CEE DP 43

The Literacy Hour

Stephen Machin

Sandra McNally

December 2004

The Literacy Hour

Stephen Machin
Sandra McNally

Acknowledgements

Stephen Machin is Head of the Department of Economics at University College London, Director of Research at the Centre for Economic Performance, London School of Economics, and Director of the Centre for the Economics of Education. Sandra McNally is a Research Fellow at the Centre for Economic Performance, London School of Economics and Centre Coordinator of the Centre for the Economics of Education.

Executive Summary

Literacy matters. One in five adults in the UK is not functionally literate and this has serious implications for their well-being and economic circumstances, as well as for national productivity. To ensure that this problem does not beset future generations, attention must be given to how best to educate the young to read and write. While economists have much to say about the influence of changing school resources on pupil attainment, there is very little economic research about the effect of changing the content and structure of teaching.

In this paper, we evaluate the effect of "the literacy hour" in English primary schools on pupil attainment. This was first introduced in the context of the National Literacy Project (NLP) in September 1996, before it was implemented nationally from September 1998 onwards in the context of the National Literacy Strategy. The central idea is to raise standards of literacy in schools by improving the quality of teaching through more focused literacy instruction and effective classroom management.

We evaluate the literacy hour for schools in the National Literacy Project (NLP), which was undertaken in about 400 English primary schools in the school years 1996-97 and 1997-98. 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 large increase in attainment in reading and English for pupils in NLP schools as compared to pupils not exposed to the literacy hour over this time period.

A further aspect of this policy is its potential impact on the gender gap in pupil attainment. For many years, the attainment of boys in literacy-related activities has been considerably lower than that of girls. We find some evidence that at age 11, boys received a greater benefit from the literacy hour than girls.

Finally, we consider the cost-effectiveness of the policy. The benefits of the policy (in terms of standard deviations) are comparable to much more expensive policies such as a class size reduction. We estimate the wage return likely to arise from the increase in reading attainment as a consequence of the literacy hour. The per-pupil cost of the NLP is only a small fraction of the estimated benefits. Hence, the policy is extremely 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.

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. The lower tail of the adult literacy skills distribution is particularly pronounced in some developed countries, notably the UK and the US.[1] 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 educational performance can be raised through government policy.

More generally, there is a substantial body of research investigating how best to raise pupil achievement. The economic literature primarily addresses the impact of changing school resources such as class size, teacher quality and measures of expenditure.[2] Educationalists address many of the same issues, but also investigate the efficacy of what is taught in particular subjects 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.[3]

---

[1] For example, in the UK, the Moser report (DfEE, 1999) identifies one in five adults as not being functionally literate. Numbers from the International Adult Literacy Survey of 1995 show countries like the UK and US have very dense lower tails of their adult literacy skill distributions (including amongst younger adults) whereas in other countries like Sweden and Germany hardly any adults are at these low levels.

[2] See, for example and amongst many others, Angrist and Lavy (1999), Card and Krueger (1992) or Hanushek (1997, 2003).

[3] See the reviews by Sammons (1999) or Teddlie and Reynolds (2000) on educational research on school effectiveness, or Stainthorp (1996) for a review of evidence on what children need to learn to become skilled readers, and Scheerens (2000) and Creemers (1994) on classroom instruction. On the other hand, there are hardly any papers in economics looking at how subjects are taught. Exceptions are Angrist and Lavy's (2002) paper addressing the role of Computer Aided Instruction; the paper by Glewwe et al. (2000), which investigates the impact of flip charts in Kenyan classrooms; and Angrist and Lavy's (2001) paper about the effect of in-service teacher training on achievement in Jerusalem elementary schools.

Some research on reading appears to have influenced public policy. For example, in the US, research findings have been used as a rationale for the reading component of the No Child Left Behind Act of 2001.[4] 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 that affect what and how a subject is taught can affect achievement in schools. This is the subject matter of this paper. We look at the introduction of a highly structured literacy hour that was introduced in English primary schools in the 1990s. A strong rationale for this policy was to try to alleviate low levels of reading and writing skills held by children in many schools.[5] We have a strong research design, exploiting the fact that some children were first exposed to the literacy hour for up to two years in a period when other children were not.

To our knowledge, this is the first economic study of whether a policy of this type actually works in raising standards of literacy. Specifically, we use a difference-in-difference framework, sometimes coupled with statistical matching methods, to carry out an analysis of what happened to reading and English achievement before and after introduction of the literacy hour to pupils in schools affected by the policy relative to that in a control group of schools.

---

[4] 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).

[5] See West and Pennell (2003) for a discussion of the low levels of achievement.

A further aspect of the literacy hour is its scope to impact on gender gaps in pupil achievement. Boys have traditionally performed considerably worse at literacy-related activities, where there has been a gender gap in favour of girls for many years.[6] For example, in the school year before the literacy hour was introduced, 50 percent of 11 year old boys achieved the government target in English at the end of primary school, whereas 65 percent of girls achieved this target. There are many (often hotly contested and much discussed) reasons for why this gender gap exists, and there is a very large literature on gender gaps in education.[7] One very pertinent 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 literacy hour could differentially benefit boys.

We report results showing significant improvements in reading and English achievement for children exposed to the literacy hour, and that there were bigger gains for boys than for girls. Moreover, the policy seems to be incredibly successful in that it was not expensive to implement and thus highly cost effective. Hence, when one considers this kind of policy against other, much more expensive, alternatives – like class size reductions and raising teacher salaries – one sees that the literacy hour fares incredibly well. Of course, one should realise that its success rests on improving literacy skills at the lower end of the achievement distribution and getting children to basic levels which, as our evidence shows, can also complement their achievement levels in other areas like Mathematics. On this basis it seems to be a highly successful and desirable education policy.

---

[6] 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.

[7] A lot of this work, again, is within the education field (see Maynard, 2002, White, 1996, or Millard, 1996), but there is some more recent work in economics (like Jacob, 2002).

The structure of the rest of the paper is as follows. In Section 2, we discuss the introduction of the literacy hour, paying a lot of attention to the way in which schools were selected and our choice of comparison schools for the evaluation approach and research design we follow. In Section 3, we describe the nature of testing in English schools, the data we use and we present some descriptive statistics. Section 4 provides treatment-control estimates of the impact of the literacy hour on primary school performance, and examines whether there is evidence of any differential impact by gender. Then in Section 5, we look at cost effectiveness, including an appraisal of possible spillovers on to other parts of the curriculum. Conclusions are presented in Section 6.

2 English Primary Schools and the Literacy Hour

2.1 The Literacy Hour

Following much discussion about poor standards of English teaching, the literacy hour was introduced into English primary schools in the mid 1990s. It was first introduced to a sub-set of schools in certain Local Education Authorities[8] (LEAs) through the National Literacy Project (NLP) in September 1996. It was then introduced into all English primary schools through the National Literacy Strategy during the school year 1998/99.

---

[8] There are 150 Local Education Authorities in England. They are responsible for the strategic management of local authority education services including planning the supply of school places, ensuring every child has access to a suitable school place, intervening where a school is failing its pupils and for allocating funding to schools.

The literacy hour is firmly based upon criteria aimed at improving standards of literacy highlighted in the National Curriculum that was introduced to schools in 1988. This sets out details of what must be taught, the standards that should be achieved at different stages of the education sequence and recommends a minimum teaching time for core subjects.

The National Literacy Strategy, and the National Literacy Project before it, were aimed at raising standards of literacy in primary schools by improving the quality of teaching through more focused literacy instruction and effective classroom management. It can also be seen as an attempt to improve school management of literacy through target-setting linked to systematic planning and monitoring and evaluation.

Key components of the policy are a *framework for teaching*, which sets out termly teaching objectives for the 5-11 age range and provides a practical structure of time and class management for a daily literacy hour.[9] This daily literacy hour 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.

Whilst the literacy hour introduction constituted a discrete, well-defined and systematic change in the teaching of literacy in primary schools, one might plausibly ask why the quality of teaching was sub-standard prior to its introduction. Looking into this, it is evident that the general standard of reading and writing was a serious

---

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

concern, particularly in some LEAs.[10] 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).[11]

2.2   The National Literacy Project

The National Literacy Project was introduced in about 400 schools during the school years 1996/97 and 1997/98.[12] The Office for Standards in Education (which published the above cited report about the teaching of literacy in inner city London) approached the then Secretary of State for Education to adopt the policy within Local Education Authorities (LEAs) where educational standards were low. However, as we will see below, not all such LEAs were selected into the NLP.[13]

---

[10] Indeed, John Stannard, the Director of the National Literacy Project, told us that 'in some LEAs teaching of literacy had fallen apart'. The person in charge of introducing NLP in Tower Hamlets LEA in East London was equally downbeat saying, prior to NLP, teachers used to 'hear' reading rather than 'teach' it and noted that there was a lot of 'quiet reading time'.

[11] These concerns about reading in particular prompt us to consider the impact of the NLP on reading in particular, as well as on overall English.

[12] The NLP was also launched in an additional 112 'first' or infant schools, which we do not consider here.

[13] According to the Director of the National Literacy Project, the LEAs chosen for the NLP were selected (fairly arbitrarily) from a longer list. The LEAs of concern were those where educational standards were relatively low, although social disadvantage (which is highly correlated with low educational achievement) and demographic factors were also relevant.

From a research design perspective the NLP is highly attractive in that it gives a setting where some children were exposed to the literacy hour for up to two years when children in similar schools were not. We can thus implement a treatment-control type evaluation comparing what happened to pupil achievement before and after NLP introduction in affected and unaffected schools.

Of course, for this evaluation to be valid, we have to be very careful to ensure that affected and unaffected schools are similar to one another in the pre-policy period, and indeed that there were no different pre-treatment trends, so as to ensure that any inference we make is unbiased. Thus, discussion of how the NLP was introduced is a vital ingredient of our analysis, as is the way we define our set of control schools for our empirical application.

2.3   How were NLP schools selected?

There was a two stage process in the selection of NLP schools. First, some Local Education Authorities were selected to participate, and second some schools within LEAs were chosen. At the first stage, the LEAs chosen were, in the main part, those perceived to have lowest pupil performance. In actuality, this turns out to be somewhat mixed, with a range of LEAs being chosen. Most, but clearly not all, are from the lower end of the national achievement distribution, and some LEAs with below average levels were not included.

It turns out that about 80 per cent of NLP schools were located in LEAs in inner city, urban areas. [14] This is where the most disadvantaged and poorly performing schools in England are concentrated (West and Pennell, 2003). Collectively, about 40

---

[14] Specifically, these inner city schools are located within several Local Education Authorities in London and also Sandwell, Liverpool, Manchester, Sheffield, Newcastle and Bristol.

per cent of primary schools within these LEAs were involved in the NLP. The remaining NLP schools were in eight LEAs in three counties.[15] Only about 7 per cent of primary schools in these counties were involved in the NLP.

In terms of which schools were selected within LEAs, advice was given to choose schools most in need of the programme, but to achieve some balance between schools that were perceived to be low achieving and those with on-going problems (e.g. poor leadership) and low achieving schools showing some signs of improvement.[16] However, since the NLP was planned to last 5 years, a rolling programme was envisaged wherein schools would enter the programme in different waves. The planned cost of the NLP was £12.5 million over five years.

In terms of practicalities and management, the NLP was introduced to school staff through attendance at an induction day on Management of Literacy by the Headmaster and Chair of Governors, and a training week for designated key teachers, including English coordinators. There was also one school inset day devoted to NLP issues. This reveals that the NLP was a fairly low cost policy to implement and we will return to this later.

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. To evaluate the policy, one needs to carefully define a counter-factual where children did not receive treatment. We have therefore spent considerable time defining the comparison group against which to benchmark the policy and in designing a methodology used to evaluate the NLP in this treatment-control setting.

---

[15] The three counties are Hampshire, Essex and Norfolk. The precise LEAs are as follows: Hampshire, Portsmouth, Southampton, Essex, Southend-on-Sea, Thurrock, Isle of Wight, Norfolk.
[16] John Stannard, Director of the National Literacy Project 1996-98, personal communication.

2.4 Definition of control group

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 evident 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. According to the Director of the NLP, John Stannard, maximum co-operation between schools and cross-fertilisation of ideas was encouraged within LEAs. Indeed there has always been an effort by LEAs to facilitate collaboration and co-operation between primary schools. Hence there would have been opportunities through meetings and informal networking for teachers working in NLP schools to impart information to other teachers working in the same LEA. Given that a 5-year rolling programme was envisaged at the start of the NLP, it seems likely that teachers in non-participating schools would have been interested. However, a deliberate effort was made to contain the effect of the NLP within the selected LEAs.[17] This was to avoid a dilution of important messages expected to come out of the project.

---

[17] For example, there was no formal transfer of information about the NLP across LEA boundaries. Schools outside the LEAs involved would not have been able to obtain a copy of the framework from the national centre.

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 fully account for with available data.[18] To start with we therefore adopt the following strategy. We first identify LEAs that are in areas geographically adjacent to 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.[19] Therefore we are forced to omit some LEAs, mainly the counties, where it proves difficult to identify a good comparison group.[20] Because of this, our focus is on the inner city LEAs, which 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.

---

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

[19] 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.

[20] The only NLP LEA in a city that is omitted from our analysis is Bristol, where all adjacent areas differ massively in their pupil achievement and population densities (as the city area of Bristol is surrounded by semi-rural areas).

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 this both within a difference-in-difference framework and using statistical matching methods. For comparative purposes, and as a robustness check, we also present results based on using all non-NLP schools in the country as an alternative control group.

Finally, it is important to note that the literacy hour represents a change in how literacy skills are taught rather than an increase in the time devoted to English. The National Curriculum (introduced in 1988) already had guidelines about the minimum number of hours to be spent on core subjects. The 'literacy hour' does not constitute an increase over and above this guideline. However, we consider potential spillover effects of the literacy hour on other subjects later in the paper.

3   Patterns of Achievement in English, Data and Initial Descriptive Statistics

3.1   Testing and patterns of achievement in English primary schools

Following the introduction of the National Curriculum in 1988, testing throughout the school years has been an important feature of the English school system. Children are administered tests at ages 7, 11, 14 and 16, in what is known as Key Stages 1, 2, 3 and 4. Key Stages 1 and 2 take place in primary school and Key Stages 3 and 4 in secondary school.

Since the literacy hour was introduced in primary schools we assess its impact on the Key Stage 2 tests that children take at the end of their time in primary school. Table 1 provides some statistics on Key Stage 2 English at age 11 between 1996 and

2002. It shows the percentage of children reaching the government target of level 4 (or above) in English and in reading.[21] 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 the rest of the paper.

In 1996, the year before the introduction of NLP, 57 percent of children achieved at least level 4 in their overall English assessment. One sees a gradual increase in this measure over time, although with something of a plateauing out from 2000 onwards. 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 20 percentage points over this time period compared to a 14 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.

3.2   Data

The empirical analysis is based on administrative records of pupil-level achievement and school-level data. The former consists of detailed test score information on students at the end of Key Stage 2.[22] The tests are set and marked externally to the school. The first available year of national Key Stage 2 data at pupil-level is in 1996,

---

[21] The maximum level attainable at Key Stage 2 is Level 6.
[22] The Key Stage 2 tests were first taken in the school year 1994/95.

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 test scores, gender of the student and school codes. 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 identified as Special Educational Needs) and other school characteristics (e.g. school type).[23]

With regard to outcome measures, we concentrate on two measures at the end of primary school: the percentile reading score and the percentage of students achieving level 4 or above in Key Stage 2 English. 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.[24]

3.3   Descriptive statistics

Descriptive statistics are provided in Table 2. The Table shows measures of primary school achievement before and after the introduction of the policy for NLP and

---

[23] A full list of the detailed set of variables used as controls is provided in the notes to Tables 3 and 4.

[24] The National Literacy Taskforce (which developed the National Literacy Strategy) advised that by 2002, 80 per cent 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.

comparison group schools. The Table shows two primary school 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. This is important since, as noted above, for valid inference to be drawn we need to standardise for baseline differences in the pre-policy period (to ensure we are comparing like with like) and we are careful to investigate this in detail in our empirical modelling below.

With this in mind, one should be careful in reading the Table, but a suggestive and 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 by 2 percentile points in the NLP schools relative to the control group, as shown in the final column showing the difference-in-difference in the period surrounding NLP introduction. The same relative pattern of improvement is seen for KS2 English, where the percentage of pupils attaining level 4 or above rises by 3 percentage points more in NLP schools (by 11 percentage points as compared to 8 in the control group).

4 The Impact of NLP on Pupil Achievement

In this section, we evaluate the NLP policy impact, looking at difference-in-difference models that control for a large number of observable factors and for unobserved school heterogeneity. This is important owing to the different levels of pre-policy achievement in the NLP and control schools already discussed. The basic estimates

are derived from the following model for pupil i in school s in year t:

(1) $\qquad A_{ist} = \alpha_1 + \beta_1 NLP_s * Policy\ On_t + \gamma_1 NLP_s + \delta_1 X_{ist} + \lambda_1 Z_{st} + \pi_1 T_t + \varepsilon_{1ist}$

where A is pupil achievement, 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_1 is the difference-in-difference estimate of the NLP policy. The γ_1 coefficient is the initial (pre-policy) baseline difference in achievement 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:

(2) $\qquad A_{ist} = \alpha_s + \beta_2 NLP_s * Policy\ On_t + \delta_2 X_{ist} + \lambda_2 Z_{st} + \pi_2 T_t + \varepsilon_{2ist}$

In (2) 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_2 measures within-school changes in achievement before and after NLP introduction in treatment schools relative to within-school changes in achievement in control schools.

4.1   Estimated policy impacts (Key Stage 2)

Estimates of equations (1) and (2) are reported in Panels A and B of Table 3 for primary school reading and overall English. 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.[25]

The results in Panels A and B of the Table are highly supportive of the hypothesis that the literacy hour, via the NLP policy, significantly improved reading and English performance. The first column specification shows a highly significant 2.1 percentile point increase in reading scores and a 2.7 percentage point increase in the percent 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 children in NLP schools was significantly worse, at -8.8 percentile points lower for reading and a lower 12.4 percentage points achievement level 4 in overall English.

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, in the regression models with controls, we are

---

[25] 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).

comparing like with like in the initial pre-policy period and thus suggests that a treatment-control methodology of the type 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 3.2 higher percentage achieving level 4 or above in English. Both of these difference-in-difference policy impacts are strongly significant in statistical terms.

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 to 3.2 for the percentage achieving level 4 or above in English. It is interesting that the effect of controlling for both observable and unobservable variables (the latter via school fixed effects) actually leads to a rise in the estimate of the NLP effect. This suggests a downward bias in 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 1.8 higher percentage achieving level 4 or above in English. Given that the difference-in-difference estimates already eliminate baseline achievement differences, we believe these matching estimates to be conservative.

Nonetheless they show the policy to have led to a marked improvement in NLP schools.

The finding of a positive and significant literacy hour effect, coupled with an increase in this effect after including observable controls and school fixed effects, makes us reasonably confident that we have pinned down an effect which is attributable to the policy. In particular, the increase in the size of the effect in the conditional models makes us think that other possible pre/post changes unique to the NLP schools are not driving the results. However, a problem would arise if earlier trends in achievement were different in NLP and comparison schools. In this regard, an important check would be to test for pre-treatment differences in trends. If a difference-in-difference estimate of the literacy hour introduction did not exist before the policy was introduced, this would give support to the argument that it is not pre-existing differences in NLP and comparison schools that are 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 have been able to obtain two years of pre-policy school-level data, for 1995 and 1996.[26] Data is only available for overall Key Stage 2 English. Hence, we compare pre-treatment changes across NLP and comparison schools using this measure.

The exercise comparing results with those of the pre-treatment period is reported in Panels C and D of Table 3. Panel C is the school-level analogue to Panel B. Reassuringly the school-level results are almost identical to the pupil-level results.[27] Panel D then reports results for the pre-treatment period wherein the difference-in-difference estimate is obtained using school-level data in 1995 and 1996. The results are strong and highly reassuring for our identification strategy as all

---

[26] This data is from the School Performance Tables in these years.
[27] Note that the dependent variable is taken from the School Performance Tables rather than pupil-level administrative data sets.

three specifications show there to be no difference whatsoever in the pre-treatment trends between NLP and comparison schools.[28] This is an important finding which shows that, whilst the period of the policy intervention saw the English skills of pupils in NLP schools improve significantly, this clearly was not happening before. This gives us some confidence that there are no pre-existing differences in the NLP and comparison schools that are affecting changes over time in outcomes.

4.2   Effects at higher and lower levels of achievement

So far, in terms of overall English, we have concentrated on the government target level, level 4. One may, however, be interested in examining other impacts, both for lower and higher levels. The rationale for looking at lower levels is clear, if one thinks the literacy hour has potential as a useful policy tool to improve basic literacy. In terms of higher levels, one may be concerned that concentrating on trying to improve lower levels could harm better performing students, and so one needs to consider the impact on higher levels as well.

To this end, Table 4 contains results looking at level 3 and level 5 KS2 English outcomes. The Table shows a very strong impact on level 3 with reading and English significantly improving at this lower level. Moreover, there is no evidence at all of a harmful effect at level 5 where the effects are insignificantly different from zero in all cases (and mostly with positive estimated coefficients).

---

[28] Notice that there are 20 fewer schools (821 rather than 841) in Panel D compared to the other panels. 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 coefficient estimates (and standard errors) comparable to columns (1) to (3) for the 821 schools sample were 2.558 (1.080), 3.112 (0.913) and 3.207 (0.934) for Panel B and 2.103 (1.077), 2.663 (0.944) and 2.596 (1.152) for Panel C.

4.3 Robustness

One might also express a concern that results are sensitive to the choice of Local Education Authorities used to define our sample of control schools. As a robustness check, we have therefore 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 Panels A and B of Table 3. 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 3.6 percentage points in level 4 or above 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.

Hence, these results strongly corroborate the spatial matching estimates and suggest that the literacy hour significantly raised pupil performance in 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.

4.4   Gender differences

As explained in the introduction, from a theoretical perspective one may be interested in the impact of the literacy hour on gender differences in achievement. Given the existence of sizable gaps in English achievement 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.[29]   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 raises the percentage achieving level 4 or above in KS2 English by between 2.7 and 4.2. These are large effects. Thus, it appears that the literacy hour was more effective for boys and as such reduced the gender gap 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 has reduced in

---

[29] The model is estimated over the pooled data. It allows for a separate literacy hour impact by gender. All variables (including fixed effects) are interacted with the gender variable.

recent years. The results we report here are entirely consistent with the literacy hour having played an important role.

5   Measuring Economic Costs and Benefits

This analysis has shown a significant impact of the literacy hour on reading and on English achievement. 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).[30] 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 estimate in terms of standard deviations. This is calculated as 0.091 standard

---

[30] This includes infant schools.

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).[31] 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 0.54, 0.42 and 0.20 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 0.091 increase in the standard deviation (sd) of reading percentiles is then calculated as 0.091*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 such calculations, certain issues that may lead one to question these economic benefits. One clear example in this case is that the beneficiaries 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

---

[31] 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).

score distribution, yet the earnings effect from age 10 reading is assumed linear in the regressions in Table 7. Therefore we have 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 0.458 (standard error = 0.100) as compared to the top half effect of 0.435 (0.084); 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 0.06 standard deviation increase), the economic benefits are measured in the range of £1375 to £3581.

The benefits of the policy (a 0.06 to 0.09 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 0.13 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 0.11 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.

5.1   Spillovers on to other subjects

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, given the guidelines in the National Curriculum, 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. Indeed, a recent report by OFSTED (2002) suggests that the National Literacy Strategy has raised the quality of teaching in the rest of the curriculum.

In Table 8 we consider the possible substitution/complementarity with Mathematics, reporting a test of whether the literacy hour helped or hindered achievement in this subject. 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 2.5 higher percentage achieving at least level 4. Also, when estimating these regressions using the matched sample of schools (column 5), these coefficients reduce to 0.838 and 1.373 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. 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, as reported in Panel D of Table 8. The effects in the pre-treatment period are small and always insignificantly different from zero. As such, looking at the impact of the literacy hour on 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 the

teaching of 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.

6   Conclusions

In this paper, we have considered the potential for a change in the content and structure of teaching to impact on pupil performance. Our analysis is facilitated by the introduction of a literacy hour to English primary schools, through the National Literacy Project (NLP), which introduced the literacy hour to around 400 English primary schools in 1997 and 1998. We adopt an explicit treatment-control group approach, investigating what happened to pupil achievement 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 levels rose by more in NLP schools between 1996 and 1998. Having subjected our identification strategy to a number of robustness checks, we are confident that this constitutes an NLP effect. In fact, the estimated policy impact rises when we control for other factors that may have changed at the same time as the policy was introduced and also following the inclusion of controls for unobserved heterogeneity (through school fixed effects). We show there to be no trend difference in pupil achievement in NLP relative to comparison schools in the pre-policy period. Since significant gender gaps in English performance exist (in favour of girls), we also consider whether the literacy hour had a differential impact by gender and report some evidence that, at age 11, boys

benefited more than girls. Finally, we show the benefits of the literacy hour to exceed the costs of the policy by a large margin.

These findings are of considerable significance when placed into the wider education debate about what works best in schools for improving pupil performance. They are also of considerable significance for education policies in countries which have problems with their levels of literacy skills. As suggested in the introduction, the research remit of economists has tended to be rather narrow in comparison with educationalists in this area. Our approach shows that one of the areas receiving much less attention from economists in the past, namely looking at the content and organisation of what is taught, matters for English and reading.[32] This is particularly important given that it will almost certainly be the case that the same teachers were teaching literacy before and after the introduction of the literacy hour.[33] Indeed, as the effects we identify come from a government policy aimed at improving literacy, the evidence we report suggests that public policy aimed at changing literary instruction can significantly raise pupil achievement and can do so in a highly cost effective manner.

---

[32] Of course, one should recognize that changing teaching methods more generally may not necessarily operate in a similar manner to the literacy hour. For example, Cohen and Hill (2000) report that changing teacher practice in mathematics teaching in the US only happens slowly and partially, and they argue that the inertia in the process is probably driven by the lack of a knowledge base amongst mathematics teachers. Understanding better whether the change was facilitated more easily because the literacy hour is a government policy, or whether it is something specific to the ability to change teaching methods for particular subjects, is an important area for future research.

[33] Moreover, we have looked at teacher turnover before and after the introduction of the literacy hour in NLP schools versus the control schools and find no significant change occurring.

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) Does teacher training affect pupil learning: Evidence from matched comparisons in Jerusalem public schools, Journal of Labor Economics, 19, 343-369.

Angrist, J., and V. Lavy (2002) 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, 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.

Cohen, D. and H. Hill (2000) Instructional Policy and Classroom Performance: The Mathematics Reform in California, Teachers College Record, 102, 294-343.

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, <u>Educational Evaluation and Policy Analysis</u>, 19, 141-64

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

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

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

Hopkins, D. (1999) Success for All – Part One, <u>Literacy Today</u>, 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, <u>Economic Journal</u>, 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 Achievement, Centre for the Economics of Education, unfinished draft.

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

Millard, E. (1996) <u>Differentially Literate: Boys, Girls and the Schooling of Literacy</u>, 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.

Office for Standards in Education (2002) 'The Curriculum in Successful Primary Schools', 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.) <u>Equity in the Classroom: Towards Effective Pedagogy for Girls and Boys</u>, Falmer Press: London.

Table 1 Primary School English and Reading 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			
	Pre-Policy, 1996	Post-Policy, 1997-98	Change	Pre-Policy, 1996	Post-Policy, 1997-98	Change	Difference-in-Difference
Percentile Reading Score	45	46	1 (p=.03)	54	53	-1 (p = .19)	2 (p =.01)
Percentage Reaching Level 4 or above in KS2 English	38	49	11 (p = .00)	50	58	8 (p = .00)	3 (p = .00)

Notes: Total sample size is 104654 pupils in 841 schools. The NLP school and control group school areas are defined in Appendix Table A1. P-values (based on standard errors clustered on school codes) in parentheses.

Table 3 NLP and Primary School Reading and English

	(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	2.423	2.631	1.715	1.791
	(0.741)	(0.647)	(0.650)	(0.631)	(0.628)
NLP	-8.808	-1.075	--	-.962	--
	(0.845)	(0.883)		(0.892)	
Number of Pupils	104654	104654	104654	96083	96083
Number of Schools	841	841	841	761	761
B. Percentage Achieving Level 4 or Above Key Stage 2 English					
NLP*Policy On	2.710	3.160	3.182	1.982	1.797
	(1.068)	(0.903)	(0.924)	(0.905)	(0.914)
NLP	-12.421	-0.795	--	-.606	--
	(1.192)	(1.288)		(1.332)	
Number of Pupils	104654	104654	104654	96083	96083
Number of Schools	841	841	841	761	761
C. Percentage Achieving Level 4 or Above Key Stage 2 English (School Level)					
NLP*Policy On	2.346	2.746	2.607	1.608	1.367
	(1.063)	(0.936)	(1.139)	(0.955)	(1.156)
NLP	-11.815	-0.691	--	-.852	--
	(1.165)	(1.202)		(1.247)	
Number of Schools	841	841	841	761	761
D. Percentage Achieving Level 4 or Above Key Stage 2 English (School Level), Pre-Treatment, 1995 and 1996					
NLP*Year=1996	.198	.244	.304	.688	.647
	(1.320)	(1.337)	(1.902)	(1.418)	(2.011)
NLP	-12.245	-1.745	--	-2.061	--
	(1.455)	(0.914)		(0.960)	
Number of Schools	821	821	821	747	747

Notes: standard errors (clustered on school) in parentheses; all specifications include year dummies; 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. In Panels C and D the dependent variable is from the School Performance Tables (this is very highly correlated with aggregated to school-level data from Panel B). In Panels C and D the estimates are weighted using pupil numbers to maintain comparability with Panel B.

Table 4 Effects at Different Key Stage 2 Levels

	(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. Percentage Achieving Level 3 or Above Key Stage 2 English					
NLP*Policy On	2.299	2.586	3.419	1.933	2.472
	(0.881)	(0.810)	(0.893)	(0.759)	(0.826)
NLP	-7.198	1.132	--	1.852	--
	(0.907)	(1.034)		(1.020)	
Number of Pupils	104654	104654	104654	96083	96083
Number of Schools	841	841	841	761	761
B. Percentage Achieving Level 5 or Above Key Stage 2 English					
NLP*Policy On	.184	0.318	0.394	-0.216	-0.068
	(0.543)	(0.511)	(0.526)	(0.499)	(0.535)
NLP	-4.700	-0.411	--	-0.620	--
	(0.533)	(0.789)		(0.807)	
Number of Pupils	104654	104654	104654	96083	96083
Number of Schools	841	841	841	761	761

Notes: as for Table 3, Panel B.

Table 5
NLP and Primary School Reading and English
(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	2.118	2.460	2.003	2.319
	(0.620)	(0.586)	(0.569)	(0.584)	(0.570)
NLP*Policy On (Counties)	-0.753	-1.022	-0.980	-1.225	-1.175
	(1.019)	(0.860)	(0.915)	(0.869)	(0.926)
NLP (Cities)	-12.803	-1.658	--	-1.728	--
	(0.651)	(0.424)		(0.422)	
NLP (Counties)	-8.239	-0.395	--	-0.242	--
	(0.946)	(0.506)		(0.506)	
Number of Pupils	1716855	1716855	1716855	1632198	1632198
Number of Schools	14670	14670	14670	13072	13072
B. Percentage Achieving Level 4 or Above in Key Stage 2 English					
NLP*Policy On (Cities)	2.893	3.249	3.639	3.067	3.399
	(0.902)	(0.821)	(0.817)	(0.821)	(0.821)
NLP*Policy On (Counties)	-0.434	-0.930	-1.299	-1.210	1.565
	(1.423)	(1.198)	(1.291)	(1.209)	(1.310)
NLP (Cities)	-18.464	-2.150	--	-2.226	--
	(0.947)	(0.544)		(0.539)	
NLP (Cities)	-13.689	-1.787	--	-1.580	--
	(1.495)	(0.756)		(0.760)	
Number of Pupils	1716855	1716855	1716855	1632198	1632198
Number of Schools	14670	14670	14670	13072	13072

Notes: standard errors (clustered on school) in parentheses; all specifications include year dummies; 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
A. Percentile Reading Scores		
Boys	3.412 (.741)	2.536 (0.738)
Girls	1.781 (.752)	0.987 (0.733)
B. Percentage Achieving Level 4 or Above in Key Stage 2 English		
Boys	4.191 (1.153)	2.749 (1.175)
Girls	2.097 (1.112)	0.756 (1.111)

Notes: as for Table 3, Panels A and B, for models including school fixed effects. Column (1) of this Table is the same specification as for column (3) of Table 3, and column (2) is the same specification as column (5) of Table 3, except that both additionally allowing for gender differences in the NLP impact.

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)	0.544 (0.024)	0.423 (0.035)	0.207 (0.036)
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	0.5

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) No controls	(2) Full set of controls	(3) Full Model (with school fixed effects)	(4) Full Set of Controls, Matching	(5) Full Model (with school fixed effects), Matching
A. Percentile Scores in Mathematics					
NLP*Policy On	1.275	1.585	1.530	0.935	0.838
	(0.789)	(0.662)	(0.633)	(0.684)	(0.660)
NLP	-7.793	-0.338	--	-0.481	--
	(0.891)	(0.994)		(1.022)	
Number of. Pupils	104654	104654	104654	96083	96083
Number of Schools	841	841	841	761	761
B. Percentage Achieving Level 4 or Above in Key Stage 2 Mathematics					
NLP*Policy On	2.419	2.849	2.461	1.789	1.373
	(1.210)	(1.027)	(0.989)	(1.067)	(1.027)
NLP	-11.845	-1.235	--	-1.265	--
	(1.317)	(1.373)		(1.419)	
Number of Pupils	104654	104654	104654	96083	96083
Number of Schools	841	841	841	761	761
C. Percentage Achieving Level 4 or Above Key Stage 2 Mathematics (School Level)					
NLP*Policy On	2.336	2.669	2.186	1.438	0.948
	(1.266)	(1.113)	(1.289)	(1.155)	(1.331)
NLP	-11.420	-1.737	--	-1.963	--
	(1.349)	(1.356)		(1.400)	
Number of Schools	841	841	841	761	761
D. Percentage Achieving Level 4 or Above Key Stage 2 Mathematics (School Level), Pre-Treatment, 1995 and 1996					
NLP*Year=1996	-0.797	-0.759	-0.861	-0.291	-0.445
	(1.514)	(1.534)	(2.153)	(1.583)	(2.228)
NLP	-9.473	2.026	--	2.037	--
	(1.696)	(1.022)		(1.072)	
Number of Schools	821	821	821	747	747

Notes: standard errors (clustered on school) in parentheses; all specifications include year dummies; 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 achievement)

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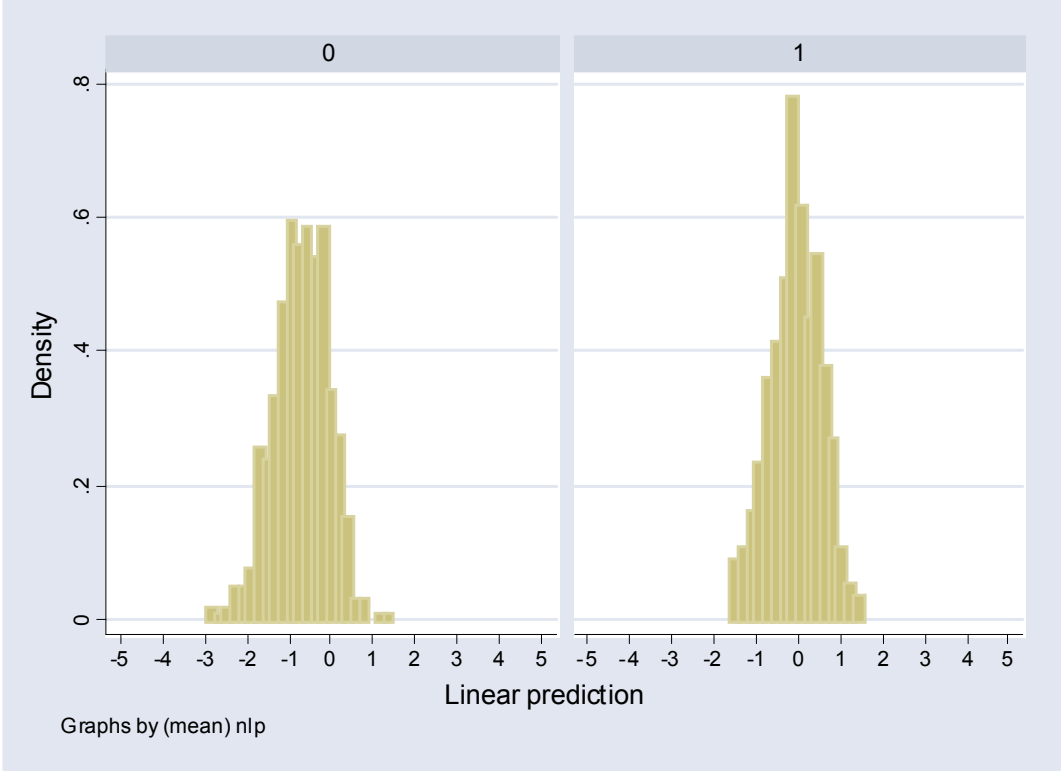**Probability of Treatment (NLP = 1),**
Estimates for Table 3 Specifications

	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

Graphs by (mean) nlp

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	-.053 (.134)
Percentile writing score: 2nd quartile	-.128(.119)
Percentile writing score: 3rd quartile	-.171(.119)
Percentile reading score: 1st quartile	.473(.181)
Percentile reading score: 2nd quartile	.371(.151)
Percentile reading score: 3rd quartile	.137(.138)
English: Proportion attaining level 4 and above .	-.289(.359)
English: Proportion with no level due to absence/disapplication	-.733(1.055)
Mathematics: Proportion attaining level 4 and above .	-.145(.290)
Mathematics: Proportion with no level due to absence/disapplication	.672(.998)
Science: Proportion attaining level 4 and above .	.101(.242)
Science: Proportion with no level due to absence/disapplication	-.657(.966)
Proportion entering KS2 English Extension test in 1996	.796(.363)
Proportion entering KS2 Mathematics Extension test in 1996	-.554(.490)
Proportion with missing results in 1996 (Performance Tables)	-.428(.221)
Proportion missing from Performance Tables in 1996	-.041(.351)
Religious school	-.002(.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	.399(.136)
Proportion of teachers unqualified	1.618(.863)
Ratio support staff to teachers	-.749(.221)
% of teachers: graduates	.052(.249)
% of teachers with - higher/1st/2nd degree	.380(.248)
Proportion female teachers	.021(.229)
Proportion with missing information on 'non-white'	.333(.463)
Proportion missing from DTR	-.072(.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
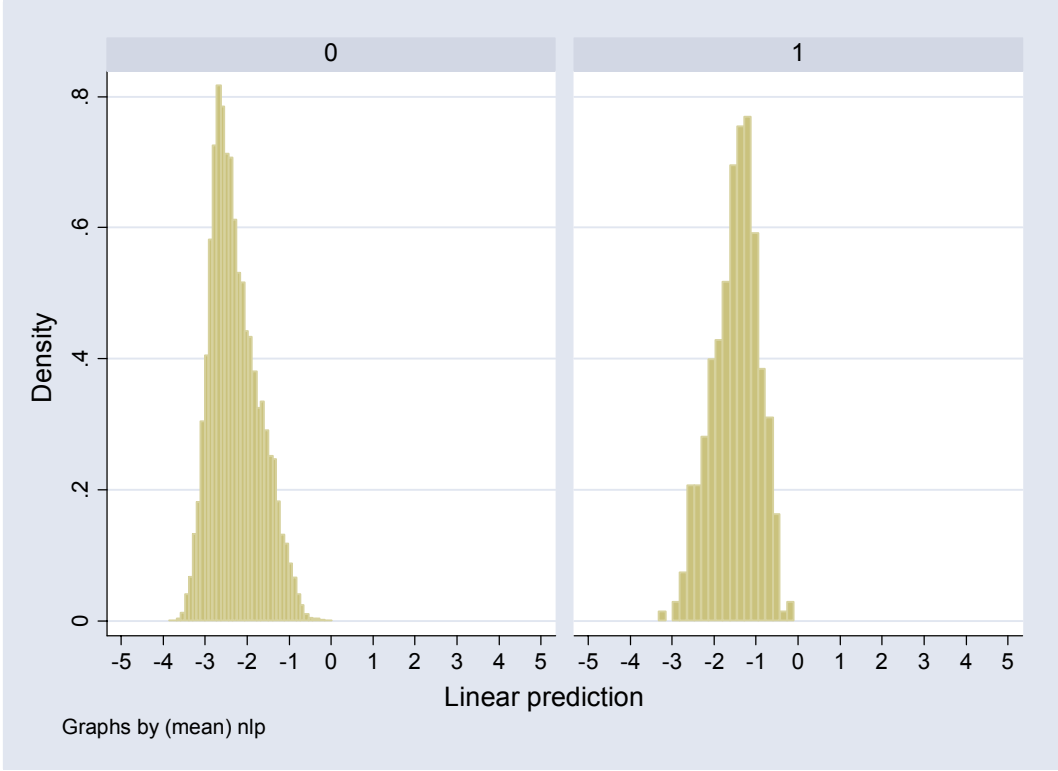
Graphs by (mean) nlp

Selected for 'matching' specification:
Schools with a predicted linear index of the propensity score between -2.94 and -0.19.

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

Kernel regression, bw = 20, k = 5
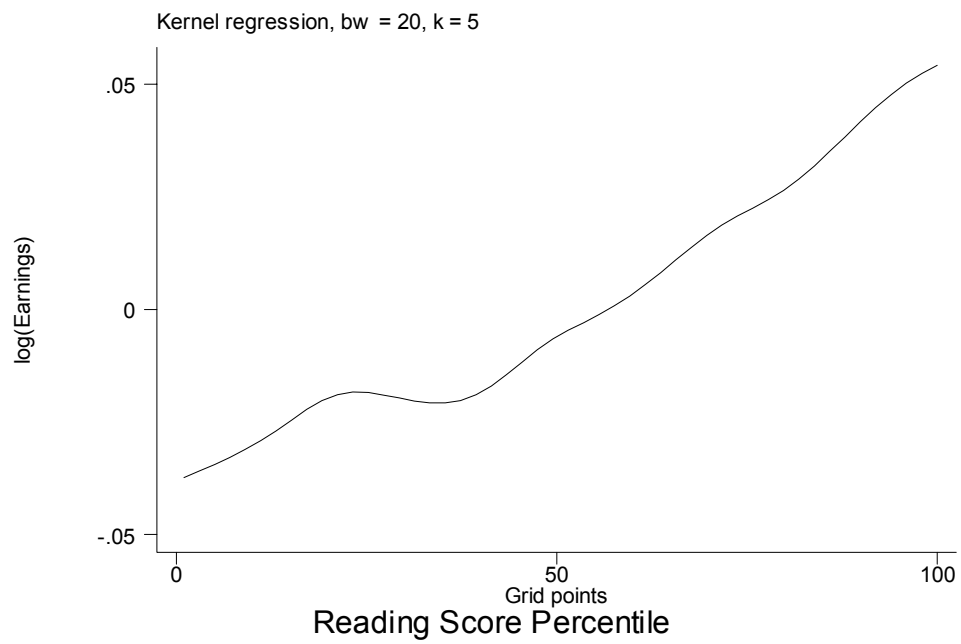
Notes: This is the log(earnings)-reading score percentile relation from a specification comparable to column (2) of Table 8.