The Literacy Hour

Stephen Machin* and Sandra McNally**

March 2004

* *Department of Economics, University College London, Centre for the Economics of Education and Centre for Economic Performance, London School of Economics*

** *Centre for the Economics of Education and Centre for Economic Performance, London School of Economics*

Abstract

In this paper, we evaluate the effect of the literacy hour in English primary schools on pupil attainment. The National Literacy Project (NLP) was undertaken in about 400 English primary schools in 1997 and 1998. We compare the reading and overall English attainment of children in NLP schools as compared to a set of control schools at the end of primary school education (age 11). We find a larger increase in attainment in reading and English for pupils in NLP schools as compared to pupils not exposed to the literacy hour between 1996 and 1998. Since gender gaps in English performance exist (in favour of girls), we consider whether the literacy hour has had a differential impact by gender. We find some evidence that at age 11, boys received a greater benefit than girls. Finally, we show the policy to be cost effective.

These findings are of strong significance when placed into the wider education debate about what works best in schools for improving pupil performance. The evidence reported here suggests that public policy aimed at changing the content and structure of teaching can significantly raise pupil attainment.

JEL Keywords: The literacy hour; English; reading. JEL Classification: I2.

Acknowledgements

We would like to thank Matthew Young at DfES for making available the codes for identifying schools that took part in the National Literacy Project. We also thank Mike Treadaway of the Fischer Trust for providing the pupil-level attainment data. We are very grateful to Sami Berlinksi, Steve Gibbons, Paul Johnson, Eric Maurin, Patrick Puhani, Anna Vignoles, Matthew Young and participants in seminars at the Institute of Education, HM Treasury, LSE, Newcastle and UCL for helpful comments.

1. Introduction

Literacy matters. Many people in different countries fail to reach even basic levels of literacy. This severely hampers their individual circumstances and lowers national productivity. For example, in the UK, the Moser report (DfEE, 1999) identifies one in five adults as not being functionally literate. How can we ensure that future generations of adults do not suffer from such problems? One way is by trying to provide a means whereby literacy levels and by association overall pupil attainment can be raised through government policy.

More generally, there is a substantial body of research investigating how best to raise pupil attainment. The economic literature primarily addresses the impact of changing school resources such as class size, teacher quality and measures of expenditure (for example, Angrist and Lavy, 1999; Card and Krueger, 1992; Hanushek, 1997, 2003). Educationalists address many of the same issues (as reviewed by Sammons, 1999, or Teddlie and Reynolds, 2000), but also investigate the efficacy of what is taught in particular subjects (for example, see Stainthorp, 1996, for a review of evidence on what children need to learn to become skilled readers) and how this is put across in the classroom in terms of such issues as time on task, frequency of monitoring pupils' work and grouping arrangements (e.g. Scheerens, 2000; Creemers, 1994).[1]

Some of this research appears to have influenced public policy. For example, with regard to reading, US research findings have been used as a rationale to underlie the

[1] One of the few papers in the economic literature addressing the potential effectiveness of how subjects are taught is Angrist and Lavy's (2002) paper addressing the role of Computer Aided Instruction. Another exception is the paper by Glewwe et al. (2000), which investigates the impact of flip charts in Kenyan classrooms.

reading component of the No Child Left Behind Act of 2001.[2] As discussed by Beard (2000), this research, as well as the experience of 'Success for All' in the US (Slavin and Madden, 2000, 2003), has been influential in the development of a national policy on literacy in England (the National Literacy Strategy). Both these policies involve the implementation of a framework stating what should be taught and specifying the structure of the lesson, as well as its duration.

An important question is whether policies like these, through affecting what and how a subject is taught, can raise attainment in schools. This is the subject matter of this paper. In order to shed light on what the literacy hour can do to pupil performance we evaluate the forerunner of the National Literacy Strategy, the National Literacy Project (NLP) which was introduced to a number of Local Education Authorities before the policy was rolled out nationally in the late 1990s. A strong rationale for this policy was to try to alleviate the very low levels of reading and writing skills held by children in many schools, but particularly in badly performing inner city schools.[3]

The main aim of the National Literacy Strategy has been to raise standards of literacy in primary schools by improving the quality of teaching through more focused literacy instruction and effective classroom management. It is also an attempt to improve school management of literacy through target-setting linked to systematic planning and monitoring and evaluation. To this end, key components of the policy are a *framework for teaching*, which sets out termly teaching objectives for the 5-11 age range and

[2] In particular the 'Success for All – Reading First' research: see the Success for All Foundation, "Success For All-Reading First: fulfilling the requirements of the Reading First Legislation" (http://successforall.net/current/ReadingFirst/index.htm).

[3] See West and Pennell (2003) for a discussion of the low levels of achievement in the inner cities.

provides a practical structure of time and class management for a daily literacy hour.[4] Training material is also provided for teachers (i.e. course booklets, videotapes etc. to support three in-service training days and some further training in after-school sessions, in the first year of implementation, with continued funded training offered in subsequent years). The policy, which was introduced to all primary schools at the start of the 1998/99 school year and is still ongoing, was preceded by a project that was implemented in about 400 schools two years earlier, during the 1996/97 and 1997/98 school years - the National Literacy Project (NLP).[5] From our perspective, the interesting feature of the NLP is that it resulted in some children being exposed to up to two years of the literacy hour in a period when other children were not.

In the United States, 'Success for All' has been described as a more intensive way of delivering the aims of the National Literacy Strategy (Hopkins, 1999). It has been suggested that this US program is, arguably, the most widely implemented, widely researched and widely critiqued educational program in the US (Borman and Hewes, 2002). However, like the National Literacy Project, it does not appear to have been evaluated by economists. Nor has it yet found its way into the economic literature on what can raise pupil attainment. On account of this, our assessment and evaluation of the National Literacy Project takes on an added significance - it potentially sheds more light on what can raise attainment in debates within the economics of education. Moreover, the NLP provides a better testing ground than the 'Success for All' program since the latter is

[4] In this paper, we refer to these set of measures as 'the literacy hour'.

[5] The National Literacy Project was not explicitly viewed as a pilot for the National Literacy Strategy at the time of its introduction. However on the recommendation of the Literacy Task Force, the approaches were incorporated into the National Literacy Strategy, which all schools were encouraged to adopt, with additional funding and support, from 1998.

subject to a potentially important selection problem - 80 per cent of a school's staff must vote to adopt the program and the school must be able to afford the program.[6]

There are existing, mostly descriptive, evaluations of the National Literacy Project (Sainsbury, 1998; OFSTED, 1998). They have not used a 'treatment' and 'control' group design. In these evaluations, the treatment group of pupils in NLP schools has been monitored over time, where progress has been measured on the basis of 'what is expected' according to the National Curriculum. Similarly to studies about 'Success for All', evaluations have been highly favourable about the impact of these policies.

However, this is the first economic study of whether a policy of this type actually works in raising standards of literacy. We set up empirical tests of whether exposure to the NLP enabled an increase in reading and English attainment, as hoped by those designing the policy. Due to the nature of the project, we have an opportunity to evaluate the impact of the literacy hour on attainment in schools that implemented the policy as compared to those that did not. Specifically, we use a difference-in-difference framework, sometimes coupled with statistical matching methods, to carry out an analysis of what happened to pupil attainment before and after introduction of the NLP in schools affected by the policy relative to that in a control group of schools.

A further aspect of this policy is its potential impact on the gender gap in pupil attainment. Boys have traditionally performed considerably worse at literacy-related activities, where there has been a gender gap in favour of girls for many years.[7] For

[6] According to Herman *et al*. (1999), Success for All is one of the most expensive whole-school reforms with estimates of first-year personnel, material and training costs of between $70,000 and $270,000 for a typical school.

[7] Machin and McNally (2003) show there to be a significant gender gap (favouring girls) in reading abilities dating at least far back as 1980 using data on reading tests administered to children of the British Cohort Study, a birth cohort of all children born in a week of April 1970.

example, in the school year before the NLP was introduced, 49 percent of 11 year old boys achieved the government target of level 4 or above in Key Stage 2 (KS2) English, whereas 64 percent of girls achieved this target (see Table 1). There are many reasons for why this gender gap exists, and there is a very large literature on gender gaps in education - again mostly within the education field (Maynard, 2002; White, 1996; Millard, 1996), with some more recent work in economics (Jacob, 2002). One very relevant issue is whether the highly structured nature of the literacy hour had the potential to make inroads into this gender gap. If one thinks boys have a greater problem with concentration and focus relative to girls, it may well be that the NLP could differentially benefit boys. We investigate this hypothesis towards the end of the paper.

The structure of the rest of the paper is as follows. In Section 2, we discuss the National Literacy Strategy in English schools, and give details about the National Literacy Project. In Section 3, we describe the data used in our empirical analysis. Section 4 provides estimates of the impact of the NLP on primary school attainment. In Section 5, we address the question as to whether NLP had a differential impact by gender. Then in Section 6, we estimate the relationship between higher literacy skills and labour market earnings and use this to convert our estimate of the NLP impact to a monetary measure. We compare the present value of the discounted benefits per pupil to the costs and thus assess the cost-effectiveness of the NLP policy. Conclusions are presented in Section 7.

2. The National Literacy Strategy and Patterns of English Attainment

The National Literacy Strategy

The National Literacy Strategy was introduced into all English primary schools during the school year 1998/99 with the aim of improving standards of literacy, in particular as measured by the percentage of 11 year olds reaching the expected standard for their age in national tests (i.e. level 4 at Key Stage 2). The Strategy involves the introduction of a daily literacy hour, which is divided between 10-15 minutes of whole class reading or writing; 10-15 minute whole-class session on word level work (e.g. phonics, spelling) and sentence level work; 25-30 minutes of directed group activities; and a plenary session at the end for pupils to revisit the objectives of the lesson, reflect on what they have learnt and consider what they need to do next.

This well-structured hour was implemented so as to address some of the criticisms of how English was taught in primary schools. For example, an OFSTED (Office for Standards in Education) report about the teaching of reading in Inner London primary schools included criticism of the following practices: free reading with little or no intervention by the teacher; too much time spent hearing individual pupils read; insufficient attention to the systematic teaching of an effective programme of phonic knowledge and skills (OFSTED, 1996). It was thought that standards in the teaching of reading varied hugely from school to school, with many primary teachers not having had the opportunity to update their skills to take account of evidence about effective methods of teaching reading and how to apply them (Literacy Task Force, 1997a).

The forerunner of the National Strategy was the National Literacy Project, which was introduced in about 400 junior schools during the school years 1996/97 and

1997/98.[8] About 80 per cent of NLP schools were located in inner cities, which is where the most disadvantaged and poorly performing schools in England are concentrated. Most of the schools entered the Project because they had weaknesses in reading and writing (OFSTED, 1999). The planned cost of the NLP was £12.5 million over five years.

This NLP policy is of considerable interest for evaluating the impact of structured literacy through the literacy hour as it exposed children in some schools to the policy whereas children in other schools did not participate. Hence, for the purposes of policy evaluation, the former group can be considered as the relevant treatment group. To evaluate the policy, one needs to define a counter-factual where children did not receive treatment. We spend considerable time below defining the comparison group against which to benchmark the policy impact and describing the methodology used to evaluate the NLP in this treatment-control setting.

Changing Patterns of English Attainment

The background to the introduction of the National Literacy Strategy was recognition that the literacy skills and attainment of primary school children in England were a problem, especially in some inner city schools. Table 1 provides some statistics on English attainment at age 11. It shows the percentage of children reaching the government target of level 4 (or above) in English and in reading.[9] Years denote the end of school year - namely when children take their tests (so, for example, 1996 refers to the 1995/6 school year). We adopt this year naming convention throughout this paper.

In 1996, the year before the introduction of NLP, 57 percent of children achieved at least level 4 in English (the overall English levels are arrived at by combining separate

[8] The NLP was also launched in an additional 112 'first' or infant schools, which we do not consider here.
[9] The maximum level attainable at Key Stage 2 is Level 6.

test scores for reading and writing). One sees a gradual increase in this measure over time. By 2002, the percentage achieving level 4 or above had risen to 75. Although sizable gender gaps are evident, it is interesting to note for the later analysis that between 1996 and 2002 boys improved their relative position, with a rise of 21 percentage points over this time period compared to a 15 percentage point rise for girls.

The same pattern is evident for reading. We only have statistics here from 1997 onwards. The percentage achieving at least level 4 increased from 67 to 80 percent between 1997 and 2002. Once again, although sizeable gender gaps are present at each point in time, over this period boys experienced an increase of 14 percentage points compared to an increase of 12 percentage points for girls.

3. The Data and Modelling Approach

Key Stage Attainment Data

The empirical analysis is based on administrative records of pupil-level attainment and on school-level data. The former consists of very detailed attainment information from when students were 11 years of age. At age 11, all pupils in England are tested at the end of what is known as Key Stage 2 (i.e. the phase of education from 7-11 years, as stipulated in the National Curriculum). The first available year of national Key Stage 2 data for pupils is in 1996, which corresponds to the school year before the National Literacy Project was introduced. We also use the national Key Stage 2 data in the 'policy on' school years of 1997 and 1998, before the National Literacy Strategy was introduced nationwide.

The pupil-level administrative files have detailed information on attainment, the gender of the student and codes for the school attended. This latter information allows the files to be matched up to national school-level data available in the School Performance Tables and the LEA and School Information Service (LEASIS). Available information includes measures of school outcomes (results, absences), inputs (e.g. pupil-teacher ratios), disadvantage (e.g. the percentage of students eligible for Free School Meals or those identified as having Special Educational Needs) and other school characteristics (e.g. school type).[10]

With regard to outcome measures, we concentrate on two measures of attainment at the end of primary school: the percentile reading score and the percentage of students attaining level 4 or above in Key Stage 2 English. Reading is one of several tests students take as part of their Key Stage 2 assessment. Since the marking scheme can change over time, we convert raw scores to percentile scores. The scores of various tests are aggregated and then converted to an overall 'level' (which is in a range of 2-6). The key indicator of policy interest is the percentage of students attaining level 4 and above at age 11, which is the standard deemed to be appropriate at this age.[11]

Specifics of the NLP and Definition of Control Schools

As noted above, the NLP was introduced in the school year 1996/97 for the first cohort of schools and 1997/98 for the second cohort. Of about 400 schools in the project (about half in each cohort), 80 per cent were located in the inner cities – within several Local Education Authorities in London and also Sandwell, Liverpool, Manchester,

[10] A full list of the detailed set of variables used as controls is provided in the notes to Tables 3 and 4.
[11] The National Literacy Taskforce (which developed the National Literacy Strategy) advised that by 2002, 80% of 11 year olds should reach Level 4 in Key Stage 2 English. In fact, as the numbers in Table 1 show, this target has not been met.

Sheffield, Newcastle and Bristol. Within England these inner city areas are characterised by problems of economic and social disadvantage. Collectively, about 40 per cent of primary schools within these LEAs were involved in the NLP. The remaining NLP schools were in eight LEAs in three counties (Hampshire, Essex and Norfolk).[12] Only about 7 per cent of primary schools in these counties were involved in the NLP.

Local Education Authorities (LEAs) were involved in the administration of the NLP within the treatment schools. The Literacy Task Force (1997b) identified LEAs as the 'first link in the chain' of implementation, being expected to provide a strong lead to all their schools, profiling literacy as a priority for the LEA with publicity, producing visible targets and so on.

This matters a lot for the definition of control areas. It seems possible that all primary schools within an LEA taking part in the NLP may have (at least indirectly) benefited from the initiative – and not just those schools taking part in the project. For example, the NLP policy was well advertised within LEAs. Therefore, it is highly questionable whether 'non-NLP' schools within NLP LEAs should be used as controls for the treatment schools.

However, there is also a good case for not using all other primary schools in the country as a control group: there are over 14,000 in England, and these are likely to be heterogeneous in ways we cannot account for with available data.[13] Thus, we adopt the following strategy. We first identify LEAs that are in areas geographically adjacent to

[12] The precise LEAs are: Hampshire, Portsmouth, Southampton, Essex, Southend-on-Sea, Thurrock, Isle of Wight, Norfolk.

[13] However, we do have very detailed school-level controls. Thus, as a robustness check, we later estimate regressions using the full sample of schools.

LEAs involved in the NLP. Then, if multiple non-NLP areas are present, we choose those with the closest educational performance indicator in the pre-policy period.

This approach has similarities to the approach adopted in evaluations of recent UK area-level initiatives (such as the Educational Maintenance Allowance in Dearden *et al.*, 2003, and the New Deal in Blundell *et al.*, 2002). However, it does place some constraints on our analysis and makes it clear that some LEAs in the NLP do not have a suitable control group.[14] Therefore we have to omit some LEAs, mainly the counties, where it proves difficult to identify a good comparison group. The only NLP LEA in a city that is omitted from our analysis is Bristol, where all adjacent areas differ massively in their pupil attainment and population densities (as the city area of Bristol is surrounded by semi-rural areas). Because of this, our focus is on the inner city London LEAs in the NLP and on the urban LEAs of Sandwell, Liverpool, Manchester, Sheffield and Newcastle. These comprise 72 percent of the NLP schools. We would argue that these are precisely the schools of interest for which to evaluate the literacy hour, particularly in the context of debate about poorly performing inner city schools in the UK.

A list of NLP LEAs and their 'matched areas' are provided in Appendix Table A1. Of course, one should be clear that this spatial matching might prove less than perfect. We need to standardise our treatment and control schools for different observable characteristics to ensure that the evaluation does compare like with like. This is the reason (described more formally below) why we estimate models that include detailed controls and school fixed effects. In our empirical work, we do so both within a difference-in-difference framework and using statistical matching methods.

[14] Again, this has some similarities with the Educational Maintenance Allowance evaluation where there was one rural treatment area (Cornwall). Being unable to define a suitable control area, the researchers dropped this area from their analysis.

Descriptive Statistics

Descriptive statistics are provided in Table 2. The Table shows measures of primary school attainment before and after the introduction of the NLP for schools in LEAs affected by the policy and for schools in the control LEAs. The Table shows two primary school attainment measures - the mean percentile reading score and the percentage reaching level 4 or above in KS2 English. It is clear that average levels are lower in NLP schools at each point in time. Hence, we need to standardise for baseline differences in the pre-policy period (so as to ensure we are comparing like with like) and we are careful to investigate this in some detail in our empirical modelling.

However, an interesting pattern emerges in terms of changes before and after NLP introduction. For reading scores, there is evidence of improvement in the NLP schools relative to the control schools. The mean reading score goes up 2.1 percentile points in the NLP schools and falls by 1.1 in the control group. The same relative pattern of improvement is seen for KS2 English, where the percentage of pupils attaining level 4 or above rises by more in NLP schools (by 12.2 percentage points as compared to 8.8 in the control group).

4. The Impact of NLP on Pupil Attainment

In this section, we consider tests of the NLP policy effect, looking at difference-in-difference models that control for a large number of factors. This, of course, is important since there are different levels of initial attainment in the NLP and control schools and we need to condition out the reasons for such differences.

The basic difference-in-difference estimates are based upon the following model for pupil i in school s in year t:

$$P_{ist} = \alpha + \beta NLP_s * Policy\ On_t + \gamma NLP_s + \delta X_{ist} + \lambda Z_{st} + \pi T_t + \varepsilon_{ist} \quad (1)$$

where P is pupil attainment, NLP is a dummy equal to one for NLP schools, X denotes a set of pupil characteristics, Z a set of school characteristics, T a set of year dummies and ε is an error term. The Policy On_t variable is a dummy variable equal to one for the time periods when the NLP policy was in place (and 0 in pre-policy periods) so the coefficient β is the difference-in-difference estimate of the NLP policy. The γ coefficient is the initial (pre-policy) baseline difference in attainment between treatment and control schools.

The structure of our data is such that we observe pupils in schools, so we are able to control for unobserved heterogeneity across schools by estimating a more stringent model incorporating school fixed effects. This allows each school to have its own intercept and adds an *s* subscript to the constant term as follows:

$$P_{ist} = \alpha_s + \beta NLP_s * Policy\ On_t + \delta X_{ist} + \lambda Z_{st} + \pi T_t + \varepsilon_{ist} \quad (2)$$

In (2) that the linear NLP term drops out as it is subsumed into the fixed effects. We still identify the policy impact through the interaction term NLP_s*Policy On_t as the estimate of β measures within-school changes in attainment before and after NLP introduction in treatment schools relative to within-school changes in attainment in control schools.

Primary School Attainment (Key Stage 2)

Estimates of equations (1) and (2) are reported in Table 3 for primary school reading and English attainment. The first three columns report difference-in-difference specifications, from the column (1) model with no controls, to column (2) which presents estimates of equation (1) and column (3) which presents estimates of equation (2). The final two columns present analogues to columns (2) and (3) where we have selected a sample of matched schools using propensity score matching techniques on the 1996, pre-policy data. The propensity score distributions and probit models used to generate them are shown in the Appendix. The basic method used is that of Heckman, Ichimura and Todd, 1997, where propensity scores are estimated and the sample then trimmed to exclude poorly matched schools.[15]

The results are supportive of the hypothesis that the literacy hour, via the NLP policy, improved reading and English performance. The first column specification shows a highly significant 2.1 percentile point increase in reading scores and a .027 increase in the proportion of children achieving level 4 or above in KS2 English. However, it also shows what was clear in the descriptive analysis in Table 2, namely that the initial performance of NLP schools was significantly worse, at -8.8 percentile points lower for reading and a –.124 lower probability of achieving level 4 in overall English attainment.

However, when we include detailed controls in column (2), it is reassuring to see that these initial baseline differences become small and are driven to statistical insignificance. This demonstrates that we are comparing like with like in the initial pre-policy period and thus a treatment-control type evaluation of the manner we have adopted is appropriate. Moreover, the magnitude of the policy impact rises slightly for the more detailed specifications, going to a 2.4 percentile point improvement in reading and a .032 higher proportion achieving level 4 or above in English. Both of these difference-in-difference policy impacts are strongly significant in statistical terms.

[15] See Rosenbaum and Rubin (1983, 1984) for the initial statements on how to use propensity score matching as a means of reducing bias in observational studies designed to compare treatments and controls (= 0 or 1, in our case being distinguished by the NLP variable).

As described above, since our data consists of different cohorts of children in the same schools over time, not only can we control for observable characteristics of schools, but we can also net out unobservable time invariant aspects of schools by incorporating school fixed effects. We implement this strong test in column (3). The NLP effect remains strong and actually rises to 2.6 percentile points in reading and .032 for the proportion achieving level 4 or above in English. It is interesting that both the observable control variables and the control for unobservables via school fixed effects actually result in a rise in the estimates, suggesting that what we do to try and net out differences between pupils and schools results in larger effects, implying if anything a downward bias from the simple difference-in-difference estimates without controls.

The final two columns of the Table use matching methods in an attempt to further standardise, in a less parametric way, the set of treatment and control schools. Doing so tends to reduce the measured impact of the NLP policy to some extent, though the impact remains significant and sizable. In the final model, using the matched sample with school fixed effects, the NLP schools have reading scores of 1.8 percentile points higher and a .018 higher proportion achieving level 4 or above in English. Given that the difference-in-difference estimates already eliminate baseline differences in attainment, we believe these matching estimates to be conservative estimates, yet nonetheless they show there to be a marked improvement in NLP schools.

Robustness

The findings of a positive and significant literacy hour effect on children affected before and after the introduction of NLP, couple with the increase in this effect on inclusion of observable controls and school fixed effects, makes us reasonably confident that we have pinned down a literacy hour effect. The latter increase in the size of the effect in the conditional models, in particular, makes us think that other pre/post changes that were unique to the NLP schools are not driving the results. However, a problem would arise if earlier trends in attainment were different in NLP and comparison schools. An important robustness check would therefore be to test for pre-treatment differences in trends in pre-treatment periods. If a difference-in-difference estimate of the literacy hour introduction did not exist before the policy was introduced this gives additional confidence that it is not pre-existing differences in NLP and comparison schools that is driving the results.

Data availability does restrict investigation of this possibility in that pupil-level Key Stage 2 data is not available prior to 1996. However, we can get two years of pre-policy school-level data, for 1995 and 1996, to look at pre-treatment differences. We can only get, from the School Performance Tables, data on Key Stage 2 English and so compare pre-treatment changes across NLP and comparison schools for this measure of attainment.

The comparison of results with the pre-treatment period exercise is thus carried out in Table 4. Panel A of the Table reproduces, for ease of comparison of results, the estimates from Panel B of Table 3. Panel B is then the school-level analogue to Panel A. Reassuringly the school-level results from the Performance Tables are almost identical to the pupil-level results. Panel C reports results from the pre-treatment period looking at difference-in-difference estimates between 1995 and 1996. The results are strong and highly reassuring for our identification strategy as all three specifications show there to be no difference whatsoever in the pre-treatment trends between NLP and comparison schools.[16] This is an important finding and shows that the period of the policy intervention saw the English skills of pupils in NLP schools improve significantly when this clearly was not happening before.

One might also express a concern that results are sensitive to the choice of Local Education Authorities we use to define our sample of control schools. As a robustness check, we have also estimated regressions where the control group consists of all other LEA maintained schools in England. Results are reported in Table 5, in an analogous structure to Table 3. In these specifications, so as to look at the full population of English primary schools, we also include the smaller number of NLP schools in counties, but estimate separate policy effects for these schools and schools in cities. This is to ensure comparability with the previous analysis. It is also because there may be a genuinely heterogeneous impact of the policy across cities and counties, since participation was much higher in the cities (as discussed above) and thus much higher profile (and more actively promoted) within these areas. The results support this hypothesis. More importantly, they show a strong similarity with results reported in Table 3 with regard to the effect of the NLP in cities. In the unmatched sample, the NLP effect is estimated as 2.5 percentiles in reading and .036 for the proportion achieving level 4 or above in English, when using the full set of controls and school fixed effects (column 3). When we use a statistical matching approach in an attempt to make the treatment and control schools more comparable (based on observable, pre-policy characteristics) estimates reduce only slightly and are very much in the same ballpark as the earlier analysis.

[16] Notice that there are 20 less schools (821 rather than 841) in the lower Panel. This is due to missing data in 1995/96. However, restricting the upper panels to the same 821 schools barely changed the results (NLP*Policy On estimates comparable to columns (1) to (3) for the 821 schools sample were .026 (.011), .031 (.009) and .032 (.009) for Panel A and .021 (.011), .027 (.009) and .026 (.012) for Panel B).

Hence, these results strongly corroborate the spatial matching estimates and suggest that the literacy hour significantly raised pupil performance in the primary schools that were exposed to the policy. This finding is of considerable significance to the economics of education literature on what can raise performance as it provides clear evidence that public policy aimed at improving pupil performance through changing the content and structure of what is taught can be effective.

5. Gender Gaps in Attainment

As explained in the introduction, from a theoretical perspective one may be interested in the impact of the literacy hour on gender differences in attainment. Given the existence of sizable gaps in attainment between boys and girls, this also matters from a policy perspective. Thus, the results reported in this section break down the NLP effect, estimating separate effects by gender.

Table 6 shows separate NLP effects for boys and girls. The estimates reported are those from the full model, which incorporates school fixed effects. As for the earlier analysis, basic difference-in-difference and matching estimates are reported. The Table reveals evidence of gender differences in the NLP policy impact at primary school. The NLP effect for boys is numerically much larger than that for girls. For reading, the literacy hour raised boys' mean percentile reading scores by somewhere between 2.5 and 3.4 percentile points and the proportion achieving level 4 or above in KS2 English by between .027 and .042. These are large effects. Thus, it appears that the literacy hour was more effective for boys and as such reduced the gender gap in attainment at primary school.

It is interesting to place this finding in the context of the national roll-out during the school year 1998/99 and with the national figures given in Table 1. It is evident that the gender gap in primary school reading and English attainment has reduced in recent years. The results we report here are entirely consistent with the literacy hour playing an important role.

6. Measuring Economic Costs and Benefits

This analysis has shown a significant impact of the literacy hour on reading and on English attainment. The question remains as to whether the policy was cost effective. Hence we now compare the per pupil costs of the policy with the economic benefits, as reflected in predicted labour market earnings.

The planned cost of the NLP was £12.5 million over 5 years. The main costs were 14 local centres (each costing about £25,000 per year) and literacy consultants in each participating Local Education Authority (about £27,000 per year for each consultant). Schools also received some funding for teacher training and resources, which was broadly the same for each school (though some account was taken of the pupil-teacher ratio). However, since the national roll out took place two years after the NLP was introduced, only the first two years are relevant. The total cost per annum was thus £2.5 million (or about £2.8 million in 2001 prices). We observe the number of students affected from pupil numbers in the schools within Cohorts 1 and 2 in 1997 and 1998 (i.e. 222,261 pupils in aggregate).[17] Hence the cost per pupil is £25.52 per annum.

To estimate benefits of the policy, we first convert the impact of the policy on reading scores (i.e. 2.63 percentiles, as shown in Table 3, column 3) to an equivalent

[17] This includes infant schools.

estimate in terms of standard deviations. This is calculated as 0.091 standard deviations. Secondly, we estimate the impact of reading scores on future labour market earnings using the British Cohort Study. This is a panel survey of all those living in Great Britain who were born between 5th and 11th April 1970. We regress the log of labour market earnings (at age 30, in 2000) on age 10 percentile reading scores (from 1980).[18] Results are shown in Table 7. We show three specifications. In column (1), controls are included for gender and region; in (2) we add controls for family background; and in (3) we include dummy variables for the participant's highest educational qualification achieved by age 30. Since the latter variable is likely to partly capture the effect of the reading score, the effect of reading on labour market earnings in column (3) should be considered as a lower bound estimate (or even an under-estimate).

The estimates are .54, .42 and .20 in each of the respective columns and are always statistically significant, showing a higher standard of reading at age 10 to be associated with higher earnings at age 30. The earnings impact of a .091 increase in the standard deviation (sd) of reading percentiles is then calculated as .091*sd*[exp(reading score coefficient/100)-1]. This amounts to an annual sum of £196.32, £154.23 and £75.40 for each specification. Assuming that labour market participation occurs between the age of 20 and 65, and using a discount rate of 3 per cent, the corresponding present discounted value of the cumulative effect of the literacy hour is estimated as £5,476, £4302 and £2103.

There are, as with all calculations like this, certain issues that may make one question these economic benefits. One clear example in this case is that the beneficiaries

[18] The reading test is a shortened version of the *Edinburgh Reading Test*, which is a test of word recognition. It examines vocabulary, syntax, sequencing, comprehension and retention (see Godfrey Thompson Unit, University of Edinburgh, 1978, or Plake and Impara, 2001, for more details).

of the NLP literacy hour tended to be children from less well performing schools. On average these children are likely to be located further down the reading score distribution, yet the earnings effect from age 10 reading is assumed linear in the regressions in Table 7. We have therefore considered economic benefits from parametric models where we allow separate effects on earnings for the top and bottom half of the reading score distribution and from non-parametric regressions (using the Nadaraya-Watson estimator). In both cases the linearity assumption seems reasonable: the bottom half effect for the column (2) specification in Table 7 is estimated to be .458 (standard error = .100) as compared to the top half effect of .435 (.084); a non-parametric regression is shown in Appendix Figure A3.

Whichever way one looks at it, the benefits of the literacy hour seem to be sizeable and the costs are much smaller. Even if we take the smallest impact estimate from our analysis (the 1.72 percentile improvement in column (4) of Table 3, which corresponds to a .06 standard deviation increase), the economic benefits are measured in the range of £1375 to £3581.

The benefits of the policy (a .06 to .09 standard deviation in reading scores) seem comparable to more expensive programs like reducing class size. For example, Krueger and Whitmore (2001) found a class size reduction in the STAR program to lead to an increase in test scores of .13 standard deviations. Rivkin, Hanushek and Kain (2002) suggest that having a teacher at the higher end of the quality distribution raises student achievement by at least .11 standard deviations. However, the costs of such programs are more substantive. They are simply not comparable to this apparent low-cost literacy hour program of changing teacher practice. Of course, the financial costs of the program may not reflect its true resource cost. For example, it might be argued that the measured costs do not reflect any extra effort the teacher might have to put in to learning and implementing the new teaching method. On the other hand, such effort may be fully accounted for in the cost of training. Furthermore, there are reports of a very positive response by teachers (e.g. Fisher and Lewis, 1999, or Smith and Whitely, 2000), who find the learning objectives and structure of the literacy hour to provide a clear focus for what they teach.

One might also argue that the literacy hour takes teaching effort and resources away from other subjects and that this indirect cost effect (via substitution) should be taken account of in a cost-benefit calculation. However, it seems likely that literacy was being taught in some form before the policy, for a commensurate time period. Therefore, the literacy hour represents a change in how reading and writing are taught, rather than merely an increase in the time devoted to the subject. One might also suspect that the literacy hour could lead to positive spillovers due to complementarities between pupil subject areas and associated teacher practice. Firstly, since the ability to read and write are important generic skills, an improvement in how these skills are taught might lead to improved performance in other subjects. Secondly, the literacy hour might have caused teachers to re-evaluate their teaching methods in other subjects. It may be that aspects of the approach are transferable to the teaching of other subjects (e.g. the highly structured nature of lessons; the balance between group work and whole class activity). This is important in English primary schools because generally pupils within a particular year group are taught every subject by the same teacher.

In Table 8 we consider the possible substitution/complementarity with other subjects and report a test of whether the literacy hour helped or hindered attainment in Mathematics. To do this, we simply re-estimate regressions reported in Table 3, where the dependent variables are now respectively, the percentile Mathematics score and whether the student obtains level 4 or above in Mathematics. The first three columns suggest a positive impact of the literacy hour. Interestingly, the magnitudes of the estimates are smaller, at around three-fifths of the impact on English (in Table 3). Including all controls and school fixed effects (column 3), NLP schools show higher scores of 1.53 percentiles and a .025 higher proportion achieving at least level 4. Also, when estimating these regressions using the matched sample of schools (column 5), these coefficients reduce to .838 and .014 respectively (about half of the effects for English seen in Table 3) and become statistically insignificant.

These results suggest, if anything, a complementary impact of the literacy hour on English and Mathematics attainment. Furthermore, and in line with the English results above there is no evidence of any pre-treatment differences in Mathematics outcomes between NLP and non-NLP schools. Estimating a difference-in-difference model for 1995 to 1996 (analogous to the English models in Table 4) produced insignificant estimates, ranging from an effect of -.008 (with associated standard error 0.015) from a specification with no controls through to -.009 (.022) for the full school effects specification for the 821 schools considered in Table 4. As such, looking at the impact of the literacy hour on attainment in reading and English appears more likely to underestimate rather than overestimate the benefits associated with the literacy hour.

Hence, the literacy hour seems to be cost effective. It represented a change in the content and organization of how literacy was taught – this enhanced pupil performance, but was not a change that involved much diversion of resources. However, to implement such practices requires knowledge of what works in how to teach literacy. As discussed above, ideas used to construct the literacy hour were based on experience in other countries and on research. The value of this information is not included in the costs, yet it is manifestly important in generating the benefits.

7. Conclusions

In this paper, we have considered the impact of the National Literacy Project (NLP), the introduction of the literacy hour that was undertaken in around 400 English primary schools in 1997 and 1998. We compare reading and overall English performance of age 11 children in the NLP schools as compared to pupils in a set of control schools. We adopt an explicit treatment-control group approach, investigating what happened to pupil attainment in schools exposed to the literacy hour before and after the policy was introduced relative to pupils in schools that were not subject to the policy.

We find that reading and English Key Stage 2 attainment levels rose by more in NLP schools between 1996 and 1998. By means of implementing our identification strategy to a number of robustness checks we are confident that this constitutes an NLP effect. This is because the estimated policy impact, in fact, rises when we control for other factors that may have changed at the same time as policy introduction, and rises on inclusion of controls for unobserved heterogeneity through school fixed effects, and because there was no trend difference in attainment in NLP relative to comparison schools in the pre-policy period. Since significant gender gaps in English performance exist (in favour of girls), we consider whether the literacy hour had a differential impact by gender. We find some evidence that, at age 11, boys benefited more than girls. Finally, we show the benefits of the literacy hour to easily exceed the costs of the policy.

These findings are of considerable significance when placed into the wider education debate about what works best in schools for improving pupil performance. As suggested in the introduction, the research remit of economists has tended to be rather narrower than that of educationalists in this area. Our approach shows that one of the areas receiving much less attention from economists in the past, namely the content and organisation of what is taught, matters for attainment in English and reading. Indeed, as the effects we identify come from a government policy aimed at improving literacy - the literacy hour - the evidence we report suggests that public policy aimed at changing the content of teaching can significantly raise pupil attainment and can do so in a cost effective manner.

References

Angrist, J. and V. Lavy (1999) Using Maimonides' rule to estimate the effect of class size on scholastic achievement, Quarterly Journal of Economics, 114, 533-75.

Angrist, J., and V. Lavy (2001) New evidence on classroom computers and pupil learning, Economic Journal, 112, 735-765.

Beard, R. (2000) Research and the National Literacy Strategy, Oxford Review of Education, 26, (Nos. 3 & 4), 421-36.

Blundell, R., M. Costa Dias, C. Meghir, and J. Van Reenen (2002), Evaluating the employment impact of a mandatory job search program, The Institute for Fiscal Studies, Working Paper 01/20.

Borman, G. and G. Hewes (2002) The long-term effects and cost-effectiveness of Success for All, Educational Evaluation and Policy Analysis, *24*, 243-266.

Card, D. and A. Krueger (1992) Does school quality matter? Returns to education and the characteristics of public schools in the United States, Journal of Political Economy, 100, 1-40.

Creemers, B. (1994) The Effective Classroom, London, Cassell.

Dearden, L., C. Emmerson, C. Frayne, and C. Meghir (2003) The impact of financial incentives on educational choice, mimeo, Institute for Fiscal Studies.

Department for Education and Employment (1999) A Fresh Start: Improving Literacy and Numeracy, DfEE: London.

Fisher, R. and M. Lewis (1999) Anticipation or Trepidation? Teachers' Views on the Literacy Hour, Reading, 33, 23-28.

Glewwe, P., M. Kremer, S. Moulin, E. Zitzewitz (2000) Retrospective vs. prospective analyses of school inputs: the case of flip charts in Kenya, National Bureau of Economic Research. Working Paper No. 8018.

Godfrey Thompson Unit, University of Edinburgh (1978) Edinburgh Reading Test. Sevenoaks: Hodder and Stoughton.

Hanushek, E. (1997) Assessing the effects of school resources on student performance: an update, Educational Evaluation and Policy Analysis, 19, 141-64

Hanushek, E. (2003) The failure of input-based schooling policies, The Economic Journal, 103, F64-F98.

Heckman, J., H. Ichimura and P. Todd (1997) Matching as an Econometric Evaluation Estimator, Review of Economic Studies, 65, 261-294.

Herman, R., D. Aladjem, P. McMahon, E. Masem, I. Mulligan, A. O'Malley, S. Quinones, A. Reeve and D. Woodruff (1999) An educators' guide to schoolwide reform, Arlington, VA: Educational Research Service.

Hopkins, D. (1999) Success for All – Part One, Literacy Today, 21.

Jacob, B. (2002) Where the Boys Aren't: Non-cognitive Skills, Returns to Schooling and the Gender Gap in Higher Education, National Bureau of Economic Research Working Paper 8964.

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, 34-63.

Literacy Task Force (1997a) A Reading Revolution: How We Can Teach Every Child to Read Well, London: The Literacy Task Force c/o University of London: Institute of Education.

Literary Task Force (1997b) The Implementation of the National Literacy Strategy, London: Department for Education and Employment.

Machin, S. and S. McNally (2003) Gender and Educational Attainment, Centre for the Economics of Education, unfinished draft.

Maynard, T. (2002) Boys and Literacy: Exploring the Issues, RoutledgeFalmer: London.

Millard, E. (1996) Differentially Literate: Boys, Girls and the Schooling of Literacy, Falmer Press: London.

Office for Standards in Education (1996) 'The Teaching of Reading in 45 Inner London Primary Schools', A report by Her Majesty's Inspectors in collaboration with the LEAs of Islington, Southwark and Tower Hamlets, London: Ofsted.

Office for Standards in Education (1998), The National Literacy Project: An HMI Evaluation, London: Ofsted.

Office for Standards in Education (1999) 'Primary Education 1994-1998: A Review of Primary Schools in England', London: Ofsted.

Plake, B. and J. Impara (Eds.). (2001) The Fourteenth Mental Measurements Yearbook, Lincoln, NE: Buros Institute of Mental Measurements.

Rivkin, S., E. Hanushek and J. Kain (2002) Teachers, schools and academic achievement, revised version of Working Paper No. 6691, National Bureau of Economic Research, available at http://edpro.stanford.edu/eah/eah.htm.

Rosenbaum, P. and D. Rubin (1983) The Central Role of the Propensity Score in Observational Studies for Causal Effects, Biometrika, 70, 41-55.

Rosenbaum, P. and D. Rubin (1984) Reducing Bias in Observational Studies Using Subclassification on the Propensity Score, Journal of the American Statistical Association, 79, 516-524.

Sammons, P. (1999) School Effectiveness: Coming of Age in the Twenty First Century, Lisses, The Netherlands: Swets and Zeitlinger.

Sainsbury, M. (1998), Evaluation of the National Literacy Strategy: Summary Report, Slough: NFER.

Scheerens, J. (2000) Improving School Effectiveness, Paris: IIEP.

Stainthorp, R. (1996) Teaching reading in the primary classroom, in P. Croll and N. Hastings (eds.) Effective Primary Teaching: Research Based Classroom Strategies, London: David Fulton.

Slavin, R. and N. Madden (2000) Research on achievement outcomes of success for all: a summary and response to critics, Phi Delta Kappan, 82, 59-66.

Slavin, R. and N. Madden (2003) Success for All/Roots and Wings: 2003 Summary of Research on Achievement Outcomes, Baltimore: Johns Hopkins University, Center for Research on the Education of Students Placed at Risk.

Smith C. and H. Whitely (2000) Developing Literacy through the Literacy Hour: A Survey of Teachers' Experiences, Reading, 34, 34 -38.

Teddlie, C., and D. Reynolds (2000) The International Handbook of School Effectiveness Research, Falmer Press: London.

West, A. and H. Pennell (2003) Underachievement in Schools, RoutledgeFalmer: London and New York.

White, J. (1996) Research on English and the Teaching of Girls, in P. Murphy and C. Gipps (eds.) Equity in the Classroom: Towards Effective Pedagogy for Girls and Boys, Falmer Press: London.

Table 1: Primary School English and Reading Attainment Over Time

Percentage of Pupils Achieving Level 4 and Above in Key Stage 2 English								
	1996	1997	1998	1999	2000	2001	2002	Change 1996-2002
All	57	63	65	71	75	75	75	18
Boys	50	57	57	65	70	70	70	20
Girls	65	70	73	76	79	80	79	14
Percentage of Pupils Achieving Level 4 and Above in Key Stage 2 Reading								
	1996	1997	1998	1999	2000	2001	2002	Change 1997-2002
All	n/a	67	71	78	83	82	80	13
Boys	n/a	63	64	75	80	78	77	14
Girls	n/a	71	79	82	86	85	83	12

Notes: Data comes from DfES national statistics. There was no official level for reading in 1996.

Table 2: Descriptive Statistics

	NLP Schools				Schools in Control LEAs			
	1996	1997	1998	Change 1996-98	1996	1997	1998	Change 1996-98
Percentile Reading Score	44.4	45.3	46.5	2.1 (.8)	53.9	53.2	52.8	-1.1 (.5)
Percentage Reaching Level 4 or above in KS2 English	37.3	47.6	49.5	12.2 (.9)	50.5	57.8	59.3	8.8 (.6)
Number of Pupils	12645	12586	12814		22461	21976	22172	
Number of Schools	283	284	283		553	551	547	

Notes: NLP school and control group school areas are defined in Appendix Table A1. Standard errors in parentheses.

Table 3: NLP and Primary School Reading and English Attainment

	(1)	(2)	(3)	(4)	(5)
	No controls	Full set of controls	Full Model (with school fixed effects)	Full Set of Controls, Matching	Full Model (with school fixed effects), Matching
A. Percentile Reading Scores					
NLP*Policy On	2.143 (.741)	2.423 (.647)	2.631 (.650)	1.715 (.631)	1.791 (.628)
NLP	-8.808 (.845)	-1.075 (.883)	--	-.962 (.892)	--
Year=1997	-.352 (.392)	-.532 (.377)	-.486 (.387)	-.533 (.391)	-.404 (.408)
Year=1998	-.732 (.458)	-1.120 (.433)	-.881 (.473)	-.901 (.448)	-.535 (.496)
Number of Pupils	104654	104654	104654	96083	96083
Number of Schools	841	841	841	761	761
B. Probability of Achieving Level 4 or Above Key Stage 2 English					
NLP*Policy On	.027 (.011)	.032 (.009)	.032 (.009)	.020 (.009)	.018 (.009)
NLP	-.124 (.012)	-.008 (.013)	--	-.006 (.013)	--
Year=1997	.079 (.006)	.077 (.006)	.077 (.006)	.079 (0.006)	.080 (0.006)
Year=1998	.090 (.007)	.086 (.007)	.089 (.007)	.091 (.007)	.094 (0.008)
Number of Pupils	104654	104654	104654	96083	96083
Number of Schools	841	841	841	761	761

Notes: standard errors (clustered on school) in parentheses; independent variables in columns (2), (3), (4) and (5) are as follows (all are at school-level apart from gender of student): average percentile reading score 1996; average percentile writing score 1996; % achieving level 4 or above in English, Mathematics and Science (at KS2) respectively; % missing due to absence or disapplication in English, Mathematics and Science (at KS2) respectively; % entering extension test in English and Mathematics respectively at KS2 in 1996. % eligible for Free School Meals; % non-white students; % students with Special Educational Needs, with statement and without statement; pupil-teacher ratio; number of pupils; whether all girls school; all boys school; religious school; % teachers who are not fully qualified; ratio of support staff to teachers; % teachers who are graduates and with particular class of degree; % female teachers; missing variable indicators; dummy for whether school is in London; % achieving level 4 or above in English within the LEA in 1996.

Table 4: Pre-Treatment Difference-in-Difference Estimates (1995 to 1996)

	Probability of Achieving Level 4 or Above Key Stage 2 English		
	(1)	(2)	(3)
	No controls	Full set of controls	Full Model (with school fixed effects)
A. Estimates from Panel B of Table 3			
NLP*Policy On	.027 (.011)	.032 (.009)	.032 (.009)
NLP	-.124 (.012)	-.008 (.013)	--
Year=1997	.079 (.006)	.077 (.006)	.077 (.006)
Year=1998	.090 (.007)	.086 (.007)	.089 (.007)
Number of Pupils	104654	104654	104654
Number of Schools	841	841	841
B. School-Level Estimates Comparable to Panel A			
NLP*Policy On	.023 (.011)	.027 (.009)	.026 (.011)
NLP	-.118 (.012)	-.007 (.012)	--
Year=1997	.077 (.006)	.073 (.006)	.074 (.008)
Year=1998	.092 (.007)	.086 (.007)	.088 (.009)
Number of Schools	841	841	841
C. Pre-Treatment School-Level Estimates, 1995 and 1996			
NLP*Year=1996	.002 (.013)	.002 (.013)	.003 (.019)
NLP	-.122 (.015)	-.017 (.009)	--
Year=1996	.117 (.008)	.115(.009)	.109 (.014)
Number of Schools	821	821	821

Notes: as for Table 3 (in upper panel); in Panels B and C the dependent variable is from the School Performance Tables (this is very highly correlated with aggregated to school-level data from Table 3 Panel B); in Panels B and C the estimates are weighted using pupil numbers to maintain comparability with Panel A.

Table 5:
NLP and Primary School Reading and English Attainment
(Control Sample of all Other LEA Maintained English Primary Schools)

	(1)	(2)	(3)	(4)	(5)
	No controls	Full set of controls	Full Model (with school fixed effects)	Full Set of Controls, Matching	Full Model (with school fixed effects), Matching
A. Percentile Reading Scores					
NLP*Policy On (Cities)	1.876 (.620)	2.118 (.586)	2.460 (.569)	1.982 (.586)	2.288 (.572)
NLP*Policy On (Counties)	-.753 (1.019)	-1.022 (.860)	-.980 (.915)	-1.270 (.869)	-1.221 (.927)
NLP (Cities)	-12.803 (.651)	-1.658 (.424)	--	-1.669 (.422)	--
NLP (Counties)	-8.239 (.946)	-.395 (.506)	--	-.222 (.505)	--
Year=1997	-.019 (.110)	-.220 (.097)	-.319 (.095)	-.038 (.101)	-.120 (.099)
Year=1998	-.036 (.115)	-.356 (.104)	-.342 (.108)	-.162 (.108)	-.116 (.113)
Number of Pupils	1716855	1716855	1716855	1607596	1607596
Number of Schools	14670	14670	14670	12781	12781
B. Probability of Achieving Level 4 or Above in Key Stage 2 English					
NLP*Policy On (Cities)	.029 (.009)	.032 (.008)	.036 (.008)	.030 (.008)	.034 (.008)
NLP*Policy On (Counties)	-.004 (.014)	-.009 (.012)	-.013 (.013)	-.013 (.012)	-.016 (.013)
NLP (Cities)	-.185 (.009)	-.021 (.005)	--	-.022 (.005)	--
NLP (Cities)	-.137 (.015)	-.018 (.008)	--	-.015 (.008)	--
Year=1997	.072 (.002)	.070 (.001)	.068 (.001)	.073 (.001)	.071 (.001)
Year=1998	.087 (.002)	.084 (.001)	.083 (.002)	.087 (.001)	.086 (.002)
Number of Pupils	1716855	1716855	1716855	1607596	1607596
Number of Schools	14670	14670	14670	12781	12781

Notes: standard errors (clustered on school) in parentheses; independent variables in columns (2), (3), (4) and (5) are as follows (all at school-level apart from gender of student): average percentile reading score 1996; average percentile writing score 1996; % achieving level 4 or above in English, Mathematics and Science (at KS2) respectively; % missing due to absence or disapplication in English, Mathematics and Science (at KS2) respectively; % entering extension test in English and Mathematics respectively at KS2 in 1996. % eligible for Free School Meals; % non-white students; % students with Special Educational Needs, with statement and without statement; pupil-teacher ratio; number of pupils; whether all girls school; all boys school; religious school; % teachers who are not fully qualified; ratio of support staff to teachers; % teachers who are graduates and with particular class of degree; % female teachers; missing variable indicators; dummy for whether school is in London; % achieving level 4 or above in English within the LEA in 1996.

Table 6: Gender Gaps

	(1) Full model (with school fixed effects)	(2) Full Model (with school fixed effects), Matching
Percentile Reading Scores		
Boys	3.412 (.741)	2.536 (.738)
Girls	1.781 (.752)	.987 (.733)
Probability of Achieving Level 4 or Above in Key Stage 2 English		
Boys	.042 (.012)	.027 (.012)
Girls	.021 (.011)	.008 (.011)

Notes: as for Table 3.

Table 7: Earnings Gains Associated With Age 10 Reading Skills, British Cohort Study

	(1) Basic Specification	(2) (1) Plus Family Background	(3) (2) Plus Highest Qualification
Reading Score at Age 10 Percentile (X 100)	.544 (.024)	.423 (.035)	.207 (.036)
Controls	Yes	Yes	Yes
Family Background	No	Yes	Yes
Highest Qualification	No	No	Yes
Sample Size	6587	3488	3488
Percent Earnings Impact of .091 Increase in Standard Deviation	1.4	1.1	.5

Notes: dependent variable is log(weekly earnings) in 2001 prices; standard errors in parentheses; controls included in all specifications for gender and region; family background variables are log(parental income at age 16), dummies for mother's and father's education; highest qualification are dummy variables for highest educational qualification achieved by age 30. The percent earnings impact of a .091 increase in the standard deviation (sd) of reading percentiles is calculated as .091*sd*[exp(reading score coefficient/100)-1].

Table 8: NLP and Primary School Mathematics

	(1)	(2)	(3)	(4)	(5)
	No controls	Full set of controls	Full Model (with school fixed effects)	Full Set of Controls, Matching	Full Model (with school fixed effects), Matching
A. Percentile Scores in Mathematics					
NLP*Policy On	1.275 (.789)	1.585 (.662)	1.530 (.633)	.935 (.684)	.838 (.660)
NLP	-7.793 (.891)	-.338 (.994)	--	-.481(1.022)	--
Year=1997	-.200 (.388)	-.237 (.378)	-.239 (.384)	-.222 (.396)	-.229 (.406)
Year=1998	-.420 (.453)	-.604 (.432)	-.359 (.456)	-.415 (.448)	-.176 (.480)
Number of. Pupils	104654	104654	104654	96083	96083
Number of Schools	841	841	841	761	761
B. Probability of Achieving Level 4 or Above in Key Stage 2 Mathematics					
NLP*Policy On	.024 (.012)	.028 (.010)	.025 (.010)	.018 (.011)	.014 (.010)
NLP	-.118 (.013)	-.012 (.014)	--	-.013 (.014)	--
Year=1997	.098 (.006)	.098 (.006)	.099 (.006)	.098 (.006)	.099 (.006)
Year=1998	.049 (.007)	.047 (.007)	.053 (.007)	.051(.007)	.057 (.008)
Number of Pupils	104654	104654	104654	96083	96083
Number of Schools	841	841	841	761	761

Notes: standard errors (clustered on school) in parentheses; independent variables in columns (2), (3), (4) and (5) are as follows: (all school-level apart from gender of student): average percentile reading score 1996; average percentile writing score 1996; % achieving level 4 or above in English, Mathematics and Science (at KS2) respectively; % missing due to absence or disapplication in English, Mathematics and Science (at KS2) respectively; % entering extension test in English and Mathematics respectively at KS2 in 1996. % eligible for Free School Meals; % non-white students; % students with Special Educational Needs, with statement and without statement; pupil-teacher ratio; number of pupils; whether all girls school; all boys school; religious school; % teachers who are not fully qualified; ratio of support staff to teachers; % teachers who are graduates and with particular class of degree; % female teachers; missing variable indicators; dummy for whether school is in London; % achieving level 4 or above in English within the LEA in 1996.

Appendix

Table A1:
NLP and 'Matched' Local Education Authorities for NLP Cities
(close geographically and with a similar level of educational attainment)

NLP LEAs	Control LEAs
Inner London: Hackney, Islington, Lambeth, Southwark, Tower Hamlets, Newham, Waltham Forest	Inner London: Camden, Haringey, Lewisham, Wandsworth,
Sandwell	Walsall
Liverpool	Knowsley
Manchester	Rochdale
Sheffield	Rotherham
Newcastle	South Tyneside

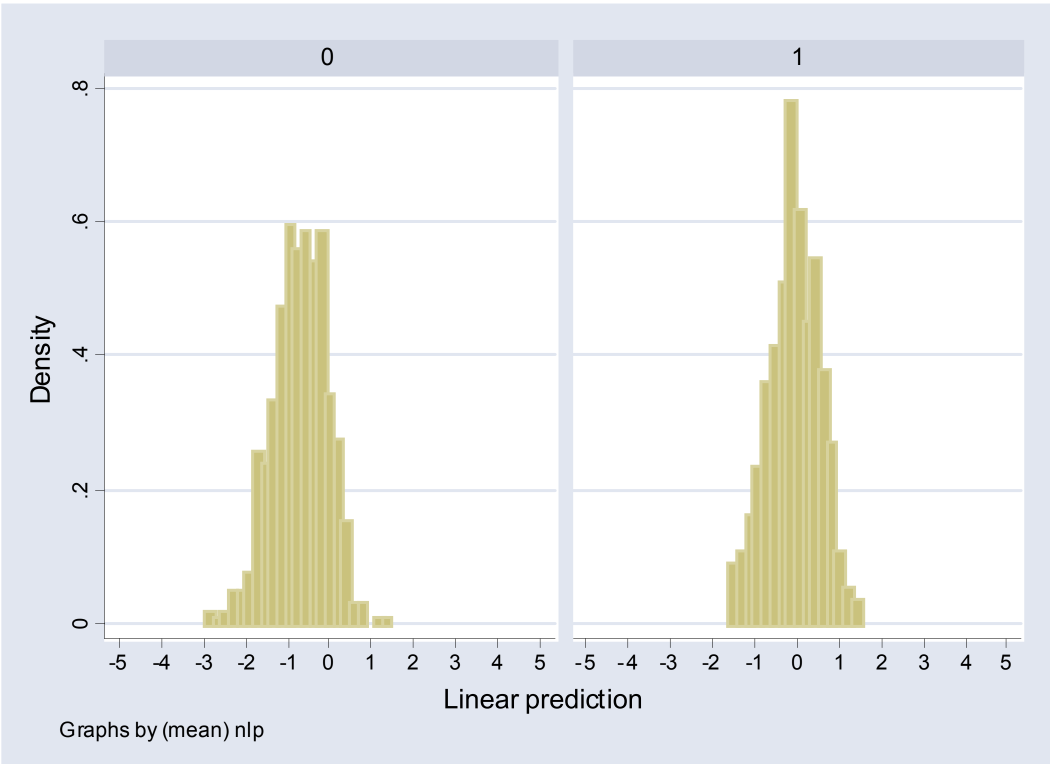
Table A2:
Probability of Treatment (NLP = 1),
Estimates for Table 3 Specifications

	Coefficient (Standard Error)
Percentile writing score: 1st quartile	.397 (.228)
Percentile writing score: 2nd quartile	.247 (.192)
Percentile writing score: 3rd quartile	.120 (.177)
Percentile reading score: 1st quartile	.118 (.299)
Percentile reading score: 2nd quartile	.078 (.230)
Percentile reading score: 3rd quartile	-.015 (.192)
English: Proportion attaining level 4 and above .	-.837 (.688)
English: Proportion with no level due to absence/disapplication	-.842 (2.011)
Mathematics: Proportion attaining level 4 and above .	-.169 (.550)
Mathematics: Proportion with no level due to absence/disapplication	2.312 (2.076)
Science: Proportion attaining level 4 and above .	.110 (.467)
Science: Proportion with no level due to absence/disapplication	-2.278 (1.982)
Proportion entering KS2 English Extension test in 1996	.862 (.686)
Proportion entering KS2 Mathematics Extension test in 1996	-2.100 (.964)
Proportion with missing results in 1996 (Performance Tables)	.205 (.464)
Religious school	.130 (.127)
% of pupils eligible for Free School Meals	.009 (.004)
% of pupils with Special Educational Needs, no statement	.017 (.006)
% of pupils with Special Educational Needs, with statement	.029 (.037)
Number pupils/100	.001 (.001)
Pupil-teacher ratio	.055 (.017)
Proportion of pupils: non-white	.654 (.269)
Proportion of teachers unqualified	2.269 (1.891)
Ratio support staff to teachers	-1.022 (.534)
% of teachers: graduates	.496 (.546)
% of teachers with - higher/1st/2nd degree	-.035 (.528)
Proportion female teachers	-1.334 (.439)
Observations	835

Notes:
Probit model; Coefficients and standard errors reported
All explanatory variables are 1996 values of school-level variables
Regression weighted by number of pupils in the school

This regression is used to predict the linear index of the propensity score, which is plotted in Figure A1 for treatment and control schools respectively. Schools within the 'common support' are then selected for the difference-in-difference analysis that is reported in Table 3 (columns 4 and 5).

Figure A1:
Propensity Scores for NLP and non-NLP Schools,
For Table 3 Specifications



Selected for 'matching' specification:
Schools with a predicted linear index of the propensity score between -1.6 and 0.8.

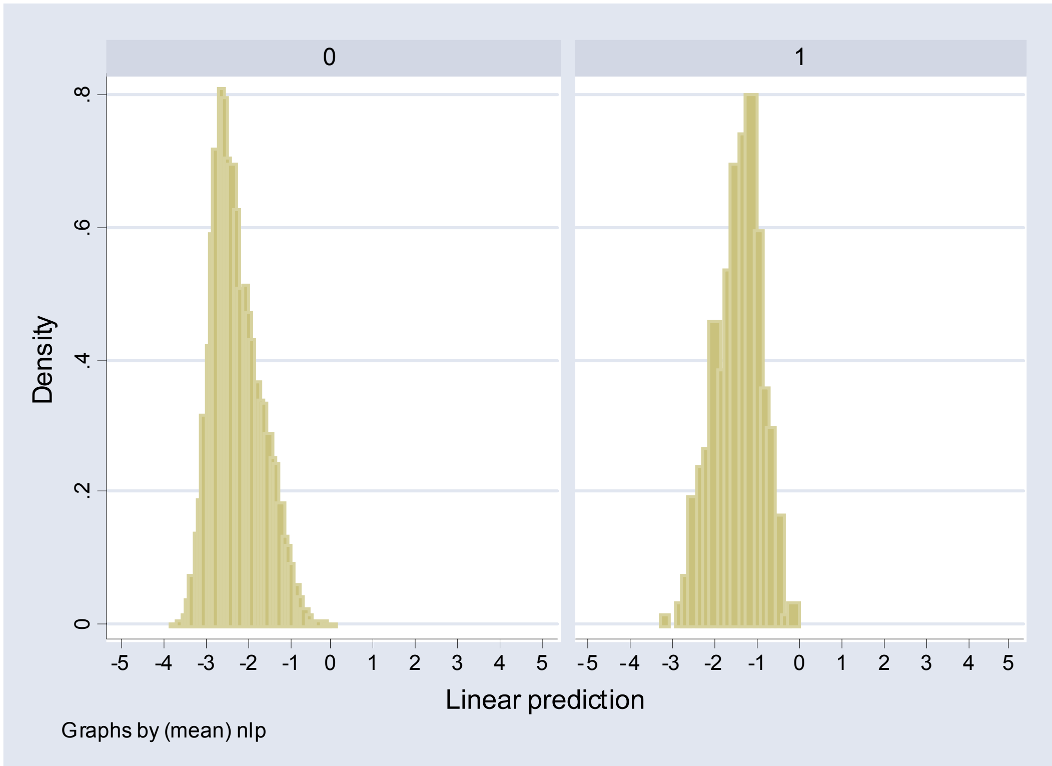
Table A3:
Probability of Treatment (NLP = 1) Estimates (All LEA Maintained Schools), For Table 5 Specifications

	Coefficient (Standard Error)
Percentile writing score: 1st quartile	-.032(.133)
Percentile writing score: 2nd quartile	-.115(.119)
Percentile writing score: 3rd quartile	-.163(.119)
Percentile reading score: 1st quartile	.433(.180)
Percentile reading score: 2nd quartile	.358(.150)
Percentile reading score: 3rd quartile	.136(.139)
English: Proportion attaining level 4 and above .	-.316(.362)
English: Proportion with no level due to absence/disapplication	-.740(1.054)
Mathematics: Proportion attaining level 4 and above .	-.159(.290)
Mathematics: Proportion with no level due to absence/disapplication	.669(1.000)
Science: Proportion attaining level 4 and above .	.101(.242)
Science: Proportion with no level due to absence/disapplication	-.650(.962)
Proportion entering KS2 English Extension test in 1996	.800(.364)
Proportion entering KS2 Mathematics Extension test in 1996	-.562(.491)
Proportion with missing results in 1996 (Performance Tables)	-.442(.221)
Proportion missing from Performance Tables in 1996	-.051(.350)
Religious school	-.003(.061)
% of pupils eligible for Free School Meals	.016(.002)
% of pupils with Special Educational Needs, no statement	.011(.002)
% of pupils with Special Educational Needs, with statement	.013(.016)
Number pupils/100	.000(.000)
Pupil-teacher ratio	.011(.008)
Proportion of pupils: non-white	.403(.136)
Proportion of teachers unqualified	1.612(.863)
Ratio support staff to teachers	-.753(.221)
% of teachers: graduates	.045(.248)
% of teachers with - higher/1st/2nd degree	.385(.247)
Proportion female teachers	.024(.228)
Proportion with missing information on 'non-white'	.326(.465)
Proportion missing from DTR	-.076(.263)
Observations	14456

Notes:
Probit model; Coefficients and standard errors reported
All explanatory variables are 1996 values of school-level variables
Regression weighted by number of pupils in the school

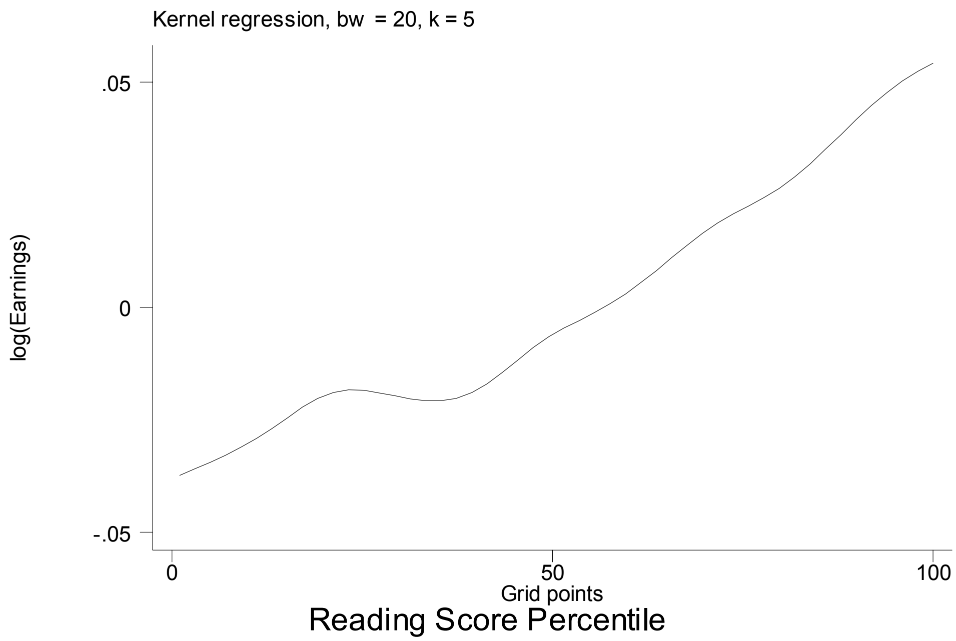
This regression is used to predict the linear index of the propensity score, which is plotted in Figure A2 for treatment and control schools respectively. Schools within the 'common support' are then selected for the difference-in-difference analysis that is reported in Table 5 (columns 4 and 5).

Figure A2:
Propensity Scores for NLP and non-NLP Schools,
for Table 5 Specifications



Selected for 'matching' specification:
Schools with a predicted linear index of the propensity score between -2.9 and -0.43.

Figure A3:
Non-Parametric Earnings Regression (Nadaraya-Watson Estimator)



Notes: This is the log(earnings)-reading score percentile relation from a specification comparable to column (2) of Table 8.