

Sánchez-Coll, Sergi

Article

Born this way: The effect of an unexpected child benefit at birth on longer-term educational outcomes

SERIEs - Journal of the Spanish Economic Association

Provided in Cooperation with:

Spanish Economic Association

Suggested Citation: Sánchez-Coll, Sergi (2023) : Born this way: The effect of an unexpected child benefit at birth on longer-term educational outcomes, SERIEs - Journal of the Spanish Economic Association, ISSN 1869-4195, Springer, Heidelberg, Vol. 14, Iss. 1, pp. 105-141, <https://doi.org/10.1007/s13209-022-00270-y>

This Version is available at:

<https://hdl.handle.net/10419/286572>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Born this way: the effect of an unexpected child benefit at birth on longer-term educational outcomes

Sergi Sánchez-Coll¹

Received: 3 May 2022 / Accepted: 24 October 2022 / Published online: 2 December 2022
© The Author(s) 2022

Abstract

Aiming to boost fertility rates, in 2007 the Spanish government implemented a universal €2500 baby bonus paid to mothers giving birth or adopting a child, leading to a short-lived increase in births. In this study, I measure the causal impact that the transfer had on the language and mathematical competencies of the children of eligible mothers at the end of primary school in the Catalonia region. I do so by taking advantage of how the policy was announced, leading to a sharp regression discontinuity design and a difference-in-discontinuities specification. The subsidy did not improve student achievement at age 12, since in the preferred diff-in-disc specification using the pooled sample of schools we can rule out grade improvements greater than 0.1 standard deviation units with 95% confidence. While some effects in the subsample of boys in disadvantaged schools are large in magnitude, of roughly 0.2–0.41 standard deviation units representing a 4–11% improvement from the average test score, they do not reach statistical significance and are likely caused by the high variability in test scores both before and after the policy implementation rather than by the subsidy itself, as suggested by robustness tests.

Keywords Income shock · Children · Education · Natural experiment · Baby bonus · Spain

JEL Classification H31 · I38 · J13 · J24

1 Introduction

Child benefits remain a popular tool to promote fertility in developed countries, as governments typically spend between 1 and 4% of GDP on family policies (Sobotka et al. 2019). Their role in fostering economic growth and development is well known, as

✉ Sergi Sánchez-Coll
ssanchezc@iese.edu

¹ IESE Business School, Barcelona, Spain

the return on investment is highest at young ages (Heckman 2008). Those policies come in a variety of forms. Universal direct cash transfers are straightforward to implement and, despite their cost and rather modest impact compared to other interventions, they help decrease the cost of having children and thus temporarily increase fertility (Milligan 2005; Sinclair et al. 2012; González 2013; Chuard and Chuard-Keller 2021).

The effects of some of those income-increasing policies, especially when targeting disadvantaged families, have been shown to go beyond pure fertility boosts, implying improvements in maternal and children health at birth and early childhood (Cooper and Stewart 2013, 2020). Longer-term effects have also been observed, leading to better educational and human capital outcomes, some even persisting into adulthood.

In this paper, I study the longer-term effects on educational outcomes of a universal subsidy implemented in Spain in 2007, consisting of a single €2500 payment given to mothers of a recently born baby or who adopted one. Using microdata containing the universe of sixth-grade students participating in the basic competency tests at the end of primary school in the Catalonia region, I take advantage of the natural experiment that the timing of the measure created, generating a discontinuity at the 1 July 2007 threshold when households became eligible to receive the payment, as opposed to parents of children born immediately before that cut-off, who did not receive it.

First, I propose a regression discontinuity design (RD) where I compare children born on both sides of the cut-off. Potential seasonality in children characteristics affecting both sides of the cut-off differently, such as early developmental differences or parental education levels, might bias the results. To overcome those issues and gain precision, I then combine the previous RD of the treatment year with a RD of the previous non-treatment years in a difference-in-discontinuities specification (diff-in-disc).

I perform a non-parametrical estimation of the models, using local linear regression with triangular kernel smoothing and data-driven bandwidth selection procedures for the pooled sample of observations. Additionally, I explore the heterogeneity of results in a number of subsamples, considering student gender and nationality, school complexity, ownership and size.

The benefit did not have significant positive effects overall on children achievement test scores at age 12, as we can rule out grade improvements greater than 0.1 standard deviation units with 95% confidence. In subsample analyses, RD results show slightly negative effects that are not significantly different from zero except for a decrease of 11% of a standard deviation in Catalan language grades in non-disadvantaged schools. Boys in disadvantaged schools seem to have benefited the most from the subsidy, with increases of up to 43% of a standard deviation in the English language test score. Diff-in-disc estimates lose almost all statistical significance while remaining roughly similar in magnitude, which is especially relevant for those attending disadvantaged public schools, with estimated improvements of 4–11% of the average test score for boys and 1–4.1% for girls depending on the assessed subject, also lacking statistical significance.

However, robustness tests suggest that the reported changes in test scores are due to chance rather than caused by the subsidy, as natural variability in the data is significantly higher than the discontinuities reported in the treatment period. Insufficient

transfers, its universal nature, the lack of earmarking, the timing and the fading out of effects over time can potentially explain the findings.

This paper contributes to the rich literature on the effects of income transfers to families on longer-term human capital outcomes. It specifically complements existing evidence evaluating the 2007 child subsidy in Spain and first estimates its effects on the acquisition of language and mathematical skills when children reach preadolescence and in a low-stakes standardised examination setting.

The remainder of the paper proceeds as follows: Sect. 2 gives an overview of the state of the literature and the institutional background of the subsidy and the competency tests. Section 3 describes the identification strategy as well as the empirical specifications. Section 4 presents the data and their main characteristics. Section 5 reports the results, the robustness checks and a discussion, while Sect. 6 concludes the paper.

2 Background

2.1 Literature review

The prominent strand of literature on the foetal origins hypothesis, according to which, contrary to previous beliefs, shocks in utero can have long-term consequences in a child's life (Barker 1990, 1998), has been expanded to account for shocks in early childhood and has been supported by a wealth of evidence in economics research.

Theoretical efforts have been made to model the relationship between parental income and children achievement, starting with intergenerational transmission models (Becker and Tomes 1979, 1986). More recently, multiple-generations models of investment in human capital have emerged building on their work, using more realistic assumptions (Cunha and Heckman 2007; Del Boca et al. 2014; Cunha et al. 2020). However, most analyses remain reduced form (e.g. measuring the effect of a shock on an observable outcome), as several environmental factors constantly interact with each other during the continuous formation of skills in a child, implying strong assumptions by researchers in their modelling.

Three main determinants of parental investments in their offspring can be distinguished, namely parental preferences, informational constraints and financial constraints, which are the most relevant for this paper (Attanasio et al. 2021). However, the role that they might play in explaining the findings might be overstated (Carneiro and Ginja 2016).

Heckman and Mosso (2014) survey the structural literature on the factors influencing human capital development and conclude that unrestricted income transfers are unlikely to improve children's skills. Almond and Currie (2011) and Almond et al. (2018) provide extensive reviews of studies on the effects of various types of shocks and circumstances experienced during early childhood, including family income, on later human capital formation outcomes. Similarly, Cooper and Stewart (2013, 2020) review studies focusing exclusively on parental income shocks and on their effects on child development, which are the most relevant for this paper. Most of the reviewed

evidence finds improvements in cognitive development and school achievement, ranging from 5 to 37% of a standard deviation in experimental and quasi-experimental research, whereas children from disadvantaged backgrounds typically see the largest returns (Cooper and Stewart 2020).

Some child development periods are more critical than others. Investments during early childhood are more productive than later remediation investments, and while IQ scores stabilise at roughly age 10, non-cognitive skills and personality traits remain malleable until early adulthood (Cunha et al. 2006, 2010).¹ Recent evidence points at the desirability of balanced parental income levels throughout the childhood period, while investments are more productive during early childhood and adolescence, and less so during late childhood (Carneiro et al. 2021).

Evidence based on natural experiments is becoming increasingly common. Akee et al. studied the intergenerational effects of unexpected substantial governmental payments to Indian-American families (Akee et al. 2010), finding positive but mostly non-significant effects on years of education, parental education and a significantly higher probability of completing high school among treated students. However, the effects become much greater in magnitude and significance when restricting the sample to poor households. Similarly, Milligan and Stabile (2011) and Dahl and Lochner (2012) find greater impacts among disadvantaged children when increasing child benefits in Canada and the Earned Income Tax Credit in the USA, respectively.

A similarly valid source of randomness in income shocks is lotteries. Cesarini et al. (2016) used data of Swedish lottery winners to estimate its impact on longer-term educational outcomes such as school test scores at age 15 and military conscription scores, as well as on health outcomes. Certain significant effects were found on the latter but not on the former.

Natural resource windfalls such as the discovery of oil in Norwegian territory during the 1970s have also been used as an instrument to retrieve the causal impact of income on educational attainment (Løken 2010). While estimates are negative or close to zero and do not reach significance, when nonlinear estimators are used the results become aligned with what economic theory predicts, showing an increasing concave relationship between family income and child outcomes and sizeable marginal effects for disadvantaged families (Løken et al. 2012). Also in Norway, the introduction of a child care subsidy significantly increased children's later educational performance (Black et al. 2014).

The studies closest to the current one take advantage of two similar exogenous income shocks, namely the introduction in 2004 of a \$3000 Baby Bonus in Australia (Gaitz and Schurer 2017; Deutscher and Breunig 2018), and the implementation of the universal €2500 child benefit in Spain in 2007, which led to studies on its impact on births and abortions (González 2013; González and Trommlerová 2021a), female

¹ Apart from cognitive skills, personality traits are also influential in achievement test scores (Borghans et al. 2011), and thus interventions aimed at adolescents can still improve test scores through this channel (Guryan et al. 2021).

labour supply (González 2013; Hernández Alemán et al. 2017; González and Trommlerová 2021a), health (González and Trommlerová 2021b) and children human capital (Borra et al. 2021), being the latter the closest to my approach.²

2.2 Institutional background

2.2.1 The 2007 Spanish child benefit

On 3 July 2007, the Spanish Prime Minister unexpectedly announced the introduction of a universal child benefit, a one-time €2500 payment towards mothers of children born on and after 1 July 2007, being passed as a law a few months later (Law 35/2007). The main goal of the policy was to help parents face the early expenses of having a child, ultimately tackling population ageing by increasing fertility rates, which remained particularly low in comparison with other European countries (Eurostat 2007; Sobotka et al. 2019).

The eligibility criteria were straightforward, as every mother who was a Spanish national or who had resided in the country during the two years preceding the birth or adoption could apply for it. The take-up rates were high, covering at least 65% of births during the first months of the policy (Borra et al. 2021), and reaching a 95% coverage in children born in 2008 (González 2013).

The subsidy was not income-dependent nor changed with the number of births of a mother, and thus the amount paid per child corresponded to a shock that increased the 2007 average household income in Catalonia by almost 9% (INE 2008), and the minimum yearly individual income by 31.3% during the year when the child was born, arising from earning the 2007 minimum salary of €665.7/month in 12 payments (Ministerio de Trabajo y Asuntos Sociales 2006).

The policy was extended in 2009 to all mothers regardless of their years of residence in Spain until budgetary cuts during the financial crisis ended the payments on 31 December 2010.

2.2.2 The assessment of children's basic competencies in Catalonia

At the end of primary school (age 11–12), students in the Catalonia region complete standardised tests that assess the basic skills that should have been learned during their time at school. The tests have been designed and run yearly since 2009 by the Catalan governmental agency in charge of evaluating the educational system (Consell Superior d'Avaluació del Sistema Educatiu, CSDA in short), which also develops a similar set of tests written at the end of secondary school.

The exams take place during April or early May and are spread along two or three days. They are written in class during lecture hours and externally graded with scores ranging from 0 to 100. The results do not affect any academic outcome of the student,

² Canada also implemented a baby bonus scheme, but with a different design and payments increasing with the number of children born. See for instance Malak et al. (2019) and Milligan (2005) for evidence on its effects and for further references. In addition, the Swiss baby bonus has only been evaluated in terms of changes in fertility and children health (Chuard and Chuard-Keller 2021).

including access to secondary school. However, they are useful for schools to improve their educational practice, and parents do receive a report with the contextualised results of their child.

The content of the tests has been evolving since their creation, and the exams currently test skills in five subjects: oral and reading comprehension and writing skills in Catalan, Spanish and foreign language,³ mathematics (numeration and calculation, space, data measurement and graphical representation of data, relationships, and change) and natural environment⁴ (Departament d'Educació 2021).

3 Identification strategy

3.1 Regression discontinuity design (RD)

The features of the implementation of the benefit create a natural experiment in which children's dates of birth around the threshold are as good as random, and so is their families' eligibility for the subsidy. The assignment rule is simple, and can be defined as

$$D_i = \begin{cases} 1 & \text{if } X_i \geq X^T \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Being born on or after the 1 July 2007 threshold makes parents eligible to receive the subsidy.⁵ This creates a sharp eligibility cut-off D_i as Eq. (1) shows, equalling one if the running variable X_i , which measures the distance between a certain date of birth and the threshold, is equal or greater than the threshold date of birth X^T .

A regression discontinuity design identifies the causal effect under the condition that potential outcomes are continuous at the cut-off. All factors determining test scores other than child benefit eligibility must evolve smoothly at the threshold, and so the date of birth must be the only discontinuity determining eligibility. Visual evidence of continuity of predetermined covariates at the threshold is found in Fig. 1, where no jump is statistically significant at the threshold. The assignment to treatment at the threshold must be as good as random, implying that individuals are unable to perfectly manipulate their assignment to treatment (i.e. parents are unable to strategically modify the date of birth of their children), and treated and untreated children born in similar dates around the cut-off can therefore be compared. This assumption is tested in Fig. 2, finding no evidence of sorting, as the null hypothesis of continuity around the July 1,

³ While a reduced number of schools might choose to offer another main foreign language, this study only focuses on the English language results. Similarly, in the Val d' Aran county, students also faced an Aranese language skills test, although those results are also excluded from the analysis.

⁴ Since this part of the test was added in 2018 and because it assesses skills from different fields that are unrelated to each other and difficult to disentangle, the results of this part are excluded from the study.

⁵ Notice the choice of wording here since, as mentioned above, the take-up of the subsidy was very high but not complete. I do not specifically observe which students benefited from the measure, and therefore the analysis is measuring a sort of intention-to-treat effect, where the probability of treatment below the cut-off is 0, and the probability of treatment above the cut-off is less than 1.

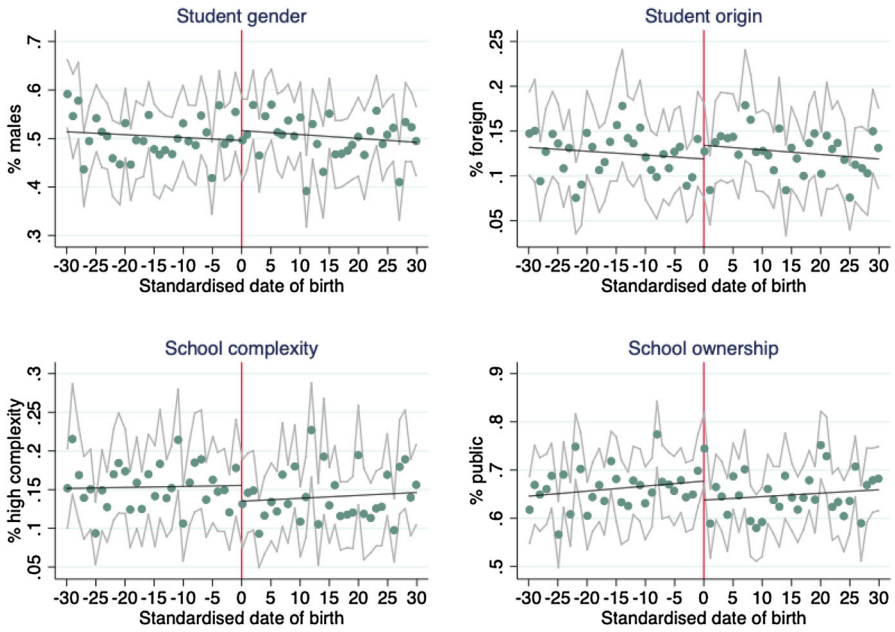


Fig. 1 Continuity test RD plots, 2007 cohort

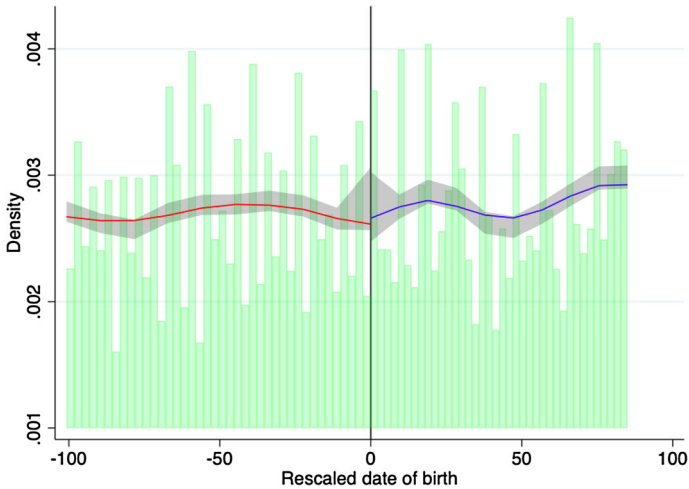


Fig. 2 Density test at the eligibility threshold, 2007 cohort

2007 cut-off cannot be rejected, with a t value of -0.90 and a p value of 0.37 . This was expected due to the timing of the announcement of the policy.

I estimate the local average treatment effect (LATE) of the subsidy for students close to the cut-off non-parametrically, using local linear regression with a triangular kernel smoothing that gives more weight to observations close to the threshold and the

tools developed by Calonico et al. (2014), including robust bias-corrected confidence intervals and data-driven bandwidth selection.

The model is defined as

$$Y_i = \theta_0 + \theta_1 D_i + \theta_2 f(X_i) + \theta_3 D_i f(X_i) + \varepsilon_i, \quad (2)$$

where θ_1 is the coefficient of interest that estimates the LATE of the subsidy on the test score Y_i , which is standardised at the subject-cohort level. X_i is the running variable, normalised at the threshold and allowed to have different functional forms, and θ_3 is the coefficient of the interaction of the subsidy with the running variable so that the regression function can differ in each side of the cut-off, as proposed by Lee and Lemieux (2010).

3.2 Difference-in-discontinuities (diff-in-disc)

If unobserved variables feature seasonality and differ at each side of the cut-off, the estimates can become biased and violate the continuity assumption. As the next section discusses, it can be the case with the test scores data. To overcome those potential issues, we can eliminate unobserved variables potentially biasing the results by comparing the 2007 RD with the previous two years without a subsidy in place, using a difference-in-discontinuities design. This method was formalised by Grembi et al. (2016). It could be seen as a two-period extension of a regression discontinuity design, where the diff-in-disc estimator takes the difference between RD discontinuities before and after the introduction of the subsidy. The assignment rule becomes:

$$D_i = \begin{cases} 1 & \text{if } X_i \geq X^T, t \geq T_1 \\ 0 & \text{otherwise} \end{cases}, \quad (3)$$

where T_1 is the treatment year 2007, while the diff-in-disc model can be written as follows:

$$Y_i = \alpha_0 + \alpha_1 X_i + D_i(\beta_0 + \beta_1 X_i) + T_i[\gamma_0 + \gamma_1 X_i + D_i(\delta_0 + \delta_1 X_i)] + v_i. \quad (4)$$

Y_i denotes student i 's standardised grade in a certain test. D_i is a dummy indicating a date of birth later than July 1. T_i equals one if the year is 2007 (the treatment year) and 0 otherwise. X_i is the normalised running variable, measuring the distance in days from the threshold date, δ_0 is the parameter of interest, measuring the interaction between D_i and T_i , while v_i is the error term.

The model is also estimated non-parametrically, following the extensions of the work by Calonico et al. (2014) to difference-in-discontinuities developed by Rafael P. Ribas (Ribas 2016; Giambona and Ribas 2021), providing the precise econometric causal reasoning behind this method.

In order to estimate the LATE of the subsidy, three identifying assumptions need to hold, borrowed from Grembi et al. (2016): (i) all potential outcomes in all time periods are continuous in X at X^T ; (ii) the effect of confounding variables at X^T in

the case of no treatment is constant over time; (iii) the effect of the subsidy at X^T does not depend on the confounding variables, they do not interact with each other. I test each assumption empirically in Sect. 5.2.1.

4 Data

4.1 Data sources

I use individual microdata from the CSDA⁶ containing the exact date of birth of each student and their results in the Catalan, Spanish, English and mathematics tests at the end of primary school, the demographics of each student (gender, county of origin, Spanish/foreign origin), as well as and characteristics of the type of school they attend (public or private ownership, degree of socioeconomic complexity⁷). The data covers the results for three years: 2019, when most children born in the treatment year, 2007, finished primary education, and 2017 and 2018, which serve as controls. The full sample contains 220,804 observations of children who took at least one exam.

The subsample chosen for analysis encompasses all students turning 12 in 2017, 2018 and 2019 who participated in the basic competency tests. Older students who might have repeated a grade (roughly 5% in each cohort) and those who are younger but were advanced to higher grades (less than 0.2% in each cohort) were excluded from the analysis as they were not eligible for the subsidy and had specific characteristics. Foreign students were also excluded from the main analyses, as it is less likely that they benefited from the subsidy. A small number of observations without a linked date of birth were also deleted.

In addition, registry microdata of the total number of sixth-grade students was obtained from the Department of Education of the Catalan government (Departament d'Educació 2019).

4.2 Descriptive statistics

Naturally, not every student took the tests. The database containing the whole universe of sixth-grade students in Catalonia allows for a certain comparison of demographics to rule out sorting.⁸ In 2017, 90.51% of all students took at least one exam. This figure grew to 93.68% in 2018 but declined to 87.01% in 2019. Still, those figures imply an improvement with respect to previous evidence on the topic in Catalonia, where the

⁶ Since they were obtained under a confidentiality agreement, the data sets are not publicly available, but can be requested to the agency for research purposes.

⁷ The degree of complexity of a school is defined according to a set of socioeconomic criteria, including origin, parental education and occupation, and the presence of students with special educational needs who increase the administrative complexity of the school. See Consell Superior d'Avaluació del Sistema Educatiu (2021b) for further details [in Catalan].

⁸ This comparison can be imprecise, as the registered students dataset includes older students retaking a grade and younger promoted students, as well as foreign pupils, while the studied sample is limited to Spanish nationals progressing normally according to their year of birth. However, the number of students in each cohort born earlier and later than expected tends to remain similar over time. Retakers are foreign to a greater extent than non-retakers.

average grades of 70% of second and third-year students were observed only in public schools (Borra et al. 2021).

Disparities in samples can be attributed to students randomly missing school on the day of the test, students not taking the tests due to parental opposition, foreign students being in the process of integration before being able to attend regular lessons with the rest of the class while still being registered as sixth graders, and reporting errors. Sorting by schools (i.e. excluding bad students to avoid reputational consequences or emphasising their poor results in order to receive extra funding) is very unlikely, as school public funds do not depend on student performance and the complexity status depends on student and parental socioeconomic variables, not on school performance. In addition, external monitoring of the testing process takes place, aiming to prevent this strategic exclusion of students. The data does not allow for further investigation of this phenomenon, but if it existed, it would mostly affect the representativeness of foreign students, which are excluded from the studied sample, and would not be expected to vary at the threshold nor over time.

Table 1 shows the main descriptive statistics of the variables used in the analysis, using the subsample of Spanish nationals progressing normally with their studies (e.g. students who have not retaken any grade nor have been promoted to higher grades due to outstanding abilities). Even though they assess the same competencies, the content of the exams is modified every year, and average grades differ slightly for each cohort. While minimum grades of zero do exist in some cases, they are found in no more than a handful of observations for each subject and therefore do not drive the results. Student demographic variables are comparable between cohorts, with the notable exception of foreign origin students, which spike in the 2019 tests. This increase in foreign students taking the test, both in public and private schools, is not explained by the proportion of foreign students registered at grade six, which increases at a similar pace than the previous year.

While this phenomenon might not be random, as it could be linked to a sharp increase in students having arrived in Catalonia very recently (Consell Superior d'Avaluació del Sistema Educatiu 2021a), but at the same time it might also imply that less local students took the tests, it should not affect the conclusions significantly since, as mentioned above, incentives by schools to overrepresent or hide poor performers do not exist. Most importantly, there is no spike in the proportion of foreign students at the 1 July 2007 threshold compared to previous years, as shown in Table 4 in Sect. 5.2.1 Still, the remaining observable variables, including school characteristics, can be compared across study periods.

Table 6 in the “Appendix” reports the average grades by subject and gender. Girls consistently outperform boys in languages across all levels of school complexity, while boys perform better in mathematics, a pattern usually found in the literature (Niederle and Vesterlund 2010). This also provides suggestive evidence that tests were not perceived as high-stakes by students, as when stakes are high, boys tend to close the gender gap and perform better (Azmat et al. 2016; Montolio and Taberner 2021).

While data on parental income and education is unavailable, the best proxy available is the socioeconomic complexity of the school. Figure 3 shows the grade distribution by school complexity. The main feature is larger tails at the left of the distribution

Table 1 Descriptive statistics by student year of birth

	N	Mean	SD	Min	Max
<i>2005 (2017 tests)</i>					
Catalan	67,693	76.73	15.376	0	100
Spanish	67,123	75.1	19.192	0	100
English	66,757	80.884	15.759	0	100
Mathematics	67,988	78.526	16.558	5.917	100
Gender (1 = male)	67,988	0.502	0.5	0	1
Origin (1 = foreign)	67,988	0.053	0.223	0	1
Low school complexity	67,988	0.182	0.386	0	1
Medium school complexity	67,988	0.683	0.465	0	1
High school complexity	67,988	0.135	0.342	0	1
School ownership (1 = public)	67,988	0.648	0.478	0	1
<i>2006 (2018 tests)</i>					
Catalan	68,794	74.162	12.701	17.29	97.727
Spanish	69,206	77.566	12.405	0	97.727
English	68,882	77.148	16.758	0	97.5
Mathematics	69,183	79.974	15.946	0	100
Gender (1 = male)	70,075	0.502	0.5	0	1
Origin (1 = foreign)	70,075	0.05	0.217	0	1
Low school complexity	70,075	0.18	0.384	0	1
Medium school complexity	70,075	0.673	0.469	0	1
High school complexity	70,075	0.147	0.354	0	1
School ownership (1 = public)	70,075	0.654	0.476	0	1
<i>2007 (2019 tests)</i>					
Catalan	69,338	76.138	14.309	0	100
Spanish	69,885	75.187	16.114	0	100
English	69,584	79.108	18.727	0	100
Mathematics	69,702	78.659	12.388	29.784	100
Gender (1 = male)	70,973	0.501	0.5	0	1
Origin (1 = foreign)	70,973	0.125	0.331	0	1
Low school complexity	70,973	0.178	0.382	0	1
Medium school complexity	70,973	0.675	0.469	0	1
High school complexity	70,973	0.148	0.355	0	1
School ownership (1 = public)	70,973	0.658	0.474	0	1

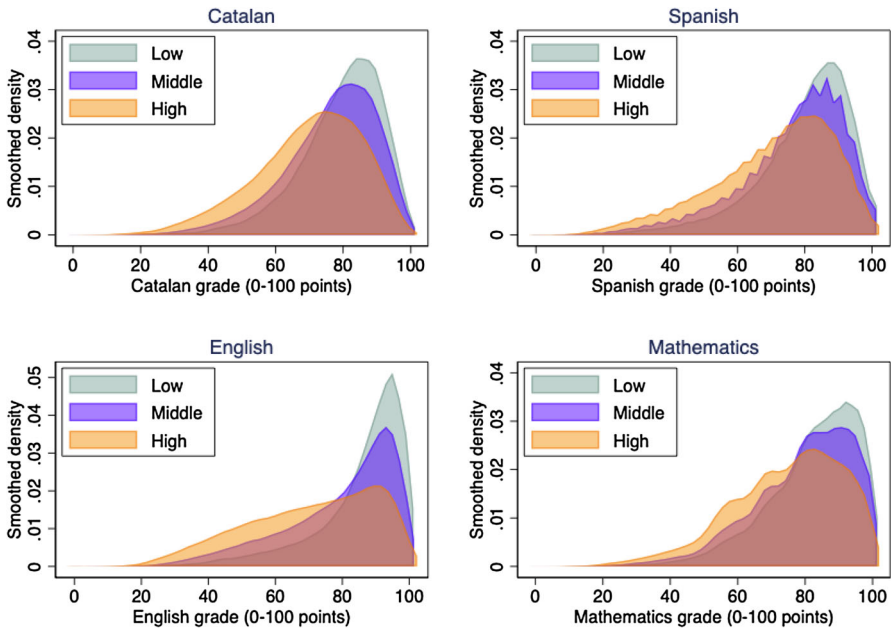


Fig. 3 Grade distributions by subject and school complexity level, pooled 2017–2019 sample

the higher the complexity. A change in the distribution shape for high-complexity schools is especially relevant in the English test, where lower-than-average results are usual. This might be explained by the common practice by many middle- and upper-class families of providing their children with extracurricular English lessons at private language schools from early ages in order to advance at a higher pace than the school, which only aims to achieve the B1 level of the Common European Framework of Reference for Languages by the end of secondary school. Private primary schools might also enjoy more resources to promote foreign-language learning.

It is well known that children born later in the year start school having developed their cognitive and non-cognitive abilities less than their peers, with a disadvantage that can be reduced but is carried on during their lives and can lead to year repetition.⁹ Figure 4 shows this clear declining pattern in the 2019 results, with up to four points separating the average test scores of students born in January and December.

A potential source of concern is the significant observed differences between average grades above and below the July cut-off (e.g. grades of children born in May or June compared to August and September). This phenomenon can bias the results downwards and could imply a violation of the continuity assumption, as the effect of being born later could be attributed to the subsidy. In addition, concerns about different parental educational levels before and after the July threshold can also be relevant (Berniell and Estrada 2020), as mechanisms such as parental time spent with children

⁹ See Bedard and Dhuey (2006). Calsamiglia and Loviglio (2020) and Berniell and Estrada (2020) provide evidence of student birth developmental differences within the Catalan educational system and how parents try to compensate them, respectively.

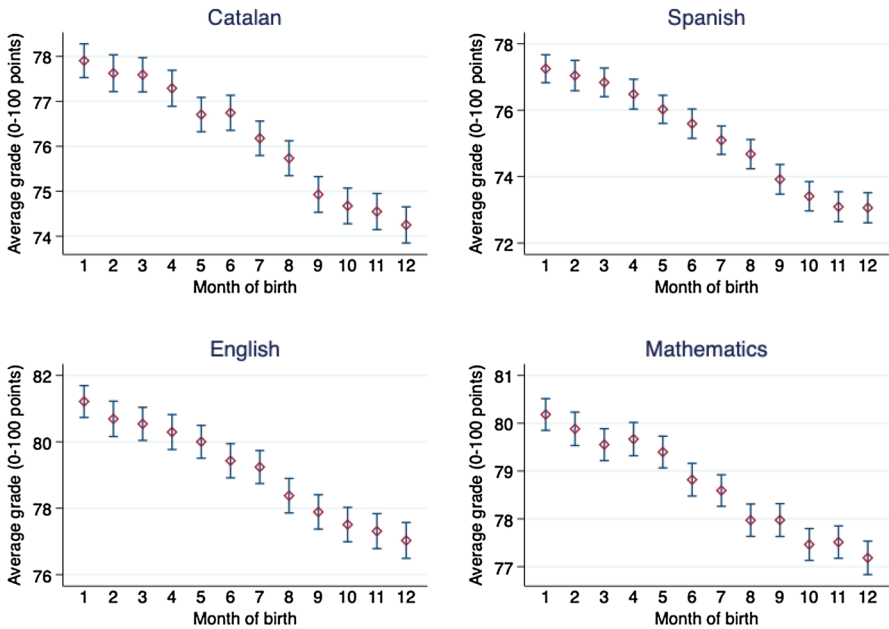


Fig. 4 Average student performance in 2019 tests by month of birth and subject, 2007 cohort. Note: 95% confidence intervals in blue

depend strongly on parental income and education (Guryan et al. 2012). Thus, this is the fundamental reason behind the use of a diff-in-disc empirical strategy.

5 Results

5.1 Regression discontinuity design

Figure 5 gives a visual overview of regression discontinuity plots for each subject using a 30-day bandwidth and a linear fit. No significant jumps are observed in any of the subjects, implying that, overall, children whose mothers were eligible to receive the subsidy did not perform better than those who did not benefit from it.

Table 2 reports RD coefficients for each subject from separate regressions, as well as its standard error, the optimal bandwidth in days and the number of effective observations used in each optimal bandwidth. It covers the treatment year 2007, using observations of the 2019 tests above and below the eligibility cut-off.

Throughout this section, regression tables are organised as follows. Column 1 considers the whole sample of schools and all students who are Spanish citizens and who have not retaken any year. Columns 2 and 3 restrict the sample to non-disadvantaged schools, that is, schools categorised as having low and middle levels of socioeconomic complexity. Column 2 reports the results for the subsample of boys attending those schools, and column 3 does the same for girls. Columns 4 and 5 focus on complex

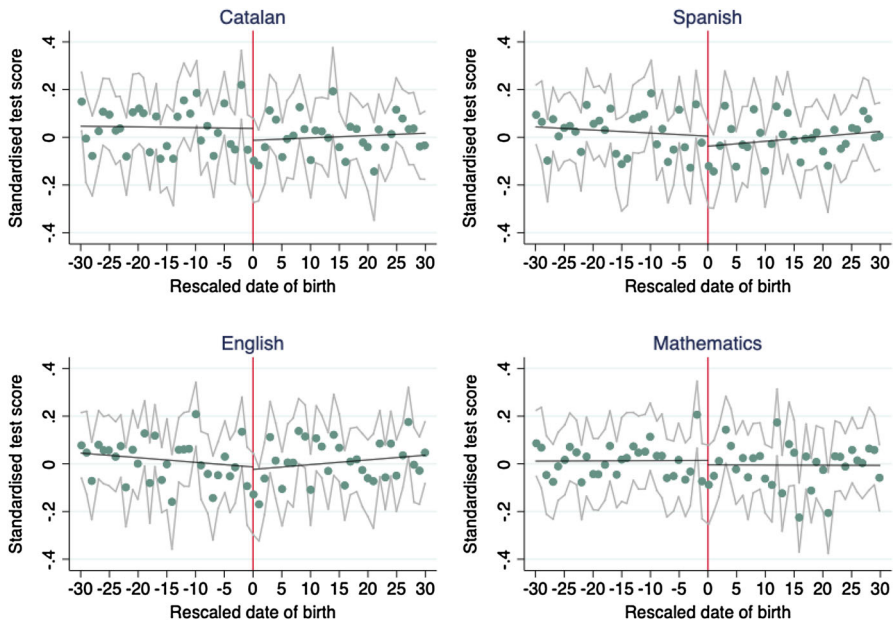


Fig. 5 RD plots by subject, 2019 tests

schools, while columns 6 and 7 on public complex schools, which typically contain the most vulnerable students and families.

The results in Table 2 indicate a modest negative effect of the subsidy on Catalan language grades overall, decreasing scores by 5% of a standard deviation, although it does not reach statistical significance. It is driven by boys attending schools with low and middle levels of complexity, who see a decline in scores of 0.113 standard deviation units, significant at the 90% level. The effect for boys becomes positive within disadvantaged schools, with a score increase of 0.207 standard deviation units, and approaching a potential improvement of up to almost 60% of a standard deviation with 95% confidence. However, none of those coefficients reaches statistical significance. In girls, coefficients remain small and negative across all levels of school complexity, although we cannot rule out coefficients as large as 0.25 standard deviation units.

Regarding Spanish language scores, we can observe a similar pattern in signs in the first three columns, ruling out positive effects greater than 0.05 standard deviation units. Estimates for boys in disadvantaged schools and public schools increase in magnitude and robustness, reaching certain statistical significance. English language estimates remain relatively close to zero overall in girls even though improvements of up to 0.26 standard deviation units cannot be ruled out, while in boys scores increase by 35% of a standard deviation in disadvantaged schools and by 43% in disadvantaged public schools. Finally, scores in mathematics remain similar across non-disadvantaged schools, ruling out effects greater than 8% of a standard deviation. In disadvantaged environments, girls' coefficients become positive, while only in the case of boys in

Table 2 RD regression results, 2019 tests

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
	Pooled		Non-disadvantaged schools		Disadvantaged schools		Disadvantaged schools		Disadvantaged schools		Disadvantaged schools		Disadvantaged public schools	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
<i>Catalan</i>	-0.0511	-0.0314	-0.113*	0.181	0.181	-0.00982	0.207	-0.0199	0.207	-0.00982	0.207	-0.0199	0.207	-0.0199
Standard error	(0.0363)	(0.0448)	(0.0531)	(0.133)	(0.133)	(0.108)	(0.144)	(0.112)	(0.144)	(0.108)	(0.144)	(0.112)	(0.144)	(0.112)
95% confidence interval	[-0.144, 0.0219]	[-0.148, 0.0559]	[-0.246, 0.00005]	[-0.113, 0.509]	[-0.113, 0.509]	[-0.266, 0.242]	[-0.089, 0.575]	[-0.28, 0.25]	[-0.089, 0.575]	[-0.266, 0.242]	[-0.089, 0.575]	[-0.28, 0.25]	[-0.089, 0.575]	[-0.28, 0.25]
Optimal bandwidth	42.814	57.841	45.790	56.421	56.421	72.056	53.636	72.727	53.636	72.056	53.636	72.727	53.636	72.727
Observations	14,175	8,128	6,473	1,312	1,312	1,770	1,126	1,638	1,126	1,770	1,126	1,638	1,126	1,638
<i>Spanish</i>	-0.0305	-0.0361	-0.0728	0.251*	0.251*	-0.0294	0.310**	-0.0263	0.310**	-0.0294	0.310**	-0.0263	0.310**	-0.0263
Standard error	(0.0359)	(0.0432)	(0.0547)	(0.141)	(0.141)	(0.106)	(0.155)	(0.11)	(0.155)	(0.106)	(0.155)	(0.11)	(0.155)	(0.11)
95% confidence interval	[-0.121, 0.046]	[-0.148, 0.0506]	[-0.204, 0.0527]	[-0.0452, 0.608]	[-0.0452, 0.608]	[-0.288, 0.21]	[0.017, 0.718]	[-0.289, 0.233]	[0.017, 0.718]	[-0.288, 0.21]	[0.017, 0.718]	[-0.289, 0.233]	[0.017, 0.718]	[-0.289, 0.233]
Optimal bandwidth	44.903	60.677	46.626	53.040	53.040	71.667	47.449	72.907	47.449	71.667	47.449	72.907	47.449	72.907
Observations	14,938	8,673	6,693	1,266	1,266	1,775	1,024	1,656	1,024	1,775	1,024	1,656	1,024	1,656
<i>English</i>	-0.00321	-0.0064	-0.00742	0.351***	0.351***	-0.0336	0.426***	-0.0138	0.426***	-0.0336	0.426***	-0.0138	0.426***	-0.0138
Standard error	(0.0365)	(0.0503)	(0.0549)	(0.128)	(0.128)	(0.127)	(0.145)	(0.123)	(0.145)	(0.127)	(0.145)	(0.123)	(0.145)	(0.123)
95% confidence interval	[-0.0947, 0.0747]	[-0.196, 0.0343]	[-0.113, 0.127]	[0.0877, 0.679]	[0.0877, 0.679]	[-0.0365, 0.229]	[0.159, 0.807]	[-0.32, 0.267]	[0.159, 0.807]	[-0.0365, 0.229]	[0.159, 0.807]	[-0.32, 0.267]	[0.159, 0.807]	[-0.32, 0.267]
Optimal bandwidth	43.972	43.667	46.047	57.203	57.203	56.684	49.424	65.640	49.424	56.684	49.424	65.640	49.424	65.640
Observations	14,520	6,206	6,635	1,349	1,349	1,436	1,056	1,522	1,056	1,436	1,056	1,522	1,056	1,522

Table 2 (continued)

	(2) Non-disadvantaged schools		(3) Disadvantaged schools		(4) Disadvantaged schools		(5) Disadvantaged schools		(6) Disadvantaged public schools		(7) Disadvantaged public schools	
	Pooled	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	
Mathematics	0.00446	-0.0373	-0.0167	0.143	0.156	0.221*	0.109	0.221*	0.109	0.221*	0.109	
Standard error	(0.0282)	(0.0455)	(0.0438)	(0.119)	(0.115)	(0.134)	(0.119)	(0.115)	(0.119)	(0.134)	(0.119)	
95% confidence interval	[-0.0568, 0.0753]	[-0.135, 0.0768]	[-0.124, 0.0829]	[-0.118, 0.439]	[-0.133, 0.41]	[-0.0402, 0.572]	[-0.184, 0.381]	[-0.133, 0.41]	[-0.184, 0.381]	[-0.0402, 0.572]	[-0.184, 0.381]	
Optimal bandwidth	68,986	68,960	72,268	74,536	57,247	74,869	56,356	57,247	56,356	74,869	56,356	
Observations	23,572	8,686	10,132	1,266	1,449	1,054	1,359	1,449	1,359	1,054	1,359	

This table shows the regression discontinuity results of the 2007 cohort as in Eq. (2), using triangular kernels and normalised grades $N(0,1)$. Each coefficient comes from a different regression, only the coefficient of interest θ_1 is reported. Optimal bandwidth chosen according to Calonico et al. (2014)

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

disadvantaged public schools does the positive coefficient become marginally significant.

5.2 Difference-in-discontinuities

To eliminate potential seasonality as a source of bias, Table 3 reports difference-in-discontinuities estimates, using 2019 exam-takers and 2017 and 2018 tests as controls, following Eq. (4). In the first column, coefficients remain similar to those reported in Table 2, and from the confidence intervals we can rule out effects greater than 0.1 standard deviation units. In the subsample of boys in non-disadvantaged schools, negative effects are found in the four subjects, including marginally significant coefficients in the Catalan and Spanish language test scores. Regarding English and mathematics, we can rule out effects greater than 0.14 and 0.12 standard deviation units, respectively. In girls, English language coefficients double in magnitude with respect to the previous table but do not reach statistical significance, and effects larger than 0.16 standard deviation units can be rejected. The magnitude of the effects is unlikely to be relevant, as they imply changes in grades of less than 1% of the average test score, only reaching 1% in the Catalan and Spanish tests for boys.

In the disadvantaged school subsamples, while benefiting from a higher sample size than before, coefficients lose all statistical significance, even though their size remains similar and we cannot rule out test score increases as high as 0.95 standard deviation units in boys in public disadvantaged schools for the English test, and slightly lower for the remaining subjects. In addition, girls' coefficients become positive in the diff-in-disc setting, and we can rule out grade increases larger than roughly half a standard deviation unit.

In terms of magnitude, the effects in the subsample of disadvantaged schools become relevant. The coefficients for boys imply increases of 5.6%, 6%, 7.9% and 2.4% of the average Catalan, Spanish, English and mathematics test score, reaching 7%, 8.5%, 11% and 4% in disadvantaged public schools. To a lesser extent, girls' grade improvements also become relevant in this subsample, increasing the average test score by 1.6%, 1%, 2.2% and 4.1%, respectively. Still, the lack of statistical significance adds caution to the causal interpretation of some of the coefficients found in Table 2.

Results using polynomial specifications are reported in Tables 7 and 8 in the “Appendix”. While coefficients remain similar to those of the linear specification in Table 3, in the first table, two coefficients reach a marginal statistical significance, translating to a decrease in Catalan grades in boys from non-disadvantaged schools of 16% of a standard deviation, and an increase in English grades for boys in disadvantaged schools of almost half of a standard deviation. While its use is discouraged (Gelman and Imbens 2019), third-order polynomials do not change the coefficients but remove all statistical significance.

Other heterogeneity analyses are also located in the “Appendix”. In Table 9, I report how the estimates differ by school size. Small schools, those with 30 students or less taking the exam, seem to drive the negative results of the pooled sample, bordering a fifth of a standard deviation decrease in scores except for mathematics.

Table 3 Difference-in-discontinuities regression results

	(1)	(2)		(3)		(4)		(5)		(6)		(7)
		Non-disadvantaged schools		Disadvantaged schools		Disadvantaged schools		Disadvantaged schools		Disadvantaged public schools		
	Pooled	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	
<i>Catalan</i>	– 0.0335 (0.0539)	– 0.149* (0.0768)	0.00568 (0.0823)	0.263 (0.225)	0.0176 (0.218)	0.314 (0.238)	0.0801 (0.241)					
95% confidence interval	[– 0.139, 0.0721]	[– 0.3, 0.001]	[– 0.156, 0.167]	[– 0.177, 0.703]	[– 0.41, 0.446]	[– 0.152, 0.78]	[– 0.391, 0.552]					
Optimal bandwidth	60.414	70.05	54.288	62.875	52.891	62.667	48.386					
Observations	63,137	32,084	24,334	4,342	3,688	3,957	3,119					
<i>Spanish</i>	– 0.0588 (0.0525)	– 0.141* (0.0851)	– 0.0393 (0.0847)	0.251 (0.245)	0.0106 (0.191)	0.340 (0.289)	0.0447 (0.214)					
95% confidence interval	[– 0.162, 0.0442]	[– 0.308, 0.026]	[– 0.205, 0.127]	[– 0.229, 0.731]	[– 0.363, 0.385]	[– 0.226, 0.905]	[– 0.375, 0.464]					
Optimal bandwidth	64.948	59.701	48.553	56.658	74.738	47.792	64.115					
Observations	67,044	26,943	21,615	3,938	5,216	3,052	4,179					
<i>English</i>	– 0.031 (0.0616)	– 0.0494 (0.101)	– 0.134 (0.0895)	0.305 (0.247)	0.0003 (0.209)	0.407 (0.277)	0.0914 (0.230)					
95% confidence interval	[– 0.152, 0.0896]	[– 0.247, 0.148]	[– 0.31, 0.0413]	[– 0.179, 0.789]	[– 0.409, 0.409]	[– 0.135, 0.95]	[– 0.359, 0.541]					
Optimal bandwidth	52.592	45.919	47.645	52.764	75.426	46.472	63.732					
Observations	54,170	20,429	21,091	3,650	5,257	2,958	4,105					

Table 3 (continued)

	(2) Non-disadvantaged schools		(3) Disadvantaged schools		(4) Disadvantaged schools		(5) Disadvantaged schools		(6) Disadvantaged public schools		(7)
	Pooled	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
<i>Mathematics</i>	0.00791	- 0.0343	0.00566	0.128	0.200	0.201	0.193	0.201	0.201	0.201	0.193
Standard error	(0.0467)	(0.0803)	(0.0819)	(0.186)	(0.187)	(0.207)	(0.197)	(0.187)	(0.207)	(0.207)	(0.197)
95% confidence interval	[- 0.0835, 0.0994]	[- 0.192, 0.123]	[- 0.155, 0.166]	[- 0.237, 0.494]	[- 0.165, 0.566]	[- 0.204, 0.606]	[- 0.192, 0.578]	[- 0.165, 0.566]	[- 0.204, 0.606]	[- 0.204, 0.606]	[- 0.192, 0.578]
Optimal bandwidth	67.8	50.934	52.765	62.279	65.818	58.289	62.577	65.818	58.289	58.289	62.577
Observations	70,691	22,834	23,442	4,405	4,643	3,746	4,062	4,643	3,746	3,746	4,062

This table shows the diff-in-disc regression results of the 2019 tests with the tests of the previous two years acting as placebo tests, following Eq. (4), using triangular kernels and normalised grades $N(0,1)$. Each coefficient comes from a different regression, only the coefficient of interest δ_0 is reported. Optimal bandwidth chosen according to Calonico et al. (2014) $***p < 0.01, **p < 0.05, * p < 0.1$

For the remaining school sizes, the effects are mildly positive and remain statistically insignificant.

I also investigate whether the results hold in the subsample of foreign students, who benefited from the subsidy to a much lesser extent, as only mothers who had resided in Spain during the two years before birth were eligible to receive it. The results for the pooled sample are presented in Table 10. No coefficients reach statistical significance, even though they are comparable in magnitude with the subsample of Spanish nationals in the main analysis except for mathematics.

5.2.1 Robustness checks

In this subsection, I challenge the results to (i) the choice of bandwidth, ii) the validity of the diff-in-disc assumptions from Sect. 3.2, and iii) the possibility that the results are due to random chance. Figure 6 shows the sensitivity of the results to the choice of bandwidth for the pooled sample. Coefficients remain very close to zero, implying no effect of the subsidy. Figures for the remaining subsamples are available upon request and generally maintain the coefficient sign or remain close to zero across bandwidths.

Next, I test the diff-in-disc assumptions. In Table 4, I test Assumption 1 using diff-in-disc specifications with a set of outcomes that should not vary at the threshold over time. The significant coefficients in the high-complexity and ownership cases suggest that the assumption of continuity in potential outcomes at the threshold might not hold over different periods. Considering this, in Table 11 in the “Appendix” report the

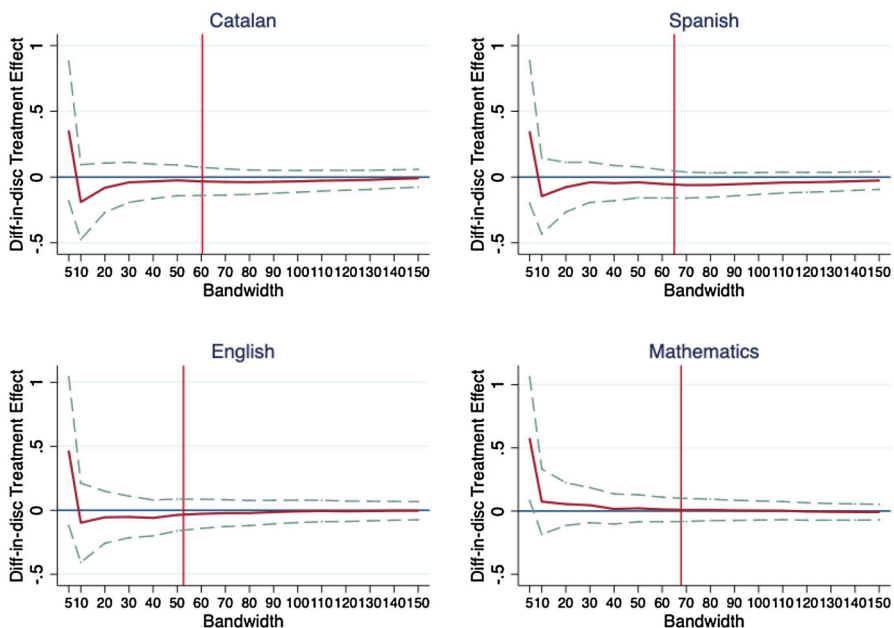


Fig. 6 Diff-in-disc sensitivity to bandwidth choice, pooled sample. *Note:* 95% confidence interval in dashed lines. The vertical line indicates the optimal bandwidth from Table 3

Table 4 Diff-in-disc continuity test

	(1) Sex	(2) Foreign	(3) High complexity	(4) Ownership
Coefficient	0.01	0.0115	− 0 .0296*	−0 .048**
Standard error	(0.0224)	(0.013)	(0.0159)	(0.0236)
Optimal bandwidth	58.198	55.954	59.593	49.379
Observations	61,565	63,339	62,790	51,878

This table shows the diff-in-disc regression results of a placebo treatment of the 2006 cohort and a control group of the 2005 cohort, following Eq. (2), using triangular kernels and normalised grades $N(0,1)$. Each coefficient comes from a different regression, only the coefficient of interest δ_0 is reported. Optimal bandwidth chosen according to Calonico et al. (2014)

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

diff-in-disc results using month of birth and school ownership as control variables, with coefficients remaining roughly similar to those in Table 3.

Assumption 2 is tested in the “Appendix” Figs. 8 and 9, which show regression discontinuity plots of the two years without treatment. The lack of significant discontinuities created by confounding variables at the cut-off in years with no treatment implies that the assumption holds.

Table 5 reports the placebo diff-in-disc estimates arising from using the 2018 tests as the placebo treatment year and using the cohort who did the exams in 2017 as a control group. This effectively tests Assumption 3 and serves to rule out sizeable changes in test scores that can happen from year to year and could bias diff-in-disc results in Table 3.

No result should be significantly different from zero. However, Spanish language scores for boys increase by more than three-quarters of a standard deviation, with grade increases within the 95% confidence level of up to more than one standard deviation. Coefficients for boys in disadvantaged schools remain very similar to the diff-in-disc analysis in the previous table, while coefficients for girls remain negative, lower in magnitude and statistically insignificant. The lack of robustness of the results to the placebo estimates cast serious doubts over the findings, as they could be the result of natural variation in scores from one year to another at both sides of the cut-off.

Finally, in Fig. 7, I combine into cumulative distribution functions for each subject the coefficients of 82 diff-in-disc regressions using placebo treatment cut-offs at birth dates away from 1 July 2007, 41 on each side of the real cut-off, in order to further confirm that the results are due to chance. To rule out that the results are due to chance, one would expect almost none of those coefficients to be greater than the actual cut-off coefficient and its opposite sign, that is, to not be located outside the range between the vertical red lines. However, the way in which a significant number of coefficients are greater in magnitude than the coefficient using the real cut-off and its opposite sign confirms the hypothesis, which is likely to apply to the other subsamples as well as to the full sample.

Table 5 Placebo diff-in-disc regression results

(1)	(2)		(3)		(4)		(5)		(6)		(7)	
	Non-disadvantaged schools		Disadvantaged schools		Disadvantaged schools		Disadvantaged schools		Disadvantaged public schools		Disadvantaged public schools	
Pooled	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
<i>Catalan</i>	0.0588	0.126	− 0.0225	− 0.159	0.450	− 0.159	0.471	− 0.172	0.471	− 0.172	0.471	− 0.172
Standard error	(0.0586)	(0.105)	(0.0756)	(0.266)	(0.314)	(0.266)	(0.338)	(0.289)	(0.338)	(0.289)	(0.338)	(0.289)
95% confidence interval	[− 0.0561, 0.174]	[− 0.0792, 0.331]	[− 0.171, 0.126]	[− 0.68, 0.362]	[− 0.166, 1.0662]	[− 0.68, 0.362]	[− 0.192, 1.135]	[− 0.738, 0.395]	[− 0.192, 1.135]	[− 0.738, 0.395]	[− 0.192, 1.135]	[− 0.738, 0.395]
Optimal bandwidth	65.218	49.779	70.238	43.147	42.914	43.147	41.377	41.051	41.377	41.051	41.377	41.051
Observations	46,381	15,248	21,687	1,988	2,013	1,988	1,793	1,735	1,793	1,735	1,793	1,735
<i>Spanish</i>	0.0162	− 0.0204	− 0.0294	− 0.184	0.766**	− 0.184	0.779**	− 0.223	0.779**	− 0.223	0.779**	− 0.223
Standard error	(0.0569)	(0.0888)	(0.0829)	(0.306)	(0.351)	(0.306)	(0.365)	(0.325)	(0.365)	(0.325)	(0.365)	(0.325)
95% confidence interval	[− 0.0952, 0.128]	[− 0.194, 0.154]	[− 0.192, 0.133]	[− 0.0785, 0.417]	[0.0782, 1.453]	[− 0.0785, 0.417]	[0.0646, 1.494]	[− 0.86, 0.413]	[0.0646, 1.494]	[− 0.86, 0.413]	[0.0646, 1.494]	[− 0.86, 0.413]
Optimal bandwidth	71.937	68.883	59.616	41.792	35.721	41.792	36.597	40.362	36.597	40.362	36.597	40.362
Observations	50,609	21,260	18,259	1,898	1,674	1,898	1,576	1,701	1,576	1,701	1,576	1,701
<i>English</i>	− 0.00254	− 0.0195	− 0.0247	− 0.107	0.255	− 0.107	0.521*	− 0.125	0.521*	− 0.125	0.521*	− 0.125
Standard error	(0.0584)	(0.0976)	(0.0878)	(0.279)	(0.267)	(0.279)	(0.306)	(0.290)	(0.306)	(0.290)	(0.306)	(0.290)
95% confidence interval	[− 0.117, 0.112]	[− 0.211, 0.172]	[− 0.197, 0.147]	[− 0.653, 0.439]	[− 0.268, 0.778]	[− 0.653, 0.439]	[− 0.0793, 1.122]	[− 0.693, 0.442]	[− 0.0793, 1.122]	[− 0.693, 0.442]	[− 0.0793, 1.122]	[− 0.693, 0.442]
Optimal bandwidth	65.129	53.968	51.685	43.003	61.876	43.003	48.801	41.942	48.801	41.942	48.801	41.942
Observations	45,994	16,435	15,508	1,977	2,833	1,977	2,049	1,726	2,049	1,726	2,049	1,726

Table 5 (continued)

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
	Pooled		Non-disadvantaged schools		Disadvantaged schools		Disadvantaged schools		Disadvantaged schools		Disadvantaged public schools		Disadvantaged public schools	
			Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
<i>Mathematics</i>	-0.0749		-0.0375	-0.0738	0.533*	-0.239	0.574*	-0.239	0.574*	-0.239	0.574*	-0.205	0.574*	-0.205
Standard error	(0.0599)		(0.0767)	(0.0845)	(0.297)	(0.304)	(0.323)	(0.304)	(0.323)	(0.304)	(0.323)	(0.318)	(0.323)	(0.318)
95% confidence interval	[-0.192, 0.0424]		[-0.188, 0.113]	[-0.239, 0.0917]	[-0.0485, 1.115]	[-0.836, 0.357]	[-0.059, 1.207]	[-0.836, 0.357]	[-0.059, 1.207]	[-0.836, 0.357]	[-0.059, 1.207]	[-0.827, 0.418]	[-0.059, 1.207]	[-0.827, 0.418]
Optimal bandwidth	63.887		79.681	75.05	42.839	44.752	39.822	44.752	39.822	44.752	39.822	44.28	39.822	44.28
Observations	45,161		24,768	22,267	2,038	2,037	1,735	2,037	1,735	2,037	1,735	1,861	1,735	1,861

This table shows the diff-in-disc regression results of a placebo treatment of the 2006 cohort and a control group of the 2005 cohort, following Eq. (4), using triangular kernels and normalised grades $N(0,1)$. Each coefficient comes from a different regression, only the coefficient of interest δ_0 is reported. Optimal bandwidth chosen according to Calonico et al. (2014) *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

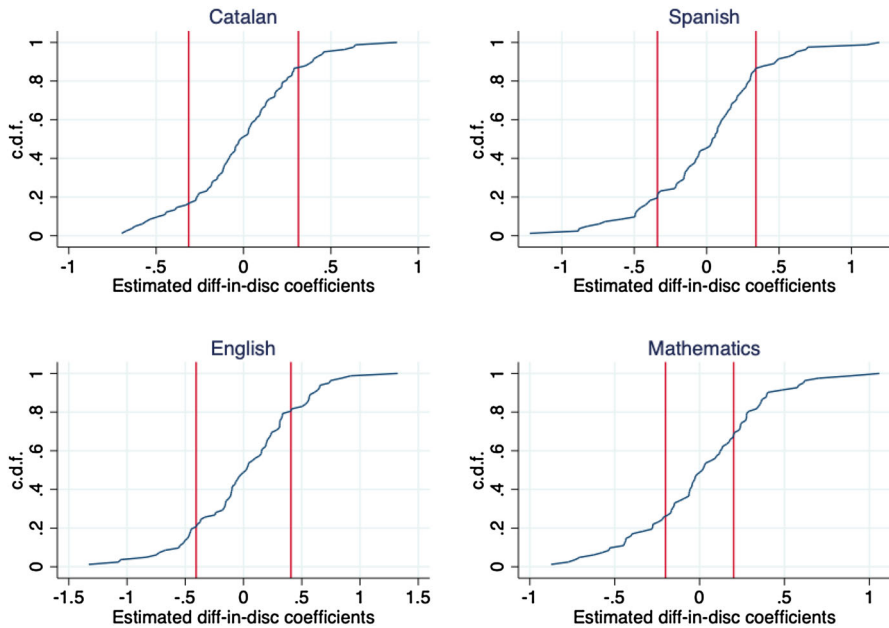


Fig. 7 Cumulative distribution functions of placebo cut-offs, boys in disadvantaged public schools. *Note:* For each subject, this figure reports the empirical cumulative distribution function of diff-in-disc estimates using fake cut-offs to the left (-80 to -40) and to the right (40 to 80) of the actual July 1 cut-off (0)

5.3 Discussion

Overall, the previous section clearly shows that the variations in test scores found are not necessarily related to the introduction of the child benefit. In the full sample, improvements in grades exceeding 10% of a standard deviation can be ruled out at the 95% confidence level and are unlikely to represent meaningful deviations from the average grade. In non-disadvantaged schools, the confidence limits roughly range between -0.31 and 0.17 standard deviation units, while in disadvantaged schools grade improvements account for 4–11% of the average test score for boys in public schools and 1–4.1% in girls, which could be greater as confidence levels allow for effects to be as large as almost a full standard deviation increase in boys' grades and nearly 60% of a standard deviation in girls'. The fact that variance in test scores is typically higher in boys might explain this persistent gender difference in effects. However, none of those effects is statistically significant at the conventional levels.

While it is challenging to assess the magnitude of the effects described earlier, they can be contextualised by comparing them to similar empirical evidence. In Spain, Borra et al. (2021) do not find any significant impact of the 2007 subsidy on the average grades of second and third-year students, reporting similar negative and statistically insignificant coefficients to those reported above, of roughly -0.03 to -0.125 standard deviation units in their full sample, while they find no relevant differences in effects by socioeconomic status.

In the Australian case, Gaitz and Schurer (2017) report an increase in learning outcomes of 0.26 standard deviations as a result of the introduction of the 2004 child benefit, without reaching any statistical significance. Later, Deutscher and Breunig (2018) also report no effects of benefit on the whole population, while in subsamples of disadvantaged families magnitudes increase and they even find a modest statistically significant increase in test scores of 4% of a standard deviation among families having completed high school education or less.

There are a number of reasons why the Spanish subsidy might not have been successful at improving educational outcomes. The first one is a potential fading out of its effects over time, as it happens with certain interventions, especially those improving cognitive outcomes (Protzko 2015). Part of these findings can be attributed to the standardisation of test scores data in analyses, making improvements appear weaker than they really are (Cascio and Staiger 2012). Still, while it is not the case, the positive effects of children interventions sometimes reappear at later ages in the form of improved human capital outcomes (Garces et al. 2002; Araujo et al. 2019).

The permanent income theory literature might provide another explanation for the findings, as permanent shocks tend to have higher longer-term impacts than one-off payments (Blau 1999; Carneiro and Ginja 2016). Similarly, the amount paid is relatively low in present value considering the long-term expenditures linked to the upbringing of a child.

The lack of earmarking of the subsidy might also have had an influence, as families might decide to spend the money on other goods (Beatty et al. 2014). In addition, the marginal utility of income is higher in disadvantaged households, and they usually benefit the most from subsidised child care programmes (Havnes and Mogstad 2015).

Two theoretical mechanisms play a role in modifying children's outcomes as a result of a parental income shock (Fernald et al. 2012). The first mechanism is related to resources, as subsidies generate income and substitution effects, leading to families being able to buy more goods, some of which may benefit their children, and to possible reductions in labour supply. The second mechanism is an improvement in parental mental health, reducing conflict within the household. Due to the reduction in working hours, this mechanism leads to better parenting and more nurturing, letting better children outcomes emerge.

Those mechanisms have already been studied in the context of the Spanish child subsidy using survey data. However, the effect on female labour supply is ambiguous. González (2013), using the Labour Force Survey and Social Security data, finds that mothers took longer to return to work after giving birth and receiving the benefit, working between 0.2 and 0.4 fewer months during the first year after giving birth and earning less. They probably used less private childcare, although only a few coefficients are marginally significant. On the other hand, Hernández Alemán et al. (2017), using EU-SILC¹⁰ data, report an average increase in labour supply of two weekly hours. Differences in findings can be attributed to how each survey is designed, although the former analysis seems more credible than the latter. In addition, Borra et al. (2021) do not find any effect on mother labour supply using longer-term LFS data. They also

¹⁰ European Union Statistics on Income and Living Conditions.

investigate the second mechanism, finding no significant changes in family conflicts in the form of divorces or partnerships as a result of the subsidy.

As a policy recommendation, the lack of results and the literature suggest that, for the child benefit cash transfers to potentially have a persistent effect on language and mathematical abilities, their value, the periodicity of payments or both should be increased, while the target of the policy should move to disadvantaged families.

Still, the estimates presented in the previous section are a lower bound, as if I could actually observe those families treated and untreated with the subsidy instead of relying on the date of birth and their eligibility, the data would allow for more precise identification of the causal effect. In that sense, the availability of data at the municipality or neighbourhood level would help with identifying disadvantaged areas. An additional limitation is the lack of linked data between parents and children, including their income, education achieved and data on siblings, as it is key to explore the role that they might play in the trade-off between children quantity and quality when benefiting or not from the subsidy (Black et al. 2005; Lee 2008).

6 Conclusion

This paper has aimed to estimate the causal effect that €2500 cash transfer implemented in Spain in 2007 and paid to mothers giving birth or adopting a child had on the achievement of basic competencies in language and mathematics at the end of primary school in the Catalonia region.

The way in which the policy was announced, three days after the eligibility began, provides a framework for causal inference. I estimate a regression discontinuity design model that compares children born right before and right after the 1 July 2007 eligibility threshold. I do so for each subject using local linear regressions and a data-driven bandwidth selection procedure. Since there are potential confounding variables affecting children at both sides of the cut-off differently, such as those originating in early seasonal cognitive development differences, I combine the 2019 observations with those from the previous two years into a difference-in-discontinuities specification.

The results show that the subsidy did not have any significant positive effect on test scores, as potentially significant coefficients in the RD specification are not found in the diff-in-disc setting. We can thus rule out, with 95% confidence, grade improvements greater than 10% of a standard deviation in the full sample, although we cannot disregard relevant grade improvements in certain small-sized disadvantaged subsamples ranging between 0.50 and 0.95 standard deviation units. Even though I only observe the eligibility to receive the subsidy and not the actual receipt, it is likely that the amount paid was too low and that potential benefits faded out quickly. Still, whether the subsidy had any long-term effects on labour market outcomes or criminal activity remains to be seen.

Acknowledgements I am thankful to Dinand Webbink, Anne C. Gielen, Núria Mas, the Co-Editor Caterina Calsamiglia and the anonymous referees for their insights, comments and encouragement, as well as to the Erasmus School of Economics in Rotterdam, where part of this research was carried out. I am also grateful to Carmina Pinya and the staff at CSDA for having provided the administrative data for this paper. All errors are my own.

Declarations

Conflict of interest The author declares that he has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by the author.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

See Tables 6 , 7 , 8 , 9 , 10, 11 and Figs. 8 and 9 .

Table 6 Average grades by gender and school complexity

	Pooled		Low complexity		Mid complexity		High complexity	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Catalan	73.904	77.445	77.698	80.835	74.213	77.858	67.431	71.231
Spanish	74.712	77.206	78.529	80.701	74.919	77.499	68.755	71.475
English	77.082	80.979	83.77	87.173	77.063	81.13	68.518	72.577
Mathematics	80.435	77.668	83.094	81.051	80.672	77.848	75.847	72.607

Table 7 Diff-in-disc regression results, second-order polynomials

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
	Pooled		Non-disadvantaged schools		Disadvantaged schools		Disadvantaged schools		Disadvantaged schools		Disadvantaged public schools		Disadvantaged public schools	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
<i>Catalan</i>	-0.0222 (0.0698)	-0.0109 (0.0998)	-0.160* (0.0960)	-0.0109 (0.0998)	0.311 (0.247)	0.0332 (0.250)	0.383 (0.259)	0.0332 (0.250)	0.383 (0.259)	0.0332 (0.250)	0.383 (0.259)	0.0632 (0.266)	0.383 (0.259)	0.0632 (0.266)
95% confidence interval	[-0.159, 0.115]	[-0.206, 0.185]	[-0.348, 0.0287]	[-0.206, 0.185]	[-0.173, 0.795]	[-0.457, 0.523]	[-0.124, 0.891]	[-0.457, 0.523]	[-0.124, 0.891]	[-0.457, 0.523]	[-0.124, 0.891]	[-0.457, 0.584]	[-0.124, 0.891]	[-0.457, 0.584]
Optimal bandwidth	63.836	63.878	77.103	63.878	89.584	70.832	91.673	70.832	89.584	70.832	91.673	67.496	89.584	67.496
Observations	66,073	28,478	35,246	28,478	6,216	4,947	5,765	4,947	6,216	4,947	5,765	4,372	6,216	4,372
<i>Spanish</i>	-0.0570 (0.0600)	-0.0259 (0.0907)	-0.150 (0.0924)	-0.0259 (0.0907)	0.205 (0.281)	-0.0518 (0.256)	0.426 (0.272)	-0.0518 (0.256)	0.426 (0.272)	-0.0518 (0.256)	0.426 (0.272)	-0.0218 (0.283)	0.426 (0.272)	-0.0218 (0.283)
95% confidence interval	[-0.175, 0.0606]	[-0.204, 0.152]	[-0.331, 0.0316]	[-0.204, 0.152]	[-0.345, 0.755]	[-0.553, 0.449]	[-0.107, 0.96]	[-0.553, 0.449]	[-0.107, 0.96]	[-0.553, 0.449]	[-0.107, 0.96]	[-0.0576, 0.532]	[-0.107, 0.96]	[-0.0576, 0.532]
Optimal bandwidth	86.637	75.695	87.415	75.695	78.109	69.543	88.445	69.543	78.109	69.543	88.445	65.214	78.109	65.214
Observations	90,175	33,885	39,922	33,885	5,441	4,898	5,612	4,898	5,441	4,898	5,612	4,245	5,441	4,245
<i>English</i>	-0.0309 (0.0594)	-0.0597 (0.103)	-0.0597 (0.101)	-0.0597 (0.103)	0.272 (0.301)	-0.0968 (0.320)	0.498* (0.269)	-0.0968 (0.320)	0.498* (0.269)	-0.0968 (0.320)	0.498* (0.269)	0.0318 (0.329)	0.498* (0.269)	0.0318 (0.329)
95% confidence interval	[-0.147, 0.0856]	[-0.258, 0.139]	[-0.258, 0.139]	[-0.258, 0.139]	[-0.318, 0.863]	[-0.725, 0.531]	[-0.0303, 1.0254]	[-0.725, 0.531]	[-0.0303, 1.0254]	[-0.725, 0.531]	[-0.0303, 1.0254]	[-0.613, 0.676]	[-0.0303, 1.0254]	[-0.613, 0.676]
Optimal bandwidth	57.001	61.858	76.838	61.858	58.818	59.054	86.226	59.054	58.818	59.054	86.226	60.496	58.818	60.496
Observations	59,330	27,533	34,651	27,533	4,051	4,212	5,430	4,212	4,051	4,212	5,430	3,914	4,051	3,914
<i>Mathematics</i>	0.0327 (0.0648)	0.0327 (0.0938)	-0.0435 (0.0804)	0.0327 (0.0938)	0.340 (0.261)	0.0285 (0.238)	0.320 (0.255)	0.0285 (0.238)	0.340 (0.261)	0.0285 (0.238)	0.320 (0.255)	0.0838 (0.246)	0.320 (0.255)	0.0838 (0.246)
95% confidence interval	[-0.0944, 0.16]	[-0.155, 0.212]	[-0.201, 0.114]	[-0.155, 0.212]	[-0.172, 0.852]	[-0.377, 0.554]	[-0.180, 0.820]	[-0.377, 0.554]	[-0.172, 0.852]	[-0.377, 0.554]	[-0.180, 0.820]	[-0.399, 0.566]	[-0.180, 0.820]	[-0.399, 0.566]
Optimal bandwidth	60.956	67.530	85.833	67.530	56.315	69.553	64.803	69.553	56.315	69.553	64.803	68.401	69.553	68.401
Observations	63,368	30,381	39,140	30,381	3,978	4,906	4,128	4,906	3,978	4,906	4,128	4,438	4,906	4,438

This table shows the diff-in-disc regression results of the 2019 tests with the tests of the previous two years, acting as placebo tests, following Eq. (4) but extending it to include quadratic polynomial terms. Triangular kernels and normalised grades. Each coefficient comes from a different regression, only the coefficient of interest δ_0 is reported. Optimal bandwidth chosen according to Calonico et al. (2014)

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 8 Diff-in-disc regression results, third-order polynomials

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
	Pooled		Non-disadvantaged schools		Boys		Girls		Disadvantaged schools		Boys		Girls	
<i>Catalan</i>	- 0.0250	- 0.141	- 0.0130	0.287	0.130	0.347	0.221	0.347	0.130	0.347	0.221	0.347	0.130	0.221
Standard error	(0.0751)	(0.113)	(0.107)	(0.316)	(0.308)	(0.352)	(0.324)	(0.316)	(0.308)	(0.352)	(0.324)	(0.316)	(0.308)	(0.324)
95% confidence interval	[- 0.172, 0.122]	[- 0.362, 0.0812]	[- 0.223, 0.197]	[- 0.332, 0.906]	[- 0.474, 0.734]	[- 0.344, 1.0373]	[- 0.415, 0.856]	[- 0.362, 0.0812]	[- 0.474, 0.734]	[- 0.344, 1.0373]	[- 0.415, 0.856]	[- 0.362, 0.0812]	[- 0.474, 0.734]	[- 0.415, 0.856]
Optimal bandwidth	84.401	86.701	85.589	82.909	67.925	76.088	70.239	82.909	67.925	76.088	70.239	82.909	67.925	70.239
Observations	88,177	39,415	38,283	5,708	4,773	4,808	4,529	5,708	4,773	4,808	4,529	5,708	4,773	4,529
<i>Spanish</i>	- 0.0452	- 0.112	- 0.0315	0.113	- 0.0961	0.364	- 0.0126	0.113	- 0.0961	0.364	- 0.0126	0.113	- 0.0961	- 0.0126
Standard error	(0.0675)	(0.111)	(0.0945)	(0.331)	(0.313)	(0.313)	(0.328)	(0.331)	(0.313)	(0.313)	(0.328)	(0.331)	(0.313)	(0.328)
95% confidence interval	[- 0.177, 0.0871]	[- 0.330, 0.106]	[- 0.217, 0.154]	[- 0.536, 0.762]	[- 0.710, 0.518]	[- 0.249, 0.977]	[- 0.656, 0.631]	[- 0.330, 0.106]	[- 0.710, 0.518]	[- 0.249, 0.977]	[- 0.656, 0.631]	[- 0.330, 0.106]	[- 0.710, 0.518]	[- 0.656, 0.631]
Optimal bandwidth	106.486	93.678	106.704	85.665	74.876	104.116	74.753	93.678	74.876	104.116	74.753	93.678	74.876	74.753
Observations	111,512	42,647	47,918	5,940	5,216	6,651	4,767	5,940	5,216	6,651	4,767	5,940	5,216	4,767
<i>English</i>	- 0.0440	- 0.0666	- 0.120	0.277	- 0.0930	0.428	0.0552	0.277	- 0.0930	0.428	0.0552	0.277	- 0.0930	0.0552
Standard error	(0.0757)	(0.108)	(0.100)	(0.297)	(0.332)	(0.309)	(0.349)	(0.297)	(0.332)	(0.309)	(0.349)	(0.297)	(0.332)	(0.349)
95% confidence interval	[- 0.192, 0.104]	[- 0.279, 0.146]	[- 0.316, 0.0766]	[- 0.304, 0.858]	[- 0.744, 0.558]	[- 0.177, 1.0334]	[- 0.629, 0.739]	[- 0.279, 0.146]	[- 0.744, 0.558]	[- 0.177, 1.0334]	[- 0.629, 0.739]	[- 0.279, 0.146]	[- 0.744, 0.558]	[- 0.629, 0.739]
Optimal bandwidth	94.111	105.407	101.734	97.918	81.190	100.314	85.094	97.918	81.190	100.314	85.094	97.918	81.190	85.094
Observations	98,503	47,867	45,553	6,783	5,686	6,342	5,445	47,867	5,686	6,342	5,445	47,867	5,686	5,445
Mathematics	0.0230	- 0.0650	0.0277	0.334	0.0142	0.391	0.0228	0.334	0.0142	0.391	0.0228	0.334	0.0142	0.0228
Standard error	(0.0637)	(0.0999)	(0.0971)	(0.268)	(0.278)	(0.299)	(0.290)	(0.0637)	(0.278)	(0.299)	(0.290)	(0.0637)	(0.278)	(0.290)
95% confidence interval	[- 0.102, 0.148]	[- 0.261, 0.131]	[- 0.163, 0.218]	[- 0.192, 0.860]	[- 0.531, 0.560]	[- 0.195, 0.978]	[- 0.545, 0.591]	[- 0.102, 0.148]	[- 0.531, 0.560]	[- 0.195, 0.978]	[- 0.545, 0.591]	[- 0.102, 0.148]	[- 0.531, 0.560]	[- 0.545, 0.591]
Optimal bandwidth	96.568	87.754	100.242	80.612	73.694	78.200	74.424	87.754	73.694	78.200	74.424	87.754	73.694	74.424
Observations	101,623	40,074	45,367	5,622	5,153	4,979	4,773	40,074	5,153	4,979	4,773	40,074	5,153	4,773

This table shows the diff-in-disc regression results of the 2019 tests with the tests of the previous two years acting as placebo tests, following Eq. (4) but extending it to include cubic polynomial terms. Triangular kernels and normalised grades. Each coefficient comes from a different regression, only the coefficient of interest δ_0 is reported. Optimal bandwidth chosen according to Calonico et al. (2014)

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 9 Diff-in-disc regression results by school size

	(1) Small	(2) Intermediate	(3) Big
<i>Catalan</i>	− 0.207**	0.0145	0.0782
Standard error	(0.0826)	(0.0641)	(0.0884)
95% confidence interval	[− 0.369, − 0.0447]	[− 0.111, 0.140]	[− 0.0952, 0.252]
Optimal bandwidth	62.242	54.123	78.891
Observations	17,622	31,740	13,761
<i>Spanish</i>	− 0.204**	− 0.0241	0.0222
Standard error	(0.0925)	(0.0583)	(0.0956)
95% confidence interval	[− 0.385, − 0.0222]	[− 0.138, 0.0901]	[− 0.165, 0.210]
Optimal bandwidth	53.148	63.307	65.052
Observations	15,009	37,005	11,494
<i>English</i>	− 0.188**	0.0290	0.0536
Standard error	(0.0924)	(0.0678)	(0.105)
95% confidence interval	[− 0.369, − 0.0676]	[− 0.104, 0.162]	[− 0.153, 0.260]
Optimal bandwidth	53.255	54.061	57.228
Observations	14,940	31,608	10,012
<i>Mathematics</i>	− 0.00897	0.0190	− 0.0594
Standard error	(0.0711)	(0.0487)	(0.0956)
95% confidence interval	[− 0.148, 0.130]	[− 0.0766, 0.114]	[− 0.247, 0.128]
Optimal bandwidth	77.236	73.469	58.664
Observations	22,044	43,045	10,322

This table shows the diff-in-disc regression results of the 2019 tests with the tests of the previous two years acting as placebo tests, following Eq. (4). Triangular kernels and normalised grades. Each coefficient comes from a different regression, only the coefficient of interest δ_0 is reported. Optimal bandwidth chosen according to Calonico et al. (2014). School size is defined by the number of observations per school: 30 or less (small), between 31 and 60 (intermediate), and more than 60 (big).

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 10 Diff-in-disc regression results, foreign students subsample

	Foreign students
<i>Catalan</i>	0.0209
Standard error	(0.172)
95% confidence interval	[− 0.316, 0.357]
Optimal bandwidth	73.717
Observations	6,188
<i>Spanish</i>	− 0.0689
Standard error	(0.171)
95% confidence interval	[− 0.403, 0.266]
Optimal bandwidth	71.199
Observations	6,153
<i>English</i>	0.112
Standard error	(0.193)
95% confidence interval	[− 0.267, 0.492]
Optimal bandwidth	56.927
Observations	4,863
<i>Mathematics</i>	− 0.359
Standard error	(0.211)
95% confidence interval	[− 0.772, 0.0541]
Optimal bandwidth	48.889
Observations	4,196

This table shows the diff-in-disc regression results of the 2019 tests with the tests of the previous two years acting as placebo tests, following Eq. (4), using the subsample of foreign students. Triangular kernels and normalised grades. Each coefficient comes from a different regression, only the coefficient of interest δ_0 is reported. Optimal bandwidth chosen according to Calonico et al. (2014)

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 11 Diff-in-disc regression results with control variables

	(1)		(2)		(3)		(4)		(5)	
	Pooled		Non-disadvantaged schools		Disadvantaged schools		Boys		Girls	
<i>Catalan</i>										
Standard error	-0.0257	(0.0481)	-0.105	(0.0658)	0.0052	(0.0601)	0.185	(0.187)	0.0799	(0.193)
95% confidence interval	[-0.12, 0.0686]		[-0.234, 0.0244]		[0.114, 0.124]		[-0.181, 0.551]		[-0.299, 0.459]	
Optimal bandwidth	52.073		64.176		65.928		60.663		48.085	
Observations	50,482		26,577		26,871		3,883		3,226	
<i>Spanish</i>										
Standard error	-0.08666*	(0.0475)	-0.155**	(0.0669)	-0.0892	(0.0702)	0.266	(0.189)	0.083	(0.151)
95% confidence interval	[-0.18, 0.00658]		[-0.287, -0.024]		[-0.227, 0.0484]		[-0.105, 0.636]		[-0.214, 0.38]	
Optimal bandwidth	47.784		60.708		46.446		62.152		77.412	
Observations	46,379		24,998		19,506		4,047		5,099	
<i>English</i>										
Standard error	-0.0327	(0.0547)	-0.0444	(0.0849)	-0.0887	(0.0571)	0.389*	(0.212)	0.144	(0.188)
95% confidence interval	[-0.14, 0.0746]		[-0.211, 0.122]		[-0.201, 0.0233]		[-0.0256, 0.805]		[-0.224, 0.512]	
Optimal bandwidth	41.976		42.728		73.889		45.847		62.884	
Observations	41,166		18,161		30,332		3,035		4,093	
<i>Mathematics</i>										
Standard error	-0.0237	(0.0387)	-0.0411	(0.0639)	-0.0610	(0.0698)	0.0583	(0.155)	0.202	(0.151)
95% confidence interval	[-0.0996, 0.0521]		[-0.166, 0.0841]		[-0.198, 0.0758]		[-0.246, 0.362]		[-0.0935, 0.498]	
Optimal bandwidth	66.305		53.02		49.228		63.064		69.431	
Observations	63,832		22,433		20,595		4,203		4,572	

This table shows the diff-in-disc regression results of the 2019 tests with the tests of the previous two years acting as placebo tests, following Table 3 but controlling for month of birth and the type of school ownership. Triangular kernels and normalised grades. Each coefficient comes from a different regression, only the coefficient of interest δ_0 is reported. Optimal bandwidth chosen according to Calonico et al. (2014)

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

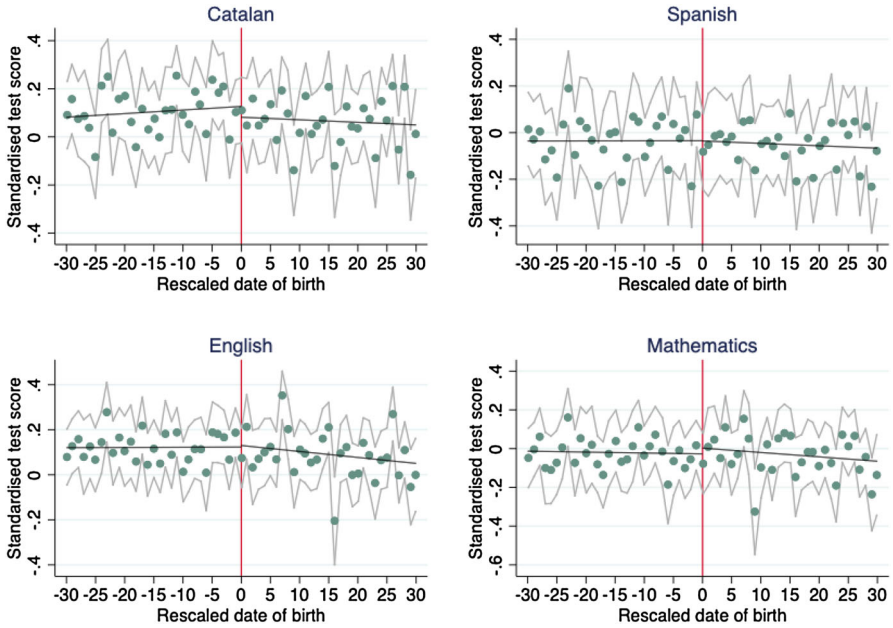


Fig. 8 Placebo RD plots by subject, 2017 tests

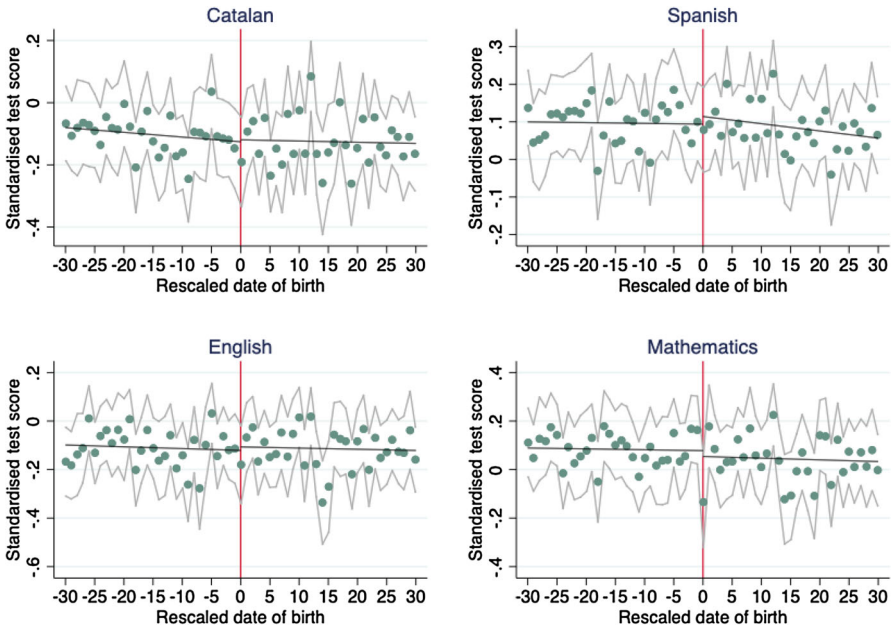


Fig. 9 Placebo RD plots by subject, 2018 tests

References

- Akee R, Copeland W, Keeler G et al (2010) Parents' incomes and children's outcomes: a quasi-experiment using transfer payments from casino profits. *Am Econ J Appl Econ* 2(1):86–115. <https://doi.org/10.1257/app.2.1.86>
- Almond D, Currie J (2011) Human capital development before age five. In: Card D, Ashenfelter O (eds) *Handbook of labor economics*. Elsevier B.V, Amsterdam, pp 1315–1486
- Almond D, Currie J, Duque V (2018) Childhood circumstances and adult outcomes: Act II. *J Econ Lit* 56(4):1360–1446. <https://doi.org/10.1257/jel.20171164>
- Araujo MC, Bosch M, Schady N (2019) Can cash transfers help households escape an intergenerational poverty trap? In: Barrett CB, Carter MR, Jean-Paul C (eds) *The economics of poverty traps*. University of Chicago Press, Chicago, pp 357–382
- Attanasio O, Cattani S, Meghir C (2021) Early childhood development, human capital and poverty. NBER Working Paper 29362. <https://doi.org/10.3386/w29362>
- Azmat G, Calsamiglia C, Iriberrri N (2016) Gender differences in response to big stakes. *J Eur Econ Assoc* 14(6):1372–1400. <https://doi.org/10.1111/jeea.12180>
- Barker DJP (1990) The fetal and infant origins of adult disease. *Br Med J Clin Res Ed* 301(6761):1111. <https://doi.org/10.1136/bmj.301.6761.1111>
- Barker DJP (1998) In utero programming of chronic disease. *Clin Sci* 95(2):115–128. <https://doi.org/10.1042/cs19980019>
- Beatty TKM, Blow L, Crossley TF, O'Dea C (2014) Cash by any other name? Evidence on labeling from the UK winter fuel payment. *J Public Econ* 118(1):86–96. <https://doi.org/10.1016/j.jpubeco.2014.06.007>
- Becker GS, Tomes N (1979) An equilibrium theory of the distribution of income and intergenerational mobility. *J Polit Econ* 87(6):1153–1189. <https://doi.org/10.1086/260831>
- Becker GS, Tomes N (1986) Human capital and the rise and fall of families. *J Labor Econ* 4(3):1–47. <https://doi.org/10.1086/298118>
- Bedard K, Dhuey E (2006) The persistence of early childhood maturity: international evidence of long-run age effects. *Q J Econ* 121(4):1437–1472. <https://doi.org/10.1093/qje/121.4.1437>
- Berniell I, Estrada R (2020) Poor little children: The socioeconomic gap in parental responses to school disadvantage. *Labour Econ*. <https://doi.org/10.1016/j.labeco.2020.101879>
- Black SE, Devereux PJ, Løken KV, Salvanes KG (2014) Care or cash? The effect of child care subsidies on student performance. *Rev Econ Stat* 96(5):824–837. https://doi.org/10.1162/REST_a_00439
- Black SE, Devereux PJ, Salvanes KG (2005) The more the merrier? The effect of family size and birth order on children's education. *Q J Econ* 120(2):669–700. <https://doi.org/10.1162/0033553053970179>
- Blau DM (1999) The effect of income on child development. *Rev Econ Stat* 81(2):176–261. <https://doi.org/10.1257/app.20140162>
- Borghans L, Golsteyn BHH, Heckman J, Humphries JE (2011) Identification problems in personality psychology. *Pers Individ Dif* 51(3):315–320. <https://doi.org/10.1016/j.paid.2011.03.029>
- Borra C, Costa-Ramón A, González L, Sevilla A (2021) The causal effect of an income shock on children's human capital. BSE Working Paper 1267
- Calonico S, Cattaneo MD, Titiunik R (2014) Robust data-driven inference in the regression-discontinuity design. *Stata J* 14(4):909–946. <https://doi.org/10.1177/1536867x1401400413>
- Calsamiglia C, Loviglio A (2020) Maturity and school outcomes in an inflexible system: evidence from Catalonia. *SERIEs* 11(1):1–49. <https://doi.org/10.1007/s13209-019-0196-6>
- Carneiro P, García IL, Salvanes KG, Tominey E (2021) Intergenerational mobility and the timing of parental income. *J Polit Econ* 129(3):757–788. <https://doi.org/10.1086/712443>
- Carneiro P, Ginja R (2016) Partial insurance and investments in children. *Econ J* 126(596):F66–F95. <https://doi.org/10.1111/eoj.12421>
- Cascio E, Staiger D (2012) Knowledge, tests, and fadeout in educational interventions. NBER Working Paper 18038. <https://doi.org/10.3386/w18038>
- Cesarini D, Lindqvist E, Östling R, Wallace B (2016) Wealth, health, and child development: evidence from administrative data on Swedish lottery players. *Q J Econ* 131(2):687–738. <https://doi.org/10.1093/qje/qjw001>
- Chuard C, Chuard-Keller P (2021) Baby bonus in Switzerland: effects on fertility, newborn health, and birth-scheduling. *Health Econ* 30(9):2092–2123. <https://doi.org/10.1002/hecc.4366>
- Consell Superior d' Avaluació del Sistema Educatiu (2021a) Estudi sociodemogràfic i lingüístic de l' alumnat de 4t d' ESO a Catalunya 2006–2013–2021. Informes d' Avaluació 29

- Consell Superior d'Avaluació del Sistema Educatiu (2021b) Com classificar els centres educatius segons la complexitat? Informes d'Avaluació 26
- Cooper K, Stewart K (2013) Does money affect children's outcomes? A systematic review. Joseph Rowntree Foundation Report
- Cooper K, Stewart K (2020) Does household income affect children's outcomes? A systematic review of the evidence. *Child Indic Res* 14(1):981–1005. <https://doi.org/10.1007/s12187-020-09782-0>
- Cunha F, Elo I, Culhane J (2020) Maternal subjective expectations about the technology of skill formation predict investments in children one year later. *J Econom.* <https://doi.org/10.1016/j.jeconom.2020.07.044>
- Cunha F, Heckman J (2007) The technology of skill formation. *Am Econ Rev* 97(2):31–47. <https://doi.org/10.1257/aer.97.2.31>
- Cunha F, Heckman J, Schennach SM (2010) Estimating the technology of cognitive and noncognitive skill formation. *Econometrica* 78(3):883–931. <https://doi.org/10.3982/ecta6511>
- Cunha F, Heckman JJ, Lochner L, Masterov DV (2006) Interpreting the evidence on life cycle skill formation. In: Hanushek EA, Welch F (eds) *Handbook of the economics of education*. Elsevier, Amsterdam, pp 697–812
- Dahl GB, Lochner L (2012) The impact of family income on child achievement: Evidence from the earned income tax credit. *Am Econ Rev* 102(5):1927–1956. <https://doi.org/10.1257/aer.20150988>
- Del Boca D, Flinn C, Wiswall M (2014) Household choices and child development. *Rev Econ Stud* 81(1):137–185. <https://doi.org/10.1093/restud/rdt026>
- Departament d'Educació (2021) Resolució EDU/3545/2021, de 26 de novembre, per la qual s'aproven les directrius per dur a terme la prova d'avaluació de sisè curs de l'educació primària i es convoca la prova per al curs 2021–2022. DOGC 8557
- Departament d'Educació (2019) Educació primària. Alumnes per nivell, sexe, edat, zona nacionalitat, nacionalitat i necessitats específiques de suport educatiu. In: *Estadística de l'ensenyament*. <https://educacio.gencat.cat/ca/departament/estadistiques/estadistiques-ensenyament/cursos-anteriors/curs-2018-2019/infantil-2cicle-primaria/>
- Deutscher N, Breunig R (2018) Baby bonuses: Natural experiments in cash transfers, birth timing and child outcomes. *Econ Rec* 94(304):1–24. <https://doi.org/10.1111/1475-4932.12382>
- Eurostat (2007) Total fertility rate. <https://ec.europa.eu/eurostat/databrowser/view/tps00199/default/table?lang=en>
- Fernald LCH, Gertler PJ, Hidrobo M (2012) Conditional cash transfer programs: Effects on growth, health, and development in young children. In: King R, Maholmes V (eds) *The Oxford handbook of poverty and child development*. Oxford University Press, Oxford, pp 569–600
- Gaitz J, Schurer S (2017) Bonus skills: examining the effect of an unconditional cash transfer on child human capital formation. IZA discussion paper 10525
- Garces E, Thomas D, Currie J (2002) Longer-term effects of head start. *Am Econ Rev* 92(4):999–1012. <https://doi.org/10.1257/00028280260344560>
- Gelman A, Imbens G (2019) Why high-order polynomials should not be used in regression discontinuity designs. *J Bus Econ Stat* 37(3):447–456. <https://doi.org/10.1080/07350015.2017.1366909>
- Giambona E, Ribas RP (2021) The price of obscenity in housing markets: evidence from closing prostitution windows. SSRN Working Paper 2994037. <https://doi.org/10.2139/ssrn.2994037>
- González L (2013) The effect of a universal child benefit on conceptions, abortions, and early maternal labor supply. *Am Econ J Econ Policy* 5(3):160–188. <https://doi.org/10.1257/pol.5.3.160>
- González L, Trommlerová S (2021a) Cash transfers and fertility: How the introduction and cancellation of a child benefit affected births and abortions. *J Hum Resour.* <https://doi.org/10.3368/jhr.59.1.0220-10725r2>
- González L, Trommlerová S (2021b) Prenatal transfers and infant health: evidence from Spain. BSE Working Paper 1261
- Grembi V, Nannicini T, Troiano U (2016) Do fiscal rules matter? *Am Econ J Appl Econ* 8(3):1–30. <https://doi.org/10.1257/app.20150076>
- Guryan J, Hurst E, Kearney M (2012) Parental education and parental time with children. *J Econ Perspect* 22(3):23–46. <https://doi.org/10.1257/jep.22.3.23>
- Guryan J, Ludwig J, Bhatt MP, et al (2021) Not too late: Improving academic outcomes among adolescents. NBER working paper series 28531. <https://doi.org/10.3386/w28531>
- Havnes T, Mogstad M (2015) Is universal child care leveling the playing field? *J Public Econ* 127(1):100–114. <https://doi.org/10.1016/j.jpubeco.2014.04.007>

- Heckman JJ (2008) Schools, skills, and synapses. *Econ Inq* 46(3):289–324. <https://doi.org/10.1111/j.1465-7295.2008.00163.x>
- Heckman JJ, Mosso S (2014) The economics of human development and social mobility. *Annu Rev Econom* 6:689–733. <https://doi.org/10.1146/annurev-economics-080213-040753>
- Hernández Alemán A, León CJ, Márquez-Ramos L (2017) The effect of the universal child care cash benefit on female labour supply in Spain. *Estud Econ Apl* 35(3):801–818. <https://doi.org/10.33119/gn/125463>
- INE (2008) Encuesta de condiciones de Vida, evolución (2003–2007) de la renta anual neta media por hogar, persona y unidad consumo por CCAA. <https://www.ine.es/jaxi/Tabla.htm?path=/t25/p453/a2008/10/&file=00002.px&L=0>
- Lee DS, Lemieux T (2010) Regression Discontinuity Designs in Economics. *J Econ Lit* 48(2):281–355
- Lee J (2008) Sibling size and investment in children's education: an Asian instrument. *J Popul Econ* 21(4):855–875. <https://doi.org/10.1007/s00>
- Løken KV (2010) Family income and children's education: using the Norwegian oil boom as a natural experiment. *Labour Econ* 17(1):118–129. <https://doi.org/10.1016/j.labeco.2009.06.002>
- Løken KV, Mogstad M, Wiswall M (2012) What linear estimators miss: the effects of family income on child outcomes. *Am Econ J Appl Econ* 4(2):1–35. <https://doi.org/10.1257/app.4.2.1>
- Malak N, Rahman MM, Yip TA (2019) Baby bonus, anyone? Examining heterogeneous responses to a pro-natalist policy. *J Popul Econ* 32(4):1205–1246. <https://doi.org/10.1007/s00148-019-00731-y>
- Milligan K (2005) Subsidizing the stork: new evidence on tax incentives and fertility. *Rev Econ Stat* 87(3):539–555. <https://doi.org/10.1162/0034653054638382>
- Milligan K, Stabile M (2011) Do child tax benefits affect the well-being of children? Evidence from Canadian child benefit expansions. *Am Econ J Econ Policy* 3(3):175–205. <https://doi.org/10.1257/pol.3.3.175>
- Ministerio de Trabajo y Asuntos Sociales (2006) Real Decreto 1632/2006, de 29 de diciembre, por el que se fija el salario mínimo interprofesional para 2007. *Boletín Oficial Del Estado* 312:46653–46654
- Montolio D, Taberner PA (2021) Gender differences under test pressure and their impact on academic performance: a quasi-experimental design. *J Econ Behav Organ* 191:1065–1090. <https://doi.org/10.1016/j.jebo.2021.09.021>
- Niederle M, Vesterlund L (2010) Explaining the gender gap in math test scores: The role of competition. *J Econ Perspect* 24(2):129–144. <https://doi.org/10.1257/jep.24.2.129>
- Protzko J (2015) The environment in raising early intelligence: a meta-analysis of the fadeout effect. *Intelligence* 53:202–210. <https://doi.org/10.1016/j.intell.2015.10.006>
- Ribas RP (2016) Multidimensional regression discontinuity and regression kink designs with differences-in-differences. In: 2016 Stata conference
- Sinclair S, Boymal J, De Silva A (2012) A re-appraisal of the fertility response to the Australian baby bonus. *Econ Rec* 88(1):78–87. <https://doi.org/10.1111/j.1475-4932.2012.00805.x>
- Sobotka T, Matysiak A, Brzozowska Z (2019) Policy responses to low fertility: How effective are they? UNFPA Working Paper Series 1

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.