Do achievement labels affect students' well-being? Evidence from discontinuities in test scores in England*

Marcello Sartarelli
Institute of Education

October 17, 2009

Abstract

In this paper I estimate the effect of using achievement labels to grade school tests, e.g. bad or good, on a measure of well-being given by the event of a police contact or visit to parents due to the behaviour of children in secondary schools. Firstly I illustrate the potential spurious correlation that arises from unobservables affecting both achievement and well-being in a reduced form model of returns to education. Then I give causal interpretation to the estimates by employing a research design that exploits discontinuities in test scores. I find that a jump from a low to a high achievement level decreases the probability of a police contact by 4 percentage points. Under the identifying assumptions OLS estimates are upward biased. An increasing concern by policy-makers for the role of education in shaping well-being suggests that the mechanism through which schools contribute to teaching core subjects, e.g. Maths and norms helping children to lead a healthy and safe life needs additional scrutiny.

JEL Classification: C21, I20, I21
Keywords: well-being, civic outcomes, education, returns to education, regression discontinuity

*Thanks to Lorraine Dearden, Alfonso Miranda, Anna Vignoles, participants to the European University Institute Micro Group, the Institute of Education seminar, the North Americann Summer Meeting of the Econometric Society in Boston, the Australasian Meeting of the Econometric Society in Canberra and to the Department of Children, School and Family for access to the data. All errors are mine. email m.sartarelli@ioe.ac.uk

1 Introduction

Individuals as well as firms and governments acknowledge the relevant role of education in building individuals' human capital and helping to signal the otherwise unobserved ability of individuals in the labour market. This is suggested among others by the great attention on returns to education[1] and training programs.[2] However, a sharp increase in interest in the role of education in shaping individuals' well-being by policy-makers such as health or self-control or civic behaviour, e.g. compliance with the law and altruism, has so far only relied on mixed evidence. For example a police warning at age 16 is correlated with an increase by 10 percentage points or 25% of the probability of assuming heavy drugs, e.g. heroine, cocaine, ecstasy at age 30 using the 1970 British Cohort Study longitudinal dataset. Moreover, similar patterns hold for employment, wage and homeless status but it is unclear whether what causes this early police warning. Then in this paper I estimate the effect of using achievement labels, e.g. bad or good on the probability of a police contact or visit to parents due to the behaviour of children in UK secondary schools. The definition of well-being that I use is that of a metric, e.g. the rule of the law, that an individual, e.g. a policeman, uses to assess the behaviour of other individuals, e.g. children. This motivates using the event of a police contact due to the behaviour of a children as outcome variable, it is common in the Psychology literature and it is also illustrated in Heckman (2008). Several and potentially competing interpretations can help to explain the variation in the civic behaviour of children. Among others are variation in the opportunity cost or relative price of leisure that children face when their skills are labelled for example as good rather than bad. Through this channel that Benabou and Tirole (2003) and Benabou and Tirole (2002) call motivation or self-confidence children may alter their behaviour thus potentially breaking the law by for example vandalising at the risk of being caught by the police. Over time children with high school attainment may either keep studying and behave more safely to achieve in the future than lower achievers do as a standard agency model, for example Lazear (2001), with high school grades as incentives would predict. Alternatively, high achievers today may respond to the ability signals that achievement labels give them in the opposite way by allocating more time to leisure which decreases the study time. This concept is known as cognitive dissonance

[1] See Angrist and Krueger (1999) for a survey of the literature on returns to education.

[2] See for example Blundell et al. (2004) and Heckman et al. (1999) for evaluations of training programs respectively in the UK and the USA.

in psychology and experimental evidence in an application to school tests in Flink et al. (1990) among others confirm its empirical relevance. Lastly peer effects may facilitate the assimilation of civic or social norms that vary by children's achievement levels and their socio-economic background as for example in Gaviria and Raphael (2001) and Duncan et al. (2005).

Table 1 shows that in survey interviews parents report that approximately 7% of children behaves in such a way to prompt the police to contact or visit them while their children report that 9% vandalised public property. Attainment as well as the broader outcomes of education than test scores seem to be of great concern in the UK due to mixed evidence on the disengagement of children in secondary schools. For example column (3) in Table 2 shows that 36% of children score level 3 in the Key Stage 3 Maths test while the Department of Children, School and Family (DCSF) expects them to reach level 4 or above at that stage of their education.[3] This leads DCSF to set up in 2008 a joint policy-research initiative "Every Child Matters"[4] that follows "No Child Left Behind" [5] that is a similar initiative in the USA. Both aim at delivering such values as stay health, safe and socialise beside teaching core competence subjects to children. While this paper aims at informing the decisions of policy-makers at DCSF, Hastings (2007) offers a policy evaluation of the No Child Left Behind program. The emphasis of these programs is on those children for whom education is marginally more relevant to influence their civic behaviour than the role of their parents is as the latter may not guarantee careful enough guidance. While the literature on the relationship between education and economic growth[6] and that on returns to education[7] have been widely explored, the civic returns to education or the effect of labelling or sorting students at school on measures of civic outcomes or their well-being have been little explored still. Gibbons and Silva (2008) find only a moderate correlation between enjoyment at school or measures of well-being and achievement by UK children in compulsory education. Chevalier and Feinstein (2006) instead find a positive effect of education and mental health using the UK National Child Development Study longitudinal dataset. Dee (2004) finds positive returns to civic education in the US by looking at such civic outcomes as support for free speech and reading newspapers.

[3]See Ray (2006) for more details on the functioning of the educational system in the UK.
[4]http://www.everychildmatters.gov.uk
[5]http://www.ncpublicschools.org/nclb/
[6]See Krueger and Lindahl (2001) for a survey.
[7]See Blundell et al. (2001) for a survey in the UK and Angrist and Krueger (1999) in the USA.

Lastly Grossman (2005) looks at the effect of education on non-market outcomes from both a theoretical and an empirical perspective. His findings suggest a positive impact of education on such outcomes as health and inputs into the production of own health, fertility, and child quality or well-being reflected by their health and cognitive development. This paper contributes in two ways to the recent policy debate and the literature on the wider returns to education than test scores can measure. Firstly it characterises the relationship between achievement labels and the civic behaviour of school children and it then offers a research design to estimate the causal effect of labels on well-being. Secondly, the positive and significant effect of labels on well-being suggests to put the role of education in fostering the civic attitudes of children under closer scrutiny. The structure of the rest of the paper is as follows. Section 2 sketches a stylised form model of education, sorting and well-being. Section 3 describes firstly the challenge in identyfying the causal effect of sorting on the well-being of students and then a solution using a regression discontinuity research design. Section 4 describes the data and Section 5 presents estimates and sensitivity analysis. Section 6 concludes.

2 Stylised model of education, sorting and well-being

Let students choose time to spend in leisure t_l and supervised learning t_s and to maximise in (1) a concave and C^2 utility u arising from leisure $l(t_l)$ and supervised learning $s(t_s, a)$ by schools/parents with varying intensity of achievement a. Students are subject to a time constraint $t_s + t_l \leq 1$ that is normalised to 1.

$$
\begin{aligned}
\max_{t_s,t_l} u(t_s, t_l, a) \quad &= \quad \underbrace{s(t_s, a)}_{\text{supervised learning}} + \underbrace{l(t_l)}_{\text{leisure}} \qquad (1) \\
\text{s.t.} \quad & \quad n_c + n_d = 1
\end{aligned}
$$

The intuition behind the utility specification is that students are endowed with total day time equal to 1 and allocate jointly with their parents' advice between two main activities. They obtain utility $s(t_s, a)$ by spending t_s units of time in supervised learning activities such as education, more broadly sports or generally games where there is a dimension of achievement a with ordered properties. This is given by the test scores they get at school as well as by match scores or ranks in sport competitions as well as titles in the hierarchical structure of youth clubs such as the Scout Movement. From the rest of their day they obtain utility $l(t_l)$ by spending t_l units of time in leisure activities that I define loosely as any non-actively supervised activities such as playing with toys, playing with other children in a park or watching TV. Children solve jointly with their parents the problem in (2) by choosing an optimal time t_l^* to allocate to leisure. The role that parents play in the allocation of time by their children suggest that the preferences of parents may influence both their time allocation and thier achievement at school as Lizzeri and Siniscalchi (2008) show.

$$t_l^* \in arg\max_{t_l} u(t_l, a) = \underbrace{s(1-t_l, a)}_{\text{supervised learning}} + \underbrace{l(t_l)}_{\text{leisure}} \quad (2)$$

$$\frac{\partial t_l^*}{\partial a} \lesseqgtr 0 \quad (3)$$

Consider the following timing: first the child and parents allocate time t_l^* optimally, then the child performs a school tests and obtains a measure a of achievement. Lastly the child and the parent will decide the optimal t_l^* for future periods. Without explicitly modelling dynamics, the comparative statics in (3) gives the effect of a change in sorting or achievement labels a from school attainment on the change to the optimal allocation of time t_l^*. Figure 1 illustrates the joint decision problem of the child and the parents. (t_s^*, t_l^*) is the bundle describing the initial optimal allocation of time. Then a positive shock a of achievement or sorting to the utility function of the child has the effect of altering the composition of the new optimal time bundle $(t_s^{*'}, t_l^{*'})$. This is alternatively illustrated in Figure 1 by considering the effect that a change in the relative price $\frac{p_{t_l^*}}{p_{t_s^*}}$ of leisure to that of supervised learning has on the optimal allocation of time the relative prices of leisure.[8] Lizzeri and Siniscalchi (2008) suggest a mechanism of learning over time by the child and the parents that influences both the child's achievement today a and the behaviour $\frac{\partial t_l^*}{\partial a}$ tomorrow. Benabou and Tirole (2003) instead use a principal-agent framework to reconcile contrasting views on the response of individuals to incentives in economics and psychology. While theory and evidence in economics suggest that individuals are responsive to incentive and for example increase their effort under pay on performance relative to a flat pay scheme as in Lazear (2001), mainly experimental evidence in psychology has shown that rewards or punishments can deter the motivation of individual as in Flink et al. (1990). As the reduced form model stands in (1) one can't tease out the potentially contrasting forces at work in determining the behaviour of a child responding to shocks to the utility function. Hence without imposing further structure on the determinants

[8]The price of time are set to unity for simplicity in the time constraint in (1).

of supervised learning whether achievement labels or sorting a (de)increases the optimal leisure time of a child is ultimately an empirical matter.

3 Reduced form model and research design

In this section I takes the stylised reduced form model that I sketched in section 2 to the data in two steps. Firstly in section 3.1 I describe the challenge in identyfying the causal effect of sorting on the well-being of students. Secondly, in section 3.2 I propose a solution to the identification challenge by using a regression discontinuity research design that exploits certain institutional features in the UK compulsory school curriculum.

3.1 Reduced form model

Monetary and non-monetary returns to education models are in their simplest specification linear models in which educational attainment A_i of children i is used to explain a generic outcome of interest T_i, e.g. wages or non-monetary outcomes such as life expectancy, measures of health, crime and attitudes towards public policy, politics and the law. In this paper I estimate a linear model of non-monetary returns to achievement labels in school tests. T_i^* is a latent outcome measure of the time that a child allocates to leisure and A_i is a discrete measure of educational attainment, e.g. good or bad, of children in school test scores. Let T_i in (4) be a binary dummy equal to 1 if the police contacts or visits a child's family as a consequence of the behaviour of the child with and 0 otherwise. The dependent variable T_i is a binary measure of the underlying and latent allocation of time T^*. If $T_i^* > 0$ this can be interpreted for example as the truant behaviour of a child at school or more generally the lack of supervised guidance by the school or the paprents. The effect of achievement labelling on the probability of a police visit can be estimated in Equation (5) using a Probit model assuming normality of the error term $U_i \sim N(0, \sigma^2)$. Alternatively Equation (6) can be estimated using a Linear Probability Model (LPM).

$$\begin{aligned} T_i^* &= \alpha + \beta A_i + U_i \\ T &= \begin{cases} 1 \text{ if } T^* \geq 0 \\ 0 \text{ otherwise} \end{cases} \qquad (4) \\ Pr(T_i = 1) &= Pr(T_i^* > 0) \qquad (5) \\ &= Pr(\alpha + \beta A_i + U_i > 0) \\ Y_i &= \alpha + \beta A_i + U_i \qquad (6) \end{aligned}$$

If no information that is unobservable to the econometrician in the error term U_i for child i, e.g. mother's education, is also contained in the achievement label A_i, then Probit or LPM yield unbiased estimates. If instead the covariate of interest A_i and the error term U_i are correlated, e.g. those children with highly educated mothers score higher in tests than those with less educated mothers, then estimates suffer from endogeneity bias. Figure 2 illustrates the timing of the Key Stage 3 Maths test, the potential threat to identification that endogeneity in achievement poses and the strategy that the research design in this paper proposes.[9] To consistently identify the effect of labelling on police contacts or visits in a non-experimental setting, one purges endogeneity by using an Instrumental Variable approach that exploits an exclusion restriction, i.e. a variable Z that shocks the endogenous treatment/covariate A but doesn't directly shock the outcome variable T.[10] Alternatively, endogeneity can be dealt with by making assumptions on the model functional form and the distributions of error terms or by relaxing model assumptions on the selection process. However, the knowledge of the econometrician about the nature of selection rule based on unobservables is by defintion limited with the exception of matching methods that rely on selection on observables as van der Klaauw (2002) points out. Then, additional knowledge about the selection rule that attaches different achievement labels to children in UK compulsory secondary schools can help to obtain unbiased estimates.
The British national school curriculum for children in the age range 5 to the compulsory school leaving age 16 is divided into four Key Stages (KS) from 1 to 4. Each is associated to programmes of study in the core subjects: English, maths and science and foundation subjects such as geography and history. From KS 2 on all students take national tests in the core subjects. Their test scripts are anonymous and they are marked by external examiners rather than by the teachers of the children who are tested. External examiners are not necessarily teachers and they are specifically trained to mark Key Stage tests by the Qualifications and Curriculum Authority (QCA). This is a governmental agency in charge of supporting the delivery of the public exam system and the developing and deliverinf high-quality national curriculum tests and assessments. Marking of scripts by external examiners rather than by students' teachers is a relevant institutional feature to minimise the strategic manipulation of test scripts and scores as a result for example of

[9] See Sartarelli (2009a) for an application exploiting the timing of collection of information in linked data to estimate the causal effect of enjoyment on achievement.

[10] See Angrist et al. (1996) for a discussion of local average treatment effect and instrumental variables.

incentives linking teachers' compensation or school funding to performance.[11] Scripts are marked using numerical scales that QCA sets and they vary over time, by subject and Key Stage. In this paper I focus on the Key Stage 3 maths test whose marks are in the range 0-99+ as column (2) in Table 2 shows. Once QCA exminers have marked scripts, marks are turned into achievement levels in column (1) and levels are disclosed to children. For example, two students who score 63 and 64 as total mark in the Key Stage 3 maths test in Tier 3-5[12] achieve respectively levels 3 and 4. The student achieving level 2 with 63 as total mark in maths is in the same achievement level as one scoring as low as 29, the minimum mark necessary to achieve level 2. In contrast the student achieving level 3 with 64 as mark is in the same achievement level of highly achieving students with marks as high as 98. Then two students whose observable ability in maths differ by one mark are placed in two achievement levels that differ considerably in the ability composition of students in each level.

The discontinuity in test scores that the institutional design of Key Stage tests in the British national school curriculum offers allows to employ a Regression Discontinuity Design (RDD). By exploiting this I estimate the causal effect of labelling similar students with different achievement levels on their civic behaviour as measured by the event of a police visit at the children house due to their behaviour.

3.2 Regression discontinuity design (RDD)

RDD is a quasi-experimental evaluation design that Thistlethwaite and Campbell (1960) first introduced. Its popularity in the evaluation literature since the 1960s[13] has only been matched in Economics starting from the 1990s.[14]

$$T_i = \alpha + \beta A_i + U_i$$

In equation (6) T_i is a binary dummy equal to one if a household receives a police visit for the behaviour of their child and A_i is an endogenous polychotomous dummy for achievement levels as unobservables U_i such mother's education influence T_i thus violating the independence assumption $E[A_i|U_i] = 0$. In a policy evaluation framework the achieve-

[11] See for example a case of potential manipulation of tests scores in the USA in Jacob and Lefgren (2004) and in the UK in Sartarelli (2009b).

[12] See Miranda and Sartarelli (2009) for a characterisation of the interaction student-teacher using an agency model and an empirical test of the standard agency model prediction against a cognitive dissonance model by exploiting the allocation of students to different test papers by Tier.

[13] See for example Cook and Campbell (1979); Trochim (1984).

[14] See Imbens and Lemieux (2008) for a review.

ment level A_i is an endogenous treatment with non-random assignment to treatment as the socio-economic background of a child likely influences her achievement and civic behaviour. In the case of RDD additional information is available as the assignment to treatment depends deterministically on the numerical score relative to the cutoffs underlying the achievement level. For example, look at Table 2 and consider the achievement of level 4 in the Maths test as an endogenous binary treatment relative to the placebo at level 3. Then one can write the deterministic decision rule governing assignment to treatment as $A_i = f(Z_i) = 1Z_i > 63$, an indicator function switching the achievement level dummy A_i from 0 to 1 if the mark Z_i is greater or equal than the mark cutoff $\bar{Z} = 64$. However, the selection rule so far is no different than the untestable exclusion restriction that is used as the identifying assumption in an experimental research design. van der Klaauw (2002) also compares RDD and identification based on selection on observables by noting that RDD violates the common support condition $0 < Pr(A = 1|Z) < 1$ in Rosenbaum and Rubin (1983) because RDD makes individuals in the treatment and control groups different at least in the average value of Z.

Identification and estimation of treatment effects in the case of RDD relies on the consideration that a sample of individuals in a very small sample in the small neighbourhood of the cutoff point $\bar{Z}$ is similar to the one of a randomised experiment with cutoff at $\bar{Z}$. If one considers instead a natural experiment to identify a Local Average Treatment Effect (LATE)[15], RDD can be thought of as a local version of LATE around the cutoff. The bias of treatment effect is likely to increase with the size of the neighbourhood around $\bar{Z}$ if assignment to treatment Z is related to the outcome T conditional on the treatment status A. To overcome this, RDD relies on an assumption on the functional form of the relationship between the average outcome and the selection variable. Thanks to this it allows to use more observations around the cutoff and also extrapolate from the neighbourhood of the cutoff to what a randomised experiment has done. More formally, RDD relies on the assumption 3.1 to identify a constant treatment effect while assumption 3.2 is used in the case of varying treatmenf effects.

Assumption 3.1 *The conditional mean function $E[U|Z]$ is continuous at $\bar{Z}$*

Assumption 3.2 *The mean treatment effect function $E[\beta_i|Z]$ is right-continuous at $\bar{Z}$*

[15] See Angrist et al. (1996) for a formal characterisation of LATE.

Then it follows that the treatment effect β can be identified under assumptions 3.1 and 3.2 by the difference in equation (7) and $E[\beta_i|Z=\bar{Z}]$ gives the average treatment effect of those individuals i at the margin between achievement levels.[16]

$$\begin{aligned} lim_{Z\downarrow\bar{Z}}E[Y|Z]-lim_{Z\uparrow\bar{Z}}E[Y|Z] &= \beta+lim_{Z\downarrow\bar{Z}}E[U|Z]-lim_{Z\uparrow\bar{Z}}E[U|Z] \qquad (7)\\ &= \beta \text{ under the constant treatment assumption 3.2}\end{aligned}$$

An assignment to treatment A_i rule that depends deterministically on Z_i with all individuals following the assignment rule is described as *sharp RDD*. This contrasts the case of a stochastic dependence of assignment to treatment on Z_i that is known as *fuzzy RDD*. In this paper the rule defining achievement levels given Key Stage 3 maths marks is deterministic and the average treatment effect $E[\beta_i|Z]$ of labelling on the probability of police visits is then estimated using a sharp RDD. van der Klaauw (2002) correctly points out that the estimation of constant treatment effect with a sharp RDD is a speacial case of selection on observables because the treatment status A_i is correlated to an observed variable Z_i that can itself be correlated with the outcome T_i. This would bias the OLS estimate of β. Instead by correctly specifying A as a polynomial function of Z in (9) to yield an estimate of the conditional mean function $E[U|Z,S]$ of the error term[17], include it in (6) to obtain (8) and re-estimating this to obtain a consistent estimate of β.

$$\begin{aligned} T_i &= \alpha+\beta A_i+\gamma\hat{E}[U|Z,A]+W_i \qquad (8)\\ A &= f(Z) \qquad (9)\end{aligned}$$

An incorrect specification of $f(Z)$ yields a biased estimate of β. Hence a strategy that minimises the likelihood of misspecification is to either adopt a semiparametric specification of the control function or to use nonparametric regression around the cutoff $\bar{Z}$ as Imbens and Lemieux (2008) suggest. In this paper I use the the first of the two methods and estimate a polynomial $k(Z)\approx\sum_{j=1}^{J}\eta_i Z^j$ of order $J=4$ in Z.

4 Data

I estimate the effect of achievement labels on the probability of police visits due to the behaviour of a child by using two linked datasets. The first is the Longitudinal Study of

[16] The case of varying treatment effects and its application to Key Stage tests requires to estimate $\beta(Z_i)$ in $T_i=\alpha+\beta(Z_i)Z_i+u_i$ and it is left for future work.

[17] $E[U|Z,S]=E[U|Z]$ in the case of a sharp RDD.

Young People in England (LSYPE). This is a survey of British households that contains detailed information on the socio-economic background, education, work experience and aspirations children and parents in each household. It has been run annually since 2004; the sample size of each wave is approximately 15000 households and it has been stratifiead so as to be representative of the population of UK households. This dataset is linked to the National Pupil Database (NPD), an administrative dataset containing detailed information on the attainment of children attending compulsory education in UK schools. Linking the two datasets yields a sample of approximately 15000 children whose summary statistics are reported in Table 1. As dependent variable I focus on a binary dummy equal to 1 in the event of a police visit or contact to parents due to the behaviour of their child as parents rather than children answer to this question in the LSYPE survey. This is preferable to otherwise interesting measures of civic behaviour that children self-report as the answers of children are more likely to be biased than their parents' are. The richness of both the administrative NPD data on students' attainment at school and the LSYPE survey allow to disregard potential threats to the research design such as classical or non-classical measurement error and missingness non at random. I focus on test scores of children who are assigned by teachers to the same Tier and hence test paper as if teachers performance is altered by incentives and they sort students based on their subjective beliefs, this represents a potential source of bias.[18] Moreover, I choose Tier 3-5 where low achievers are because it is such that one would intuitively not expect an effect of labelling. This makes it a good candidate to then interpret estimates as lower bounds relative to other tiers in which potential manipulation of test scripts and scores cannot be ruled out as Sartarelli (2009b) shows.

5 Estimation and sensitivity analysis

I estimate the causal effect of achievement labels on the probability of police visits due to children behaviour by carrying out a RDD and following 3 steps in Imbens and Lemieux (2008). Firstly in subsection 5.1 I estimate local polynomials of the probability of a police visit separately below and above achievement cutoffs and plot the discontinuity. Secondly in subsection 5.2 I obtain RDD estimates of the effect of sorting on well-being by using a local linear regression approach. Thirdly in subsection 5.3 I perform a sensitivty analysis

[18]See footnote [12].

of the estimates and focus on two potential sources bias in the estimates. The first can arise from unobservables that affect both pre-treatment variables and outcome variables, e.g. earlier achievement of children at Key Stage 2 affecting both their achievement at Key stage 3 and the probability of police visits. The second instead may be due to either children or teachers sorting around the achievement discontinuity.

5.1 Graphical analysis

I obtain preliminary RDD estimates by fitting local polynomials of the probability of a police visit T in the total mark in Key Stage 3 maths Z separately below and above the mark cutoff at 64. Estimates are little sensitive to the choice of the kernel of the polynomial, the degree of polynomial or the bandwidth choice.[19] Figure 3 plots RDD estimates of the the effect of achievement labels in 2004 KS3 Maths total mark on the probability of police visits in 2005. An increase in 2004 KS3 maths total mark that switches the achievement label from left to right of the cutoff decreases the probability of a police visit by approximately 4 percentage points. I obtain the estimates by averaging values of the KS3 Maths total mark. Doing so helps to correct for measurement error that would otherwise arise as the marks differ by a unit rather than being continuous as Lee and Card (2008) suggest.

5.2 RDD using local linear regression approach

Column (2) in Table 4 reports RDD estimates obtained by using local linear regression approach. A jump in achievement labels due to the discontinuity in marks decreases the probability of a police visit by 4.4 percentage points and the estimate is significant at 5% level. By comparing the RDD estimates in column (2) to the OLS estimates in columns (3)-(5) and under the weak and testable RDD assumptions OLS estimates yield a spurious zero correlation between sorting and well-being that is likely driven by unobservables and they are then upward biased. This suggests that the unobservables in the socio-economic background of a child is likely to impact both her achievement at school and her civic behaviour outside school. In contrast, an exogenous and positive shock in the achievement group where a child is placed has a statistically significant effect on the civic behaviour of the child by reducing the likelihood that she behaves in such a way to potentially act

[19] See Imbens and Lemieux (2008) and Lee and Lemieux (2009) for guidance over the the degree of polynomial or the bandwidth choice and the sensitivity analysis.

unlawfully and cause the intervention of police. The magnitude of the RDD estimate is robust to changes in the bandwidth and degree of polynomials.

5.3 Sensitivity analysis

The first threat to a RDD is posed by the joint effect of unobservables on pre-treatment covariates that are determined e.g. the year before the Key Stage 3 maths test and the Key Stage 3 maths mark around the cutoff at 64. I test for this by estimating the local polynomials in the Key Stage 3 maths test total mark using an array of pre-treatment covariates and I plot two sets of estimates. Figure 4 reports estimates of the effect of unobservables on the cutoff mark and pre-treatment covariates of the educational attainment of a child. Figure 5 reports the same type of estimates but this time using pre-treatment covariates of the main parent of a child. Neither of the figures show suspicious jumps in pre-treatment variables on the Key Stage 3 maths total mark of a child and this rules out unobservables. Moreover, mean values of the covariates to the left of the cutoff and to the right are not statistically different as Table 3 reports.

The second threat to a RDD is the potential sorting of individuals around the discontinuity cutoff that may occur if they know the cutoff before the treatment and can influence either the assignment to treatment or the magnitude of the treatment effect. McCrary (2008) proposes a density-based test on the continuity of the running variable Z, i.e. Key Stage 3 Maths test total mark in this paper, around the cutoff $\bar{Z}$. Figure 3 plots an undersmoothed histogram of the forcing variable, the total maths score measured continuously, and shows no jump in the size of the bins of the histogram at the cutoff at 64. However, jumps occur elsewhere and Sartarelli (2009a) exploits these jumps to interpret the behavioural response of students and schools to the institutional features in compulsory scholing in England. Moreover, I perform the McCrary (2008) test of the null hypothesis of no sorting that is not rejected at 5% significance level.

6 Conclusions

In this paper I look at the role of education in shaping individuals' well-being by estimating the causal effect of using achievement labels, e.g. good or bad, on the probability of a police visit due to the behaviour of a child. I do so by exploiting discontinuities in school test scores and using data on children in secondary schools in the UK. I find that a jump from low to high achievement decreases by 4 percentage points the probability that the parents of a child receive a police visit or contact due to the behaviour of the child. The estimates that I obtain are robust to a sensitivity analysis that tests for the presence of confounders invalidating the research design. The results suggests that the mechanism through which education contributes to teaching core subjects such as Maths as well as norms helping children to lead a healthy and safe life needs additional scrutiny by policy-makers. This confirms the recent evidence in the UK of disengagement, truancy and crime by school children correlating with poor attainment and a potentially dangerous behaviour of school children. The next steps to uncover in future work the determinants of school attainment and the social behaviour of children lie in extending the analysis to behavioural outcomes post-compulsory schooling, explicitly modelling school supply and potentially disentangling the contribution of parenting and schooling, for example teachers and peers, in shaping the civic and labour market outcomes of a child.

References

Angrist, J., G. Imbens and D. Rubin (1996) *Identification of Causal effects Using Instrumental Variables*, Journal of Econometrics 71, 145-160

Angrist, J. and A. Krueger, *Empirical Strategies in Labor Economics*, Chapter 23 in O. Ashenfelter and D. Card, eds., The Handbook of Labor Economics 3 (North Holland), 1278-1357

Benabou, R. and J. Tirole (2003) *Intrinsic and Extrinsic Motivation*, Review of Economic Studies 70, 489-520

Benabou, R. and J. Tirole (2003) *Self-Confidence and Personal Motivation*, Quarterly Journal of Economics 117 (3), 871-915

Blundell, R., L. Dearden and B. Sianesi (2001) *Estimating the Returns to Education: Models, Methods and Results*, LSE Centre for the Economics of Education Discussion Papers 16

Blundell, R., M. Dias, C. Meghir and J. Van Reenen (2004) *Evaluating the Employment*

Impact of a Mandatory Job Search Program, Journal of the European Economic Association 2, 569-606

Chevalier, A. and L. Feinstein (2006) *Sheepskin or Prozac: The Causal Effect of Education on Mental Health*, LSE Centre for the Economics of Education Discussion Paper 71

Campbell, D. and J. Stanley (1963) *Experimental and Quasi-Experimental Designs for Research*, Rand McNally, Chicago

Dee, T. (2004) *Are there any returns to civic education?*, Journal of Public Economics 88, 1697-1720

Duncan, G., J. Boisjoly, M. Kremer, D. Levy and J. Eccles (2005) *Peer Effects in Drug Use and Sex Among College Students*, Journal of Abnormal Child Psychology 33(3), 375-385

Flink, C., A. Boggiano and M. Barrett (1990) *Controlling teaching strategies: Undermining children's self-determination and performance* Journal of Personality and Social Psychology 59, 916-924

Gaviria, A. and S. Raphael, (2001) *School-based peer effects and juvenile behavior*, Review of Economics and Statistics, 83 (2), 257-268

Gibbons, S. and O. Silva (2008) *School Quality, Child Wellbeing and Parents' Satisfaction*, London School of Economics mimeo

Grossman, M. (2005) *Education and Non Market Outcomes*, Chapter 10 in the *Handbook of the Economics of Education*, E. Hanushek and F. Welch (eds.), North Holland, Amsterdam

Jacob, B. and L. Lefgren (2004) *Remedial Education and Student Achievement: A Regression-Discontinuity Analysis*, Review of Economics and Statistics 86 (1), 226-244

Hahn, J., P. Todd and W. Van der Klaauw (2001) *Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design*, Econometrica 69, 201-09

Hastings, J. (2007) *No Child Left Behind: Estimating the Impact on Choices and Student Outcomes*, NBER Working Paper 13009

Heckman, J. (2008) *Econometric Causality*, International Statistical Review 76 (1), 1-27

Heckman, J., R. LaLonde and J. Smith (1999) *The Economics and Econometrics of Active Labor Market Programs*, Chapter 31 in O. Ashenfelter and D. Card, (eds), Handbook of Labor Economics 3 (North Holland), 1865-2086

Imbens, G. and T. Lemieux (2008) *Regression discontinuity esigns: A guide to practice*, Journal of Econometrics 142, 615-635

Krueger, A. and M. Lindahl (2001) *Education for Growth: Why and for Whom?* Journal of Economic Literature 39, 1101-1136

Lazear, E. (2000), *Performance, Pay and Productivity*, American Economic Review 90, 1346-1361

Lee, D. and D. Card (2008) *Regression Discontinuity Inference with Specification Error*, Journal of Econometrics 142 (2), 655674

Lee, D. and T. Lemieux (2009), *Regression Discontinuity Designs in Economics*, NBER Working Paper 14723

Lizzeri, A. and M. Siniscalchi (2008) *Parental Guidance and Supervised Learning*, Quarterly Journal of Economics 123, 1161-1195

McCrary, J. (2008) *Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test*, Journal of Econometrics 142, 698-714

Miranda, A. and M. Sartarelli (2009) *Agency theory in schools: do incentives affect performance? Evidence from the assignment to tests in schools*, Institute of Education mimeo

Ray, A.(2006) *School Value Added Measures in England*, Department of Children, School and Family

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

Sartarelli, M. (2009a) *A note on incentives and the manipulation of test scores. Evidence from secondary school tests in the UK*, Institute of Education mimeo

Sartarelli, M. (200b) *If you enjoy what you do, do you perform better? Evidence from matched panel data of UK school children*, Institute of Education mimeo

Thistlethwaite, D. and D. Campbell (1960) *Regression-Discontinuity Analysis: An alternative to the ex post facto experiment*, Journal of Educational Psychology 51, 309317

Trochim, W. (1984) *Research Design for Program Evaluation: The Regression-Discontinuity Approach*, Beverly Hills, CA: Sage Publications

van der Klaauw, W. (2002) *Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach*, International Economic Review 43, 1249-1287

Table 1: Summary statistics

Variable	Mean	Std. Dev.
Young person / child		
Key Stage 2 English Total Mark	59.18	14.31
Key Stage 2 Science Total Mark	56.39	13.29
Key Stage 2 Maths Total Mark	61.44	20.96
Key Stage 2 Maths Marks in Paper A	23.86	8.92
Key Stage 3 Maths Total Mark	75.96	22.16
Whether ever vandalised public property	0.08	
Whether taken part in fighting or public disturbance	0.17	
Share of males	0.51	
Whether ever in special education	0.19	
Year of birth	1990	0.48
Main parent		
Whether police got in touch due to the behaviour	0.07	
Age at which first left school	16	2.09
Age at which left school altogether	24	13.09

Sources: Longitudinal Survey of Young People in England Wave 2, 12344 observations and National Assessment Agency

Table 2: Tables of levels and marks in the Key Stage 3 Maths in 2004, Tier 3-5

(1) Level	(2) Mark range	(3) Percentage of students
2	23-28	2.97
3	29-63	36.31
4	64-98	47.79
5	99+	12.95

Source: Longitudinal Survey of Young People in England Wave 2, 12344 observations

Table 3: Mean value of covariates in the neighbourhood of the cutoff

	Left of cutoff	Right of cutoff	P-value of difference in covariates
Key Stage 2 English total mark	55.97	57.53	0.00
Key Stage 2 Science total mark	53.15	54.33	0.00
Key Stage 3 English total mark	39.64	40.69	0.00
Key Stage 3 Science total mark	84.40	87.11	0.00
Share of males	0.47	0.50	0.00
Absence in last 12 months	0.27	0.26	0.07
Use of education or social services	0.09	0.09	0.02
Fighting or public disturbance	0.18	0.20	0.00
Happy when at school	0.84	0.85	0.01
Special education needs ever	0.19	0.20	0.03
Special education needs ever	0.19	0.20	0.03
Age at which main parent left school	23.63	23.44	0.12

Table 4: RDD estimates using local linear regression and OLS estimates of the effect of achievement labels on the probability of police visits due the behaviour of a child

(1)	(2)	(3)	(4)	(5)
	RDD	OLS		
		Covariates	Achievement quadratic	Achievement quadratic and covariates
Estimate	-.044	-.036	.003	.013
Standard error	.0176	.047	.031	.071

Figure 1: Illustration of the model comparative statics: the effect of achievement labels a on the change in the optimal allocation of leisure time t_l^*

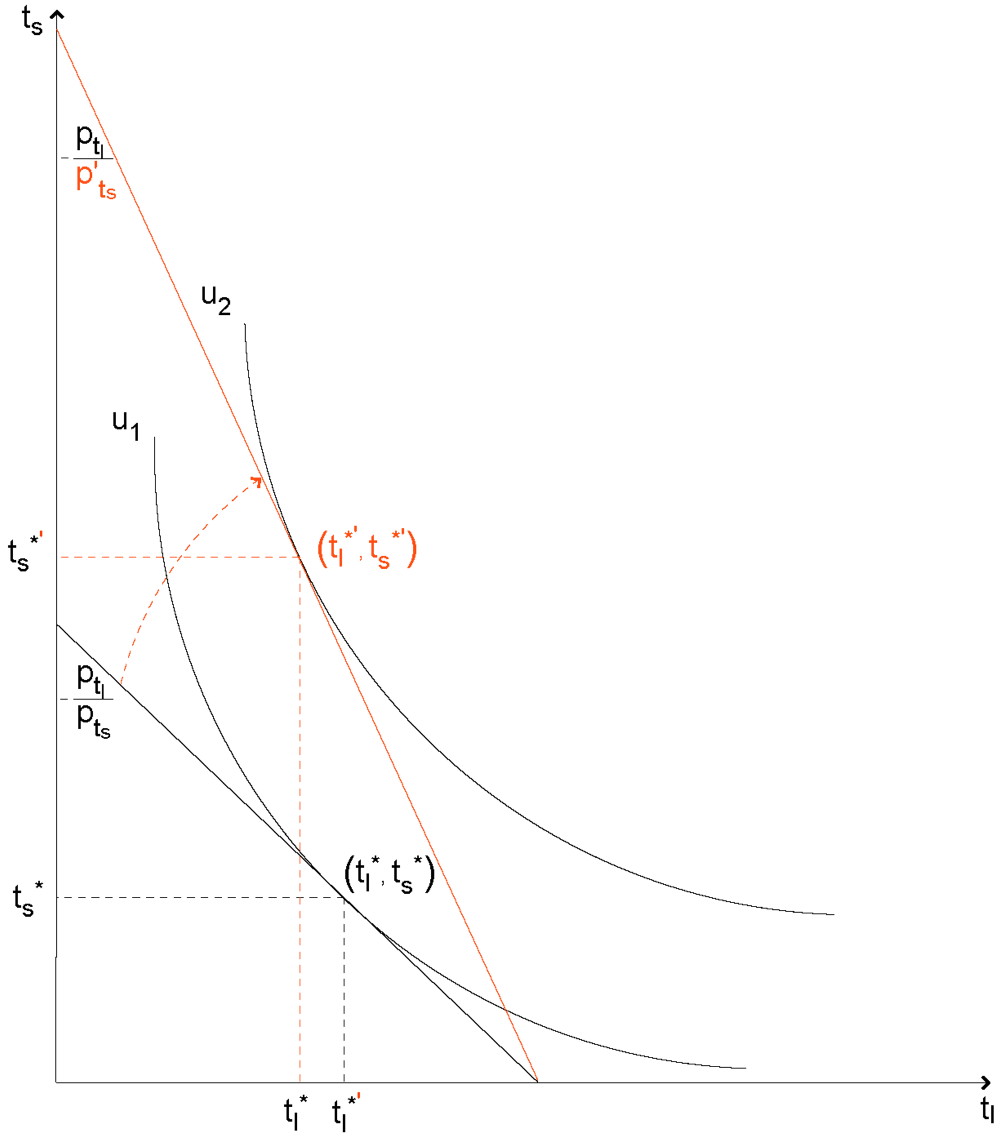

Figure 2: Illustration of timing, potential endogeneity and identification strategy of the effect of achievement labels on the probability of police visits due the behaviour of a child

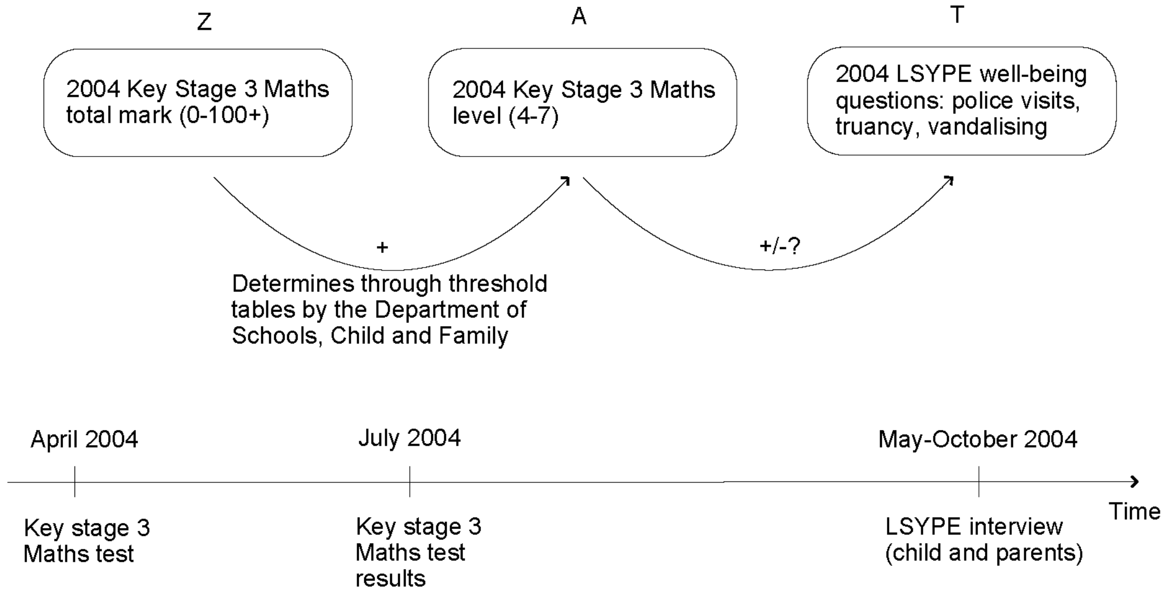

Figure 3: RDD estimate: graphical analysis of the effect of achievement labels on the probability of police visits due the behaviour of a child

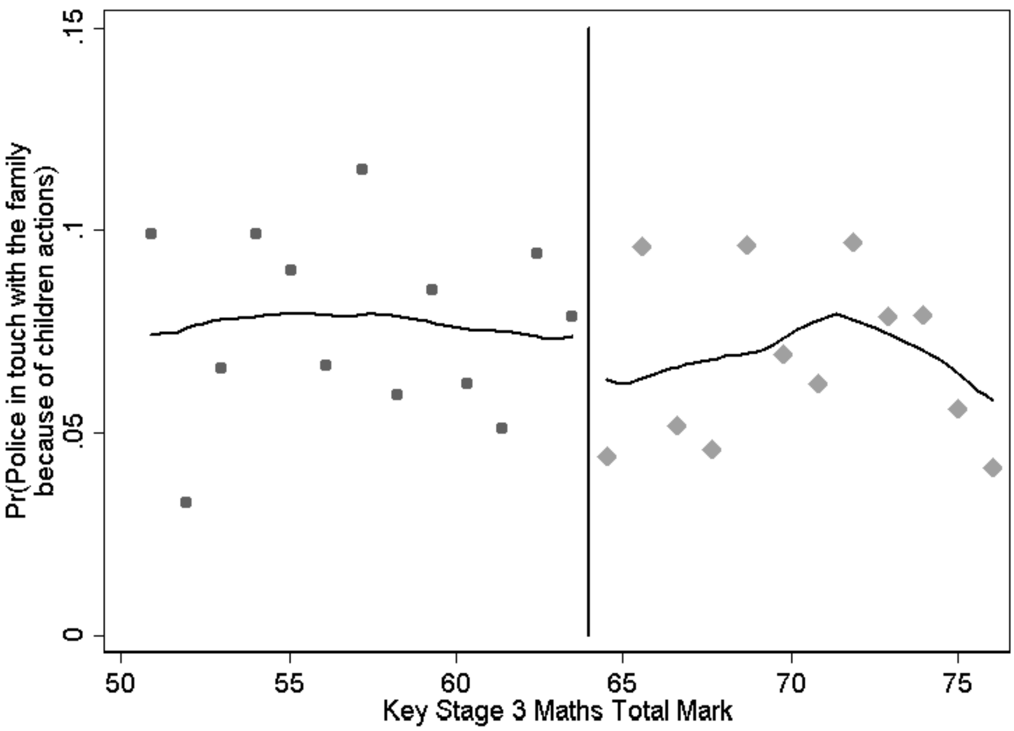

Figure 4: RDD estimate: graphical analysis of the effect of unobservables on the cutoff mark and pre-treatment covariates of the school attainment of a child

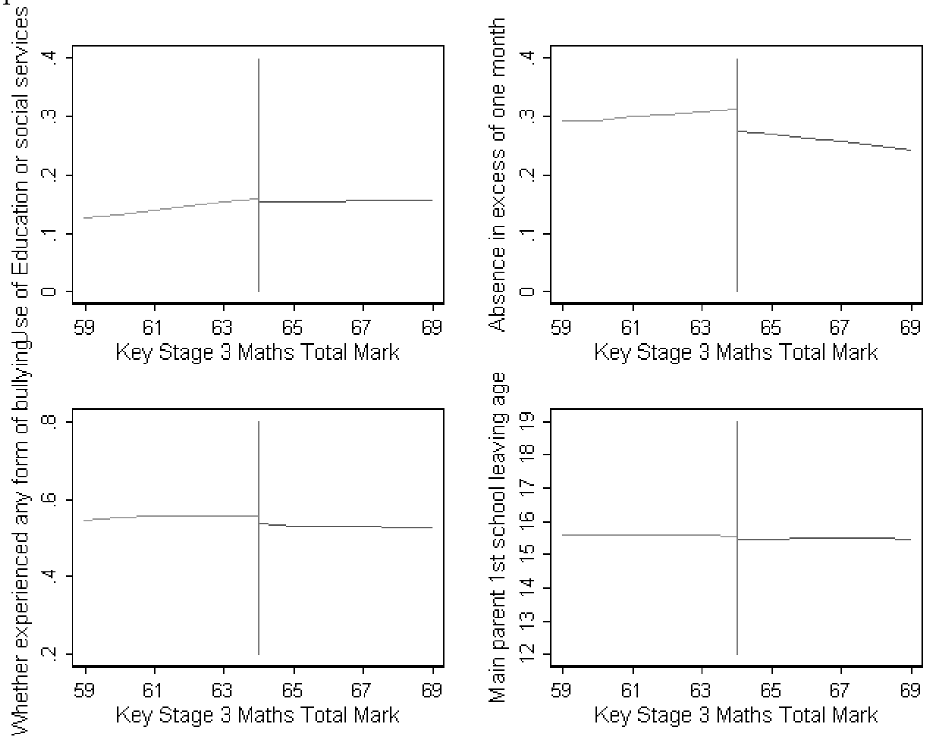

Figure 5: RDD estimate: graphical analysis of the effect of unobservables on the cutoff mark and pre-treatment covariates of the main parent of a child

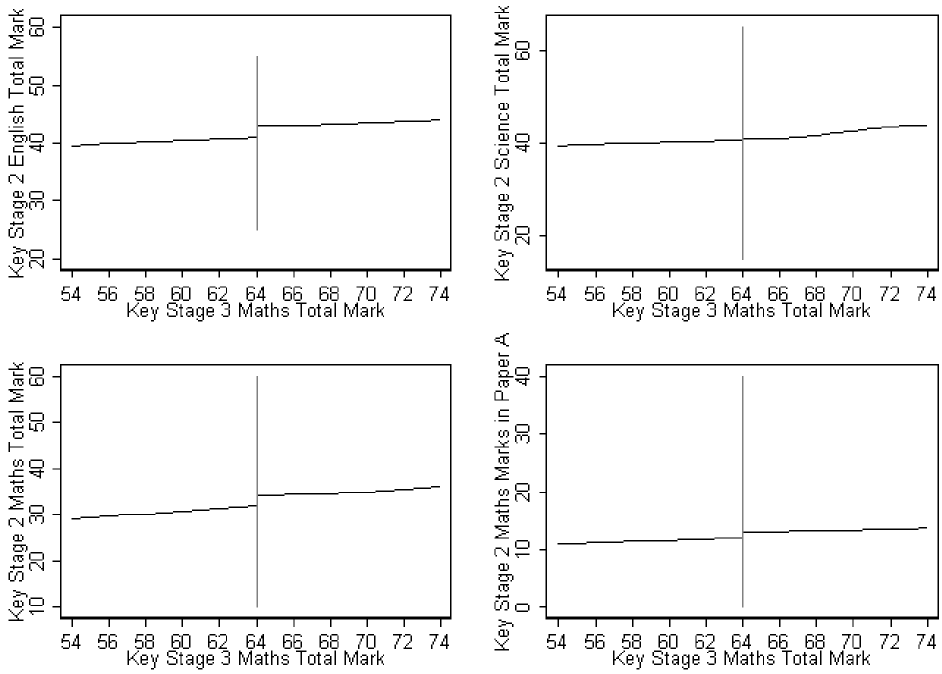

Figure 6: RDD estimate: undersmoothed histogram plot of the of Key Stage 3 Maths test total mark around the mark cutoff at 64