CHICAGO JOURNALS

Funding, School Specialization, and Test Scores: An Evaluation of the Specialist Schools Policy Using Matching Models
Author(s): Steve Bradley, Giuseppe Migali, and Jim Taylor
Source: *Journal of Human Capital*, Vol. 7, No. 1 (Spring 2013), pp. 76-106
Published by: The University of Chicago Press
Stable URL: http://www.jstor.org/stable/10.1086/669203
Accessed: 17/04/2013 12:31

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

The University of Chicago Press is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Human Capital.*

http://www.jstor.org

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

Funding, School Specialization, and Test Scores: An Evaluation of the Specialist Schools Policy Using Matching Models

Steve Bradley
Lancaster University

Giuseppe Migali
Lancaster University and Universita' Magna Graecia

Jim Taylor
Lancaster University

We evaluate the causal association between the specialist schools policy, a UK reform that has increased funding and encouraged secondary school specialization in particular subjects, and pupils' test score outcomes. Using the National Pupil Database, we estimate difference-in-difference matching models. We find a small, positive, and statistically significant causal effect on test scores at age 16. Pupils from poorer social backgrounds benefited more than pupils from richer backgrounds; pupils from ethnic minority backgrounds benefited less. We disentangle the funding effect from a specialization effect, which yields a relatively large proportionate improvement in test scores in particular subjects.

I. Introduction

In the United Kingdom a number of educational policy reforms have been introduced, such as the 1988 Education Reform Act, which led to the creation of a quasi market in education, at the heart of which is enhanced parental choice and competition between schools for pupils.

We would like to thank the Nuffield Foundation for its financial support; the University of Pennsylvania for providing Migali excellent research facilities; and Petra Todd, Ian Walker, Colin Green, Geraint Johnes, and seven anonymous referees for their useful comments. We would also like to thank the participants at the 2009 European Economic Association annual conference, the 2009 Scottish Economic Association annual conference, and seminar participants at the Institute of Education and at the Department of Economics of the University of Sheffield and Lancaster University.

[*Journal of Human Capital,* 2013, vol. 7, no. 1]
© 2013 by The University of Chicago. All rights reserved. 1932-8575/2013/0701-0001$10.00

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

However, funding for secondary schools has also been increased substantially since 1997, rising from £9.9 billion to £15.8 billion in 2006–7. Over the same period real expenditure per pupil increased by over 50 percent, from £3,206 to £4,836 (in 2005–6 prices). One of the key policy initiatives that has led, in part, to this increase in funding is the specialist schools policy, which was introduced in 1994. To obtain specialist status, state-maintained schools are required to raise sponsorship from the private sector of £50,000, which the school decides how to spend, and to have a development plan. Selected schools then received a capital grant of £100,000 from the central government and around £130 per pupil over a 4-year period.[1] This amounts to an approximate increase in funding per pupil of 5 percent. By 2010–11 over 95 percent of secondary schools were specialist, and although the policy initiative is no longer directly funded, the annual exchequer costs of the specialist schools initiative had grown to approximately £0.5 billion in 2010–11, according to the Department of Education.[2] In addition to increasing the funding of schools, the specialist schools policy also simultaneously enhanced parental choice of school and competition between schools for pupils because schools were encouraged to specialize in particular subjects stimulating greater "product differentiation."[3] The earliest specialist schools were technology schools, starting in 1994, now constituting approximately 20 percent of all schools, with significant proportions of schools focusing on arts, sports, and science. Other specialisms, such as business and math, were introduced more recently in 2002.[4]

The key objective of the specialist schools policy was to improve the test score performance of secondary school pupils in all subjects as a result of increased funding, but especially in the areas of specialization, because, for instance, schools would attract "better" teachers in their specialist subjects. Evaluating whether this policy has had the desired effect provides important guidance for policy makers contemplating whether, and how, to spend increasingly scarce resources on schooling. However, there are very few studies that focus on the evaluation of the specialist school policy, and the evidence from this literature is mixed (see Sec. II).

[1] The capital grant has been reduced to £25,000 in recent years but was higher for the time period covered by this study.

[2] See http://www.education.gov.uk/schools for more details.

[3] It is worth noting that although specialist schools are encouraged to focus on particular subjects, all schools are also required to deliver a national curriculum. Thus, most pupils will typically study around 10 subjects in their final 2 years of compulsory schooling between the ages of 14 and 16. They then sit for nationally recognized tests, the General Certificate of Secondary Education (GCSE), in each subject. The GCSE is a norm-based examination taken by almost all pupils, and the grades range from A* to G. Grades A*–C are considered acceptable for entry to university, together with the acquisition of advanced qualifications obtained 2 years later. Pupils of lower ability may also take General National Vocational Qualifications instead of GCSEs.

[4] See the Department for Children, Schools and Families website for more details: http://www.standards.dcsf.gov.uk/specialistschools.

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

Our paper makes two contributions. This is the first paper that we are aware of to evaluate the impact of the specialist school policy on test score outcomes combining matching methods at the pupil level with a difference-in-differences (DiD) analysis.[5] This approach enables us to deal with various forms of selection bias that we identify below, as well as the effect of unobserved pupil and school heterogeneity. Two measures of test score outcomes are analyzed, that is, the pupil's average test score across all subjects (*GCSEscore*) and a binary measure indicating whether a pupil obtained five or more passes at a high level (grades A*–C; *GCSEbin*). The latter is the focus of policy makers and is also a prerequisite for subsequent entry to university. The second contribution of this paper is that we provide an exploratory analysis of the relative importance of two different mechanisms by which the policy could affect test scores—*funding* and *specialization* effects. The increase in resources to specialist schools creates a funding effect whereby increased spending on books and equipment, for instance, improves the quality of the educational experience throughout the school and hence may improve test scores in all subjects. However, by allowing greater subject specialization, parents can select those schools that "match" the aptitudes and skills of their children, thereby increasing allocative efficiency.[6] "Better" subject specialist teachers may also move to schools that specialize in their subject area. Hence test scores in particular subjects may increase—a specialization effect.

There are likely to be several sources of bias in an evaluation of the specialist schools policy, arising primarily from selection on unobservables (so-called hidden bias), that must be mitigated. First, there is the nonrandom selection of schools into the program (hereafter *school selection bias*). A closely related source of bias is the nonrandom selection of pupils into specialist schools (*pupil selection bias*), insofar as unobservably more able pupils are "cream-skimmed" by "good" (specialist) schools. In principle, it would be helpful to disentangle these two sources of bias; however, this is not possible with our data. In the evaluation of the specialist schools policy that follows we therefore treat these two sources of bias as observationally equivalent.

Our approach is to exploit the availability of repeated data from the National Pupil Database (NPD), to which we append school-level data

[5] Previous studies using matching methods are mainly focused on the estimated effects of training programs on the unemployed. For example, Blundell et al. (2004) study the effects of the New Deal for Young People in the United Kingdom. Aakvik (2001) evaluates the Norwegian vocational rehabilitation program by comparing employment outcomes of participants and nonparticipants. DiPrete and Gangl (2004) analyze the impact of unemployment insurance on several outcomes such as postunemployment wage or probability of relocation. Machin, McNally, and Meghir (2004) adopt an approach similar to ours in evaluating an educational policy targeted at disadvantaged schools in urban areas known as the Excellence in Cities program.

[6] In the education economics literature, "allocative efficiency is getting the amount of education right. Productive efficiency is getting it at the least cost" (Hoxby 1996, 54).

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

from the annual School Performance Tables and the annual Schools' Census, and estimate DiD matching models following Heckman, Ichimura, and Todd (1997), Blundell and Dias (2002), Machin et al. (2004), and Smith and Todd (2005). Machin et al. do, in fact, find heterogeneous effects in their evaluation of the Excellence in Cities policy. We explore the possibility of heterogeneous effects with respect to the duration of time the school has been exposed to the specialist schools policy but also with respect to pupil socioeconomic background, gender, and ethnicity and with regard to the degree of competition in the local education market.

Our main finding is that the specialist schools policy has had a small, positive, and statistically significant causal effect on the test score outcomes of secondary school pupils in England. This is around 0.4–0.9 GCSE points and is a quantitatively small effect. There is little or no effect on test scores at the upper end of the attainment distribution. There is also little evidence that the duration of specialist school status matters, whereas pupils from poorer social backgrounds have benefited more than pupils from richer backgrounds; in contrast, pupils from ethnic minority backgrounds benefited less. Models that attempt to disentangle the funding effect from a specialization effect suggest that there is a specialization effect. This amounts to between 21 percent and 50 percent of the total effect depending on the matching estimator used.

The remainder of this paper is structured as follows. In Section II, we briefly review the literature, especially that which focuses on the specialist schools policy. In Section III, we explain the econometric approach. This is followed in Section IV by a discussion of our data and dependent variables, as well as how we select the treatment and comparison groups. Section V discusses the findings from the DiD matching models. Section VI draws some conclusions.

II. Literature

Gorard (2002), Jesson (2002), the Office for Standards in Education (2005), and Jesson and Crossley (2007) find a positive effect on test scores. Most of this early work uses school-level or pupil-level data to estimate cross-sectional ordinary least squares models. For instance, Jesson finds that specialist schools increased the percentage of pupils obtaining five or more GCSE grades A*–C by between 4.5 percentage points and 5 percentage points, and the total points score is increased by 4.2 percent. These are large effects, but no allowance is made for the different types of selection bias. Schagen and Goldstein (2002) do raise some issues regarding the methodological approach of this early work, arguing that pupil-level data and multilevel modeling techniques should be used. Estimated effects of the policy arising from this study are considerably smaller at between 0.02 and 0.11 of a GCSE point with some variation by subject: math (0.04–0.06), English (0.03–0.09), and science

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

(−0.03–0.07). Benton et al. (2003), again using multilevel modeling techniques on cross-sectional pupil-level data, find that the specialist schools policy raised GCSE grades, or points, by 1.1. The effect found by Levacic and Jenkins (2004) for 2001 was very similar at 1.4 GCSE points. In addition, these authors also found some variation by subject of specialization and the duration of the policy.

Taylor (2007) shows that the specialist schools policy has had very little impact on average test scores, though there is evidence of more substantial impacts for specific areas of specialization, for example, business and technology. Bradley and Taylor (2010) estimate the impact of the specialist schools policy, as well as other educational policies, using school-level panel data and find a small positive effect of specialist schools on test scores.

The existing literature fails to allow for the selection bias referred to in the introduction that often arises in program evaluation settings, which calls into question whether they have been able to identify a causal effect of the specialist schools policy. Furthermore, many of these studies do not explicitly consider the mechanisms by which the specialist schools policy could affect the test score outcomes of pupils.

There is a broader literature (see Hanushek [1998] for a survey) that focuses on the effectiveness of school resources, which is also relevant to this paper. For example, using US data on expenditure and National Assessment of Education Progress test scores, Krueger (1998) finds modest gains in test scores due to an increase in expenditures whereas Hanushek (1998) does not find a strong relationship between resources and student performance. More recent evidence can be found in Holmlund, McNally, and Viarengo (2010) and Lavy (2012). Holmlund et al. show that the change in funding formulas can generate an exogenous variation in school expenditure, which is exploited to estimate a causal effect of expenditure on pupils outcomes. They find a positive effect of school expenditure on national tests at the end of primary school in England, especially for pupils from disadvantaged socioeconomic backgrounds. Lavy finds for Israeli primary schools that increasing funding at the class level has positive and statistically significant effects on the average test scores of students. This effect is symmetric insofar as those schools that witnessed a reduction of resources observed a negative effect on average test scores. The positive effects of the funding change are larger for pupils from low socioeconomic backgrounds.

III. Econometric Approach: Matching Methods

Our approach is based on the concept of the education production function wherein test scores are a function of personal, family, and school inputs as well as specialist school status (Todd and Wolpin 2003). However, to estimate the effect of the specialist schools policy on the test scores of pupils requires a solution to the counterfactual question of

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

how pupils would have performed had they not attended a specialist school. We adopt the nonparametric matching method, which does not require an exclusion restriction or a particular specification of the model for the effect of the policy on test score outcomes. Thus, the main purpose of matching is to find a group of nontreated pupils who are similar to the treated in all relevant pretreatment characteristics, **x**, the only remaining difference being that one group attended a specialist school and another group did not. In the first stage we therefore estimate the propensity score using a discrete response (probit) model of attendance at a specialist school.

One assumption of the matching method is the *common support* or overlap condition, which ensures that pupils with the same **x** values have a positive probability of attending a specialist school. A second and key assumption is the *conditional independence assumption* (*CIA*), which implies that selection into treatment is based solely on observable characteristics.[7]

Given these two assumptions, the matching method allows us to estimate the average treatment effect on the treated (ATT). The ATT estimator is the mean difference in outcomes over the common support, weighted by the propensity score distribution of participants.

All matching estimators are weighted estimators, derived from the following general formula (see Blundell and Dias 2002):

$$\tau_{\mathrm{ATT}} = \sum_{i \in T} \left(Y_{1i} - \sum_{j \in C} W_{ij} Y_{0j} \right) w_i, \tag{1}$$

where T and C represent treatment and control groups, respectively; W_{ij} is the weight placed on the jth observation in constructing the counterfactual for the ith treated observation; Y_1 is the outcome of participants and Y_0 that of nonparticipants; and w_i is the reweighting that reconstructs the outcome distribution for the treated sample. A number of well-known matching estimators exist that differ in the way they construct the weights, W_{ij}. We use the nearest neighbor method, and we provide analytical standard errors (Abadie and Imbens 2008).

The main issue with cross-sectional matching analysis is that there may be a problem of hidden bias due to the effect of unobserved heterogeneity, and any positive association between a pupil's treatment status and test score outcomes may not therefore represent a causal effect. If the assumption of ignorability (i.e., no hidden bias) fails, the treatment is endogenous and the matching estimates will be biased (Heckman et al. 1998). To deal with the problem of unobserved heterogeneity, we perform our analysis using DiD models and showing the cross-sectional results only for comparative purposes. Following Heckman et al. (1997),

[7] Conditional on a set of pretreatment observable variables **x**, potential outcomes are independent of assignment to treatment.

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

Blundell and Dias (2002), Machin et al. (2004), and Smith and Todd (2005), the DiD approach allows for temporally invariant differences in outcomes between pupils in specialist and nonspecialist schools.

The DiD matching can be seen as an extension of simple matching because the bias is not required to vanish for any covariates but just to be the same before and after treatment. The DiD matching estimator for repeated cross-section data is given by

$$\tau_{\mathrm{ATT}}^{\mathrm{DiD}} = \sum_{i \in T_t} \left(Y_{1ti} - \sum_{j \in C_t} W_{ij} Y_{0tj} \right) w_{it} - \sum_{i \in T_{t'}} \left(Y_{0t'i} - \sum_{j \in C_{t'}} W_{ij} Y_{0t'j} \right) w_{it'}, \tag{2}$$

where t' and t are time periods before and after the acquisition of specialist school status. Specifically, $T_{t'}$ is formed by students in schools nonspecialist in t' that will be specialist in t; $C_{t'}$ is formed by students in schools nonspecialist in t' that will remain nonspecialist in t; T_t includes students in schools specialist in t that were nonspecialist in t'; and C_t includes students in schools nonspecialist in t that were also nonspecialist in t'.

In our analysis we consider only pupils from the treated and control groups that are on the common support. This implies that all the observations that are outside are automatically dropped. To recap, our dependent variables, Y, are GCSEscore and GCSEbin.

IV. Data and Dependent Variables

Before we describe the pupil-level data used in the econometric analysis, we present some evidence on the selection effects described in the introduction. Figure 1 disaggregates the average test score performance of schools into quintiles and plots the proportion of specialist schools in the lowest (quintile 1) and highest (quintile 5) categories. For example, in 2003 the proportion of specialist schools in the fifth quintile of test score performance is 60 percent, whereas it reduces to just 26 percent in the first quintile. What is immediately clear is that specialist schools are increasingly likely to have test scores in the highest quintile, which is strongly suggestive of nonrandom assignment of certain types of schools into the specialist schools initiative. Figure 2 provides some evidence of cream-skimming based on observable characteristics that are correlated with test score performance. Panel A of figure 2 shows that pupils from the poorest social backgrounds, as reflected by their eligibility for free school meals, are less likely to attend specialist schools. Specifically, specialist schools are more likely to have a higher percentage of pupils in the first quintile than in the fifth quintile, suggesting that pupils from poorer social backgrounds are less likely to attend specialist schools. In panel B we show the distribution of pupils from an ethnic minority background, where we plot percentage differences for the first and fifth quintiles between nonspecialist and specialist schools.

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

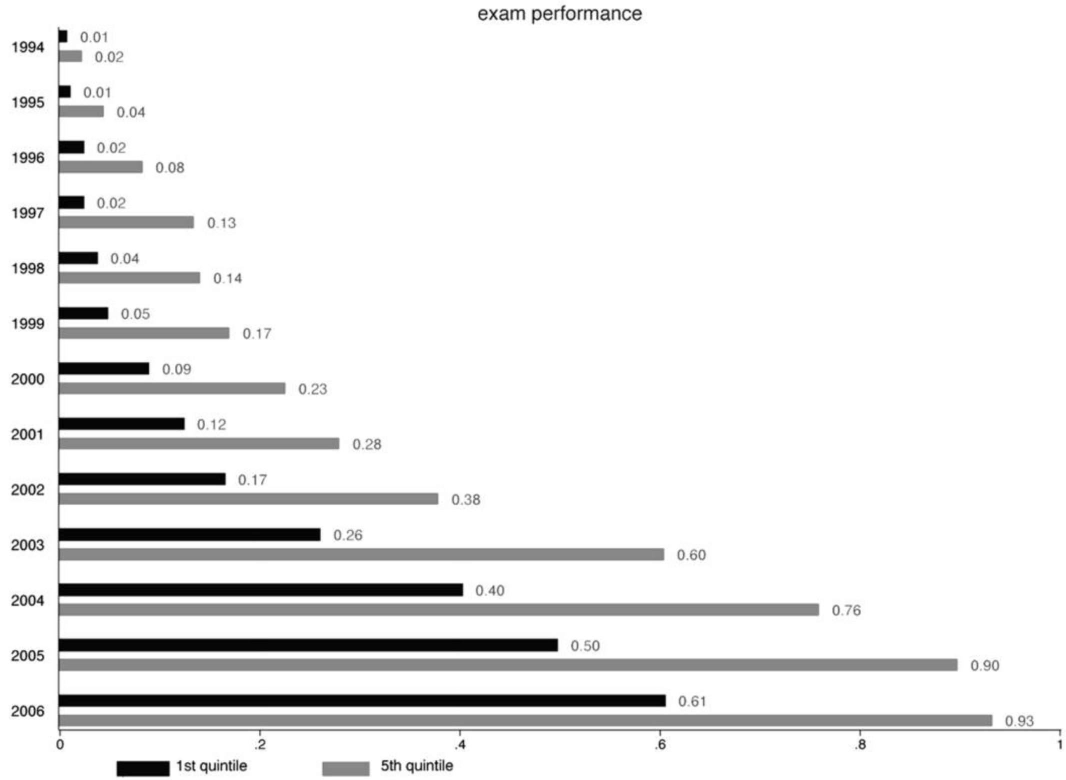

Figure 1.—Specialist schools exam performance. We consider quintiles of the distribution of school test scores performance.

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

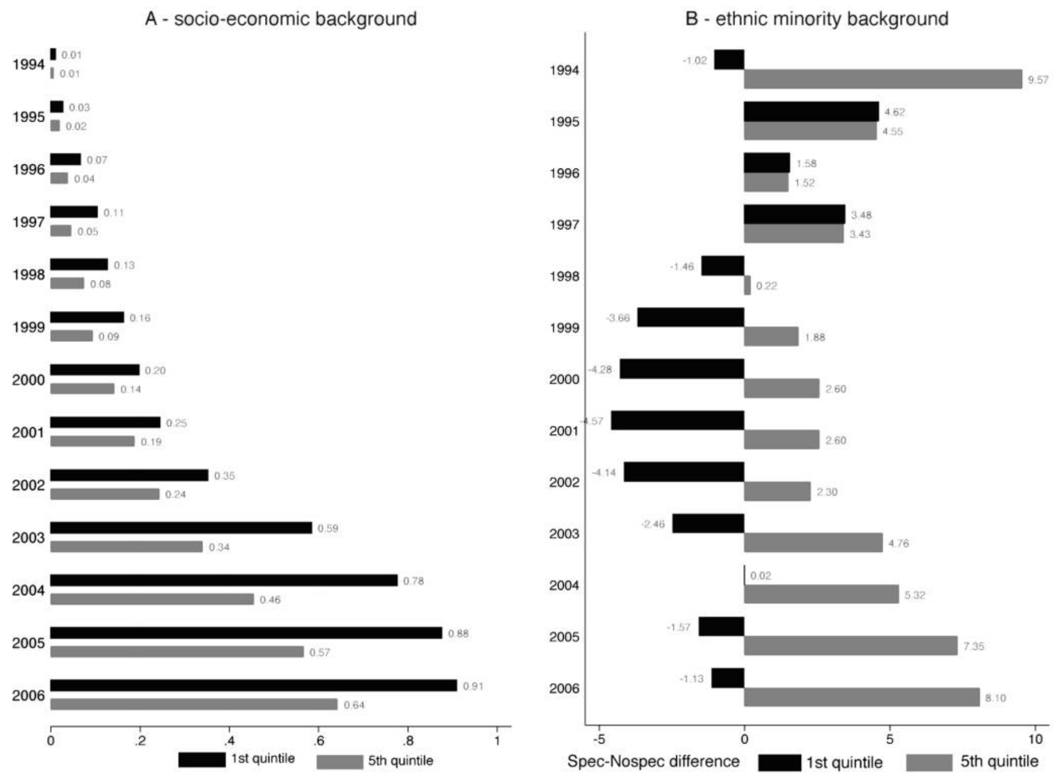

Figure 2.—Specialist schools pupil composition and cream-skimming. In the first graph we consider quintiles of the distribution of eligibility for free school meals and in the second graph quintiles of the distribution of ethnic composition.

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

Specialist schools are more likely to have a lower percentage of ethnic minority pupils. Figures 1 and 2 suggest that school selection bias and pupil selection bias may be very real.

We use data from the NPD. The NPD refers to the population of pupils attending maintained, state-funded, schools in England. The primary advantages of the NPD are that it refers to the population of pupils in secondary schooling, hence providing a large number of observations, and there are different measures of test scores. We consider several versions of the NPD in which pupils were in their final year of compulsory education in one of the years between 2002 and 2005. The original sample sizes, which arise by merging data sets containing key stage 2, 3, and 4 test scores, are 504,555, 523,658, 560,493, and 557,142 observations, respectively. We pool separately the NPD cohorts for 2002–3, 2002–4, and 2002–5. We choose these four cohorts of the NPD because this is the period in which many secondary schools acquired specialist school status and because before 2002 the information on key stage 4 test scores is not available. The percentage of pupils in specialist schools in 2002 was 30 percent, in 2003 it was 50 percent, and in 2004 it was 70 percent, and so on. This allows us to select representative treatment and control groups.

Our dependent variables are constructed from national test scores obtained by pupils at key stage 4 or GCSE tests, typically taken at the end of compulsory secondary schooling (i.e., at ages 15 and 16). The first of these variables refers to the total GCSE score and is the number of points achieved in all GCSE subjects, where grades are ranked from A* = 8 points to fail = 0. A one-point improvement would therefore be equivalent to a one-grade increase; however, this measure of test scores ignores the fact that it is more difficult to achieve at higher grades. Therefore, a second measure (GCSEbin) is a binary variable that takes the value of one if the pupil achieves five or more GCSE grades A*–C, zero otherwise. This level of achievement is seen as the minimum level of attainment necessary for entry to university, conditional also on achieving certain grades at A level at age 18.[8] Another important advantage of the NPD is that it also includes a measure of pupil attainment prior to entry into secondary schooling, that is, the key stage 2 tests.

Next we combine the pupil-level data from the NPD with school-level panel data from the annual School Performance Tables and the annual Schools' Census. We have a sample of 2,645 schools observed from 1994 to 2006, which includes variables on the characteristics of the school (e.g., average class size, type of school, etc.), as well as the year in which the school acquired specialist school status. For the analysis in 2002–3 and 2003–4, we first identify schools that were nonspecialist in

[8] When aggregated to the school level the proportion of pupils obtaining five or more GCSE grades A*–C is used to rank schools in league tables, which are available to parents to assist in school choice.

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

2002 but may become specialist in 2005 (348 schools) and 2006 (100).[9] For the 2002–5 analysis and to increase the sample size, our control group includes nonspecialist schools in 2002 that may become specialist in 2006 (100) together with schools that will never become specialist in our data (515).[10] Similarly, we identify schools that are also nonspecialist in 2002 but become specialist from 2003 onward.[11] In essence, the 2002 cohort is used as our pretreatment period whereas the 2003–5 cohorts are separately used as posttreatment periods.

There are several reasons why the full school sample may not deliver comparable treated and control groups. For instance, in figures 1 and 2 we showed that schools that have a higher proportion of pupils from poorer socioeconomic backgrounds are less likely to become specialist schools, whereas those schools with a higher proportion of pupils from ethnic backgrounds are more likely to become specialist. Therefore, to obtain more homogeneous comparison groups, we use a probit model to estimate the probability (propensity score) of a school becoming a participant in the specialist schools policy, conditional on pretreatment school characteristics, that is, the characteristics of the school in 1994, before the specialist school policy began.[12] To ensure that the treatment and control groups of schools are as alike as possible, we include in our analysis those schools that lie on the common support.[13] This implies that we are selecting only schools similar in pretreatment characteristics.[14] To assess the appropriateness of the common support condition, in figure 3 we show the distribution of the propensity score, separately for treated and untreated groups. We also report the region of common support and the number of schools included on it. Note from figure 3 that the propensity score of most observations lies on the intersection of the support of the treated and control groups.

The full and the restricted school samples are then separately merged to the NPD cohorts for 2002–3, 2002–4, and 2002–5. In this way, four categories of pupils can be identified: for example, for the 2002–3 cohort, (1) pupils in 2002 who attended nonspecialist schools that will remain so in 2003, (2) pupils in 2002 in schools that were nonspecialist but became specialist in 2003, (3) pupils in 2003 in schools that are nonspecialist in both 2002 and 2003, and (4) pupils in 2003 in schools

[9] This is important because it means that we use as a control group schools that are simply further down the "queue" for participation in the policy.

[10] As mentioned in the introduction, by 2010 over 95 percent of the 2,645 secondary schools had become specialist. Therefore, out of 515 schools that are observed in our data as nonspecialist, just over 100 had not acquired specialist status by the time the policy had been terminated in 2010–11.

[11] Precisely, 407 in 2003, 851 in 2004, and 1,199 in 2005.

[12] The covariates included in this model are the proportion of pupils eligible for free school meals, the proportion of non-UK pupils, the pupil/teacher ratio, school type (modern, comprehensive), school size, and an indicator for whether the school was single-sex (girls only).

[13] The sample size considered is that resulting from the probit estimation.

[14] For brevity, we do not report school-level descriptives, which are available on request.

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

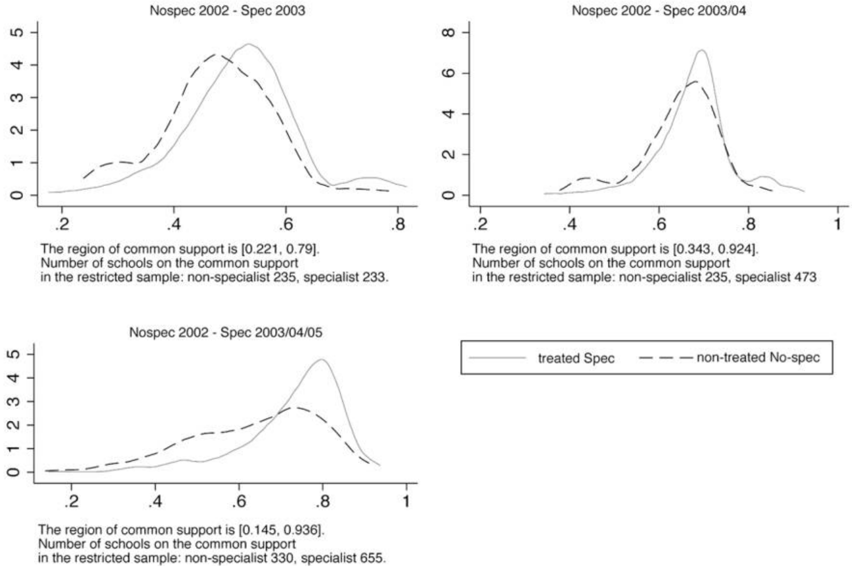

Figure 3.—Propensity score distributions for specialist and nonspecialist schools (school level).

that became specialist in 2003 but were nonspecialist in 2002. We repeat this approach for the 2002–4 and 2002–5 cohorts.

A. *Descriptives*

Table 1 provides some descriptive statistics, after combining school-level data with the pupil-level data. Table 1 suggests that there are roughly 50 percent fewer pupils in the restricted sample, which implies that we are excluding from our analysis those schools that are very different from the treated schools (pupils). It is also clear from table 1 that for both GCSEscore and GCSEbin, the raw DiD estimates are only slightly different than for the full sample, except for 2002–4 cohort comparisons. Pupil-level propensity score DiD models are estimated using both the full and the restricted samples since some readers may think that we have been overcautious in our attempts to get comparable treated and control groups.

Focusing on the findings for the restricted sample in table 1, we report the means and standard errors for GCSEscore and GCSEbin for each cohort by treated and control groups. For instance, it is clear that the treated groups of pupils in both 2002 and 2003 had higher average test scores by about 3.3 points; subtracting these differences, we derive the raw DiD of 0.09, which is highly statistically significant and suggests that the specialist schools policy had a small positive effect on average GCSE scores. The raw DiD increases for the 2002–4 cohort comparisons (to 0.25) and rises again to 0.71 for the 2002–5 comparisons. Note that for

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

TABLE 1
Descriptive Statistics for GCSEscore and GCSEbin by Cohort Comparison Groups

	GCSEscore				GCSEbin			
	Treated	Control	Difference	*N*	Treated	Control	Difference	*N*
A. 2002–3 Cohort								
Full sample:								
2003	43.671	40.212	3.459	138,772	.569	.503	.066	138,772
	(19.364)	(19.132)	(.011)		(.495)	(.500)	(.000)	
2002	44.287	40.972	3.315	125,541	.582	.516	.065	125,539
	(18.163)	(18.027)	(.010)		(.493)	(.500)	(.000)	
DiD			.144	264,313			.000	264,311
			(.011)				(.000)	
Restricted sample:								
2003	43.010	39.600	3.409	75,600	.556	.490	.066	75,600
	(19.321)	(19.046)	(.019)		(.497)	(.500)	(.000)	
2002	43.676	40.363	3.313	68,823	.568	.504	.064	68,821
	(18.111)	(18.027)	(.019)		(.495)	(.500)	(.000)	
DiD			.097	144,423			.002	144,421
			(.019)				(.000)	
B. 2002–4 Cohort								
Full sample:								
2004	44.166	41.499	2.667	216,779	.543	.497	.046	216,779
	(20.012)	(19.871)	(.008)		(.498)	(.500)	(.000)	
2002	43.585	40.972	2.613	191,780	.570	.516	.054	191,777
	(18.168)	(18.027)	(.008)		(.495)	(.500)	(.000)	
DiD			.054	408,559			−.008	408,556
			(.008)				(.000)	

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

Restricted sample:								
2004	43.755	40.804	2.951	116,304	.537	.484	.052	116,304
	(19.882)	(19.780)	(.015)		(.499)	(.500)	(.000)	
2002	43.065	40.363	2.702	103,021	.560	.504	.056	103,018
	(18.067)	(18.027)	(.014)		(.496)	(.500)	(.000)	
DiD			.249	219,325			−.003	219,322
			(.015)				(.000)	
	C. 2002–5 Cohort							
Full sample:								
2005	44.922	38.593	6.329	270,359	.546	.411	.134	270,359
	(20.205)	(20.528)	(.008)		(.498)	(.492)	(.000)	
2002	42.980	37.488	5.492	243,642	.557	.431	.125	243,638
	(18.161)	(18.458)	(.007)		(.497)	(.495)	(.000)	
DiD			.837	514,001			.009	513,997
			(.007)				(.000)	
Restricted sample:								
2005	44.458	38.333	6.125	146,038	.537	.405	.132	146,038
	(20.198)	(20.619)	(.014)		(.499)	(.491)	(.000)	
2002	42.448	37.030	5.418	131,653	.545	.420	.125	131,650
	(18.218)	(18.571)	(.013)		(.498)	(.494)	(.000)	
DiD			.707	277,691			.007	277,688
			(.014)				(.000)	

Note.—Standard errors are in parentheses.

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

these latter cohort comparisons some schools will have been specialist for more than 1 year, implying that there may well be heterogeneous policy effects. Turning to GCSEbin, the cross-sectional differences in the proportion of pupils obtaining five or more grades A*–C show that, on average, pupils in treated schools are 5–13 percentage points more likely to achieve at this level (again focusing on the restricted sample). However, there is very little difference between average outcomes for each cohort; hence the raw DiDs are very small.

B. *Individual Propensity Score Results*

For each of the cohort comparisons we use a probit model to estimate the probability of a pupil being enrolled in a specialist school, or not, conditional on pretreatment pupil characteristics and pretreatment school characteristics.[15] In the estimation of the propensity score models, the choice of the covariates to be included is an issue (Heckman et al. 1997; Bryson, Dorsett, and Purdon 2002). There is some discussion in the literature that emphasizes that a balancing test should be satisfied, hence reducing the influence of confounding variables (e.g., Dehejia and Wahba 1999; DiPrete and Gangl 2004). However, given the very large sample size that we deal with here, it is very difficult to pass this test. In this case it is important to include in the propensity score model all variables that are strong predictors of attendance at a specialist school and test score outcomes.

In table 2 we show the estimates from the propensity score models for the restricted sample only; the estimates for the full sample are very similar. In each model we use the same set of covariates, dropping only those that are irrelevant for the model estimated (e.g., we drop eligibility for free school meals when we estimate the effects of socioeconomic background). The estimates in table 2 tell a very consistent story: girls and pupils from poorer socioeconomic backgrounds are less likely to attend specialist schools whereas pupils from nonwhite ethnic backgrounds and those with higher prior attainment at the age of 11 are more likely to attend. Note that our measure of prior attainment (key stage 2 test score) captures the cumulative effect of the history of family, pupil, and school inputs that determined test scores up to age 11 (Todd and Wolpin 2003). In terms of school characteristics, larger schools and those with larger average class sizes are more likely to be a part of the specialist schools policy, whereas those pupils in more competitive local education markets are less likely to be in a specialist school, presumably because there is greater school choice in those areas. Secondary modern schools, which typically recruit less able students from poorer socioeconomic backgrounds, are less likely to be part of the specialist schools policy, which is consistent

[15] Note that because our pupil-level data relate to cohorts of pupils who entered secondary schooling between 1997 and 2002 (and completed between 2002 and 2005), we consider school characteristics in 1997.

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

TABLE 2

Estimated Effects for the Pupil-Level Propensity Score Models: Restricted Sample Only

	2002–3			2002–4			2002–5		
		% Bias			% Bias			% Bias	
Controls	Coefficient	Unmatched	Matched	Coefficient	Unmatched	Matched	Coefficient	Unmatched	Matched
Gender	−.038	−2.3	.4	−.032	−2.2	2.9	−.018	2.4	.9
	(.007)			(.006)			(.005)		
Ethnic minority	.079	2.1	5.7	.084	2.4	−3.4	−.066	−10.4	−4.1
	(.008)			(.007)			(.006)		
Free school meal	−.156	−11.2	−2.4	−.118	−5.3	−3.2	−.219	−19.2	−1.5
	(.010)			(.008)			(.007)		
Key stage 2	.052	9.8	2.1	.045	7	3.7	.085	20.5	.5
	(.003)			(.003)			(.003)		
Class size	.096	22.9	−5.4	.060	2.4	−.3	.095	27.3	−1.8
	(.003)			(.002)			(.002)		
School size	.046	26.2	0	.026	13.5	−3.1	.034	21.6	−3.1
	(.001)			(.001)			(.001)		
Comprehensive	−.425	−3.1	−.6	−.339	4.2	−1.8	−.068	−6.4	−5.9
	(.024)			(.019)			(.016)		
Modern	−.538	−4.1	−2.6	−.405	−16.3	0	.018	1.8	3.9
	(.032)			(.026)			(.023)		
Girls' school	−.036	.6	6.3	−.018	1.6	4.7	.233	7.6	3.7
	(.014)			(.012)			(.012)		
Constant	−1.423			−.331			−.943		
	(.052)			(.043)			(.038)		
Observations	144,423			219,554			277,750		
Log likelihood	−9.77e+04			−1.37e+05			−1.58e+05		

Note.—School variables are measured in 1997, before pupils' enrollment. Standard errors are in parentheses.

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

with the findings on the pupil-level covariates. Interestingly, girls-only schools have become more likely to participate in the specialist schools policy.

In sum, the covariates in the propensity score models work in the expected direction and confirm some of the descriptive evidence presented in the introduction regarding the sorting of schools and pupils into the specialist schools program. It is also worth noting in passing that there is a reduction in the standardized bias for almost all covariates, and in some cases this is quite substantial.[16] This is reassuring because it suggests that the matching procedure is able to balance characteristics in the treated and the matched comparison groups.

The other assumption underlying the propensity score approach is the common support condition. We examine the appropriateness of this condition by comparing the distributions of the propensity scores for the treated and the control groups for each of the cohorts studied. These distributions are shown in figure 4, for both the full and the restricted samples. It is clear that we lose very few observations at the boundaries of the common support, and the sample size is only slightly reduced.

V. Findings

A. *Estimation of Cross-Sectional Models*

For the purposes of comparison with the existing literature and with our DiD matching models, we estimate several "cross-sectional" models by simply looking at the estimated effect of the specialist schools policy separately for each cohort of pupils. Hence it is necessary to compare these findings with those from the DiD analysis. Table 3 shows the results of this analysis. It is clear that for the full sample the unmatched effects of the policy are quite large for both GCSEscore and GCSEbin and are consistent with the existing literature: the policy raises GCSE score by about 3–6 points, or grades, and the probability of obtaining five or more GCSEs grades A*–C by between 5 and 13 percentage points. After matching and for both the GCSEscore and GCSEbin, the estimated effects are roughly half the magnitude of the unmatched estimates. A very similar picture emerges for the restricted sample. Hence, if one was to end the analysis at this point, the conclusion would be that the specialist schools policy increased GCSEscore by about 2–4 points and GCSEbin by around 3–8 percentage points. These findings are consistent with the evidence from the existing literature and in some cases are slightly larger. However, the problem with this type of analysis is that it is subject to the biases

[16] This is computed before and after matching as $100(\bar{x}_{\text{nonsp}} - \bar{x}_{\text{spec}})/\sqrt{(s^2_{\text{nonsp}} + s^2_{\text{spec}})/2}$, where $\bar{x}_{\text{nonsp}}$ and $\bar{x}_{\text{spec}}$ are the covariate's sample mean for specialist and nonspecialist schools; s^2_{nonsp} and s^2_{spec} are the corresponding variances (Caliendo and Kopeining 2008).

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

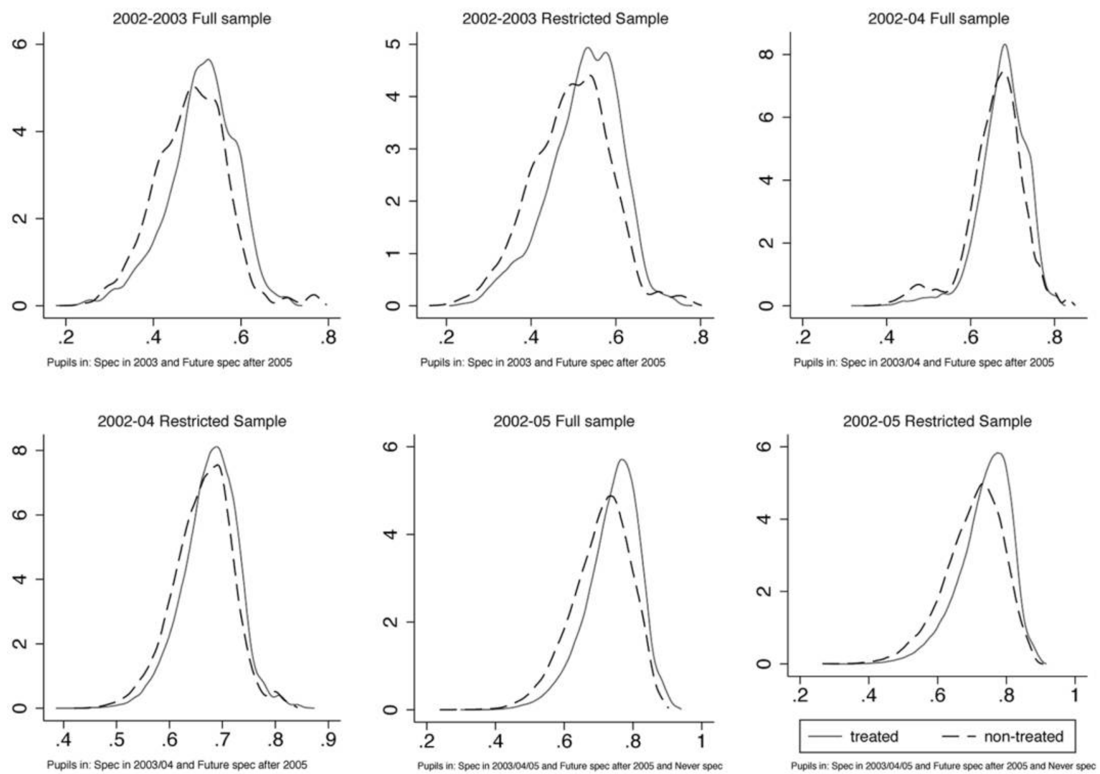

Figure 4.—Propensity score distributions for specialist and nonspecialist schools (individual level)

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

TABLE 3
THE ESTIMATED EFFECT OF THE SPECIALIST SCHOOLS POLICY: CROSS-SECTION MATCHING MODELS

	2003		2004		2005	
	GCSEscore	GCSEbin	GCSEscore	GCSEbin	GCSEscore	GCSEbin
Full sample:						
Unmatched	3.459	.066	2.667	.046	6.329	.134
	(.103)	(.003)	(.091)	(.002)	(.088)	(.002)
ATT	2.120	.036	1.503	.022	3.385	.070
	(.161)	(.004)	(.136)	(.003)	(.137)	(.003)
N (on common support)	138,772		146,185		270,504	
Restricted sample:						
Unmatched	3.409	.066	2.951	.052	6.125	.132
	(.140)	(.004)	(.124)	(.003)	(.120)	(.003)
ATT	2.221	.044	1.807	.030	3.606	.077
	(.214)	(.006)	(.182)	(.005)	(.185)	(.004)
N (on common support)	75,600		116,325		146,046	

Note.—Schools that specialized from 2002 are included. The number of observations off the common support in the full sample is 0 in 2003, 5 in 2004, and 924 in 2005. The number of observations off the common support in the restricted sample is 0 in 2003, 107 in 2004, and 4 in 2005. Standard errors are in parentheses.

outlined in the introduction and does not control for unobserved (school-level) heterogeneity.

B. *The Estimated Effect of the Specialist Schools Policy on Test Scores*

In table 4 we report the estimates of the DiD matching estimator (eq. [2]) for the full sample and for the restricted sample. The results for the full sample with respect to GCSEscore are sometimes positive and otherwise negative, and rarely statistically significant. However, a far more consistent picture emerges with respect to the restricted sample.[17] In this case the estimates are always positive and statistically significant and show evidence that the specialist schools policy has had a larger effect for later cohorts. For instance, for the 2002–3 cohorts the policy increased GCSE scores by around 0.4 of a GCSE point, whereas for the 2002–5 cohorts the effect had more than doubled to 0.9 of a GCSE point. In contrast, there is little systematic evidence that the policy has had any effect on the probability of a pupil obtaining five or more GCSE grades A*–C. In fact, the only positive and statistically significant effect arises in the 2002–5 cohort analysis, where the policy is shown to raise this probability by about 1.5 percentage points. The specialist schools policy has therefore had almost no effect at the higher end of the attainment distribution, a finding that contrasts sharply with the aggregate school-level analyses reviewed in Section II. This

[17] Note that the standard errors in the restricted sample should be considered as indicative because they do not take into account the fact that we also estimate a school-level propensity score model as a precursor to the pupil-level models.

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

TABLE 4
THE ESTIMATED EFFECT OF THE SPECIALIST SCHOOLS POLICY: DIFFERENCE-IN-DIFFERENCES MATCHING MODELS

	2002–3		2002–4		2002–5	
	GCSEscore	GCSEbin	GCSEscore	GCSEbin	GCSEscore	GCSEbin
Full sample:						
ATT	−.275	−.008	−.002	−.012	.664	.009
	(.159)	(.004)	(.135)	(.004)	(.132)	(.003)
N (on common support)	264,313		408,586		515,792	
Restricted sample:						
ATT	.355	.008	.637	.006	.892	.015
	(.212)	(.006)	(.136)	(.004)	(.185)	(.005)
N (on common support)	144,423		219,554		277,750	

Note.—The number of observations off the common support in the full sample is 0 in 2003, 27 in 2004, and 1,791 in 2005. The number of observations off the common support in the restricted sample is 0 in 2003, 229 in 2004, and 59 in 2005. Standard errors are in parentheses.

analysis therefore suggests that there was a small positive, causal, impact of the specialist schools policy on GCSE test scores at age 16.

We also investigated whether there was evidence of heterogeneity in the effects of the policy. One way of thinking about this is whether the effect of the policy varies with respect to the duration of specialist school status. Do pupils in schools that have been specialist for longer have better test scores? This might arise because it takes time for the school to fully implement the policy, and hence there is a disruption within the school during the transition from nonspecialist to specialist status.[18] For the pupils in the 2002–3 cohorts, schools that are a part of the policy have been specialist for only 1 year, whereas for the 2002–4 cohorts, this increases to 2 years; similarly, for the 2002–5 cohorts, schools have been specialist for 3 years. Table 5 shows our results. There is limited evidence that schools that have been specialist for longer have better test scores: when we compare the 2002–4 cohorts, those schools that have been specialist for 2 years have twice the effect (0.6 of a GCSE point) of those that have been specialist for 1 year (0.3). However, for the 2002–5 cohorts, the effect of the 2-year group falls back to 0.3 of a GCSE point, and this effect is more or less constant for those schools that have been specialist for 3 years. One can therefore conclude that the policy effects observed above are more like "one-off" effects: there is little or no evidence that the duration of the policy matters. This finding contrasts again with the existing literature on specialist schools and with that on other educational policies, such as the Excellence in Cities initiative, where Machin et al. (2004) do observe some effects of policy duration.

[18] Note that specialist school status will begin in September, the start of the school year, with inevitable preparations for this change prior to that date, whereas students sit for their GCSE exams in June of the following year.

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

TABLE 5
The Estimated Effects of the Specialist Schools Policy by the Duration of Specialization

	2002–3	2002–4		2002–5		
	1 Year	1 Year	2 Years	1 Year	2 Years	3 Years
	A. GCSEscore					
Full sample:						
ATT	−.275	−.093	−.113	.510	.937	.318
	(.212)	(.155)	(.16)	(.172)	(.157)	(.163)
N (on common support)	264,313	275,017	266,654	272,117	307,191	300,254
Restricted sample:						
ATT	.355	.317	.637	.223	.334	.345
	(.212)	(.208)	(.219)	(.226)	(.213)	(.223)
N (on common support)	144,423	142,093	145,544	144,798	160,734	166,053
	B. GCSEbin					
Full sample:						
ATT	−.008	−.009	−.014	.020	.011	−.002
	(.004)	(.004)	(.004)	(.004)	(.004)	(.004)
N (on common support)	264,313	275,017	266,654	272,117	307,191	300,254
Restricted sample:						
ATT	.008	−.004	.007	.026	−.003	.008
	(.006)	(.005)	(.006)	(.006)	(.005)	(.006)
N (on common support)	144,423	142,093	145,544	144,798	160,734	166,053

Note.—Standard errors are in parentheses.

In Section IV, we showed that specialist schools and nonspecialist schools differ with respect to the socioeconomic composition of the school. One measure of socioeconomic composition of the school is the proportion of pupils eligible for free school meals. Using school-level data, we first stratify the schools into quartiles of the distribution of eligibility for free school meals. We then estimate propensity score models at the pupil level for each quartile to see if the policy effects vary between "wealthy" pupils and "disadvantaged" pupils. Table 6 shows that for the earlier cohorts (2002–3) there is some evidence that the policy favored wealthy pupils, raising GCSE scores by around half of a GCSE point, whereas this effect reverses for later cohorts (2002–5). Here the policy raises the performance of more disadvantaged pupils by between 0.7 and 1.8 GCSE points. Thus, as more and more schools joined the specialist schools initiative and later arrivals tended to be worse-performing schools, our evidence suggests that it had a much greater effect for more disadvantaged pupils. This is consistent with other evidence and implies that the focus of the policy should have been on poorer-performing schools rather than on high performers as was the case in practice (Bradley and Taylor 2010).

Furthermore, we also present additional evidence in table 6 of the effects of the specialist schools policy on pupils from ethnic minority backgrounds, nonwhites, which can be compared to the effects of the policy for white pupils. For the full sample there is evidence that the

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

policy reduced the test scores of ethnic minority pupils when compared to their counterparts who went to nonspecialist schools (except for the 2002–3 cohort); however, this evidence is less robust for the restricted sample. However, there is some evidence of a differential effect of the policy for the 2002–5 comparison groups. For the white pupils the findings suggest that those who attended a specialist school achieved higher average GCSE scores (about 1.3 points) and were 3 percentage points more likely to achieve five or more grades in the A*–C range, that is, when compared to white pupils in nonspecialist schools. The corresponding results for the nonwhite groups suggest a half-point reduction in average GCSE scores and a 2 percentage point reduction in the probability of achieving high grades. Taken together this evidence implies that white pupils have benefited more from the policy than nonwhite pupils from ethnic minority backgrounds. In contrast, there is very little evidence of any gender differences in the effect of the policy.

A further issue that we address is whether the effect of the policy varies with the intensity of competition in the local education market. The 1988 Education Reform Act created a quasi market in secondary education such that schools competed for pupils and resources followed those pupils (Bradley et al. 2000). It is possible that pupils in more competitive education markets faced greater school choice, and hence more able pupils may have been less likely to sort into the typically better-performing specialist schools. One might therefore expect that the effect of the policy should be smaller in more competitive markets. To assess this we compute the Herfindahl index for each local education district, split this into quartiles, and then reestimate our models.[19] Table 7 shows the results of our analysis. No consistent pattern emerges; however, there is evidence for the 2002–5 cohorts that the policy effect is lower for pupils in schools in more competitive markets. This evidence is at best fragmentary, and taken as whole, these findings suggest that our methodology has been successful in removing the potential hidden bias from the nonrandom assignment of pupils to specialist and nonspecialist schools.

C. The Relative Importance of the Funding and Specialization Effects

So far we have considered the total impact of the specialist schools policy by simply looking at the test score outcomes in all subjects for pupils in specialist schools compared to various control groups. In this subsection we construct a test to try to disentangle the funding and specialization effects of the specialist schools policy using the NPD. We do this for the 2003 cohort only because the percentage of specialist and nonspecialist

[19] The Herfindahl is an index of the extent to which pupils are concentrated in a school's district. The index is the sum over all schools in a district of s_i^2, where s_i is each school's share of the district's pupils.

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

TABLE 6

The Estimated Effect of the Specialist Schools Policy by Socioeconomic Status, Gender, and Ethnic Background

	2002–3		2002–4		2002–5	
	GCSEscore	GCSEbin	GCSEscore	GCSEbin	GCSEscore	GCSEbin
A. Socioeconomic Status						
1st quartile: wealthy:						
ATT	.591	.013	.007	−.002	.277	.011
	(.145)	(.004)	(.281)	(.007)	(.608)	(.015)
N (on common support)	33,181		49,253		54,350	
2nd quartile:						
ATT	.249	.003	3.371	.009	.100	−.004
	(.108)	(.003)	(.840)	(.023)	(.391)	(.011)
N (on common support)	34,057		57,765		65,000	
3rd quartile:						
ATT	−.360	−.008	−.416	−.021	1.813	.032
	(.109)	(.003)	(.272)	(.007)	(.316)	(.008)
N (on common support)	41,023		61,555		78,852	
4th quartile: disadvantaged:						
ATT	−.983	−.014	−.935	−.019	.697	.000
	(.121)	(.003)	(.412)	(.010)	(.305)	(.008)
N (on common support)	35,527		51,281		80,114	
B. Gender						
Full sample:						
Boys:						
ATT	−.046	−.001	.172	−.009	.521	.006
	(.224)	(.006)	(.191)	(.005)	(.182)	(.005)
N (on common support)	132,951		205,347		259,871	
Girls:						
ATT	−.561	−.021	.064	−.009	.615	.007
	(.223)	(.006)	(.189)	(.005)	(.187)	(.005)
N (on common support)	131,362		203,239		255,921	

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

Restricted sample:						
Boys:						
ATT	.365	.003	.225	−.008	.315	−.001
	(.297)	(.008)	(.251)	(.007)	(.243)	(.006)
N (on common support)	73,210		111,090		142,528	
Girls:						
ATT	−.301	−.012	.307	−.004	.769	.013
	(.309)	(.008)	(.258)	(.007)	(.256)	(.006)
N (on common support)	71,213		108,464		135,222	
	C. Ethnicity					
Full sample:						
White:						
ATT	−.346	−.009	.322	−.005	1.156	.021
	(.180)	(.005)	(.149)	(.004)	(.146)	(.004)
N (on common support)	209,754		341,254		428,171	
Ethnic minority:						
ATT	1.404	.026	−.964	−.025	−.172	−.021
	(.341)	(.009)	(.326)	(.009)	(.299)	(.008)
N (on common support)	54,559		67,332		87,621	
Restricted sample:						
White						
ATT	−.224	−.003	.307	−.003	1.277	.027
	(.245)	(.007)	(.206)	(.005)	(.215)	(.006)
N (on common support)	109,065		172,489		216,232	
Ethnic minority:						
ATT	1.315	.013	.247	.000	−.576	−.017
	(.433)	(.012)	(.381)	(.010)	(.340)	(.009)
N (on common support)	35,358		47,065		61,518	

Note.—Standard errors are in parentheses.

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

TABLE 7

The Variation in the Effect of the Specialist Schools Policy by the Level of Competition in the Local Education Market

	2002–3		2002–4		2002–5	
	GCSEscore	GCSEbin	GCSEscore	GCSEbin	GCSEscore	GCSEbin
1st quartile: high:						
ATT	.685	.011	.289	.002	1.073	.017
	(.323)	(.008)	(.268)	(.006)	(.259)	(.007)
N (on common support)	67,279		100,177		127,909	
2nd quartile:						
ATT	−1.148	−.030	.519	−.001	−.679	.021
	(.414)	(.011)	(.363)	(.009)	(.378)	(.009)
N (on common support)	33,710		46,425		59,896	
3rd quartile:						
ATT	4.673	.081	2.134	.027	1.635	.028
	(1.122)	(.029)	(.593)	(.015)	(.488)	(.012)
N (on common support)	20,367		32,439		42,311	
4th quartile: low:						
ATT	−.687	.011	.712	.022	2.224	.026
	(.580)	(.015)	(.428)	(.011)	(.477)	(.012)
N (on common support)	21,857		38,163		47,634	

Note.—Standard errors are in parentheses.

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

schools is similar and we have more homogeneous treatment and control groups. If we use a data set with a high percentage of specialist schools, we risk having in the comparison group unusual nonspecialist schools. The 2003 cohort also provides sufficient observations in the treated and control groups when disaggregated by subject of specialization. This analysis is likely to be indicative since we cannot remove the different forms of bias referred to in the introduction. To assess the potential effects of unobserved heterogeneity, we introduce a confounder variable into our estimation.[20]

We focus on test score differences solely for the subjects in which the schools specialized. We restrict our analysis to schools that had become specialist in one of the following subject areas: languages, with and without English, and technology, which include the majority of pupils.[21]

To disentangle the specialization effect from the funding effect, we compare the estimates from panel A with those from panel B in table 8. Panel A compares the test score outcome in, say, languages of pupils in a specialist school that specializes in that particular subject (the treatment group) with the test score outcome of pupils in languages in specialist schools that do not specialize in that subject (the control group). Since both schools are specialist, they receive the same funding, and so the funding effect is constant. In contrast, panel B compares our treatment group with a different control group: pupils' test scores in languages in nonspecialist schools. Since the latter do not receive extra funding, any difference in test score outcomes in panel B must arise from both the funding and specialization effects. The difference in the estimates from panel A and panel B gives the specialization effect.

Panels A and B of table 8 show that after matching, GCSEscores fall substantially, and they are robust to a confounder variable. However, what is of most interest is the fact that the estimates from panel B are higher than those for panel A, and the magnitude of this difference

[20] In the cross-section data it is important to assess the CIA assumption. It is not directly testable because the data are uninformative about the distribution of $Y_i(0)$ for the treated and of $Y_i(1)$ for the control group. However, we use an approach developed by Imbens (2004), who proposes an indirect way of assessing the CIA. This is based on the estimation of a "pseudo" confounding factor that should, if the CIA holds, have zero effect. We adopt the method proposed by Ichino, Mealli, and Nannicini (2008). This is based on the prediction of a confounding factor, A, by simulating its distribution for each treated and control unit. Then, estimates of the ATT are derived by including the confounding factor in the set of matching variables. Different assumptions on the distribution of A imply different possible scenarios of deviation from the CIA. For simplicity, let A be a binary variable. Then its distribution is given by fixing the following parameters:

$$P(A = 1 | T = i,\ Y = j) = p_{ij}, \quad i, j = 0,\ 1,$$

where Y is a binary test score outcome (e.g., high ability = 1 and low ability = 0) and T is the pupil's treatment status. In this way, we can define the probability of $A = 1$ in each of the four groups identified by the treatment and the outcome.

[21] We tried to include more subjects, but we did not have enough observations to perform a matching analysis.

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

TABLE 8
The Relative Importance of Funding and School Specialization on GCSE Score

	English	Technology	Languages
	A. Pupils in Schools Specializing in Subject *m* vs. Pupils in Schools Specializing in Subject *n*		
Unmatched	.383	.237	.180
	(.045)	(.047)	(.045)
Nearest neighbor	.181	.181	.144
	(.034)	(.037)	(.029)
Nearest neighbor with confounder	.128	.135	.083
	(.042)	(.051)	(.037)
$\text{Kernel}_{0.1}$	.180	.161	.101
	(.045)	(.058)	(.042)
	B. Pupils in Schools Specializing in Subject *m* vs. Pupils in Nonspecialist Schools Taking Subject *m*		
Unmatched	.432	.460	.334
	(.039)	(.033)	(.049)
Nearest neighbor	.226	.231	.180
	(.032)	(.034)	(.035)
Nearest neighbor with confounder	.215	.214	.101
	(.037)	(.041)	(.045)
$\text{Kernel}_{0.1}$	.288	.318	.146
	(.050)	(.051)	(.042)

Note.—The balancing property is satisfied, and we consider only observations on the common support. Standard errors are in parentheses. Analytical standard errors are used with the nearest neighbor method; bootstrapped standard errors (500 repetitions) are used with kernel (bandwidth 0.1). The confounder is generated using the key stage 2 test score distribution. The sample sizes are the following: English literature, 48,804 observations in panel A and 166,321 in panel B; technology, 89,062 observations in panel A and 195,219 in panel B; foreign languages, 55,762 observations in panel A and 155,138 in panel B.

depends on the matching estimator. For instance, compare the nearest neighbor estimates for technology of 0.23 (panel B) and 0.18 (panel A) with the equivalent for the kernel matching method, that is, 0.32 and 0.16, respectively. The difference between the estimates in panels A and B for the nearest neighbor method is roughly 0.05 for all subjects, which implies that the specialization effect constitutes around 22 percent of the total effect of the specialist schools policy. For the kernel method the implied percentage contribution of the specialization effect varies from 38 percent for English to 50 percent for technology. Thus, although the actual magnitude of the impact of the specialist schools policy on test scores in the subjects analyzed is modest, when compared to the findings in earlier sections, the contribution of the specialization effect is quite large when measured in percentage terms.

A word of caution is necessary, however, insofar as the specialization effect might be picking up the fact that "good" specialist schools with the brightest pupils, typically the early entrants to the specialist schools initiative, simply applied for specialist status in their strongest subject. If this view is correct, then we should observe that a specialist school's

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

performance in tests in subjects other than that in which it specializes is higher when compared to that of a nonspecialist school. To assess whether this is the case, we examine the raw data and make pairwise comparisons between specialist schools in specialisms m (English, technology, or languages) versus nonspecialist schools for subjects n, where $m \neq n$.[22] For schools specializing in English, the raw data suggest superior performance in all n subjects by around 0.3 of a GCSE point; for language and technology schools, the differential falls to 0.1–0.2 of a GCSE point. This evidence implies that specialist school pupils perform well in all subjects compared to pupils in nonspecialist schools. However, it is important to note that these findings are drawn from an analysis of the raw data, and until further data become available to perform a more rigorous investigation, we conclude that there is some evidence that a specialization effect exists.

VI. Conclusions

In this paper we evaluate whether there is a causal association between the specialist schools policy, which can be regarded as a significant UK education policy beginning in 1994, and the test score outcomes of secondary school pupils in England. Our approach has been to use matching methods combined with a difference-in-differences analysis, which have become popular in the context of program evaluation, especially with respect to the effectiveness of training schemes and programs for the unemployed. We use several versions of the National Pupil Database. To our knowledge there has been no previous attempt to apply such methods to an evaluation of the specialist schools policy. By adopting this approach we explicitly confronted the twin problems of the choice of suitable control groups to answer the counterfactual question of what would have happened in the absence of treatment and the potential bias arising from a correlation between the treatment status and observed and unobserved covariates.

We find a small positive, causal, impact of the specialist schools policy on GCSE test scores at age 16. We do not find strong evidence that the duration of the policy matters. However, we do find that white pupils and those from a disadvantaged background have benefited more from the policy than nonwhite pupils and those from wealthy backgrounds. The findings with respect to pupils from disadvantaged backgrounds are consistent with evidence cited in the literature review on the effects of changes in funding on pupil test scores. There is very little evidence of any gender differences in the effect of the policy. Finally, the impact of the specialist schools policy on test scores could arise from a funding

[22] The n subjects refer to math, science, history, and business studies; we also examine languages, English, and technology, but only where $m \neq n$.

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

effect or a specialization effect. We attempted to disentangle these effects. Our findings for GCSE points scores suggested that between 21 and 50 percent of the total effect in particular subjects arises from the specialization effect.

These findings pose the question as to whether the specialist schools policy was cost effective. It is beyond the scope of this paper to undertake a full cost-benefit analysis. However, we note from the introduction that the exchequer costs of the policy were quite substantial, and these costs understate the indirect costs associated with the implementation and operation of the policy at the school level. These could also have been quite substantial involving shifts of emphasis in the curriculum and hiring subject specialist teachers, for instance. On the benefit side of the equation, one of the direct benefits of the policy was to raise pupil test scores. However, as we have shown, these effects are both quantitatively small and one-off. Furthermore, in other work we have estimated the effect of the specialist schools policy on the probability of pupils finding employment following school and their wages (Bradley and Migali 2012). We found no evidence of a direct effect of the policy on the employment outcomes of pupils from specialist schools versus their counterparts from nonspecialist schools. Nor did we find evidence of a wage premium for those pupils who had attended a specialist school.

We conclude that the specialist schools policy was an ineffective and costly means of trying to raise pupil test scores. Furthermore, it is interesting to note that the policy has now been scrapped, and although the funding has been retained, the specialist schools policy has rightly been relegated to the scrap heap of failed experiments in British secondary education policy making.

References

Aakvik, A. 2001. "Bounding a Matching Estimator: The Case of a Norwegian Training Program." *Oxford Bull. Econ. and Statis.* 63 (1): 115–43.

Abadie, A., and G. W. Imbens. 2008. "Notes and Comments on the Failure of the Bootstrap for Matching Estimators." *Econometrica* 76 (6): 1537–57.

Benton, T., D. Hutchinson, I. Schagen, and E. Scott. 2003. "Study of the Performance of Maintained Secondary Schools in England." Report for National Audit Office, Nat. Found. Educ. Res., London.

Blundell, R., and M. Dias. 2002. "Alternative Approaches to Evaluation in Empirical Microeconomics." *Portuguese J. Econ.* 1:91–115.

Blundell, R., M. Dias, C. Meghir, and J. van Reenen. 2004. "Evaluating the Employment Impacts of a Mandatory Job Search Program." *J. European Econ. Assoc.* 2:569–606.

Bradley, S., R. Crouchley, J. Millington, and J. Taylor. 2000. "Testing for Quasi Market Forces in Secondary Education." *Oxford Bull. Econ. and Statis.* 62:357–90.

Bradley, S., and G. Migali. 2012. "The Direct and Indirect Effects of Education Policy on School and Post School Outcomes." Working paper, Management School, Lancaster Univ., http://www.lums.lancs.ac.uk/economics/profiles/giuseppe-migali/.

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

Bradley, S., and J. Taylor. 2010. "Diversity, Choice and the Quasi-Market: An Empirical Analysis of Secondary Education Policy in England." *Oxford Bull. Econ. and Statis.* 72 (1): 1–26.

Bryson, A., R. Dorsett, and S. Purdon. 2002. "The Use of Propensity Score Matching in the Evaluation of Active Labour Market Policies." Working Paper no. 4, Dept. Work and Pensions, London.

Caliendo, M., and S. Kopeining. 2008. "Some Practical Guidance for the Implementation of Propensity Score Matching." *J. Econ. Surveys* 22 (1): 31–72.

Dehejia, R. H., and S. Wahba. 1999. "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs." *J. American Statis. Assoc.* 94 (448): 1053–62.

DiPrete, T., and M. Gangl. 2004. "Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments." *Sociological Methodology* 34:271–310.

Gorard, S. 2002. "'Let's Keep It Simple': The Multi-levelling Model Debate." *Res. Intelligence* (81): 24–25.

Hanushek, E. A. 1998. "Conclusion and Controversies about the Effectiveness of School Resources." *Fed. Reserve Bank New York Econ. Policy Rev.* 4 (March): 11–28.

Heckman, J., H. Ichimura, J. Smith, and P. Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66 (5): 1017–98.

Heckman, J., H. Ichimura, and P. Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *Rev. Econ. Studies* 64:605–54.

Holmlund, H., S. McNally, and M. Viarengo. 2010. "Does Money Matter for Schools?" *Econ. Educ. Rev.* 29 (6): 1154–64.

Hoxby, C. M. 1996. "Are Efficiency and Equity in School Finance Substitutes or Complements?" *J. Econ. Perspectives* 10 (Fall): 51–72.

Ichino, A., F. Mealli, and T. Nannicini. 2008. "From Temporary Help Jobs to Permanent Employment: What Can We Learn from Matching Estimators and Their Sensitivity?" *J. Appl. Econometrics* 23:305–27.

Imbens, G. 2004. "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Survey." *Rev. Econ. and Statis.* 86:4–30.

Jesson, D. 2002. "Response to Schagen and Goldestein." *Res. Intelligence* (80): 16–17.

Jesson, D., and D. Crossley. 2007. *Educational Outcomes and Value Added by Specialist Schools: 2006 Analysis.* London: Specialist Schools and Academies Trust.

Krueger, A. 1998. "Reassessing the View That American Schools Are Broken." *Fed. Reserve Bank New York Econ. Policy Rev.* 4 (March): 29–46.

Lavy, V. 2012. "Expanding School Resources and Increasing Time on Task: Effects of a Policy Experiment in Israel on Student Academic Behaviour." Working Paper no. 18369, NBER, Cambridge, MA.

Levacic, R., and A. Jenkins. 2004. "Evaluating the Effectiveness of Specialist Schools in England: Rhetoric and Reality." Discussion paper, Centre Econ. Educ., London.

Machin, S., S. McNally, and C. Meghir. 2004. "Improving Pupil Performance in English Secondary Schools: Excellence in Cities." *J. European Econ. Assoc.* 2:396–405.

Office for Standards in Education. 2005. "Specialist Schools: A Second Evaluation." Ref. HMI 2362, OFSTED, London.

Schagen, I., and H. Goldstein. 2002. "Do Specialist Schools Add Value? Some Methodological Problems." *Res. Intelligence* (80): 12–15.

Smith, J., and P. Todd. 2005. "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?" *J. Econometrics* 125 (1–2): 305–53.

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions

Taylor, J. 2007. "Estimating the Impact of the Specialist Schools Programme on Secondary School Examination Results in England." *Oxford Bull. Econ. and Statis.* 69:445–71.

Todd, P. E., and K. I. Wolpin. 2003. "On the Specification and Estimation of the Production Function for Cognitive Achievement." *Econ. J.* 113 (485): F3–F33.

This content downloaded from 129.119.32.27 on Wed, 17 Apr 2013 12:31:47 PM
All use subject to JSTOR Terms and Conditions