatment: Age 6-7)	(0.081)	(0.025)	(0.027)
Treatment*2001 Birth Cohort	0.298*	0.075*	-0.018
(Full treatment: Age 5-7)	(0.094)	(0.024)	(0.031)
Additional Controls	Yes	Yes	Yes
R^2	0.273	0.121	0.123
Sample Size	17,279	191,342	163,272
Number of Schools	1185	1217	1217

Table 4b: CLLD Treatment on Reading Assessments at Ages 5, 7 and 11

	(1)	(2)	(3)
	Age 5 Score	Age 7 Score	Age 11 Score
*Treatment*1999 Birth Cohort*	0.009	-0.015	-0.024
(No Treatment: Placebo)	(0.050)	(0.015)	(0.019)
Treatment*2000 Birth Cohort	0.015	0.031**	-0.016
(Treatment: Only Age 7)	(0.053)	(0.016)	(0.018)
Treatment*2001 Birth Cohort	0.033	0.046*	0.021
(Treatment: Age 6-7))	(0.054)	(0.017)	(0.019)
Treatment*2002 Birth Cohort	0.217*	0.073*	0.019
(Full Treatment: Age 5-7)	(0.047)	(0.017)	(0.019)
Additional Controls	Yes	Yes	Yes
R^2	0.230	0.164	0.111
Sample Size	82,495	309,769	268,565
Number of Schools	1568	1598	1598

Notes: Baseline is the 1998 birth cohort (who undertook the Age 5, 7 and 11 assessments in 2003, 2005 and 2009 respectively). The outcome for age 5 is the (teacher assessed) standardised score in Communication, Language and Literacy. The outcome for age 7 is the (teacher assessed) standardised score in Key Stage 1 reading. The outcome for age 11 is the pupil's (externally assessed) standardised test score in reading. The 2001 and 2002 birth cohorts (in bold) are the first cohorts to have received the treatment throughout their education for the EDRp and CLLD respectively. For the EDRp, the 2000 birth cohort received the treatment in Year 1 (at age 6). The 1999 birth cohort received the treatment in Year 2 (at age 7). For the CLLD, the 2001 cohort received the treatment in Year 1 (at age 6). The 2000 birth cohort received the treatment in Year 2 (at age 7). Controls are: year dummies; school fixed effects, student gender, ethnicity; whether speaks English as an additional language; whether eligible to receive free school meals, whether receives a statement of Special Educational Needs; % of students in the year group by: gender, ethnicity, whether speaks English as an additional language, whether eligible to receive free school meals, whether receives a statement of Special Educational Needs. Standard errors clustered by school. Untreated groups are in italics. ***: p<0.10; ** p<.05; * p<0.01.

Table 5: Age 5 Test Score Differences,
Millennium Cohort Study Children in England

	Naming Vocabulary			Pattern Construction			Pattern Similarity		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
English Not First Language At Home	-0.978* (0.031)		-0.931* (0.030)	-0.283* (0.034)		-0.249* (0.034)	-0.117* (0.034)		-0.091* (0.034)
Free School Meals		-0.596* (0.028)	-0.529* (0.027)		-0.398* (0.030)	-0.380* (0.030)		-0.301* (0.030)	-0.294* (0.030)
Age and Gender Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Size	9706	9706	9706	9674	9674	9674	9718	9718	9718

Notes: The dependent variable is the relevant test score standardised to have mean zero and a unit standard deviation. Standard errors in parentheses. Weighted using MCS country-specific weights.
***: p<0.10; ** p<.05; * p<0.01.

Table 6: Heterogeneity in Estimated Treatment Effects by Non-Native Speaker Status and Free School Meals Eligibility

	EDRp v Control (Cohorts 1998 and 2001)			CLLD v Control (Cohorts 1998 and 2002)		
	Age 5	Age 7	Age 11	Age 5	Age 7	Age 11
	(1)	(2)	(3)	(4)	(5)	(6)
A. Speech Nativity						
Native Speaker	0.225*	0.052**	-0.045	0.211*	0.061*	0.001
	(0.083)	(0.024)	(0.031)	(0.050)	(0.018)	(0.020)
Non-Native Speaker	0.567**	0.134**	0.045	0.201**	0.113*	0.068**
	(0.277)	(0.051)	(0.063)	(0.081)	(0.028)	(0.032)
P-value	0.194	0.115	0.155	0.906	0.055	0.035
B. Free School Meals						
Free School Meals	0.290	0.135*	0.064	0.207*	0.136*	0.062**
	(0.182)	(0.019)	(0.050)	(0.067)	(0.023)	(0.026)
Non-Free School Meals	0.306*	0.042***	-0.061***	0.221*	0.045**	-0.002
	(0.107)	(0.023)	(0.031)	(0.051)	(0.018)	(0.020)
P-value	0.934	0.024	0.009	0.833	0.000	0.000
C. Speech Nativity and Free School Meals						
Native Speaker and Free	0.270	0.096**	0.011	0.182**	0.104*	0.042
School Meals	(0.183)	(0.046)	(0.052)	(0.078)	(0.025)	(0.028)
Native Speaker and Non-	0.217**	0.038	-0.061***	0.222*	0.042**	-0.017
Free School Meals	(0.088)	(0.024)	(0.032)	(0.054)	(0.020)	(0.021)
Non-Native Speaker and	0.300	0.216*	0.181**	0.221**	0.195*	0.099**
Free School Meals	(0.406)	(0.077)	(0.087)	(0.108)	(0.038)	(0.041)
Non-Native Speaker and	0.671**	0.093***	-0.031	0.205**	0.095*	0.070**
Non-Free School Meals	(0.272)	(0.054)	(0.066)	(0.100)	(0.030)	(0.035)
P-value : Native, FSM=Native, Non-FSM	0.781	0.217	0.167	0.628	0.013	0.032
P-value: Non-Native FSM=Non-Native, non-FSM	0.350	0.122	0.014	0.904	0.012	0.464

Notes: Under each heading, results are shown from separate regressions where personal characteristics of pupils are interacted with birth cohort dummies and treatment status. The reported coefficients show the interaction between treatment, birth cohort and personal characteristic of the student. The comparison group is 'non-treated'.
***: p<0.10; ** p<.05; * p<0.01.

Appendix

Table A1: Communication, Language and Literacy at Age 5

	EDRp		**CLLD**	
	Original Pilot and Control Schools		Post-Report Programme (First Wave) and Control Schools	
	(1)	(2)	(3)	(4)
Treatment*2004	0.021	0.009	0.022	0.010
[Birth Cohort: 1999]	(0.076)	(0.075)	(0.051)	(0.050)
Treatment*2005	0.027	0.019	0.041	0.025
[Birth Cohort: 2000]	(0.080)	(0.079)	(0.053)	(0.052)
Treatment*2006	**0.246***	**0.258***	0.053	0.040
[Birth Cohort: 2001]	**(0.086)**	**(0.089)**	(0.053)	(0.053)
Treatment*2007	**0.191***	**0.183***	**0.242***	**0.229***
[Birth Cohort: 2002]	**(0.068)**	**(0.068)**	**(0.047)**	**(0.046)**
Treatment*2008	**0.197***	**0.182****	**0.299***	**0.281***
[Birth Cohort: 2003]	**(0.072)**	**(0.073)**	**(0.048)**	**(0.048)**
Treatment*2009	0.100	0.091	0.253*	0.23**
[Birth Cohort: 2004]	(0.067)	(0.069)	(0.047)	(0.047)
Treatment*2010	-0.007	-0.014	0.139*	0.120**
[Birth Cohort: 2005]	(0.068)	(0.069)	(0.047)	(0.047)
Treatment*2011	0.026	0.015	0.163*	0.142*
[Birth Cohort: 2006]	(0.068)	(0.070)	(0.047)	(0.047)
Additional Controls	No	Yes	No	Yes
R^2	0.107	0.182	0.102	0.174
Sample Size	267,094	267,093	346,410	346,409
Number of Schools	1234	1234	1603	1603

Notes: The outcome is the (teacher assessed) standardised score in Communication, Language and Literacy. Baseline is the treatment year 2003 (or 1998 birth cohort). Controls are: year dummies; school fixed effects. Standard errors clustered by school. Additional controls: student gender, ethnicity; whether speaks English as an additional language; whether eligible to receive free school meals, whether receives a statement of Special Educational Needs % of students in the year group by: gender, ethnicity, whether speaks English as an additional language, whether eligible to receive free school meals, whether receives a statement of Special Educational Needs. Control schools come into the programme in either 2009 or 2010. Highlighted cells show when the programme was operational in treated schools, but not in *any* of the control schools.
***: p<0.10; ** p<.05; * p<0.01.

Table A2: Local Authorities in Treatment and Control Groups

Groups	Phonics Programme	LA	Entry	LA names
Treatment Group 1	EDRp	Schools in 18 LAs	2005/06	Barnsley, Cheshire, Coventry, Hertfordshire, Islington, Leeds, Liverpool, Luton, Manchester, Medway, Nottingham, Peterborough, Redcar and Cleveland, Stoke-on-Trent, Tameside, Tower Hamlets, Waltham Forest, Wiltshire
Treatment Group 2	CLLD	Schools in same 18 LAs + 32 new LAs	2006/07	18 LAs above AND Bath and North East Somerset, Birmingham, Blackburn with Darwen, Bury, Dorset, Ealing, East Sussex, Essex, Gloucestershire, Greenwich, Hackney, Hammersmith and Fulham, Haringey, Hartlepool, Kent, Knowsley, Lambeth, Lewisham, Middlesbrough, North Tyneside, Oldham, Sandwell, Sefton, Sheffield, Shropshire, Southampton, Southwark, Surrey, Swindon, Thurrock, Torbay. Kingston-upon-Hull*
Control group		Schools in next 50 LAs	2008/09 and 2009/10	All remaining Local Authorities represented in control group (for schools that came into the treatment in 2008/09 and 2009/10)
		Schools in next 50 LAs	2009/10	

Notes: *Kingston-upon-Hull withdrawn due to floods (and no data available)

Table A3: Heterogeneity in Estimated Treatment Effects by Non-Native Speaker Status and Free School Meals Eligibility - Different Cohorts for the EDRp v Control

	EDRp v Control (Cohorts 1998 and 2001)			**EDRp v Control (Cohorts 1998 and 2001)**		
	Age 5	Age 7	Age 11	Age 5	Age 7	Age 11
	(1)	(2)	(3)	(4)	(5)	(6)
A. Speech Nativity						
Native Speaker	0.225*	0.052**	-0.045	0.149**	0.069*	0.021
	(0.083)	(0.024)	(0.031)	(0.064)	(0.026)	(0.033)
Non-Native Speaker	0.567**	0.134**	0.045	0.107	0.055	0.039
	(0.277)	(0.051)	(0.063)	(0.145)	(0.048)	(0.056)
P-value	0.194	0.115	0.155	0.767	0.768	0.754
B. Free School Meals						
Free School Meals	0.290	0.135*	0.064	0.108	0.103**	0.094***
	(0.182)	(0.019)	(0.050)	(0.124)	(0.043)	(0.049)
Non-Free School Meals	0.306*	0.042***	-0.061***	0.158**	0.043***	-0.007
	(0.107)	(0.023)	(0.031)	(0.069)	(0.026)	(0.032)
P-value	0.934	0.024	0.009	0.711	0.133	0.030
C. Speech Nativity and Free School Meals						
Native Speaker and Free School Meals	0.270	0.096**	0.011	0.122	0.065	0.053
	(0.183)	(0.046)	(0.052)	(0.122)	(0.049)	(0.053)
Native Speaker and Non-Free School Meals	0.217**	0.038	-0.061***	0.160**	0.069*	0.012
	(0.088)	(0.024)	(0.032)	(0.070)	(0.027)	(0.036)
Non-Native Speaker and Free School Meals	0.300	0.216*	0.181**	0.103	0.193*	0.184**
	(0.406)	(0.077)	(0.087)	(0.290)	(0.066)	(0.082)
Non-Native Speaker and Non-Free School Meals	0.671**	0.093***	-0.031	0.121	0.000	-0.026
	(0.272)	(0.054)	(0.066)	(0.151)	(0.054)	(0.056)
P-value : Native, FSM=Native, Non-FSM	0.781	0.217	0.167	0.776	0.924	0.445
P-value: Non-Native FSM=Non-Native, Non-FSM	0.350	0.122	0.014	0.957	0.010	0.005

Notes: As for Table 6. Columns (1)-(3) are reproduced from Table 6. Columns (4)-(6) report the same specifications for the 1998 and 2002 cohorts.

Table A4: Age 11 Results With and Without Imputation

A. Baseline Results (Table 4)	**EDRp**		**CLLD**	
	(1)	(2)	(3)	(4)
	Age 11 (Table 4a)	Age 11 (With Imputation)	Age 11 (Table 4b)	Age 11 (With Imputation)
Treatment*1999 Birth Cohort	0.003	-0.002	-0.024	-0.031
	(0.028)	(0.026)	(0.019)	(0.018)
Treatment*2000 Birth Cohort	-0.001	-0.010	-0.016	-0.019
	(0.027)	(0.026)	(0.018)	(0.017)
Treatment*2001 Birth Cohort	-0.018	-0.028	0.021	0.013
	(0.031)	(0.029)	(0.019)	(0.018)
Treatment*2002 Birth Cohort			0.019	0.013
			(0.019)	(0.018)
Additional Controls	Yes	Yes	Yes	Yes
R^2	0.123	0.143	0.111	0.130
Sample Size	163,272	168,689	268,565	277,474
Number of Schools	1217	1217	1598	1598

B. Heterogeneity Results (Table 6)	**EDRp v Control (Cohorts 1998 and 2001)**		**CLLD v Control (Cohorts 1998 and 2002)**	
	(5)	(6)	(7)	(8)
	Age 11 (Table 6)	Age 11 (With Imputation)	Age 11 (Table 6)	Age 11 (With Imputation)
Native and Free School Meals	0.011	0.013	0.042	0.045***
	(0.052)	(0.054)	(0.028)	(0.027)
Native and Non-Free School Meals	-0.061***	-0.066**	-0.017	-0.022
	(0.032)	(0.031)	(0.021)	(0.021)
Non-Native and Free School Meals	0.181**	0.132***	0.099**	0.097**
	(0.087)	(0.080)	(0.041)	(0.039)
Non-Native and Non-Free School Meals	-0.031	-0.045	0.070**	0.058***
	(0.066)	(0.064)	(0.035)	(0.034)
P-value : Native, FSM=Native, Non-FSM	0.0167	0.142	0.032	0.011
P-value: Non-Native FSM=Non-Native, Non-FSM	0.014	0.032	0.464	0.300
Sample Size	87,985	90,885	114,592	118,207
Number of Schools	1217	1217	1598	1598

Notes: As for Table 4 for (1)-(4). As for Table 6 for (5)-(8). Columns (1) and (3) are reproduced from Table 4, and columns (2) and (4) are the same for the extended sample with imputation. Columns (5) and (7) are reproduced from Table 6, and columns (6) and (8) are the same for the extended sample with imputation. The test score is imputed for students who were not entered into the test because they were working below the level of the English test.

Table A5: Heterogeneity in Estimated Treatment Effects by Language Type (i.e. Latin Script v Non-Latin script) and Free School Meals Eligibility

	EDRp v Control (Cohorts 1998 and 2001)	**CLLD v Control (Cohorts 1998 and 2002)**
	Age 11	Age 11
	(1)	(2)
Latin Script and Free School Meals	0.011	0.031
	(0.053)	(0.027)
Latin Script and Non-Free School Meals	-0.064**	-0.016
	(0.033)	(0.021)
Non-Latin Script and Free School Meals	0.210**	0.130**
	(0.093)	(0.048)
Non-Latin Script and Non-Free School Meals	0.006	0.089**
	(0.072)	(0.040)
P-value : Native, FSM=Native, Non-FSM	0.150	0.068
P-value: Non-Native FSM=Non-Native, Non-FSM	0.017	0.385

Notes: As for Table 6

Table A6: Heterogeneous Effects for Boys and Girls

	EDRp v Control (Cohorts 1998 and 2001)			EDRp v Control (Cohorts 1998 and 2001)		
	Age 5	Age 7	Age 11	Age 5	Age 7	Age 11
	(1)	(2)	(3)	(4)	(5)	(6)
A. Boys						
Native Speaker and Free School Meals	0.294	0.120***	0.061	0.196***	0.144*	0.106*
	(0.293)	(0.064)	(0.065)	(0.108)	(0.037)	(0.039)
Native Speaker and Non-Free School Meals	0.271**	0.034	-0.075***	0.217*	0.066**	-0.023
	(0.133)	(0.035)	(0.043)	(0.070)	(0.026)	(0.028)
Non-Native Speaker and Free School Meals	0.731	0.250*	0.262*	0.190	0.167*	0.086
	(0.481)	(0.090)	(0.087)	(0.145)	(0.052)	(0.054)
Non-Native Speaker and Non-Free School Meals	0.748**	0.153**	-0.042	0.204	0.111*	0.073
	(0.334)	(0.071)	(0.105)	(0.138)	(0.039)	(0.046)
P-value : Native, FSM=Native, Non-FSM	0.941	0.204	0.056	0.852	0.039	0.002
P-value: Non-Native FSM=Non-Native, Non-FSM	0.978	0.367	0.004	0.938	0.335	0.828
B. Girls						
Native Speaker and Free School Meals	0.087	0.071	-0.045	0.122	0.060***	-0.020
	(0.292)	(0.060)	(0.073)	(0.099)	(0.033)	(0.037)
Native Speaker and Non-Free School Meals	0.203	0.045	-0.049	0.254*	0.009	-0.014
	(0.138)	(0.029)	(0.037)	(0.068)	(0.024)	(0.027)
Non-Native Speaker and Free School Meals	-0.135	0.177	0.099	0.199	0.232*	0.121**
	(0.583)	(0.109)	(0.120)	(0.144)	(0.049)	(0.052)
Non-Native Speaker and Non-Free School Meals	0.710***	0.028	-0.019	0.191	0.073***	0.066
	(0.402)	(0.073)	(0.075)	(0.128)	(0.037)	(0.043)
P-value : Native, FSM=Native, Non-FSM	0.716	0.662	0.963	0.216	0.143	0.867
P-value: Non-Native FSM=Non-Native, Non-FSM	0.148	0.180	0.366	0.961	0.003	0.343

Notes: As for Table 6

Table A7. Heterogeneity by Subject

For EDRp, 2001 cohort receives treatment for 3 years; 2000 for 2 years; 1999 for 1 year.
For CLLD, 2002 cohort receives treatment for 3 years, 2001 for 2 years; 2000 for 1 year

A.7.1. EDRp and CLLD Treatments at Age 7, Reading, Writing and Maths

	EDRp			**CLLD**		
	(1)	(2)	(3)	(4)	(5)	(6)
	Reading	Writing	Maths	Reading	Writing	Maths
Treatment*1999 Birth Cohort	0.037***	0.052**	0.043***	-0.015	-0.016	0.014
	(0.021)	(0.024)	(0.022)	(0.015)	(0.016)	(0.016)
Treatment*2000 Birth Cohort	0.040	0.057**	0.045	0.031**	0.052*	0.045*
	(0.025)	(0.027)	(0.027)	(0.016)	(0.017)	(0.017)
Treatment*2001 Birth Cohort	0.075*	0.093*	0.056**	0.046*	0.055*	0.052*
	(0.024)	(0.027)	(0.027)	(0.017)	(0.019)	(0.018)
Treatment*2002 Birth Cohort				0.073*	0.092*	0.061*
				(0.017)	(0.019)	(0.019)
Additional Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.178	0.202	0.155	0.164	0.184	0.142
Sample Size	191,342	191,325	191,330	309,769	309,751	309,737
Number of Schools	1217	1217	1217	1598	1598	1598

Table A.7.2: EDRp and CLLD Treatments at Age 11, Reading, English and Maths

	EDRp			CLLD		
	(1)	(2)	(3)	(4)	(5)	(6)
	Reading	English	Maths	Reading	English	Maths
Treatment*1999 Birth Cohort	0.003	0.032	0.002	-0.024	-	0.022
	(0.028)	(0.030)	(0.025)	(0.019)	-	(0.018)
Treatment*2000 Birth Cohort	-0.001	0.009	-0.006	-0.016	-	0.006
	(0.027)	(0.030)	(0.024)	(0.018)		(0.017)
Treatment*2001 Birth Cohort	-0.018	0.010	-0.028	0.022	-	0.017
	(0.031)	(0.028)	(0.026)	(0.019)		(0.018)
Treatment*2002 Birth Cohort				0.019	-	0.026
				(0.019)		(0.019)
Additional Controls	Yes	Yes	Yes	Yes		Yes
R^2	0.123	0.137	0.104	0.111	-	0.097
Sample Size	163,270	162,448	163,293	268,563	-	269,018
Number of Schools	1217	1217	1202	1598		1598

Notes: As for Table 4. Columns (1) and (4) are reproduced from Table 4, and columns (2), (3), (5) and (6) use alternative subject scores.

Table A8. Heterogeneity by Subject: Threshold effects

For EDRp, 2001 cohort receives treatment for 3 years; 2000 for 2 years; 1999 for 1 year.
For CLLD, 2002 cohort receives treatment for 3 years, 2001 for 2 years; 2000 for 1 year

A.8.1. EDRp and CLLD Treatments at Age 7, Reading, Writing and Maths, Probability of Achieving 'Expected Standard' (Level 2) or More

	EDRp			CLLD		
	(1)	(2)	(3)	(4)	(5)	(6)
	Reading	Writing	Maths	Reading	Writing	Maths
Treatment*1999 Birth Cohort	0.007	0.020**	0.008	-0.006	-0.008	-0.000
	(0.008)	(0.009)	(0.007)	(0.006)	(0.007)	(0.005)
Treatment*2000 Birth Cohort	0.016***	0.029*	0.014	0.011***	0.019*	0.013**
	(0.009)	(0.011)	(0.009)	(0.006)	(0.007)	(0.005)
Treatment*2001 Birth Cohort	0.029*	0.036*	0.019**	0.014**	0.024*	0.010***
	(0.009)	(0.011)	(0.008)	(0.006)	(0.008)	(0.006)
Treatment*2002 Birth Cohort				0.031*	0.036*	0.020*
				(0.006)	(0.007)	(0.006)
Additional Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.121	0.140	0.105	0.111	0.126	0.096
Sample Size	191,675	191,675	191,675	310,304	310,302	310,307
Number of Schools	1217	1217	1217	1598	1598	1598

Table A.8.2: EDRp and CLLD Treatments at Age 11, Reading, Writing and Maths, Probability of Achieving 'Expected Standard' (Level 4) or More

	EDRp			CLLD		
	(1)	(2)	(3)	(4)	(5)	(6)
	Reading	English	Maths	Reading	English	Maths
Treatment*1999 Birth Cohort	0.002	0.003	0.009	-0.006	-0.007	0.002
	(0.008)	(0.010)	(0.009)	(0.006)	(0.007)	(0.007)
Treatment*2000 Birth Cohort	-0.009	-0.008	-0.003	-0.001	0.000	0.009
	(0.007)	(0.009)	(0.009)	(0.005)	(0.006)	(0.006)
Treatment*2001 Birth Cohort	-0.002	0.000	-0.009	0.001	0.005	0.002
	(0.007)	(0.009)	(0.009)	(0.005)	(0.006)	(0.006)
Treatment*2002 Birth Cohort				0.006	-0.007	0.015**
				(0.005)	(0.133)	(0.006)
Additional Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.091	0.105	0.080	0.082	0.102	0.076
Sample Size	164,372	169,188	168,685	269,905	219,301	277,065
Number of Schools	1217	1217	1217	1598	1598	1598

Notes: As for Table A7. Threshold dependent variables are used in each case in place of standardised test scores.