n developed countries like the United States and Britain, the continuing challenge for educators is to sort through the choices of an all-you-can-eat school system and teach the basic skills. Despite so-called universal education, an alarming number of people still fail to reach even basic levels of literacy. According to Sir Claus Moser, chairman of the Basic Skills Agency, one in five adults in Britain is functionally illiterate. The International Adult Literacy Survey shows that Britain is only slightly behind the United States, where 21–24 percent of adults have the lowest level of literacy skills. The problems in the United States and Britain are notably worse than in other developed countries.

How to ensure that future generations of adults do not suffer from such problems is, of course, education’s $64,000 question (though the price tag is considerably higher these days) and there have been as many proposals for increasing literacy as there are illiterates. Some of the more prominent initiatives—like the Reading First component of No Child Left Behind and the “Success for All—Reading First” program begun at Johns Hopkins in the late 1970s—involve the implementation of a highly structured classroom framework that spells out what should be taught, how it should be taught, and for how long.

The “literacy hour” was introduced in a select group of primary schools in September, 1996, as part of England’s National Literacy Project (NLP). It provides us with a unique opportunity to study the impact of such highly structured programs on learning. Aimed at children from 5 to 11 years of age, the literacy hour spurned the passive (or quiet) approach to reading used in many classrooms in the United States and Britain and brought a great deal of precision to the task of instruction, mainly with a tightly organized and strictly managed program.

Do such formal and structured reading programs work? Will they improve reading abilities, and will they do so at a reasonable cost? That is what we asked of the literacy hour. To answer the question, we took advantage of the fact that children in 400 schools were in the program for up to two years before it was rolled out in all of England’s primary schools.

BY STEPHEN MACHIN AND SANDRA MCNALLY
Experiment

AN HOUR
A DAY
KEEPS
ILLITERACY
AT BAY
in the fall of 1998. And we were also able to explore the program's impact on gender gaps in pupil achievement, an important issue since in England, as in other countries, girls have traditionally outperformed boys in literacy-related activities.

In fact, we found that exposure to the literacy hour significantly improved students' reading and English achievement, with bigger gains for boys than for girls. Moreover, the program proved to be a highly cost-effective means of improving reading scores, especially when compared with the common alternatives, like class size reductions and raising teachers' salaries.

A Change of Technique, Not Time
The National Literacy Project, of which the literacy hour was a key component, was meant to beef up the National Curriculum, a detailed course of studies that had been introduced in England and Wales in 1988. The curriculum had specific benchmarks at each grade level, recommended minimum teaching times for core subjects, and a full complement of tests. All schoolchildren in England aged 7, 11, and 14 (known as Key Stages 1, 2, and 3) were tested in core subjects, including English. There was a final examination in a range of subjects at age 16, at the end of compulsory schooling. There are various components of the test in English at Key Stages 1–3, including a test for reading, writing, and spelling.

At the selected schools the literacy hour was first introduced to staff by the headmaster and chair of governors, then at a training week for designated key teachers and program coordinators. There was also one in-school professional development day devoted to NLP issues.

The prospective literacy-hour teachers were given instructional material that laid out the specifics of the program. The hour was divided into two, 10–15 minute segments consisting of whole-class reading or writing and whole-class word-level (phonics, spelling) and sentence-level work; one 25–30-minute session of directed group activity; and a whole-class summary meeting at the end (5–10 minutes) for pupils to revisit the objectives of the lesson, reflect on what they had learned, and consider what they needed to do next. Guidance on content was set out in the “framework for teaching,” given to all literacy hour instructors.

Part of the rationale for this approach was the belief, as a government report showed, 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. Though we consider potential spillover effects on other subjects later in this story, it is important to note that the
Choosing the Right Control Group

Though the selection of schools to be included in the pilot program was fairly arbitrary, it was not random. The selected Local Education Authorities (LEA), the rough equivalent of an American school district, tended to have low test scores and high social disadvantage. At the end of the selection process, some 80 percent of NLP schools were located in LEAs in inner-city, urban areas, where the most disadvantaged and poorly performing schools in England are concentrated. About 40 percent of primary schools within these LEAs were involved in the NLP.

The local administrators charged with implementing the new program were expected to provide a strong lead to all their schools. Cooperation and the sharing of ideas between schools within LEAs was also actively encouraged. For these reasons the schools within the LEA that were not participating in the NLP could have been affected by the program, even if indirectly, and are thus probably not good candidates for comparison.

To find a good comparison group, it was instead necessary to turn to nonparticipating LEAs. Fortunately, for our purposes as evaluators, project administrators made a concerted effort to contain the effect of the NLP within the selected LEAs, and no information about the NLP formally crossed LEA boundaries; schools outside the LEAs involved would not have been able to obtain a copy of the framework from the national center.

To provide as fair a test as possible, we also adopted an additional strategy to restrict our analysis to those LEAs that were truly comparable with the NLP participants. To use the remaining 13,600 primary schools in the country would be to implicitly compare schools in inner-London LEAs like Hackney to schools in the Isle of Wight. (American readers might imagine comparing schools in the Bronx with schools in Wyoming.) Even with detailed information on the characteristics of the schools and the students who attend them, we may not be able to account for all the many factors that could affect student achievement.

Thus we identified LEAs that were geographically adjacent to LEAs involved in the NLP. Then, if there were multiple adjacent non-NLP LEAs, we chose the one most similar in student achievement before the start of the NLP.

In the end, we were forced to omit some LEAs, mainly in rural counties, where we could not identify a good comparison LEA. As a result, our study is essentially limited to inner-city LEAs, which compose 80 percent of the NLP schools, and are precisely the schools of interest to us, particularly in the context of the debate about poorly performing inner-city schools in the United Kingdom.

Basic Trends in Achievement

The English performance of British primary-school students has improved considerably since the literacy hour was introduced (see Figure 1). In 1995, the year before initial implementation, 57 percent of children at Key Stage 2 (what would be the end of 6th grade in the United States), achieved at least a level 4 in their overall English assessment. (There are 6 levels altogether; 4 is considered proficient, or grade-level appropriate.) Scores increased gradually over
time and, by 2002, the percentage achieving level 4 or above had risen to 75. Although sizable gender gaps remain, between 1996 and 2002, boys improved their relative position, with an increase of 20 percentage points compared to a 14-percentage-point increase for girls.

A similar pattern is evident for reading, where we have data only from 1997 onward. Those achieving at least level 4 increased from 67 to 80 percent between 1997 and 2002. Once again, although sizable gender gaps are present at each point in time, over this period boys experienced an increase of 14 percentage points compared with an increase of 12 percentage points for girls. In the detailed analysis, we are primarily interested in two main outcome measures: the percentage of children reaching the expected standard for their age in English, (“level 4” which takes account of tests in reading, writing, and spelling) and the percentile score in the reading test (as low standards in reading were of particular concern).

A simple comparison of the scores of schools in the program and the comparison LEAs suggests the literacy hour may have played a role in this general improvement. In 1996, before the program was implemented in the program schools, only 38 percent of students in NLP schools achieved level 4 in English, compared with 50 percent of students in the comparison LEAs. Over the following two years, the percentage of students achieving level 4 increased by 11 percentage points in schools already using the literacy hour, against only 8 percentage points in the comparison schools. Average reading scores increased by 1 percentile point in the NLP schools, while in the comparison schools it dropped by the same amount.

While these relative gains made by schools using the literacy hour are suggestive, it is important to consider whether they may have reflected other differences between the two groups of schools.

The Impact of the NLP
To provide a more rigorous evaluation of the program’s impact, we compare the reading and English performance of individual students attending NLP and comparison schools in 1997 and 1998, while taking into account a wide variety of school characteristics that could also influence student achievement.

These characteristics include, in addition to a variety of measures of student achievement as of 1996, the percentages of students in the school that are eligible for free school meals, those who are nonwhite, and those with special educational needs; the pupil-teacher ratio and the number of students enrolled; whether the school is all girls, all boys, a religious school, or in London; and several measures of the qualifications of the teaching staff. After taking into account all these characteristics, schools using the literacy hour outperformed comparison schools by 2.4 percentile points in reading and by 3.2 percentage points in the share of students achieving level 4 in English at Key Stage 2.

Because we observe schools over several time periods, we can subject the program to an even stricter test by controlling for all characteristics of schools that remain constant over time (by “differencing out” the effect of attending a particular school on exam scores). This effectively eliminates not only the differences between schools in the measurable characteristics listed above, but also the effects of any unobserved characteristics that are stable over time. Under this stricter test, the impact of the NLP actually

---

**Ready to Read (Figure 2)**

Eleven-year-olds in schools that participated in the literacy hour program between 1996 and 1998 were much more likely to score at or above level 4 (on a scale from 1 to 6, where 6 is the highest and 4 proficient) on the English exam than students in nonparticipating comparison schools.

<table>
<thead>
<tr>
<th>Impact of Literacy Program on 11-Year-Olds Achieving Proficiency on the English Exam</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All students</strong></td>
</tr>
<tr>
<td>0%</td>
</tr>
<tr>
<td>1%</td>
</tr>
<tr>
<td>2%</td>
</tr>
<tr>
<td>3%</td>
</tr>
<tr>
<td>4%</td>
</tr>
<tr>
<td>5%</td>
</tr>
</tbody>
</table>

* significant at the 0.05 level  ** significant at the 0.01 level

Note: Impact after adjusting for school demographic characteristics and the year the test was administered.

**SOURCE:** Authors’ calculations from data supplied by the Department for Education and Skills.
increases slightly, going to a 2.6 percentile point improvement in reading and a 3.2 higher percentage achieving level 4 or above in English (see Figure 2). This is our best estimate of the program’s true impact.

The higher performance of students in schools using the literacy hour, coupled with the fact that this difference continues to be observed even after taking into account other differences among schools, makes us reasonably confident that we have pinned down the effect attributable to the policy. However, a problem could arise if achievement had already been increasing more quickly in NLP schools than in comparison schools even before the policy’s implementation.

To rule out this possibility, we rely on school-level data on the percentage of students achieving level 4 in Key Stage 2 English, as the more detailed student-level test scores examined above are not available before 1996. Still, a careful analysis of these aggregate data reveals no difference whatsoever in the pretreatment trends between NLP and comparison schools. This confirms that the stronger performance of NLP schools after the literacy hour’s adoption was not attributable to preexisting differences in achievement.

Though we have concentrated our analysis of the NLP’s impact on English performance on the percentage of students performing at grade level at Key Stage 2, it is also important to ask whether the literacy hour improved literacy for students performing at lower levels or, conversely, whether attempts to improve basic literacy might have harmed better-performing students. We therefore also analyzed the effect of the literacy hour on the percentage of students achieving level 3 and level 5. The results confirm a strong positive impact on the share of students reaching level 3 in English. While the program did not increase the share of students achieving level 5, there is no evidence of a harmful effect on high-performing students.

Gender Differences
Given the existence of sizable gaps in English achievement between boys and girls, the impact of the program by gender is also of policy interest. In 1996 there was a 15-percentage-point gap nationally in the percentage of boys and girls achieving level 4 or above. Even though the gap has partially closed, it still was 9 percentage points in 2002. So, our analysis estimates separate NLP effects by gender.

We found that the NLP effect for boys is 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 raised the percentage achieving level 4 or above in Key Stage 2 English by between 2.7 and 4.2. These are large effects. For girls, only small effects were observed. 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 trend. As mentioned, the gender gap in primary school reading and English has been reduced in recent years. The results we report here are consistent with the literacy hour’s having played an important role.

Measuring Costs and Benefits
Our analysis has identified a significant impact of the literacy hour on reading and English achievement. Was it also cost-effective? To find out, we compared the per-pupil costs of the policy with the economic benefits, as reflected in predicted labor market earnings. (Of course, this is a narrow definition of economic benefit. A higher level of literacy may also, for example, increase the probability of employment and reduce the probability of criminal activity.)

The total annual cost of the literacy hour was £2.5 million (about £2.8 million in 2001 prices), or £25.52 per pupil in the participating schools. Most of these expenditures went to establishing 14 local centers (each costing about £25,000 per year) and providing literacy consultants in each participating LEA (about £27,000 per year for each consultant). Schools also received some funding for teacher training and resources.

To estimate benefits of the policy, we converted our best estimate of the program’s impact on reading scores (2.63 percentiles) to its equivalent in standard deviation terms, calculated as 0.09. We then use data from the British Cohort Study, which regularly surveys all those living in Great Britain born in the United Kingdom between April 5 and 11 in 1970, to estimate the impact of an improvement in reading scores of this magnitude on future labor market earnings. And there we found the

We found that exposure to the literacy hour significantly improved students’ reading and English achievement, with bigger gains for boys than for girls.
difference in percentile reading scores was associated with additional earnings at age 30 of between £75.40 and £196.32 per year. The smallest estimate controls for differences in education attainment. Because education attainment is determined in part by one’s ability to read, this estimate almost certainly understates the true benefits associated with improved literacy skills.

From any perspective, the earnings effect of boosting age 10 reading scores is considerable—and the costs of the literacy hour minimal. Even if we take the smallest impact estimate from our analysis (a 1.72 percentile improvement in reading scores, which corresponds to a 0.06 standard deviation increase), the economic benefits measure in the range of £1,375 to £3,581 over the course of a recipient’s lifetime.

These cost-benefit calculations make the literacy hour an attractive alternative to several other popular policy proposals that have been subjected to rigorous analysis. While reducing class sizes and increasing teacher quality have also been estimated to increase student achievement by roughly 0.1 standard deviation, the costs of such programs far exceed those of the literacy hour program, which focuses only on changing teachers’ practices.

Spillover Effects
Finally, one might worry that the literacy hour takes teaching effort and resources away from other subjects and that this indirect cost effect (via substitution) should be taken into account 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. As mentioned, the literacy hour represents a change in how reading and writing are taught, rather than an increase in the time devoted to the subject.

One might instead suspect that the literacy hour could lead to positive spillovers due to complementarities between pupil subject areas and associated teacher practice. Reading and writing, after all, are important generic skills, and an improvement in these skills might lead to improved performance in other subjects. The literacy hour might also have caused teachers to reevaluate their teaching methods in other subjects and change their approach in those other subjects. This is especially important in English primary schools because generally pupils within a particular year group are taught every subject by the same teacher.

To examine the possible effects of the literacy hour on mathematics, we simply repeat our main analysis, but focus this time on the percentile mathematics score and whether the student obtains level 4 or above in mathematics. There is evidence of a positive effect, though it is about three-fifths that of the impact that we saw on English. Our strictest test of the program’s impact indicates that NLP schools show higher scores of 1.5 percentile points and a 2.5 higher percentage achieving at least level 4 in mathematics. These results suggest, if anything, a complementary impact of the literacy hour on English and mathematics.

Conclusions
Does a change in the content and structure of teaching affect pupil performance? Our study of England’s literacy hour suggests that it does. Student test scores in reading, English—even math—improved significantly in the schools that implemented this highly structured approach to the teaching of reading. One of the more interesting findings from our analysis of the NLP data was the effect the program had on the so-called gender gap: boys benefited more than girls from the literacy hour. Finally, we show that the long-term benefits of the literacy hour exceed its costs by a large margin.

These findings are of considerable significance in the wider education debate about what works best in schools for improving pupil performance—especially in countries that face urgent problems in basic literacy. They are also particularly notable since almost certainly the same teachers were teaching literacy before and after the introduction of the literacy hour. Indeed, the evidence we report strongly suggests that public policy focused on the content and organization of what is taught is a relatively desirable means of preventing illiteracy and its associated ills.

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