Universitat Autònoma de Barcelona
Department of Applied Economics
PhD Program in Research of Applied Economics

Three Empirical Essays on Gender Equality and Education

PhD dissertation

Author: Natalia Nollenberger Castro

Director: Núria Rodríguez-Planas (IZA, UPF and IAE-CSIC)
Tutor: Xavier Ramos-Molilla (UAB)

March, 2013

Table of Contents

Acknowledgements

I would like to give special thanks to my advisor Núria Rodríguez-Planas for her excellent advise and support. Her academic insight, encouragement and patience has been invaluable. I would also like to thank Xavi Ramos for his support and feedback on my research. I also gained from interesting and motivating discussions with Christina Felfe.

Special thank has to go to my partner Diego for sharing this experience with me and for his love and patience. I will also be eternally thankful to my parents, who despite not finishing high school themselves, were able to transmit to their sons and daughters the importance of further education.

My uruguayan friends in Barcelona, Pierina and Eugenia, gave me permanent emotional support. I shared interesting academic discussions and memorable moments with my peers Darío Judzik, Paula Herrera, Paola Rocchi and Luciana Méndez throughout the Ph.D. program.

I would also like to acknowledge the financial support from the Spanish Agency for International Development Cooperation at the Ministry of Foreign Affairs and Cooperation.

1. Introduction

Over the past few decades both developed and developing countries have made substantial progress in closing the gender gap in opportunities. However, even in developed countries, there are still significant gaps in the job opportunities for women and also in the wages paid to them compared with their male counterparts. In particular, although the employment of women has increased substantially, labor participation rates of mothers with small children is still very low in a substantial number of countries. In addition, although the share of women among college graduated has equaled or even surpassed the share of men, women are still underrepresented in science, technology, engineering, and mathematics (STEM) careers, in which workers earn considerably more than their non-STEM counterparts. The main aim of this dissertation is to provide new evidence that help us to improve our understanding of these two relevant topics on gender equality issues.

In addition, policies aims to narrow the gender gap in opportunities also have unintended effects on other relevant fields. For instance, the provision of free childcare aims to allow the primary caregiver (usually the mother) the access to the labor market is debated because of the effect it may has on children' development. In that sense, this dissertation adopts a more comprehensive view of such a policy and also analyzes its consequences on children' cognitive outcomes. Doing this, it also provides evidence about a topic of high interest in education issues. In fact, the effects of universal provision of preschool on children' development have been a matter of growing attention in the last years with the recent contributions of Heckman and colleagues in the human capital theories that make a strong case for early investment in education.

The dissertation consists of three essays with a marked empirical orientation. The first two essays provide empirical evidence about the effect of universal preschool provision on both maternal employment (first essay) and children' cognitive outcomes (second essay). The empirical questions to be answered by these essays are: 1) To what extend the universal provision of free childcare increases the labor participation of women with small children? 2) What happens to children's long-run cognitive development when introducing universal childcare mainly crowds out maternal care? Recent studies using quasi-experimental methods focus mostly on countries with already high female labor force participation rates, high childcare coverage rates, or with many family-friendly policies. In such a context, the introduction of public childcare crowds out informal or private care arrangements. As consequence, they find modest effects on maternal employment and positive effects on children's skills, particularly among disadvantaged children. In contrast few studies analyze the effects of an expansion of full-time high-quality public childcare in a context of low female labor force participation and insufficient childcare supply, in which the policy mainly crowds out maternal care. Understanding the effects of introducing universal childcare in such a setting is the main contribution of the first two essays of this PhD thesis. To this, we use a quasi-experimental framework that arise from an educational reform carried out in Spain during the 1990s, which expanded the pre-school education to 3 years old.

The third essay provides empirical evidence about the role of culture in explaining the gender educational gaps. There is an extensive literature documenting gender differences in the performance on different educational domains, from which arise the following stylized fact: whereas girls perform slightly better than boys at school at younger ages in most subjects, they start to diverge when they become teenagers, with boys performing better in math, and girls performing better in readings. These gender differences in educational achievement are likely to affect the choice of professional careers, and therefore they may to explain the persistence of gender inequality at older ages. Contributing to understand the sources of such gender differences in tests performance is the aim of the third essay of this dissertation. The question to be answered is: To what extend culture matters to explain gender differences in educational attainment? Recently, some researchers have found a correlation between the cross-country differences in the gender gap in mathematics and reading scores and several measures of gender inequality. They interpret this evidence as suggestive that the underperformance of girls in math relative to boys is eliminated in more gender-equal cultures. However, the interrelationship between institutions and norms makes it difficult to rigorously disentangle the two on this way. Moreover, since the effect is likely to work both ways, they do not capture a causal relationship. Does more gender equality lead to smaller intra-gender performance difference? Or is it smaller girls' underperformance in math relative to boys where the society becomes more gender equal? This essay advances in these two fronts by examining second-generation immigrants in a host country to investigate whether culture determines behavior. This empirical strategy, known as "epidemiological" approach, exploits the fact that immigrants' children have lived under the institutions and markets of the host country but their culture are likely to be influenced by the culture of their parents who grew up under a different institutional framework.

In summary, this research provides new relevant evidence on three broad issues within the fields of gender economics and the economics of education: the effect of free childcare on maternal employment, the effect of universal preschool on long-run children's cognitive outcomes and the role of culture in explaining the gender gap in tests scores.

The dissertation is organized as follows. Sections two, three and four presents the essay one, two and three respectively. General conclusions and further extensions of this research are discussed in section five.

2. <u>Essay 1: Childcare and Maternal Employment</u>[1]

2.1 Introduction

Earlier studies have found that maternal employment is very elastic with respect to price of childcare (with elasticities of around -1). However, recent studies using quasi-experimental methods suggest that these earlier estimates may have been overstated due to misspecifications of functional forms and violations of the exclusion restrictions. These studies uniformly find much smaller effects of public childcare on maternal employment as the introduction of public childcare crowds out informal or private care arrangements. Nonetheless, they focus mostly on countries with already high female labor force participation rates (such as the US, Canada, Israel, Sweden, and France), high childcare coverage rates (such as Argentina, Germany, and the US), or with many family-friendly policies (Sweden and Norway).[2]

In contrast few studies analyze the effects of an expansion of full-time public childcare on maternal employment in a context of low female labor force participation and insufficient childcare supply.[3] The latter scenario, however, includes but is not restrictive to Greece, Ireland, Italy, Japan, Spain, Switzerland, and Turkey in the OECD alone. Understanding the effects of introducing full-time universal childcare in such a setting is the main objective of this paper. We argue that understanding the effects of such reform on maternal employment is highly policy relevant in countries with low female participation as the difficulties to reconcile motherhood and work are among the explanations offered to explain the low levels of female presence in the labor force (Feyrer *et al.* 2008, and De Laat and Sevilla-Sanz 2011). Moreover, this study is

---

[1] A different version of this essay was published as a discussion paper. Reference: Nollenberger, N. and Rodríguez-Planas, N (2011) "Child Care, Maternal Employment and Persistence: A Natural Experiment from Spain". IZA DP 5888.

[2] Gelbach (2002), Schlosser (2006), Lefebvre and Merrigan (2008), Baker *et al.* (2008), Lundin *et al.* (2008), Goux and Maurin (2010), and Fitzpatrick (2010), analyze such type of reform in countries where the 25- to 54-year old female labor participation ranges between 67 and 85 percent. And Gelbach (2002) Berlinski and Galiani (2007), Lefebvre and Merrigan (2008), Baker *et al.* (2008), Lundin *et al.* (2008), Goux and Maurin (2010), Cascio (2009a), Fitzpatrick (2010), and Bauernschuster and Schlotter (2012), analyze the effects of public childcare expansion in a context in which childcare enrollment prior to the reform was between 40 and 80 percent.

[3] To the best of our knowledge the only paper to do this is that of Havenss and Mogstad (2011a), which analyzes a 1970s staged expansion of subsidized childcare in Norway when the maternal employment rate was about 30 percent and subsidized child coverage below 10 percent. These authors find hardly any causal effect of subsidized childcare on the employment rate of married mothers because public childcare crowded out informal care.

particularly relevant for countries, in which female labor force participation is relatively low (Boeri *et al*. 2005), access to proper childcare provisions is limited (Del Boca, 2002), and traditional family roles are deeply rooted (Bettio and Villa 1998; Sevilla-Sanz *et al.* 2010).

We focus on an early 1990s reform in Spain, which led to a sizeable expansion of publicly subsidized full-time childcare for 3-year olds. Following the reform, overall enrollment rate in public childcare among 3-year olds increased from 8.5 percent in 1990 to 42.9 percent in 1997 and to 67.1 percent in 2002. Although the reform was national, the responsibility of implementing its preschool component was transferred to the states. The timing of such implementation expanded over ten years and varied considerably across states. Our analysis exploits this variation across time and states to isolate the reform's impact on the employment decisions of mothers of age-eligible (3-year-old) children. As the Differences-in-Differences (DD) approach may be biased if shocks specific to the treatment areas coincide with the policy changes (such as changes in state labor-market conditions), we apply a Differences-in-Differences-in-Differences (DDD) approach that exploits that the supply shocks to formal public childcare began at different points in time across different states and affected 3-year olds but not 2-year olds. We measure the effect of universal childcare for 3-year olds on maternal employment both at the time the child was eligible and as the child aged (up until the child is 7 years old). By shortening the time mothers spend outside the labor force, universal childcare can reduce mothers' human capital depreciation and preserve their job-search skills (as well as their social and professional networks), implying higher expected wages. Thus, women who would have chosen to stay out of the labor force in the absence of the policy, may decide to enter employment. If this is the case, the effects of this policy ought to persist as the youngest child ages. This may be particularly relevant in a context of low female labor force participation.[4, 5]

The analysis uses cross-sectional data from the 1987 to 1997 Spanish Labor Force Survey. The reason for focusing on the pre-1998 period is that the Spanish Government

[4] To the best of our knowledge, the only other paper to examine persistence is that of Lefebvre et al., 2009, who analyze the effect of Québec's universal childcare policy during a period of economic expansion and with 53 percent maternal employment rate.

[5] When measuring the effects of this reform on maternal employment when the youngest child is 4- to 7-years old, we continue to use mothers of 2-year olds as a comparison group since these children are not affected by the reform. Thus, when estimating the effect one year later we compare mothers whose youngest child is 4 years old with mothers whose youngest child is 2 years old in the same quarter. Whereas children of 2 years old are never affected, when we observe children of 4 years old one year after the LOGSE was implemented in each state, they were affected when they were 3 years old.

subsequently implemented new reforms that may have also potentially affected maternal employment. Our results are robust to the use of alternative specifications (including DD) and control groups (such as using mothers of older children as a comparison group). Moreover, placebo estimates using a pre-reform period support the hypothesis that our findings on the effects of the legislation are not spurious. Finally, we also rule out that endogeneity of fertility is a concern as the reform had no effect on fertility (at least during the period under analysis).

We find that the current effect of offering 9 am to 5 pm public childcare on maternal employment is much smaller than the increase in the supply and enrollment in childcare. Our estimates indicate that a 1 percentage point increase in public childcare coverage for 3-year olds in Spain led to a 0.18 percentage point increase in maternal employment rate and an increase of 7.53 hours worked per week. While this result is consistent with earlier quasi-experimental studies, it comes as surprise given the meager pre-reform level of both female labor force participation and supply of childcare spaces. Especially because, in contrast with other studies, the expansion in public childcare did not lead to a crowding out of private childcare enrollment. We then explore whether these modest effects persist over time as the child ages and whether there is heterogeneity by maternal age and education, and number of children.

We find evidence that these effects persist over time as the child ages and that these longer-term effects of the reform are driven by mothers with a high-school degree and older mothers, for whom the effects of the reform last up to four years later. The lack of persistence for mothers without a high-school degree or younger mothers suggest that the program reduces the depreciation of human capital accumulated in school and in former jobs, and that it also permits the accumulation of new human capital acquired on the job. These results contrast with those of Lefebvre and Merrigan (2009), who find that that the effects of the reform in Quebec only persist among the least educated workers. The lack of results among college educated mothers, who represented less than one tenth of mothers in the early 1990s in Spain, is most likely due to the fact that they are able to pay for daycare (even when it is mainly privately supplied), and that most of them are already strongly attached to the labor market.

The essay is organized as follows. The next section presents a literature review. Section 2.3 presents an overview of the Spanish public childcare system before and after the reform. Sections 2.4 and 2.5 present the empirical strategy and the data, respectively.

Sections 2.6 and 2.7 present the results and several specification checks. Section 2.8 concludes.

2.2 Literature on childcare costs and female labor force participation

Although the theoretical implications of child-care costs on childbearing and work decisions may seem straightforward, Blau and Robins (1989) point out that the implications derived from even a simple economic model of simultaneous decision-making are actually quite complicated. A decrease in child-care costs is expected to lead to an increase in desired fertility due to a standard price effect. Similarly, cheaper child-care services would increase desired labor supply due to a lower opportunity cost of market work. However, the baseline time costs associated with childbearing might offset the increase in desired labor supply, effectively reducing labor force participation. It is also possible that the increase in desired labor supply is sufficient to induce a lower likelihood of childbearing. Thus, the net effects on fertility and labor force participation are ambiguous. However, a common effect is a reduction in role incompatibility: the decrease in the cost of child-care should reduce the likelihood and duration of labor force exit among women who bear children and should increase the likelihood of fertility among women in the workforce. More succinctly: there should unambiguously be an increase in the joint likelihood of work and fertility.

Blau and Robins (1989) represent essentially the only other work that looks at the effects of child-care costs on simultaneous employment and fertility decisions. The analysis is conditional on initial employment status and uses geographic variation in average per-child weekly expenditures as the main measure of child-care costs. Blau and Robins find that higher local child-care expenditures are associated with lower rates of employment among all women and with decreases in childbearing among the non-employed. However, their measure of child-care costs is potentially endogenous as higher local expenditures might be the product of a greater local demand for child-care due to preferences or unobserved labor market conditions.

There is a substantial number of studies that show that young children have a strong negative impact on their mother's labor supply. Heckman (1974) was among the first ones to show that an increase in childcare costs reduces the mother's labor supply and the number of hours worked (conditional on employment). However, examining the U.S. empirical literature, Anderson and Levine (2000) and Blau and Currie (2006), report that estimates for the elasticity of employment with respect to the price of childcare range from 0 to values greater than -1, and that this discrepancy between the estimates can be explained by differences in the population studies, the data sources used, the model specification and the econometric methodology. In Canada, the estimates range from -0.156 to -0.388, indicating a more modest response (Cleveland *et al.* 1996; Powell 1997; and Michalopoulos and Robins 2000).

Traditionally, most studies have used non-experimental data, and various bias correction methods to address the two key selection problems in this literature: the female labor supply participation, and, given labor supply participation, the use of formal child-care as opposed to informal or relative care. One common approach has been to estimate the effects of child-care costs on women labor supply with a sample of mothers who were employed and who paid for childcare, applying corrections for selectivity (see, for instance, Connelly 1992; Kimmel 1995; and Ribar 1992). Alternatively, others have used structural models to identify the effects of child-care costs on female labor supply (Michalopoulous *et al.* 1992, and Ribar 1995). Finally, others have exploited geographic variation in child-care costs or nonlinearities of child-care tax credit for the identification (see Blau and Robins 1989, for the former approach; and Averett *et al.* 1997, for the latter approach).

An alternative approach has been to use an experimental or quasi-experimental approach to identify the effect of child-care costs on mothers labor supply. Gennetian *et al.* (2001), provide a good review of several recent demonstration programs for low-income families that randomly offered childcare subsidies to welfare recipients. Unfortunately, because these programs typically offered other services in addition to child-care subsidies, it is difficult to isolate the effect of child-care on labor force participation per se. In contrast, the quasi-experimental approach has been a good way to identify the problem at hand. Berger and Black (1992), were amongst the first to use this approach by comparing women receiving subsidized childcare to otherwise similar women who are on a waiting list for this care. Their estimates indicate that elasticity of employment with respect to subsidy rates ranges from 0.094 to 0.35. Alternatively, Gelbach (2002), uses the quarter of birth of the child as an instrument for public school enrollment of five-year-old children. He finds that that access to free public school increases the employment probability of (single and married) mothers whose youngest child is aged five, with an implied elasticity of labor supply with respect to childcare costs of –0.13 to –0.36.

More recently, several studies have examined how public preschool availability affects maternal labor supply in different countries using a similar identification strategy to the one applied in this study. Cascio (2009a), examines how the introduction of kindergarten programs into public schools starting in the mid-1960s in the United States affected the labor supply of women with five-year-old children. She finds that single mothers with a kindergarten eligible child and no younger children experienced a 30-percentage increase in the likelihood of employment, which represents a child-care elasticity of employment for this group between -0.2 and –0.55. No significant labor supply responses are detected among other mothers with eligible children. Lefebvre and Merrigan (2008), and Baker *et al.* (2008), both use the introduction of universal and highly subsidized childcare in Quebec to identify the impact of such a policy on female labor supply (the former), and female labor supply and subsidized child-care utilization (the latter). Lefebvre and Merrigan (2008), find that the policy increased the participation rate of mothers with at least one child aged 1 to 5 years by 8 percentage points and that hours and weeks worked (per year) increased by 231 and 5.17, respectively. Using different data and a more narrowly defined treatment population, Baker *et al.* (2008), find similar results on the labor force participation of married mothers. In addition, they find that the use of subsidized child-care also implied a sizeable increase in the use in child-care, which they interpret as "crowding out" informal child-care arrangements. Berlinski and Galiani (2007) analyze the impact of a large-scale construction of preschool facilities for children three to five years old in Argentina during the late 1990s. They find that the construction program has a sizeable impact on school enrollment among eligible children, and that it increases maternal employment (of single mothers with no other children under the age eligibility threshold). Schlosser (2006), studies the effect of a gradual introduction of free preschools for all children aged 3 and 4 in the late 1990s in Israel beginning in 1999. She finds that the intervention increased preschool enrollment and the labor supply of the more educated mothers. Finally, Havnes and Mogstad (2011a) analyze a staged expansion of subsidized childcare in Norway. Given that the subsidized childcare mostly crowds out informal childcare arrangements, they find little causal effect on maternal employment.

2.3 Overview of the Spanish Public Childcare System

School and Preschool Prior to the Reform

Mandatory schooling in Spain begins at age 6. However, preschool for 4- and 5-year olds is also offered in the premises of primary schools from 9 am to 5 pm (regardless of school ownership status). Once a primary school offers places for 4-year olds, parents who wish to enroll their children to that particular school will do so when the child turns 4 years old as the chance of being accepted in the school may decrease considerably a year later (because priority of enrollment of 5- or 6-year olds is given to those children already enrolled in that particular school when they were 4 years old). As a consequence, enrollment rates for 4- and 5-year olds in the late 1980s were 94 and 100 percent, respectively.

Primary and secondary schooling is either public or private.[6] Public schools are free of charge, except for school lunch, which costs about 100 € per child per month. Private school costs are higher - between 250 and 350 € per child per month (including lunch).[7]

At the beginning of the 1990s, childcare for children 0- to 3-years old was rather scarce, predominantly private, and also quite expensive (on average it costs between 300 and 400 € per child per month--including lunch). In contrast with Scandinavian countries and the US or Canada, family day care, in which a reduced number of children are under the care of a certified carer in her house, is practically non-existent. In Spain, children under 4 are either in formal (public or private) childcare or with their mother (or grand-mother). Unfortunately, information on grand-mother's care is unavailable. As a consequence, this paper considers motherly care as equivalent to care provided by the nuclear family.

The Reform

In 1990, Spain underwent a major national education reform (named LOGSE) that affected preschool, primary and secondary school.[8] The focus of our study lies on the preschool component of this reform, which consisted of a regulation of the supply and the quality of preschool, and its implementation began in the school year 1991/92. The

[6] About one third of children in primary school in Spain are enrolled in private schools.

[7] In this paper, private schools refer to "escuelas concertadas" for which the government subsidizes the staff costs (including teachers). There are a very small number of private schools, which tend to be foreign schools (such as the French, Swiss or American schools), and cost two to three times more than the average "escuela concertada".

[8] The primary and middle school component of the reform was first introduced in the school year 1997, which is basically outside of our period of analysis, consequently having no potential impact on our results.

LOGSE divided preschool in two levels: the first level included children up to 3 years old, and the second level included children 3 to 5 years old. While not introducing mandatory attendance, the government began regulating the supply of places for the 3 year olds. Prior to the LOGSE, free universal preschool education had only been offered to children 4 to 5 years old in Spain. After the LOGSE, preschool places for 3 year olds were offered within the premises of primary schools and were run by the same team of teachers. This implied that childcare for 3 year olds operated full-day (9 am to 5 pm) during the five working days and followed a homogeneous and well thought program. With the introduction of the LOGSE schools also had to accept children in September of the year the child turned 3 whenever parents asked for admission if places were available. Available preschool places were allocated to those who had requested admission by lottery (regardless of parents' employment, marital status, or income). As explained earlier, although enrollment was not mandatory, it was necessary to ensure a place in the parents' preferred school choice. As a consequence, child-care enrollment among 3-years-old children went from meager to universal in a matter of a decade.[9] Between the academic years 1990/91 and 1997/98 the number of 3-years-old children enrolled in *public* preschool centers quintupled from 33,128 to 154,063.[10] Federal funding for preschool and primary education increased from an average expenditure of € 1,769 per child in 1990/91 to €2,405 in 1996/97 (both measured in 1997 constant euros), implying a 35.9 percent increase in education expenditures per child.[11]

Besides regulating the supply of public child-care, the LOGSE also provided for the first time in Spain federal provisions on educational content, group size, and staff skill composition regardless of ownership status for children 3 to 5 years old. Psycho-educational theories such as constructivism (put forward by Jean Piaget, and Lev Vygotsky) and progressive education (based on Célestin Freinet and Ovide Decroly) served as a guideline for the design of the curriculum. There was a strong emphasis on play-based education, group play, learning through experiencing the environment, problem solving and critical thinking (LOGSE; 3rd October 1990). While the pedagogical movements behind the LOGSE are close to those in Scandinavian countries, they have

[9] Unfortunately we only have information on enrollment rates and not on actual supply rates for 3-year olds. In the context of rationed supply, enrollment rates should, however, resemble coverage rates quite closely.
[10] These figures exclude Basque Country, Navarra and Ceuta and Melilla as they are not included in our analysis.
[11] Unfortunately data disaggregated at the preschool level is not available.

been viewed as an alternative to the test-oriented instruction legislated by the *No Child Left Behind* educational funding act in the US or the reception class in the UK. In addition, the LOGSE established the maximum number of students per class to be 20 for 3-year olds and 25 for 4- and 5-years old. It is important to point out that classes are grouped based on the year children were born and thus, are not mixed in ages. Finally, the LOGSE required preschool teachers to have a college degree in pedagogy – a requirement previously only enforced for teachers of 4- and 5-year olds.

Despite being a national law and being financed nationally, the responsibility of implementing the expansion of public preschool slots was transferred to the states. The timing of such implementation expanded over ten years and varied considerably across states frequently for arbitrary reasons. Implementation lags arose largely due to a scarcity of qualified teachers and constraints on classroom space in existing primary schools – as mentioned above, childcare for 3-year olds were integrated in existing primary schools (El País, October 3rd 2005).

Figure 1 draws the proportion of public preschools places offered to children 3- to 5-years old for school years 1986/87 to 1998/99 by the timing of the implementation.[12] States are grouped based on the year implementation of the reform began (shown in Appendix Table A.1). As is apparent from the figure, there has been a strong growth in childcare coverage since the implementation of the reform, particularly in the early years after the implementation of the LOGSE began. For instance, among the early implementing states, the childcare coverage went from 46 percent in school year 1990/91 to around 65 percent in the school year 1993/94 and 71 percent in school year 1996/97. As enrollment rate of 4- and 5-years old was already above 90 percent in the late 1980s and fertility remained stable over that period (and began its decline in the year 1995), most of the observed increase in public childcare coverage is driven by 3-year-old children. Indeed the enrollment rate of 3-year olds in publicly-funded schools in the early implementing states went from 11 percent in school year 1990/91 to around 40 percent in the school year 1993/94 and over 50 percent in school year 1996/97 (shown in Figure A.1 in the Appendix). It is also interesting to note that in Spain the increased supply of

[12] Following Berlinski and Galiani (2007), we estimate the proportion of public preschool places offered in each state as the number of public preschool units in each region times the average size of the classroom divided by the population of 3- to 5-years old in each state. Unfortunately, these data are not available by children's age. It is important to note, however, that the increase in the proportion of places offered to children 3- to 5-years old is a weighted average of increases across the three age groups and thus underestimates the growth in public places offered to 3-year-old children, which was considerably more dramatic.

childcare for 3-year olds did not crowd out private enrollment (also shown in Figure A.1). Following Havnes and Mogstad (2011a), we focus our analysis on this early expansion. The reason is that the early expansion is more likely to reflect the sudden increase in public preschool places for 3-year olds than an increase in the regional demand for daycare. In Sections 2.5 and 2.6, we discuss potential endogeneity of the reform and the timing of the implementation of the legislation across states.

2.4 Empirical Strategy

Current effect of the reform

Most of the studies using the natural experiment framework apply the Differences-in-Differences (DD) approach, which (as explained by Cascio 2009a) may be biased if shocks specific to the treatment areas coincide with the policy changes (such as changes in state labor-market conditions) or if there are permanent unobserved differences between mothers residing in treatment and comparison areas. To address these concerns, we apply a Differences-in-Differences-in-Differences (DDD) approach that exploits that the supply shocks to public childcare began at different points in time across different states and affected 3-year olds but not 2-year olds.[13] Sensitivity analysis also presents estimates from a DD approach only using 3-year olds and exploiting regional variation across time. In addition, as one may be concerned that the reform may have also affected mothers of 2-year olds by changing the expected cost of work for these mothers when their child turns 3 years old, we also used a different control group, mothers whose children were up to two years older than our treatment group to avoid this potential "anticipation effect". Finally, using a DD approach we estimated whether there was an "anticipation effect" of the reform on mothers of 2-year olds and did not find any evidence of it in the short- or medium-run.

Our basic DDD model, estimated by OLS over the sample of mothers whose youngest child is 2 and 3 years old, can be expressed as:[14]

$$Y_{it}=\alpha_0+ \alpha_1 Post\_reform_t + \alpha_2 Treated_i + \alpha_3(Post\_reform_t * Treated_i)+$$
$$+ \alpha_4 t + \alpha_5(t * Treated_i) + X'_{it}\beta + \varepsilon_{it} \qquad (1)$$

[13] Public and private enrollment for 2 year olds over the period remained well below 7 percent over the period.

[14] We use linear probability models in all specifications. However, we replicated our analysis using logit models and find very similar results.

where Y_{it} is the employment outcome of interest for woman i in quarter t. We present estimates of employment at survey date and weekly hours worked. $Post\_reform_t$ takes value of 1 if the period is *after* the beginning of implementation of the reform in each state, and 0 otherwise. We follow the classification of states presented in Table A.1. For instance, in Madrid $Post\_reform_t$ takes value of 1 beginning in the fourth quarter of 1992 and forward, and 0 otherwise. $Treated_i$ takes value of 1 if the mother's youngest child is 3-years old, and 0 if her youngest child is 2-years old. We used the year of birth of the child (instead of the child's age reported at the time of the survey) to define the treatment group . The reason for this is that the Spanish enrollment rule is such that, in order to begin the academic year $t/(t+1)$, which starts each September, the child must have turned the mandatory age (3 years in this case) on or prior to December 31st of the calendar year t. Since the Spanish LFS is a quarterly cross-sectional dataset, this implies that our "treatment" group is defined as mothers whose youngest child is 3-years old during calendar year $t-1$ for LFS quarters one through three of year t, and as mothers whose youngest child is 3-years old during calendar year t for the fourth quarter of year t. Following the same rule, we define mothers whose youngest child is 2-years old as those whose youngest child has turned 2 in the previous (current) calendar year if we observe them in quarters one through three (four). Moreover, we eliminate from our "control" sample mothers who had a 3-year old (in addition to a 2-year old). The reason for this is that these mothers are eligible to benefit from the universal childcare by enrolling their 3-year olds and this may affect their employment decisions. This implies losing 2,024 observations (less than 2 percent of our sample). However, results are robust to relaxing this restriction.

The vector X_{it} includes individual-level variables expected to be correlated with employment: age, age squared, dummies indicating the number of other children, a dummy for being foreign-born, educational attainment dummies (high-school dropout, high-school graduate, and college), a dummy for being married or cohabitating. In addition, we include state level unemployment rate to control for possible differences across regional labor markets. We also include state and year fixed effects controlling for permanent differences in maternal employment across states and for the general Spanish economy business cycle, respectively. In addition, to control for possible pre-period trends that could bias the results (Meyer 1995), we also include a quarterly linear time trend, t, which differs for the treatment and control group, so that we can control for systematic differences in the behavior between the two groups over time. The time trends and the individual and state characteristics should control for differences in the characteristics of the treatment and control groups that affect the level of employment. In the Robustness Section, we present alternative specifications, including one that fully interacts the treatment with all the covariates. As explained below, our results are robust to these alternative specifications.

We decided against clustering the standard errors because we were concerned that clustering would under-estimate standard errors as the number of states is limited. As a robustness check, we have estimated the models with standard errors clustered at the state level, with results similar to those shown in the paper. Alternatively, we follow Bertrand *et al.* (2004), and estimate a DD approach that drops the geographical variation in the timing of the implementation ignoring the time-series information. The results are again robust to this alternative specification. Both robustness checks are presented below.

Identification Threats

The coefficient α_3 on the interaction between the $Post\_reform_t$ and $Treated_i$ captures the impact of the reform on the employment outcome measured at different points in time depending on our choice of treatment group and post-reform period. Identification rests on the assumption that (in the absence of the policy change) the average difference between the employment rates (or hours worked) of mothers of 3-year olds and mothers of 2-year olds would have changed similarly in the treatment and control states. One potential threat to our estimation strategy is that at the same time other policies affecting maternal employment are implemented in Spain. To best of our knowledge we are not aware of the existence of such policies until the end of the 1990s, when the Government introduced two major changes: (1) the 1998 and 2003 tax reforms, which substantially altered the child deduction benefits—analyzed by Sánchez and Sánchez (2008); and Azmat and González (2010);[15] and (2) the 1999 family-friendly legislation, which granted mothers with children less than 7-years old the right to reduce working hours—including to work part-time but also to resume their full-time job—and (most importantly) protected them against a layoff—analyzed by Fernández-Kranz and Rodríguez-Planas (2011b). As these policies were important and the evidence shows that

[15]Tax credits per children were small until 1997, but they were substantially increased in 1998, and then again in 1999. Finally, in 2003 an additional tax credit of €1,200 a year was granted to working mothers with children less than 3-years old.

they affected mothers' employment decision, our analysis focuses on the years 1987 to 1997 to avoid potential policy interactions.

Furthermore, migration across states in Spain is surprisingly low (Jimeno and Bentolilla 1998, Bentolilla 2001). Thus, there is little concern that the policy may have induced families to move from slow implementing states to fast implementing states. Finally, in the specification tests section we evaluate whether endogeneity of fertility is a concern.

2.5 The Data and Descriptive Statistics

We use data from the second quarter of 1987 through the last quarter of 1997 of the Spanish Labor Force Survey (LFS). The reason for not using data prior to the second quarter of 1987 is that information on the year of birth of the children is not available. As explained earlier, we focus our analysis on the years prior to 1998 to minimize concerns regarding potential policy interactions.

The Spanish LFS is a quarterly cross-sectional dataset gathering information on socio-demographic characteristics (such as, age, years of education, marital status, state of residence, marital status), employment (including weekly hours worked), and fertility (births, number of children living in the household, and their birth year). Unfortunately, we do not observe children's day care enrollment precluding us from analyzing a "first-stage" model as in Cascio (2009a) and Berlinsky and Galiani (2007), with a dummy for public day care enrollment of the mother's youngest child as dependent variable.

We restrict our sample to mothers between 18 and 45 years old at survey date. Moreover, we exclude the Basque Country and Navarra from the analysis because of their greater fiscal and political autonomy since the mid-1970s, implying that their educational policy differed from that of Spain as a whole. The final sample size is of 105,748 observations.

Unfortunately, the LFS has no information on wages. Optimally, we would have liked to use a recently available longitudinal dataset from Social Security records that contains information on wages, the Continuous Survey of Work Histories (CSWH). However, we decided against the longitudinal dataset for the following reason. The CSWH provides the complete labor market history for those women registered in the Social Security Administration in 2004. This implies that if a woman worked in the early 1990s and after having a child she decided to leave the labor force, she is *not* included in the CSWH. As most of our analysis focuses on the early- and mid-1990s, and labor force participation among mothers of young children at that time was low (around 35 percent prior to the reform), we are concerned that the data from Social Security records will provide estimates of the reform biased towards those women who are strongly attached to the labor force. As we consider that the relevant question here is the employment decision, we prefer focusing on the LFS, which is a representative sample of the Spanish working-age population.

Descriptive Statistics

Table 1 presents baseline summary statistics for the main variables that may affect employment decisions for the treated group for each group of implementing states. These pre-reform means are estimated during the years prior to the implementation of the reform in each of the states, as explained at the bottom of Table 1. In addition, Table 1 also presents treatment and control pre-reform baseline differences. Treated and control mothers are quite similar within and between each group of implementers. However, there are small but statistically significant differences: treated mothers are somewhat older, have a slightly higher number of children, are slightly less likely to be married or cohabitating, and are less educated than those whose youngest child is 2 years old. As explained earlier, our specifications control for these observable differences.

One concern is the potential endogeneity of our policy. For example, we may worry that the increase in public preschool places for 3-years old in a particular state was a response to the increasing incidence of working mothers. We may also be concerned if short-term falls in employment immediately before 1990 triggered the reform. To address these concerns, Figure 2 shows maternal employment rates and weekly hours worked for mothers whose youngest child is 2, compared to those whose youngest child is 3. Each outcome series was calculated by setting $t = 0$ as the quarter in which implementation began in each state (for instance, fourth quarter of 1991 for Cataluña, fourth quarter of 1992 for Madrid, fourth quarter of 1994 for the Canary Islands, and so on), and estimating a weighted average across states at each point in time. Figure 2 shows that both the employment rate and weekly hours worked of *all* mothers with young children increased quite steadily in the quarters preceding the implementation of the reform.[16] The policy change may have been a response, at least in part, to (long-term)

[16] The average hours worked is low because our sample includes both employed and not employed women.

low employment levels, but the period prior to the reform does not appear "special" in either outcome. Moreover, we observe that prior to the implementation of the reform the employment and hours worked of mothers whose youngest child is 2 matches quite well with those of mothers whose youngest child was 3 years old. However, *after* the implementation of the reform, there is a widening of the employment outcomes between the treatment groups and the control group.

While it is not necessary for our estimation strategy because of the inclusion of state fixed effects, it would be useful if the timing of the implementation of the legislation across states were uncorrelated with the employment outcomes of interest. In the robustness section, we test whether the timing of implementation across states can predict maternal employment outcomes. Overall, our findings indicate that this is not the case. In Appendix Table A.2, we display characteristics of the different groups of implementing states to better understand the determinants of the expansion across states. Overall differences across states are small and do not seem to follow a monotonic pattern in relation to the timing of implementation. Differences worth mentioning follow. In general, states implementing after 1993 are poorer and have higher unemployment rate than those implementing in 1991 or 1992. As a robustness test, we estimated a DDD using *only* early states implementers (that is those that implemented the reform in 1991/92 or 1992/93), which were very similar in terms of these observables. Alternatively, as we mentioned earlier, we also estimated a DD model exploiting *only* the timing of the LOGSE (and thus, omitting any regional variation in its implementation). Both of these alternative specifications present results robust to those presented in our main specification, as explained later.

2.6 Current Effects of the Reform

Table 2 presents the main results from estimating equation (1) using two alternative outcome variables: employment (panel A), and weekly hours worked (panel B).[17] Column 1 shows our preferred estimate of the coefficient of interest, α_3. It measures the effect of the legislation on employment for mothers whose youngest child is 3-years old (treated group) relative to mothers whose youngest child is 2-years old (control group) in states that implemented the reform relative to those that did not (net of any trends across the two groups). According to these estimates, after the legislation was passed mothers

---

[17] We also estimated the effects of the reform on mothers' labor force participation and found results consistent with those presented in the main text. Estimates available from authors upon request.

of 3-year olds were 2.3 percentage points more likely to work than mothers of 2-years old and they worked, on average, 0.95 hours more per week. Since prior to the reform, their average employment rate was 29.3 percent, this implies a relative increase of 7.9 percent. In terms of hours, since they worked on average 10.9 hours per week, the reform implied an 8.7 percent increase in hours worked. When compared to pre-initiative means, these results are similar in relative magnitude to those found by Schlosser (2006), for Arab mothers of 2- to 4-years old in Israel for the years 1999 and 2000; Cascio (2009a), for single mothers of 5-years old in the US from the mid-1960s through the mid-1980s; and Lefebvre and Merrigan (2008), and Baker *et al.* (2008), for mothers in Quebec in the late 1990s. And they double the size of the effects found by Havnes and Mogstad (2011a), in Norway in the early 1970s.

Table 3 estimates the effect of the legislation on public and private childcare enrollment rates for 3- and 2-year olds.[18] It shows that 3-year old children residing in treatment states were offered substantially more public childcare than children residing in control states. Column 1 in Table 3 shows that the policy change led to a differential increase of 12.6 percentage points in the public childcare enrollment rate of 3 year-old children. It is important to note that there was no differential effect on the public enrollment rate for 2-year olds (shown in column 2).[19] Using estimates from Table 2 and 3 and assuming 100 percent take-up of new public day-care places, the early 1990s reform in Spain led to a 0.18 percentage point increase in maternal employment rate per percentage point increase in public childcare coverage.[20] These estimates are again in line with earlier findings in that the effect of maternal employment is much smaller than the increase in the supply and enrollment in childcare.[21]

---

[18] Using data at state level from the Ministry of Education we regress the enrollment rate for 3- and 2-year olds respectively (in public and private schools) versus a variable that take the value of one when the implementation of the reform began in each state. We include state fixed effects and control for state characteristics, as GDP per capita and unemployment rate.

[19] Public childcare enrollment rate for 3-year olds prior to the reform was 11.4 percent in states implementing in 1991/92, 11.0 percent in states implementing in 1992/93 and 2.0 percent in states implementing in 1994/95, or 1997/98.

[20] This estimate is the ratio between the percentage point increase in maternal employment rate and the percentage point increase in 3-year olds' public childcare enrollment due to the reform, that is: 0.023/0.126 = 0.183.

[21] Gelbach (2002), uses the quarter of birth of the child as an instrument for publicly-funded school enrollment of 5-year-old children. He finds that access to free publicly-funded school increases the employment probability of (single and married) mothers whose youngest child is aged 5, with an implied elasticity of labor supply with respect to childcare costs of –0.13 to –0.36. Cascio (2009a), reports elasticity estimates in the order of -0.22 to -0.79. In contrast, Havnes and Mogstad (2011a), only find a 0.06 percentage point increase in maternal employment rate per percentage point increase in childcare coverage.

As we also observe weekly hours worked, we proceed to estimate the effect of the legislation at the intensive margin. When we do so we estimate that the childcare expansion led to an increase of 7.53 hours worked per week per percentage point increase in childcare coverage.[22] Note that if all mothers of 3-year-old children affected by the legislation were in a full-time job of 40 hours per week, a 0.18 percentage point increase in maternal employment rate per percentage point increase in public childcare coverage corresponds to an increase of 7.2 hours worked per week, not far from our estimate.[23]

Overall, the evidence presented thus far indicates that the effect of universal childcare on maternal employment is much smaller than the increase in the enrollment in childcare. While this result is consistent with earlier quasi-experimental studies, it comes as a surprise because both female labor force participation and the supply of childcare spaces were meager prior to the policy reform in Spain. Thus, one would have expected a larger impact of offering 9 am to 5 pm public childcare on maternal employment, especially because, in contrast with other studies, the expansion in public childcare did *not* lead to a crowding out of private childcare enrollment as shown in column 2 of Table 3. Indeed the preschool reform for 3-year olds was implemented within primary school regardless of school ownership. As a consequence parents who wished to enroll their children in private school would now enroll their 3-year old to the private school as soon as preschool for that age group was offered (to guarantee a space thereafter).

It is also important to note that most of the mothers of 3-year olds who worked prior to the reform in Spain had their child already enrolled in either public or private childcare. Prior to the reform, 29 percent of mothers of 3-year olds worked in treated states while 24 percent of 3-year olds were enrolled in formal care (11 percent in public childcare and 13 percent in private childcare). Taken together, the Spanish reform mainly implied that many mothers took their children to full-time (9 am to 5 pm) childcare even though they continued *not* to work. It is important to note that these results contrast with those of Havnes and Mogstad (2011a), who found meager effects of the reform on maternal employment in a context of low female labor force participation, in at least two ways. First, in their study, childcare coverage (10 percent) was half the size of maternal employment (20 percent) prior to the reform. Second, the authors find that the public

[22] This estimate is the ratio between the percentage point increase in maternal weekly hours and the percentage point increase in 3-year olds' public childcare enrollment due to the reform, (0.949/0.126 = 7.53).

[23] Spain is among the countries with a lower incidence of PT work. During the 1990s, less than 8 percent of the labor force worked part-time (Fernández-Kranz and Rodríguez-Planas 2011a).

childcare expansion had no effect on maternal employment in Norway as it mainly crowded out alternative childcare modes.

So why did *not* the majority of Spanish women enter employment once they were offered full-time free childcare? In the late 1980s and early 1990s Spain was *not* a family-friendly policy for working parents (and especially working mothers) as reflected by its low levels of social assistance to families (Adserà 2004), one the of the shortest maternity leaves in Europe (Ruhm 1998), an extremely low incidence of part-time work (only 8 percent of all jobs in 1990), as well as a rigid labor market with many jobs in the service sectors having a split shift from 9 am to 2 pm and from 5 to 8 pm (Amuedo-Dorante and de la Rica 2009). Moreover, Spain was a traditional country with low participation of men in household production (Bettio and Villa 1998; de Laat and Sevilla-Sanz, 2011). Consistent with this, only half of all married women aged 18 to 45 were working or looking for a job in 1992, and among those not participating in the labor market, 45 percent reported family responsibilities as their main reason.[24]

Sensitivity Analysis

Columns 2 to 7 of Table 2 present estimates of the coefficient of interest, α_3, under alternative specifications of equation (1). Column 2 displays estimates clustered at the state level. Column 3 presents a specification that includes time trends interacted by state dummies. Column 4 presents a more flexible specification that also includes interactions between the treatment dummy and the states fixed-effects and between the treatment dummy and the covariates. Column 5 estimates equation (1) using as comparison group mothers with children 4 and 5 years old (instead of mothers of 2 years old). Column 6 shows the results from estimating equation (1) only using states implementing in 1991/92 and 19992/93, which were very similar in terms of socio-demographic and economic characteristics before the reform. Column 7 displays the results from estimating a DD approach assuming that all states start the implementation in 1991/92 (that is, ignoring regional variation in the policy implementation). Results are robust to those presented in column 1.

Methodologically, we have relied on the DDD assumption that—in the absence of the reform—the employment and hours gap (net of the trends) between the treatment and control groups would have remained constant. As this assumption is not testable, we

24 Estimated by the authors based on micro data from the 1992 Spanish Labor Force Survey.

proceed to carry out a placebo estimate (shown in column 8 of Table 2). This is to say that we estimate the same DDD models for a period in which no reform was implemented in any state. In each state, we only use the years *before* the LOGSE was implemented. We then define as pre-LOGSE period the period that begins two years before the LOGSE was actually implemented in each state. None of the placebo estimates are statistically significant. Moreover, the coefficients are considerably smaller in size and have the wrong sign. This supports the hypothesis that our previous results on the effects of the family-friendly legislation were *not* spurious.

One may wonder what the source of identification is. To explore this, Table 4 presents the conventional DD estimates of the current and subsequent effects of the reform for mothers of 3-year olds. Focusing first on column 1 in Table 4, we observe that the effect of the reform on employment or hours worked is small and not statistically significant when no linear trend is included in the specification. To explore whether this lack of statistically significant results is due to some unobserved correlate between implementation of the reform and employment, column 3 in Table 4 presents similar estimates for the comparison group used earlier (that is, mothers of 2-year olds). These mothers were 1.5 percentage points *less* likely to work and worked about half an hour *less* per week after the reform. Estimates are statistically significant at the 90 percent level.[25] Indeed, the implementation of the LOGSE ended up occurring at a time of increasing unemployment rate during which female unemployment rate rose from 25 percent in 1990 to 31 percent at the end of 1994 to go back down to 28 percent at the end of 1997. These estimates are similar to those found in a very different context by Cascio, 2009a, and indicate that the DD estimates of the effects of universal childcare are downward biased. As in Cascio, 2009a, there is a time- and state-varying trend that is positively correlated with the implementation of the reform but negatively correlated with maternal employment (or vice-versa). When we add to the DD specification a linear trend interacted by state the current effect of the reform on treated mothers becomes a 4.2 percentage points (or a 15 percent) increase in employment—shown in column 2 in Table 4. Similarly, the DD effect of the reform on *treated* mothers' hours work implies an increase of 1.5 hours per week (or 14 percent) increase. Both of these estimates are

[25] We have conducted similar robustness checks using mothers with children 4- and 5-years older (estimates available in upon request). These estimates are similar in magnitude and sign to those estimated with mothers of 2-year olds suggesting that *all* mothers of young children experienced a negative employment shock after the reform in Spain.

statistically significant at the 99 percent level and are slightly larger in size to DDD estimates presented earlier. Note that the DD estimates are considerably smaller and not statistically significant when they are estimated using the sample of mothers *not* affected by the reform, that is mothers of 2-year olds--shown in column 4 in Table 4. This lack of result among this group is important as it reveals that the reform led to *no* "anticipation effect" among mothers of 2-year olds.

A final concern is that the timing of the implementation might be endogenous. To address this, we estimated a similar specification as in equation (1) but our $Post\_reform_t$ variable is now a dummy equal 1 one year *earlier* and zero otherwise. The coefficients on the interaction between our pre-reform variable and the treatment groups are not statistically significant indicating that endogeneity of the implementation of the reform does not seem to be a concern.[26]

Fertility Effects

One concern with this methodology is that fertility may also be affected by the reform, leading to a change in the composition of our treatment and comparison groups before and after the legislation, which would bias our estimates on the effects of the legislation on employment. To evaluate if the potential endogeneity of fertility is a concern, we analyze whether there were any effects of the reform on fertility.

The childcare cost reduction derived from the free preschool expansion could affect childbearing decisions either positively, because the direct reduction in the cost of having a child, or negatively through its effect on female labor participation. We therefore explore the net effect on fertility. As all childbearing-age women living in early implementers' states are potentially affected, we estimate the following equation:

$$Y_{it}=\alpha_0+\alpha_1 Post\_reform_t+\alpha_2 t+\alpha_3 t^2+X'_{it}\beta+Z_t'\lambda+\varepsilon_{it} \qquad (2)$$

where Y_{it} takes the value one if a woman i gave birth during the last 12 months and zero otherwise in quarter t. $Post\_reform_t$ takes value of 1 if the period is after the preschool component of the LOGSE has been implemented in each state, 0 otherwise. Thus, the α_1 coefficient captures any breaks in the fertility trend corresponding with the timing of the free preschool expansion in each state. The vector X_{it} includes individual-level variables expected to be associated with childbearing decisions (the covariates used in the previous models plus age cube and interactions terms between age, age squared and age cube and

[26] The coefficients are -0.0210 (s.e. 0.0141) in the employment equation and -0.6949 (s.e. 0.5392) in the hours equation.

the education dummies). We also include states and year fixed-effects. Finally, the vector Z_t includes aggregate controls: the state unemployment rate and the average hourly wages. Results (shown in Appendix Table A.3) reveal that, despite the increase in maternal labor supply, we do not find any significant effect on childbearing decisions. As a consequence, potential biases in our employment estimates due to endogeneity of fertility are unlikely to be a source of concern.

2.7 Persistent Effects of the Reform and Heterogeneity

Although the current effect of the reform are modest, it is important to know whether they persisted over time or whether there was heterogeneity effects. This section first estimates the effects of the reform on mothers of older children (up to 7-years old) who had been affected by the policy change when their child was 3. Then, it conducts heterogeneity analysis by maternal education level, age, and number of children.

Persistence Analysis

Why would the policy have any effect on the labor supply of mothers with 4- (5-, 6-, or 7-) year olds? This may have happened if the reform led these mothers to enter the labor market when the child was 3 but, in the absence of the reform, they would *not* have entered employment even when the child turned 4 (5, 6, or 7). To put it differently, if by reducing the time mothers spent outside employment from 3 to 2 years, the reform led to a reduction in the number of women who permanently exit the labor market after birth, we would expect to find persistent effects of this legislation.

Prior to the reform, labor force participation of mothers whose youngest child was 4- to 7-years old was below 35 percent in Spain compared to almost 60 percent in Quebec.[27] Moreover, in Spain, maternal employment does *not* increase a great deal with the age of the youngest child (contrary to what is observed in most developed countries). Indeed, Figure 3 shows that, prior to the reform, the employment rate of mothers was slightly over one half the average employment rate of childless women. Most importantly, this figure shows that the employment rate of mothers of children 8 to 18 years old is not much higher than that of mothers of 3 to 7 year olds. Consistent with this, Gutierrez-Domenech (2005), finds that the proportion of women with paid work falls from 43 percent to 33 percent after a first birth in Spain and remains around 35

[27] Quebec being the reference since this is where the other study analyzes persistence.

percent ten years after they gave birth, providing evidence consistent with permanent (rather than temporary) exits from the labor force.

Within this context, universal childcare for 3-year olds may have an effect on maternal employment once the child has turned 4 and began preschool by enhancing mothers' human capital. Shortening the span of time mothers spend outside the labor market (even if it is just by one year) stops the depreciation of human capital accumulated in school and in former jobs and allows for the accumulation of new human capital acquired in the job. As Lefebvre *et al*. (2009), explain "*this changes the expected evolution of future wages so that women who never expected to work while raising children re-evaluate their life-time utility and return to work or start working*". Indeed, Fernández-Kranz *et al.* (2013), analyze the family gap in Spain and find that a far from negligible amount of the earnings differential is explained through experience and the amount of hours worked. Alternatively, the fact that a mother spends less time outside of the labor force may also affect her cognitive and non-cognitive job-search skills (as well as her social and professional networks) in such a way that it may shorten the time it takes her to find a job. Note that this mechanism may be particularly relevant in a context such as the Spanish one with rigid labor markets where the unemployment rate in the early 1990s was above 20 percent, and where finding a job takes a long time--according to the 1987-1990 Spanish LFS, 46 percent of women spend on average two years to find a job in Spain (compared to 35 percent of men).

To analyze whether any effects of universal childcare on maternal employment persist over time, we estimate the same specification as the one in equation (1) but changing *both* the "treatment" and the "$Post\_reform_t$" variable as follows. When we estimate the effects of the reform one year later, the treatment group is defined as mothers whose youngest child is 4-years old. To guarantee that her child was eligible for universal childcare when he or she was 3, the $Post\_reform_t$ variable takes value of 1 *one year after* the state began implementation of the reform, and 0 otherwise. Similarly, when we estimate the effects of the reform two (three and four) years later, the treatment group is defined as mothers whose youngest child is 5- (6- and 7-) years old, respectively. In these cases, the $Post\_reform_t$ variable takes value of 1 two (or three or four) years *after* the state began implementation of the reform, and 0 otherwise.

We continue to use mothers whose youngest child was 2 as a comparison group, as they were not affected by this reform at that point in time. As a robustness check, we also use an alternative comparison group of mothers whose youngest child is older (up to two years older) but who was *not* eligible for the universal day care program when the child was three. Results were robust to those shown below.[28]

In the persistence analysis, one may be concerned that selection bias may arise because of how the treatment groups are constructed. As we select for our treatment groups mothers whose *youngest* child is 4 (through 7) years old, we are evaluating the effect of the LOGSE on mothers who were affected by this legislation when their child was 3 and who did *not* have any additional children thereafter. If the legislation affects the fertility decisions of these mothers, this may lead to selection bias. Despite we find no effect of the policy on fertility, in the sensitivity section we explore how robust are our results to alternative definitions of treatment and control groups.

In what follows we focus our analysis on weekly hours worked (estimates of the effects of the reform on employment are shown in the Appendix Table A.4). Panel A in Table 5 shows that the effect of universal preschool for 3-year olds on maternal weekly hours worked persists, on average, for at least two more years. Indeed, our estimates show that the positive effect of the reform on maternal weekly hours worked remains statistically significant and of similar magnitude until the child is 5-years old. We find that the reform increased weekly hours worked for mothers of 3-year olds by 1.09 hours per week one and two years after the child had been eligible for childcare (or 9.7 percent and 9.5 percent, respectively).[29] Panel B in Table 5 shows that these results are robust to clustering the standard errors at the state level.[30]

One concern is that our results from the persistence analysis emerge because we are restricting our treatment groups to mothers whose *youngest* child is 4 (or 5, or 6 or 7) years old, excluding those who decided to have another child after the focal child turns 3 years old. Panel C displays persistence estimates using as treatment group those mothers

---

[28] The reason for not using older children as our main comparison group is that the restriction that the control group are mothers whose youngest child was not affected by the LOGSE implies that we would be restricted to using *only* up to two years after the implementation in each state. Nonetheless, when we did use mothers of older children as a comparison group, the results were similar to those presented in the main text.

[29] Estimates of the reform on the likelihood of employment show that the reform led to a relative increase of 2.35 percentage points (7.6 percent) and 2.21 percentage points (7.1 percent) in the employment of mothers whose youngest child was 3 one and two years *after* the child had been eligible to participate, respectively. However, we find that, thereafter, these effects fade away as the coefficients become smaller and are no longer statistically significant (estimates shown in Appendix table A.4).

[30] We estimated DD estimates using our comparison group, that is, mothers whose youngest child is 2, to explore whether we ought to be concern of a potential "anticipation" effect among this group. We found no evidence of this. In the contrary, we found that in some cases, mothers of 2-year olds are less likely to work after the reform, an effect likely due to the timing of the reform with the recession, as explained earlier.

who decided to have another child after the focal child turned 3 years old. For instance when we analyze the effect of the policy two years later, we use as treatment group mothers who have a child of 5 years old and who may also have children of 0- or 1- or 2-years old.[31] Although the persistence estimates are slightly smaller (as one would expect), overall they remain positive and statistically significant. Finally, panel D in Table 5 shows the placebo test using pre-reform data. The size of the coefficient is considerably smaller and non-significant and, when significant, it has the wrong sign.

Heterogeneity Effects

If human capital and job-search skills matter, one would expect persistence to be strongest among higher skilled workers as they are those who hold jobs in which their human capital depreciates faster. In the early 1990s, less than 10 percent of mothers in Spain held a university degree (shown in Table 1). Thus, within this context, higher skilled workers are those with a high-school or college degree versus high-school dropouts who are likely to hold jobs requiring little qualification. Alternatively, older women are also more likely to have more experience and hold more qualified jobs than younger ones.[32] Table 6 reports the policy effects on hours worked by mother's educational attainment (panel A), and mothers' age (panel B). In addition, Table 6 also explores heterogeneity effects of this policy by mother's number of children (panel C).[33] Finally, it is worth highlighting that evidence that the persistence effects are driven by the same subgroups than when the child is eligible for the program further supports the result of persistence.

Panel A in Table 6 shows that the overall effect of the reform on mothers' hours worked is mainly driven by a significant effect among high-school graduates. For this group of mothers, we find that the reform increased hours worked when the child was eligible for public childcare. Moreover, we find that this effect persisted for at least four years after the child was eligible for the public childcare program. For instance, we find that the reform led to an average increase of 1.77 hours per week (or 12 percent) four years after the child was eligible for the public childcare program. The size of this coefficient doubles the one for the average population found earlier. Assuming that the

[31] Note that in this case, we only include in our control group mothers whose youngest is 2 years old and who do not have other children who may have been affected by the legislation, such as a 5 year old.

[32] In 1998 Spanish women had a first child on average at age 29.1 (de la Rica and Iza, 2005).

[33] As 97 percent of our sample is married or cohabitating, we are unable to estimate the analysis for single mothers.

take up is homogeneous across mothers' skill level, this would imply that a 1 percentage point increase in childcare coverage for 3-year olds in Spain led to an increase of 13.97 hours worked per week four years after the child was affected by this legislation.[34]

The effect of the reform among mothers without a high-school degree is of similar magnitude for the year the child is eligible for the program and up to one more year. However, the effects of the program on maternal employment measured three and four years after the child was eligible are negative (albeit not significant). Thus, the heterogeneity analysis reveals that the fading away of the average effect is driven by the low-skilled mothers. These results contrast with those of Lefebvre and Merrigan (2009), who find that that the effects of the reform in Quebec are driven by the least educated workers.[35] This paper cannot identify which mechanisms are at play behind our persistence results but the fact that the effects of the reform are particularly strong and persistent among mothers with a high-school degree (but not among high-school dropouts) suggests that by shortening the time span mothers of small children stay out of the labor force, the childcare program reduces the depreciation of human capital accumulated in school and in former jobs, and it also permits the accumulation of new human capital acquired on the job. As high-school dropouts tend to be concentrated in non-qualified jobs, the accumulation of human capital is less relevant, explaining the milder persistence among this group.

Among mothers with a college degree, we find no effect of the reform. The lack of results for this population is not infrequent in this literature for the following two reasons.[36] First, these women are usually in jobs that pay relatively well and thus are able to pay day care (even when it is mainly privately supplied). As a consequence, we would expect them to be less responsive to a large subsidy of day care, such as the one under analysis. Second, as many of these highly educated women are strongly involved in the labor market (as many as 70 percent of them were employed prior to the reform), it is difficult to observe large effects of this reform (or any other similar reform).[37]

[34] Unfortunately information on the child enrollment rate by their mothers' characteristics is unavailable.

[35] The definition of low- and high-educated differs across studies. Lefebvre and Merrigan, 2009, classify workers by whether they have a post-secondary diploma or not as over 40 percent of their mothers have a high-school degree.

[36] Lefebvre *et al.*, 2009, find that the policy effects are strong and persist among the low-skilled (defined as those without a college degree), but not among college educated mothers. The authors do not present outcomes by whether the mother is a high-school graduate or not.

[37] For instance, both Sánchez-Mangas and Sánchez-Marcos (2008), and Azmat and González (2010), find no effect of 1998 and 2003 tax reforms on maternal employment among college graduates.

Consistent with our finding that most of the effects of this reform are among the higher skilled workers, Panel B in Table 6 shows that both the current and persistent effects of the reform are driven by mothers 29 years old or older. Clearly these women have accumulated more experience in the labor market, and thus, the reform has made more of a difference for them (in terms of less depreciation of human capital). Finally, we also observe that most of the current and persistent effect of this reform is driven among women with two or more children. A possible explanation for this is that women with two or more children are likely to have achieved their optimal family size, and thus, they may be more responsive to the introduction of universal childcare for their youngest child. Instead those mothers who only have one child and wish to have another may prefer to postpone labor market involvement to engage (again) in motherhood.

2.8 Conclusion

Using a natural experiment framework and the introduction of universal full-time childcare for 3-year olds in Spain, this paper aims at tackling the effect of childcare on maternal employment in a context of low female labor force participation and meager supply of childcare spaces. Our results reveal that even under such scenario this policy alone will not have large effects on maternal employment even when public childcare does *not* crowd out private childcare arrangements. Nonetheless, our study reveals that although the current effects of the reform in Spain on maternal employment were modest (given the expansion in childcare), they persisted over time as the child aged and they were driven by mothers with a high-school degree and older mothers, suggesting that this policy reduced the depreciation of human capital accumulated in school and in former jobs, and that it also encouraged the accumulation of new human capital acquired on the job.

When comparing these estimates to those of the 2003 Spanish tax credit of €1,200 Euros for working mothers of children 0 to 3 years old, we observe that the provision of universal preschool is (at least) as effective in terms of getting mothers to work as the direct subsidy. While we find that the introduction of free childcare for 3-year olds led to a relative increase of 7.9 percent in maternal employment, Azmat and Gonzalez, 2010, find that the "direct" effect of the tax credit led to a relative increase of 6.5 percent in maternal employment. As the tax credit was a cash payment equivalent to about one third of (public) childcare costs that was *only* paid to working mothers, if we *only* consider the current effect of policy, the costs of universal childcare coverage for 3-year olds well exceeded those of the alternative policy. While we did find persistence in our results, it is likely that the tax credit also had similar dynamic effects--as the policy effects were driven by high-school graduates and older workers as in our study.

A related important question is whether universal childcare has short and long-term beneficial or detrimental effects on the cognitive or non-cognitive development of children relative to other forms of early childhood care, such as parental or relative care. If the Spanish policy had significant long-term beneficial effects on children's development (as found in other countries by Berlinski *et al.* 2009, Fitzpatrick 2008, Cascio 2009b, and Havnes and Mogstad 2011b), equity and efficiency reasons may justify the larger costs of implementing universal childcare instead of offering direct cash payment to working mothers. The next essay study the same childcare reform, but address the impact on the cognitive development at the end of mandatory schooling, when the children are 15 years old.

REFERENCES

Adserá, A. (2004) "Changing fertility rates in developed markets. The impact of labor market institutions". *Journal of Population Economics*, 17, 17–43.

Amuedo-Dorantes, C., and S. de la Rica (2009) "The Timing of Work and Work-Family Conflicts in Spain: Who Has a Split Work Schedule and Why?" IZA Discussion Paper 4542.

Anderson, P y Levine, P (2000) "Childcare and Mothers' Employment Decisions," in David E. Card and Rebecca M. Blank (eds.) Finding Jobs: Work and Welfare Reform. New York: Russell Sage Foundation.

Averett, S.L. and J. L. Hotchkiss (1997) "Female Labor Supply with a Discontinuous, Non-Convex Budget Constraint: Incorporation of a Part-Time/Full-Time Wage Differential" *The Review of Economics and Statistics* 79, 461-470.

Azmat G., and L. González (2010) Targeting Fertility and Female Participation through the Income Tax. *Labour Economics* 17, 487-502.

Baker, M., Gruber, J. and K. Milligan (2008) "Universal Childcare, Maternal Labor Supply and Family Well-Being." *Journal of Political Economy*, vol. 116 (4) 709-745.

Bentolilla, S. (2001) Las migraciones interiores en España, in J.A. Herce and J.F. Jimeno (eds.), Mercado de Trabajo, Inmigración y Estado del Bienestar, Madrid, Fedea and CEA.

Berlinski, S and S. Galiani (2007) "The effect of a large expansion of pre-primary school facilities on preschool attendance and maternal employment." *Labour Economics* 14, 665-680.

Berlinski, S, Galiani, S and P. Gertler (2009) "The Effect of Pre-Primary Education on Primary School Performance" *Journal of Public Economics*, 93(1-2), 219-234.

Bertrand, M., Duflo, E. and S.Mullainathan (2004) "How Much Should We Trust Differences-in-Differences Estimates?"*Quarterly Journal of Economics* 119, 249–275.

Berger, M. C. and D. A. Black (1992) "Child Care Subsidies, Quality of Care, and the Labor Supply of Low-Income Single Mothers" *Review of Economics and Statistics*, 74(4): 635-42.

Bettio, F and P. Villa (1998) "A Mediterranean perspective on the breakdown of the relationship between participation and fertility". *Cambridge Journal of Economics*22 (2), 137-171.

Blau, D. and Currie, J. (2006) "Who's Minding the Kids?: Preschool, Day Care, and After School Care," The Handbook of Education Economics, Finis Welch and Eric Hanushek (eds). New York: North Holland, v.2 1163-1278.

Blau, D. and Robins, Ph. (1989) "Fertility, Employment, and Child-Care Costs". *Demography*, Vol. 26, No. 2 (May, 1989), pp. 287-299.

Boeri, T., D. Del Boca, and C. Pissarides (eds.) (2005) *Labor Market Participation and Fertility of Women: the Effect of Social Policies*. Oxford University Press, UK.

Cascio E. (2009a) "Maternal Labor Supply and the Introduction of Kindergartens into American Public Schools," *Journal of Human Resources*.44 (1), 140-170.

Cascio E. (2009b) "Do Investments in Universal Early Education Pay Off? Long-term Effects of Introducing into Public Schools", NBER WP 14951.

Cleveland, G., M. Gunderson, and D. Hyatt (1996) "Childcare Costs and the Employment Decision of Women: Canadian Evidence." *Canadian Journal of Economics* Vol. 29, No. 1, pp. 132-148.

Connelly, R. (1992) "The effect of child care costs on married women's labor force participation" *Review of Economics and Statistics* 74, 83-90.

Del Boca, D. (2002) "The effect of childcare and part time opportunities on participation and fertility decisions in Italy," *Journal of Population Economics*, Springer, 15(3), 549-573

De Laat, J and A. Sevilla-Sanz (2011) "The Fertility and Women's Labor Force Participation puzzle in OECD Countries: The Role of Men's Home Production". *Feminist Economics*, 17 (2), 87-119.

De la Rica, S., and Iza (2005) "Career Planning in Spain: Do Fixed-Term Contracts Delay Marriage and Parenthood?" *Review of Economics of the Household,* 3, 49-73.

El País, October 3 (2005) "La LOGSE, 15 años después", by Elena Martín Ortega.

Fernández-Kranz, D., and N. Rodríguez-Planas (2011a) "The Part-Time Penalty in a Segmented Labor Market." *Labour Economics,* Volume 18, 591–606.

Fernández-Kranz, D., and N. Rodríguez-Planas (2011b) "Unintended Effects of a Family-Friendly Law in a Segmented Labor Market", IZA DP 5709.

Fernández-Kranz, D., Lacuesta A., and N. Rodríguez-Planas "Motherhood Earnings Dip: Evidence from Administrative Data." Forthcoming in *Journal of Human Resources*.

Feyrer, J., Sacerdote, B. and A.D. Stern (2008) "Will the Stork Return to Europe and Japan? Understanding Fertility within Developed Nations" *Journal of Economic Perspectives* (American Economic Association), 22 (3), 3-22.

Fitzpatrick, MD (2008) “Starting School at Four: The Effect of Universal Pre-Kindergarten on Children's Academic Achievement” *The B.E. Journal of Economic Analysis & Policy*, Berkeley Electronic Press, vol. 8(1), pages 46.

Fitzpatrick, MD (2010) “Preschoolers Enrolled and Mothers at Work? The Effects of Universal Pre-Kindergarten” *Journal of Labor Economics*, 28(1), 51-85, 01.

Gelbach, J. (2002) “Public schooling for Young children and Maternal Labor Supply”. *American Economic Review* 92, 307–322.

Gennetian, L. A., A. Gassman-Pines, A. C. Huston, D. A. Crosby, Y. E. Chang and E. D. Lowe (2001) "A Review of Child Care Policies in Experimental Welfare and Employment Program" New York: Manpower Demonstration Research Corporation.

Goux, D. and E. Maurin (2010) “Public school availability for two-year olds and mothers' labour supply” *Labour Economics*, 17, Issue 6, 951-962.

Gutiérrez-Domènech, M. (2005) “Employment Transitions after Motherhood in Spain”.*Review of Labour Economics and Industrial Relations*, 19, 123–148.

Havnes, T. and M. Mogstad. (2011a) "Money for nothing? Universal childcare and maternal employment". *Journal of Public Economics*, 95(11-12), 1455-1465.

Havnes, T. and M. Mogstad. (2011b) "No child left behind: Subsidized childcare and children's long-run outcomes". *American Economic Journal: Economic Policy*, 3(2),97-129.

Heckman, J. (1974) “Shadow Prices, Market Wages and Labor Supply”. *Econométrica*. Vol. 42 (4): 679-694.

Jimeno, JF and S. Bentolilla (1998) “Regional Unemployment Persistence (Spain, 1976-1994)”*Labour Economics*. Vol. 5, Issue 1, March 1998, Pages 25-51.

Kimmel, J. (1995) “The Effectiveness of Child-Care Subsidies in Encouraging the Welfare-to-Work Transition of Low-Income Single Mothers” *The American Economic Review.* Vol. 85 (2):

Lefebvre, P. and Ph. Merrigan (2008) “Childcare Policy and the Labor Supply of Mothers with Young Children: A Natural Experiment from Canada,” *Journal of Labor Economics*, 23(3).

Lefebvre, P., Merrigan, Ph and M. Verstraete (2009) “Dynamic labour supply effects of childcare subsidies: Evidence from a Canadian natural experiment on low-fee universal childcare”. *Labour Economics,* 16, 490-502.

LOGSE, October 3 (1990) Boletín Oficial del Estado (BOE) number 238.

Lundin, D.,Mörk, E. and B. Öckert (2008) “How far can reduced childcare prices push female labour supply?,” *Labour Economics*, 15(4), 647–659.

Meyer, B. D. (1995) “Natural and quasi-experiments in economics”, *Journal of Business and Economic Statistics*, 13, 151-161.

Michalopoulos, Ch. and Ph. K. Robins (2000) “Employment and Child-Care Choices in Canada and the United States,” *Canadian Journal of Economics* Vol. 33 (2): 435-470.

Michalopoulos, Ch., Ph. K. Robins and I. Garfinkel (1992) “A Structural Model of Labor Supply and Childcare Demand” *Journal of Human Resources*, Vol. 27 (1): 166-203.

OECD (2001) "Starting Strong. Early Childhood Education and Care"

OECD (2007) “Babies and Bosses - Reconciling Work and Family Life: A Synthesis of Findings for OECD Countries”.

Powell, L. M. (1997) “The Impact of Childcare Costs on the Labor Supply of Married Mothers: Evidence from Canada.” *Canadian Journal of Economics* Vol. 30 (3): 577-594

Ribar, D.C. (1992) "Childcare and the labor supply of married woman" *Journal of Human Resources.* Vol. 27: 134-65.

Ribar, D.C. (1995) "A Structural Model of Child Care and the Labor Supply of Married Women" *Journal of Labor Economics*. Vol. 13 (3): 558-597.

Ruhm, Ch. J. (1998) "The Economic Consequences of Parental Leave Mandates: Lessons from Europe." *The Quarterly Journal of Economics*, 113(1), 285-317.

Sánchez-Mangas, R. and V. Sánchez-Marcos (2008) “Balancing family and work: the effect of cash benefits for working mothers”. *Labour Economics*,15, 1127-1142.

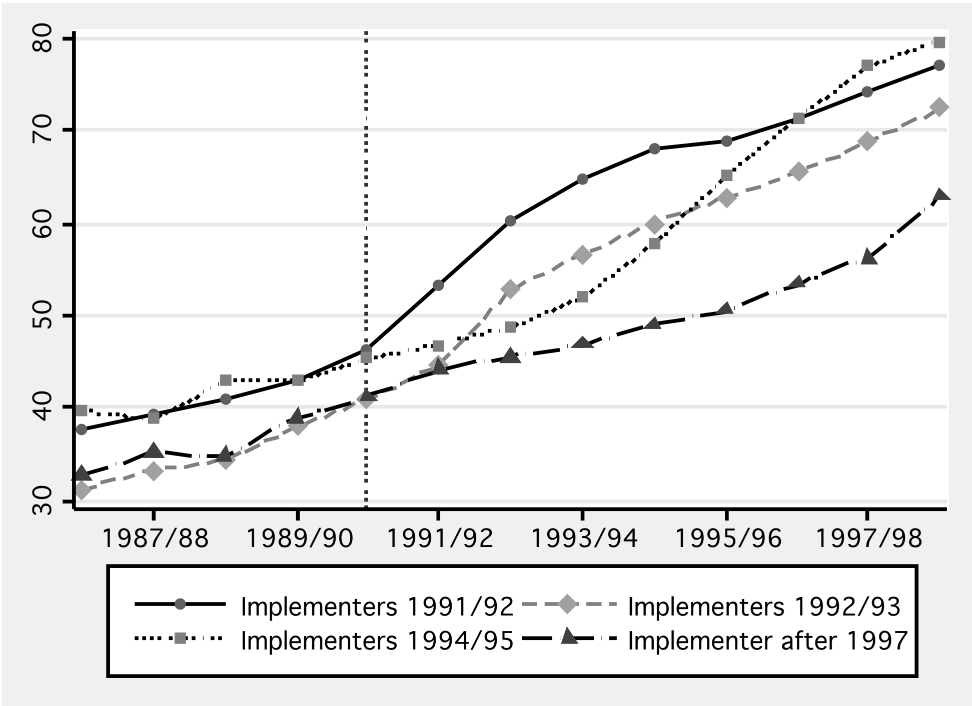
Schlosser, A. (2006) “Public Preschool and the Labor Supply of Arab Mothers: Evidence from a Natural Experiment”, *Economic Quarterly*, Vol. 3.

Sevilla-Sanz A., J. I. Gimenez-Nadal, and C. Fernandez (2010) "Gender Roles and the Division of Unpaid Work in Spanish Households." *Feminist Economics*, Taylor and Francis Journals, vol. 16(4), pages 137-184.

TABLES AND FIGURES

Figure 1. Public Child-Care Coverage Rates for Children 3- to 5-Years Old, by the Timing of the Implementation

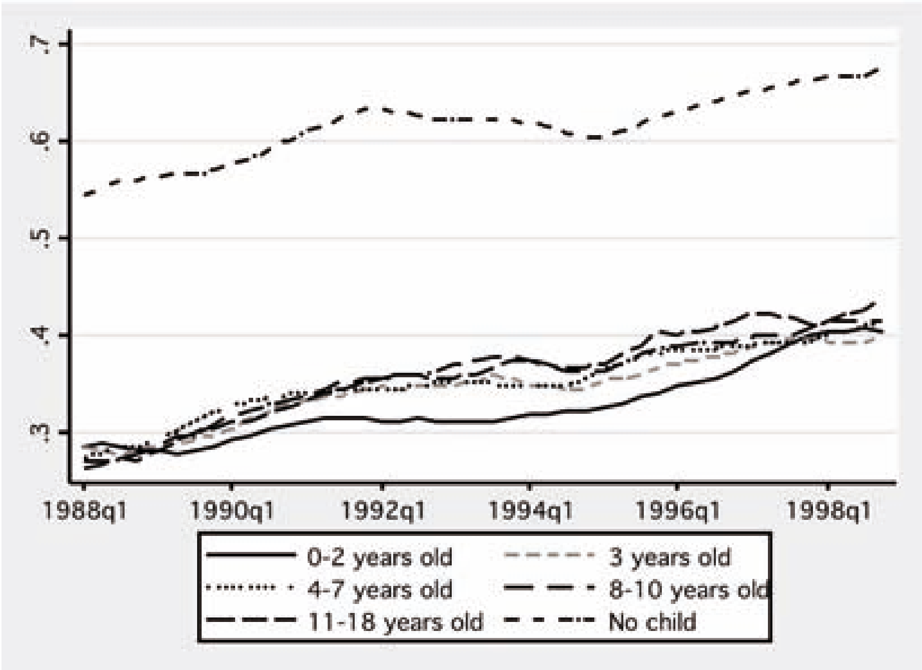
Share of public places supplied



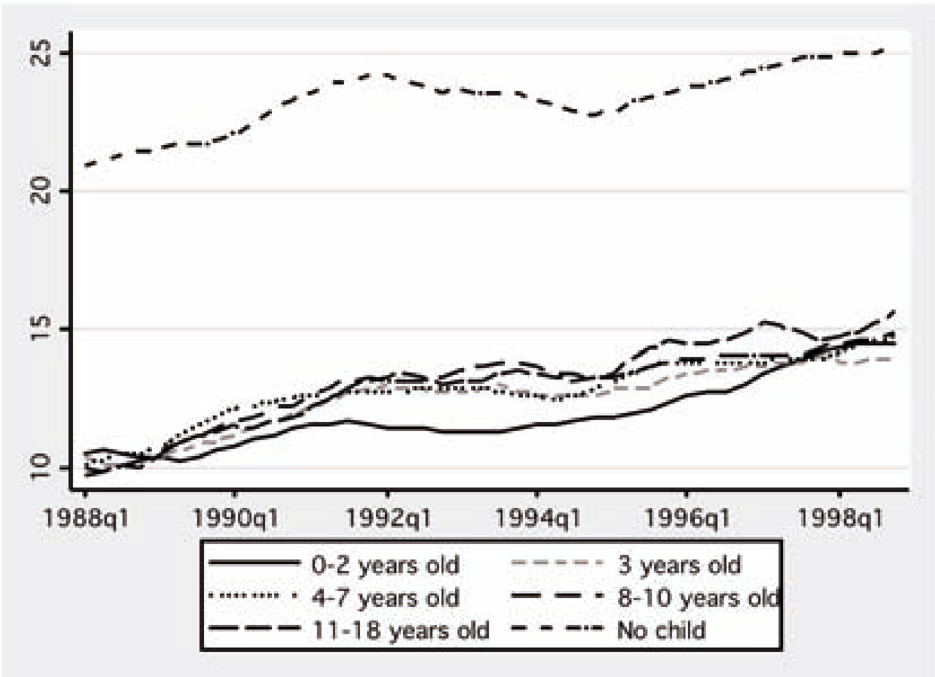
Notes: Elaborated by the authors based on statistics by state from the Ministry of Education. The figure displays the proportion of public preschool places offered in each group of state. It was estimated as the number of public preschool units in each region times the average size of the classroom divided by the population of 3- to 5-years old in each state.

Figure 2. Maternal Employment Rates and Weekly Hours Worked, by Age of the Youngest Child

Maternal employment rates



Weekly hours worked



Notes: Elaborated by the authors based on Spanish Labor Force Survey micro data.

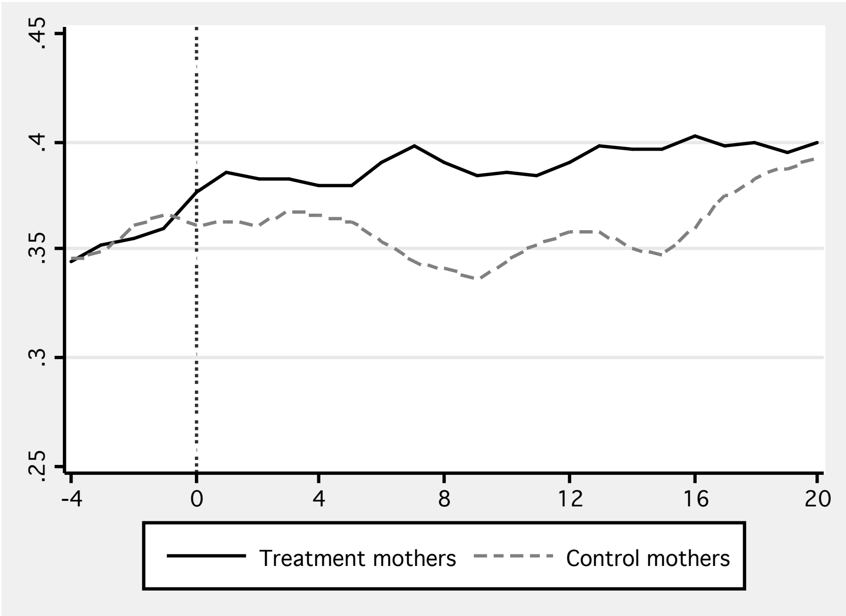
Table 1. Baseline Descriptive Statistics

	Impl 1991/92		Impl 1992/93		Impl 1994/95		Impl after 1997	
Age								
Mothers of 3-year olds	32.07	(5.21)	32.32	(5.28)	32.21	(5.53)	32.30	(5.35)
Diff Treat-Control	0.91+		0.61+		1.12+		0.79+	
Number of children								
Mothers of 3-year olds	2.07	(1.10)	2.10	(1.15)	2.29	(1.25)	2.24	(1.18)
Diff Treat-Control	0.09+		0.08+		0.04		0.10+	
Immigrants								
Mothers of 3-year olds	0.00	(0.06)	0.01	(0.11)	0.02	(0.15)	0.01	(0.10)
Diff Treat-Control	0.00		0.00		-0.01+		0.00	
Cohabiting								
Mothers of 3-year olds	0.98	(0.13)	0.98	(0.14)	0.96	(0.19)	0.98	(0.14)
Diff Treat-Control	-0.01+		-0.01+		0.00		0.00	
HS dropout								
Mothers of 3-year olds	0.54	(0.50)	0.47	(0.50)	0.52	(0.50)	0.56	(0.50)
Diff Treat-Control	0.05+		0.03+		0.05+		0.05+	
HS graduated								
Mothers of 3-year olds	0.37	(0.48)	0.43	(0.50)	0.40	(0.49)	0.37	(0.48)
Diff Treat-Control	-0.04+		-0.01		-0.03		-0.04+	
College								
Mothers of 3-year olds	0.09	(0.29)	0.11	(0.31)	0.08	(0.28)	0.07	(0.26)
Diff Treat-Control	-0.01+		-0.02+		-0.02+		-0.01+	

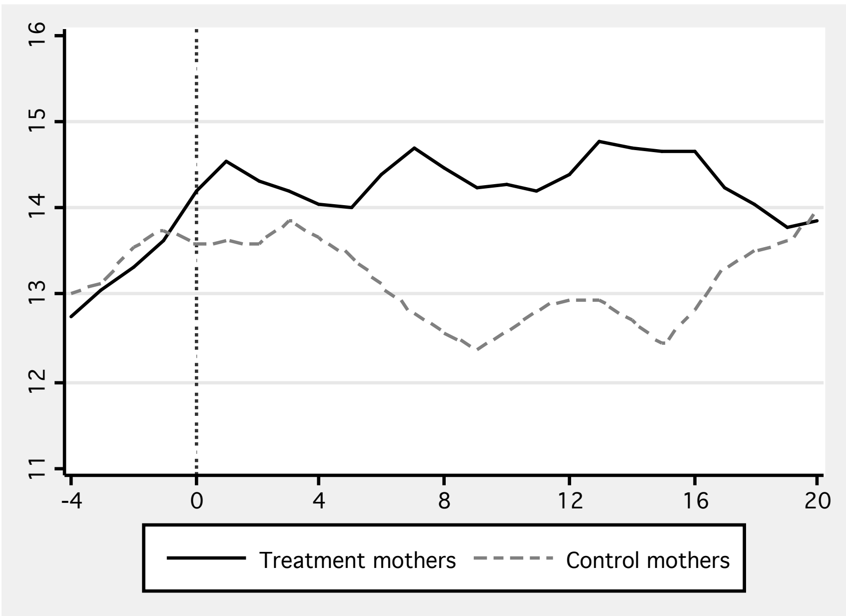
Notes: We present the mean and standard deviations (in parenthesis) in the pre-reform period for each group of implementing states : 1987-1991 (third quarter) for the implementers in 1991/92, 1987-1992 (third quarter) for the implementers in 1992/93 and so on.
+ indicates that the mean between the treatment and control group of mothers (that is, mothers of 3- and 2-year olds, respectively) during the pre-reform period is significantly different from zero at least at 95 percent of confidence level.

Figure 3. Treatment and Control Groups' Employment Rates and Weekly Hours Worked, Before and After the Implementation of the Legislation

Employment



Weekly hours worked



Notes: Treatment mothers are those whose youngest child is 3 years old and control mothers are those whose youngest child is 2 years old. We set the point 0 at the quarter of implementation in each state (for instance, fourth quarter of 1991 for Catalunya, fourth quarter of 1992 for Madrid, fourth quarter of 1994 for the Canary Islands, and so on). We then estimate a weighted average across states at each point in time. We display annual moving average of quarterly data, therefore axis labels refer to quarters.

Table 2. DDD Estimates of Universal Childcare on Maternal Employment

	Preferred specification	Clustered std. errors	Linear Trend by State	Flexible	Alternative control group: 4- and 5-year olds	Balanced panel	DD without regional variation	Placebo test
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Employment								
DDD	0.023	0.023	0.032	0.028	0.033	0.036	0.026	-0.005
	[0.008]***	[0.011]***	[0.015]**	[0.012]**	[0.010]***	[0.016]**	[0.012]**	[0.010]
Pre-average	0.293	0.293	0.293	0.293	0.293	0.329	0.302	0.275
% effect	7.9%	7.9%	11.0%	9.4%	11.4%	11.0%	8.8%	-2.0%
Panel B. Weekly Hours Worked								
DDD	0.949	0.949	1.078	0.959	1.214	1.345	0.931	-0.233
	[0.332]***	[0.499]*	[0.587]*	[0.476]**	[0.380]***	[0.6260]**	[0.480]*	[0.385]
Pre-average	10.907	10.907	10.907	10.907	10.907	12.33	11.280	10.907
% effect	8.7%	8.7%	9.9%	8.8%	11.1%	10.9%	8.3%	-2.1%
N	*105,748*	*105,748*	*105,748*	*105,748*	*109,717*	*74,653*	*105,748*	*64,769*

Notes: Robust standard errors in brackets; ***, **, * denote statistical significance at 0.01, 0.05 and 0.10 levels, respectively. DDD specification in column (1) includes year and states fixed-effects and a linear trend that differs for the treatment and control groups. It also includes individual (age; dummies indicating the level of education: HS dropout (omitted), HS graduated or college; immigration status; number of other children; marital status) and regional (unemployment rate) characteristics. Column (2) displays the same specification but clustering the standard errors at state level and column (3) presents a specification that includes a linear trend interacted by state dummies. In Column (4) presents a more flexible specification that also includes interactions between the treatment dummy and states fixed-effects and between the treatment dummy and covariates. Column (5) displays the results from estimating the same equation as in (1) but using as a control group mothers with children 4 and 5 years old instead mothers of children of 2 years old. Column (6) shows the results from estimating the same equation only using states implementing in 1991/92 and 19992/93. Column (7) displays the results from estimating a DD approach assuming that all states start the implementation in 1991/92 (that is, ignoring regional variation in the policy implementation). Finally, column (8) shows the results from estimating the same equation in (1) but using only the pre-reform data in each state and assuming the reform was implemented two years before.

Table 3. Effect of the LOGSE on Enrollment Rates of 2- and 3-Year Olds

	Enrollment rate of 3 years old		***Enrollment rate of 2 years old***	
	Publicly-funded schools	***Private Schools***	***Publicly-funded schools***	***Private Schools***
Post_reform	0.126	0.017	0.004	0.009*
	[0.026]***	[0.006]***	[0.005]	[0.005]
Linear trend	0.045	0.008	0.000	-0.001
	[0.005]***	[0.001]***	[0.001]	[0.001]
GDP per capita	0.000	0.000	0.000**	0.000**
	[0.000]	[0.000]	[0.000]	[0.000]
Growth of GPD per capita	-0.069	-0.031	-0.032*	-0.058*
	[0.073]	[0.023]	[0.018]	[0.025]
Unemployment rate	-0.004	0.001	0.001	0.000
	[0.003]	[0.001]	[0.001]	[0.001]
R-squared	0.886	0.894	0.561	0.765
N	150	150	150	150

Notes: Robust standard errors in brackets; ***, **, * denote statistical significance at 0.01, 0.05 and 0.10 levels, respectively. The variable *Implement* takes the value of one if the LOGSE was implemented in the state *s* at the time *t* and zero otherwise. We use data of enrollment rates at state level from the Ministry of Education.

Table 4. DD Estimates of Universal Childcare on Mothers of 3-Year Olds

	Mothers of 3 years old		***PLACEBO TEST: Mothers of 2 years old***	
	No trend	***With trend***	***No trend***	***With trend***
Panel A. Employment	0.0127	0.0420	-0.0152	0.0066
	[0.0084]	[0.0129]***	[0.0083]*	[0.0128]
Panel B. Weekly hours worked	0.2959	1.5172	-0.6491	0.5518
	[0.3328]	[0.5156]***	[0.3335]*	[0.5186]
Observations:	*53,012*	*53,012*	*54,760*	*54,760*

*Note*s: Robust standard errors in brackets; ***, **, * denote statistical significance at 0.01, 0.05 and 0.10 levels, respectively. DD specification includes year and states fixed-effects and individual and regional controls (see Table 2). Specification with trend includes a linear trend by state.

Table 5. Persistence Effects on Weekly Hours Worked, DDD Estimates

	1 year later	2 years later	3 years later	4 years later
Panel A. Preferred specification				
DDD	1.0877	1.0905	0.4494	0.5082
	[0.3563]***	[0.3881]***	[0.4301]	[0.4970]
Pre-average	11.492	11.489	11.954	12.358
% effect	9.7%	9.5%	3.7%	4.0%
Panel B. Clustering standard errors				
DDD	1.0877	1.0905	0.4494	0.5082
	[0.5755]*	[0.3782]**	[0.4509]	[0.4113]
Observations	105,036	102,404	100,340	98,109
Panel C. Treatment Group Mothers Who May Also Have Younger Children				
DDD	1.0811	0.8624	0.5197	0.2938
	[0.3495]***	[0.3848]**	[0.4338]	[0.5115]
Observations	109,686	106,899	103,741	100,522
Panel D. PLACEBO TEST: pre-legislation data				
DDD	0.0441	-0.9169	0.2404	-0.4668
	[0.4037]	[0.4150]**	[0.4398]	[0.4564]
Observations	71,658	76,783	82,008	86,252

Notes: Robust standard errors in brackets; ***, **, * denote statistical significance at 0.01, 0.05 and 0.10 levels, respectively. DDD model includes year and states fixed-effects, and a linear trend that differs for the treatment and control group, and individual and regional controls (see details in Table 2). Panel A presents our preferred specification with a linear trend interacted by treatment dummy. Panel B displays the same specification but clustering the standard errors at state level. In Panel C, we allow treatment mothers have younger children. Finally, Panel D shows the results from estimating the same equation in Panel A but using only the pre-reform data in each state and assuming the reform was implemented two years before.

Table 6. Heterogeneous Effects By Subgroups, Weekly Hours Worked

DDD Estimates

	Current effect	**1 year later**	**2 years later**	**3 years later**	**4 years later**
Panel A. By education level					
HS dropout	1.0426	1.0757	0.1604	-1.1077	-0.7277
	[0.5174]**	[0.5498]*	[0.6322]	[0.7094]	[0.8379]
N	44,073	45,637	46,099	45,962	45,872
HS graduate	0.9250	1.3748	1.4586	1.2388	1.777
	[0.4843]*	[0.5239]***	[0.5670]**	[0.6287]**	[0.7156]**
N	49,771	48,004	45,325	43,637	41,783
College	0.3946	-0.8221	1.7279	1.1651	-0.356
	[0.9727]	[1.0558]	[1.1166]	[1.1655]	[1.3406]
N	11,899	11,389	10,975	10,736	10,450
Panel B. By mothers age					
Younger than 30	0.0952	0.4521	-0.2626	0.154	0.1851
	[0.5803]	[0.6816]	[0.8685]	[1.1214]	[1.5102]
N	34,677	31,368	27,683	24,973	23,128
Older than 29	1.360	1.3805	1.4608	0.6075	0.9205
	[0.4039]***	[0.4227]***	[0.4471]***	[0.4853]	[0.5478]*
N	71,066	73,662	74,716	75,362	74,977
Panel C. By number of kids					
One child	0.1865	-0.4577	0.0107	0.5926	0.4102
	[0.5726]	[0.6373]	[0.7226]	[0.8202]	[0.9563]
N	38,122	35,293	32,354	30,287	29,014
Two or more	1.4496	1.9167	1.8293	0.4099	0.5957
children	[0.4055]***	[0.4292]***	[0.4633]***	[0.5126]	[0.5938]
N	67,621	69,737	70,045	70,048	69,091

Notes: Robust standard errors in brackets; ***, **, * denote statistical significance at 0.01, 0.05 and 0.10 levels, respectively. DDD model includes year and states fixed-effects, and a linear trend that differs for the treatment and control group, and individual and regional controls (see details in Table 2).

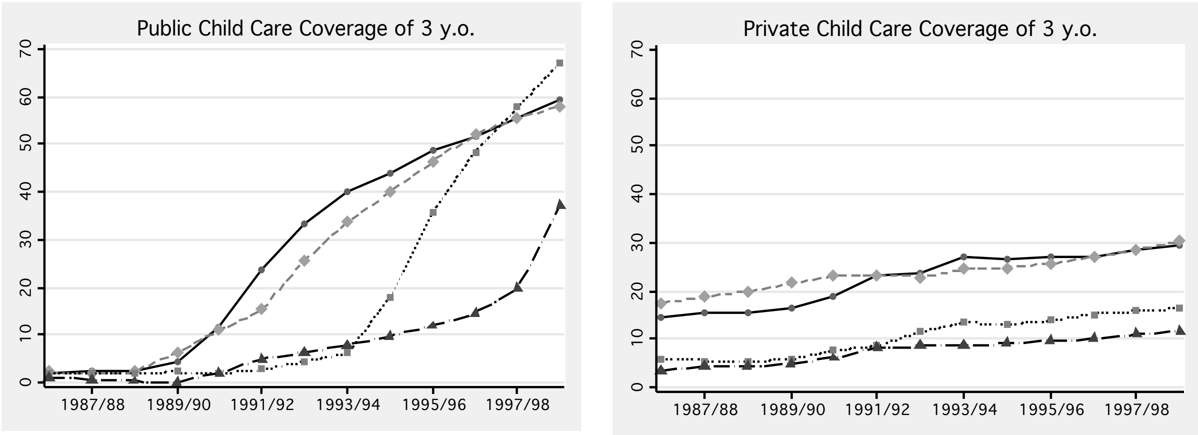
APPENDIX

Table A.1. Year of First State Funding for Three-Year Olds' Public Preschool

School year 1991/92	Asturias, Aragón, Baleares, Cantabria, Castilla-La Mancha, Catalunya, Comunitat Valenciana, Extremadura, and Galicia
School year 1992/93	Castilla y León, Madrid, Murcia, and La Rioja
School year 1994/95	The Canary Islands
School year 1998/99	Andalusia

Notes: Elaborated by the authors based on information of Ministry of Education.

Figure A.1. Public and Private Enrollment Rates of 3-Year Olds, by the Timing of the Implementation



Notes: Elaborated by the authors based on statistics by state from the Ministry of Education. Average by group of implementer weighted by population.

Table A.2. Descriptive Statistics for Groups of Implementers Before the Policy Implementation Began (1987-1990)

	Implementers 1991/92	***Implementers 1992/93***	***Implementer 1994/95***	***Implementer after 1997***
GDP growth (average annual rate, in %)	4.70	5.30	3.60	4.90
	(2.69)	(4.07)	(3.25)	(1.59)
GDP per capita (€)	9,404	10,374	9,757	7,528
	(1,790)	(1,898)	(355)	(393)†
Unemployment Rate (in %)	15.889	15.325	22.507	27.923
	(1.917)	(1.672)	(1.351)†	(2.046)†
Men	11.762	11.078	17.978	23.711
	(2.159)	(1.831)	(1.479)†	(2.827)†
Women	23.421	23.749	31.427	37.318
	(1.84)	(2.017)	(1.914)†	(1.489)†
Women Characteristics (18-45 years old)				
Age	35.152	35.164	34.788	34.710
	(6.403)	(6.235)	(6.562)†	(6.519)†
Number of kids	1.892	1.934	2.222	2.26
	(1.156)	(1.199)†	(1.428)†	(1.294)†
Immigrant	0.005	0.007	0.012	0.003
	(0.070)	(0.083)†	(0.110)†	(0.058)†
Cohabiting	0.939	0.928	0.921	0.954
	(0.239)	(0.259)†	(0.270)†	(0.210)†
HS dropout	0.590	0.544	0.592	0.684
	(0.492)	(0.498)†	(0.492)	(0.465)†
HS graduated	0.319	0.344	0.310	0.250
	(0.466)	(0.475)†	(0.463)†	(0.433)†
College	0.091	0.111	0.098	0.066
	(0.288)	(0.314)†	(0.297)†	(0.248)†
Active	0.479	0.407	0.455	0.337
	(0.500)	(0.491)†	(0.498)†	(0.473)†
Employed	0.377	0.332	0.333	0.233
	(0.485)	(0.471)†	(0.471)†	(0.423)†
Part-time (in % of employed)	0.051	0.035	0.051	0.029
	(0.220)	(0.184)	(0.220)†	(0.168)†
Fixed-term contracts (in % of employed)	0.068	0.043	0.089	0.050
	(0.252)	(0.202)†	(0.285)†	(0.219)†
Average weekly hours worked	14.299	12.593	11.921	8.838
	(19.467)	(18.595)†	(17.814)†	(16.804)†

Notes: Mean and (Standard Deviation). † indicates a mean significantly different between the first implementer states and the later groups at 95 percent of confidence level. Average by group is weighted by population.

Table A.3. Fertility Effects

	Pre-average	Births	% effect
Linear trend	0.068	0.0015	2.2%
		[0.0013]	
Linear and squared trend	0.068	0.0015	2.2%
		[0.0013]	
Linear trend*dummy by region	0.068	-0.0003	-0.4%
		[0.0017]	
Linear and squared trend* dummy by region	0.068	0.0012	1.8%
		[0.0020]	
N		*773,985*	

Notes: Results of estimating equation (2) using different specifications for trends. Dependent variable: proportion of married women aged from 18 to 45 who gave birth during the past 12 months. Robust standard errors in brackets. The Pre-average level is calculated as a weighted average of pre-LOGSE birth rates in each state. For instance, if implementation in Catalunya is the academic year 1991-92, the pre-LOGSE period for Catalunya is from 1987 up the third quarter of 1991.

Table A.4. Persistence Effects on Employment. DDD Estimator

Overall and Heterogeneous Effects by Subgroups.

	Current effect	**1 year later**	**2 years later**	**3 years later**	**4 years later**
Panel A. Overall effect (Preferred specification)					
DDD	0.023	0.024	0.022	0.003	0.007
	[0.008]***	[0.009]***	[0.010]**	[0.011]	[0.013]
Pre-average	0.293	0.310	0.309	0.322	0.332
% effect	7.9%	7.6%	7.1%	0.8%	2.1%
Observations	105,748	105,036	102,404	100,340	98,109
Panel B. By mothers' education level					
HS dropout	0.027	0.029	0.012	-0.031	-0.026
	[0.013]**	[0.014]**	[0.016]	[0.018]*	[0.022]
N	44,076	45,642	46,101	45,965	45,874
HS graduate	0.021	0.0292	0.0247	0.0203	0.042
	[0.0122]*	[0.0132]**	[0.0144]*	[0.0161]	[0.0183]**
N	49,772	48,005	45,327	43,638	41,784
College	0.022	-0.019	0.028	0.016	-0.020
	[0.025]	[0.027]	[0.028]	[0.030]	[0.034]
N	11,900	11,389	10,976	10,737	10,451
Panel C. By mothers' age					
Younger than 30	-0.003	0.016	-0.021	-0.007	0.013
	[0.015]	[0.017]	[0.022]	[0.029]	[0.039]
N	34,679	31,369	27,685	24,974	23,129
Older than 29	0.035	0.028	0.031	0.006	0.018
	[0.010]***	[0.011]***	[0.011]***	[0.012]	[0.014]
N	71,069	73,667	74,719	75,366	74,980
Panel D. By number of kids					
One child	0.006	-0.017	-0.013	0.006	0.005
	[0.014]	[0.016]	[0.018]	[0.021]	[0.024]
N	38,125	35,295	32,357	30,289	29,018
Two or more	0.035	0.046	0.045	0.002	0.009
children	[0.010]***	[0.011]***	[0.012]***	[0.013]	[0.015]
N	67,623	69,741	70,047	70,051	69,091

Notes: Robust standard errors in brackets; ***, **, * denote statistical significance at 0.01, 0.05 and 0.10 levels, respectively. DDD model includes year and states fixed-effects, and a linear trend that differs for the treatment and control group, and individual and regional controls (see details in Table 2).

3. Essay 2: Childcare and Children's Cognitive Development[1]

3.1. Introduction

What are the consequences of childcare provision for children's long-run educational achievements? When trying to answer this question, it is crucial to consider the counterfactual mode of care. Blau and Currie, (2006) distinguish between three different care modes: maternal care, informal care (subsuming relatives, unlicensed care givers, or other irregular care givers) and formal care, being either private or public.

The evidence on the effects of public childcare is scarce and focuses mostly on countries with high female labor force participation rates (such as the US and Canada), and with many family-friendly policies (such as Scandinavian countries). Specifically, recent quasi-experimental evidence on universal childcare and child development includes Cascio (2009), Fitzpatrick (2008), Gormley and Gayer (2005) for the US; Baker *et al.* (2008), for Canada, and Datta Gupta and Simonsen (2010), Drange *et al.* (2012) and Havnes and Mogstad (2011) for Scandinavian countries.[2,3] In these countries, the introduction of universal childcare mainly led to a crowding out of private or informal (non-maternal) care. The effects of such expansions on children's cognitive development are found to be mostly positive either in the short or in the long-run[4], particularly among disadvantaged children. In contrast, a few studies find rather negative effects on children's non-cognitive skills in the short-run (Baker *et al.* 2008; Loeb *et al.* 2007; Magnuson *et al.* 2007).

---

[1]A version of this essay was published as a discussion paper. Reference: Felfe, Ch., N. Nollenberger and N. Rodríguez-Planas (2012) "Can't Buy Mommy's Love? Universal Childcare and Children's Long-Term Cognitive Development" IZA DP 7053.

[2]To the best of our knowledge, Berlinski *et al.* (2008), Berlinski *et al.* (2009), and Dustmann *et al.* (2012) are the exception as they investigate such questions for Uruguay, Argentina and Germany, respectively.

[3]This literature complements substantial experimental or quasi-experimental research on the effects of childhood educational programs targeted explicitly at disadvantaged children (for an overview please refer to Blau and Currie 2006).

[4]See Berlinski *et al.* (2009), Gormley and Gayer (2005), and Fitzpatrick (2008) for effects measured during preschool or primary school, Berlinski *et al.* (2008) and Drange *et al.* (2012) for effects measured at the end of mandatory schooling, and Cascio (2009), and Havnes and Mogstad (2011) for effects measured during early adulthood.

However not much is known when the expansion of high-quality public childcare crowds out maternal care. In fact, despite several recent studies evaluate the impact of maternal care on children's development exploiting parental leave expansions[5], they focus on substituting alternative modes of care by maternal care at a much earlier age (usually within the first 15 months of the child) and the care mode crowded out can be manifold. Given that the direction and magnitude of the effects depend crucially on the relative quality of the counterfactual care, the effects of introducing public childcare when it crowds out private or informal care might not necessarily coincide with the effects of introducing public childcare when it crowds out maternal care. Understanding the effects of introducing universal childcare in such a context is therefore the main objective of this essay.

We exploits a natural experiment to analyze whether universal childcare for 3 year olds has an effect on children's educational achievement at the end of compulsory education (age 15). We focus on an early 1990s reform in Spain, which led to an expansion of publicly subsidized childcare for *all* 3-year olds. Prior to this reform, universal preschool had only been offered to children age 4 and 5 years, and the available childcare for 3 year old children was either informal or provided by the private sector. Due to the reform, public childcare for 3 year olds increased from 8.5 percent in 1990 to 42.9 percent in 1997. As showed in the first essay, this reform had a modest effect on maternal employment, did not affect fertility, and did not lead to a crowding out of private childcare. Moreover, a crowding out of informal care was unlikely as most 3-year olds whose mothers worked prior to the reform were already enrolled in either public or private childcare. As a result, our effects have to be mainly interpreted as the effects of substituting maternal care by public high-quality care. The income effect – either due to a reduction in childcare costs (by crowding out private childcare arrangements) or due to an increase in maternal earnings (by increasing maternal employment) are rather negligible.

The Spanish reform also included a federal provision regarding several quality aspects (such as curriculum, group size, and staff skill composition). While the quality improvements were not exclusive to the children who were directly affected

[5] See for instance, Carneiro *et al.* 2010, Liu and Skans 2010, Rasmussen 2010, Baker and Milligan 2012, Dustmann and Schönberg 2012. They mostly do not find any significant effect on children's long-run development, with the exception of Carneiro *et al.* (2010), who detect some positive effects on educational and labor market outcomes at age 25 in Norway.

by the expansion of public childcare, it is important to keep in mind that our findings have to be interpreted as the consequences of introducing regulated high-quality care, which can also be compared to preschool targeted at 3-year olds. Thus, the reform under study stands in stark contrast to other reforms such as the reforms in Canada, which implied moving middle class children from home care to relatively poor quality care (Baker *et al.* 2008), or in Norway, which did not occur parallel to an overall improvement in childcare quality (Havnes and Mogstad 2011).

In contrast to reforms taking place in other countries, for instance in Canada Baker *et al.* (2008) or in Norway Havnes and Mogstad (2011), the Spanish reform did not lead to a decrease in private childcare slots, but rather to a crowding out of maternal care or informal care.

Although the reform was national, the responsibility of implementing its preschool component was transferred to the states. The timing of such implementation expanded over ten years and varied considerably across states. We exploit this variation to isolate the impact of public childcare on children's long-run educational achievements. We apply a difference-in-difference (henceforth, DiD) approach. Our main empirical strategy is thus the following: we compare educational outcomes of children (at age 15) who were 3 years old before and after the reform from states where public childcare expanded substantially (treated states) and states with little or no increase in public childcare coverage (control states).

Our analysis uses data from the 2003, 2006, and 2009 *Programme for International Student Assessment* (PISA). Children from PISA 2003 were born in 1987 and hence, they were 4 years old when the childcare reform was first implemented in 1991. As a consequence, they were unaffected by the expansion of publicly subsidized childcare for 3 year old children. Children from PISA 2006 and 2009 were born in 1990 and 1993, respectively, and thus they were fully affected by the expansion of childcare at age 3 if they lived in a treated state.

We find that universal child-care for 3 year olds leads to a sizable increase in reading and math test scores, and sizable decreases in the likelihood of falling behind a grade during primary school and high-school. More specifically, we find that the reform improved reading test scores at age 15 by 0.10-0.13 standard deviations and math test scores by 0.08 standard deviations. In addition, the reform reduced the incidence of falling behind a grade by 2.5 percentage points (or 50 percent) in primary-school and by 3.1 percentage points (or 13 percent) in high-school.

Stratification with respect to gender reveals furthermore that the effects are mainly driven by girls. The observed gender asymmetry is in line with existing research reporting larger benefits of public childcare for girls (Havnes and Mogstad 2011, Gathmann and Sass 2012). Moreover, stratification with respect to parental education supports the findings of previous studies that public childcare is particularly beneficial for children from disadvantaged backgrounds (Currie and Thomas 1995 and 1999, Datta-Gupta and Simonsen 2010).

Our results are robust to the use of: (i) alternative specifications, (ii) alternative definitions of the treatment and control group, and (iii) alternative identification strategies. Moreover, placebo estimates using month of birth as the dependent variable support the hypothesis that our findings are *not* spurious.

Our contribution to the existing literature is threefold. First, we provide evidence for the impact of public childcare on children's long-run educational achievements, both in terms of cognitive skills and grade retention. Our results thus add to the existing literature on the effects of large-scale, publicly subsidized pre-school programs. Several studies focus on the short-term effects of large expansions in universal pre-primary education and conclude that increasing the supply of public childcare leads to improvements in children's educational achievements (Fitzpatrick 2008; Berlinski *et al.* 2009). To the best of our knowledge, there are only two studies which focus on long term effects of expansions of public childcare. Cascio (2009), studying the introduction of public pre-schools in the US, finds evidence for reduced high-school dropouts and decreased institutionalization; Havnes and Mogstad (2011) analyze a major expansion in public childcare in Norway and reveal positive effects on children's educational attainment and labor force participation during adulthood.

Second, we study a framework where public childcare leads to mainly a crowding out of maternal care. Thus, our findings are relevant for understanding the role of the mother versus alternative care arrangements for children's cognitive development.

Finally, we argue that understanding the effects of universal day care is particularly relevant in countries with meager labor market prospects of young people. With widespread job precariousness and high youth unemployment rate, Spain is a prime example for such a country. Thus analyzing and understanding the consequences of universal care provision on children's educational achievements and thus prospects for their labor market career under such circumstances is of highest policy relevance.

The essay is organized as follows. The next section provides an overview of the Spanish institutional background, including a description of the public childcare system before and after the reform. Sections 3.3 and 3.4 present the empirical strategy and the data, respectively. Section 3.5 presents the main results and the heterogeneity analysis. Sensitivity analysis is discussed in Section 3.6. Section 3.7 finally concludes.

3.2. Background Information[6]

Institutional Background

In Spain, female labor force participation rates are among the lowest in the OECD. In 1990, before the reform under analysis came into force, the Spanish female labor force participation rate was 43 percent, far behind the 70 percent of the US, 69 percent of Canada, 73 percent in Norway and 78 percent of Denmark (the countries on which other studies analyzing childcare expansions have focused). In addition, the employment gap due to motherhood amounted to 10 percentage points (Gutierrez-Domenech, 2005). Reasons for such a scenario are manifold. In the late 1980s and early 1990s, Spain was *not* a family-friendly country for working parents (and especially working mothers) as reflected by its low levels of social assistance to families (Adserà, 2004), one of the shortest maternity leaves in Europe (Ruhm, 1998), an extremely low incidence of part-time work (only 8 percent of all jobs in 1990), as well as a rigid labor market with many jobs in the service sectors having a split shift from 9 am to 2 pm and from 5 to 8 pm (Amuedo-Dorantes and de la Rica, 2009). Moreover, Spain was a traditional country with low participation of men in household production (Bettio and Villa, 1998; de Laat and Sevilla-Sanz, 2011). Consistent with this, only half of all married women aged 18 to 45 were working or looking for a job in 1992, and among those not participating in the labor market, 45 percent reported family responsibilities as their main reason.[7]

---

[6] This essay was written to be read independently from the first one. Therefore, those who read the first essay can skip the explanation of the educational reform in this essay.

[7] Estimated by the authors based on micro data from the 1992 Spanish Labor Force Survey.

School and Preschool Prior to the Reform

Mandatory schooling in Spain begins at age 6. However, preschool for 4- and 5-year olds is also offered at the premises of primary schools from 9 am to 5 pm (regardless of school ownership status). Once a primary school offers places for 4-year olds, parents who wish to enroll their children in that particular school will do so when the child turns 4 years old as the chance of being accepted by the school may decrease considerably a year later (as priority for enrollment of 5- and 6-year olds is given to those children already enrolled in that particular school when they were 4 years old). As a consequence, enrollment rates for 4- and 5-year olds in the late 1980s were 94 and 100 percent, respectively.

Primary and secondary schooling is either public or private.[8] Public schools are free of charge, except for school lunch, which costs about € 100 per child per month. Private school costs are higher - between 250 and 350 euros per child per month (including lunch).

At the beginning of the 1990s, childcare for children 0- to 3-years old was rather scarce, predominantly private, and also quite expensive (on average it cost between 300 and 400 euros per child per month - including lunch). In contrast with some Scandinavian countries and the US, family day care, where a reduced number of children are under the care of a certified carer in her house, is practically non-existent. In Spain, children under 4 are either in formal (public or private) childcare or with their mother (or grandmother). Unfortunately, information on grandmother care is unavailable. As a consequence, this paper considers motherly care as equivalent to care provided by the nuclear family.

The Reform

In 1990, Spain underwent a major national education reform (named LOGSE) that affected preschool, primary and high school. The focus of our study is on the preschool component of this reform, which consisted of regulating the supply and the quality of preschool, and its implementation began in the school year 1991/92. The

---

[8]About one third of children in primary school in Spain are enrolled in private schools. In this paper, private schools refer to "escuelas concertadas" for which the government subsidizes the staff costs (including teachers). There are a very small number of private schools, which tend to be foreign schools (such as the French, Swiss or American schools), and cost two to three times more than the average "escuela concertada".

primary and secondary school component of the reform increased mandatory schooling by two years (from age 14 to age 16) starting school year 1996/97. In addition, it established that primary school would end at age 12 (instead of age 14). Our analysis isolates the effect of the preschool component by focusing on children born between 1987 and 1993 who were all affected by the primary and secondary school component but not necessarily by the preschool component.

The LOGSE divided preschool into two levels: the first level included children up to 3 years old, and the second level included children 3 to 5 years old. While not introducing mandatory attendance, the government began regulating the supply of places for 3 year olds. Prior to the LOGSE, free universal preschool education had only been offered to children 4 to 5 years old in Spain. After the LOGSE, preschool places for 3 year olds were offered within the premises of primary schools and were run by the same team of teachers. This implied that childcare for 3 year olds operated full-day (9 am to 5 pm) during the five working days and followed a homogeneous and well-designed program. With the introduction of the LOGSE, schools had to accept children in September of the year the child turned 3 whenever parents asked for admission if places were available. Available preschool places were allocated to those who had requested admission by lottery (regardless of parents' employment, marital status, or income). As explained earlier, although enrollment was not mandatory, it was necessary to ensure a place in the parents' preferred school choice. As a consequence, childcare enrollment among 3-years-old children went from meager to universal in a matter of a decade.[9] Between the academic years 1990/91 and 2002/03 the number of 3-years-old children enrolled in *public* preschool centers increased extraordinarily from 33,128 to 238,709. This represented an increase in the public enrollment rate of 3-year olds by more than 58.6 percentage points, from 8.5 percent to 67.1 percent.[10] Federal funding for preschool and primary education increased from an average expenditure of €1,769 per child in 1990/91 to €2,405 in 1996/97 (both measured in 1997 constant euros), implying a 35.9 percent increase in education expenditure per child.[11]

---

[9]Unfortunately we only have information on enrollment rates and not on actual supply rates for 3-year olds. In the context of rationed supply, enrollment rates should, however, resemble coverage rates quite closely.

[10]These figures exclude the Basque Country, Navarra and Ceuta and Melilla as they are not included in our analysis.

[11]Unfortunately data disaggregated at the preschool level is not available.

Apart from regulating the supply of public childcare, the LOGSE also provided federal provisions for the first time in Spain regarding educational content, group size, and staff skill composition regardless of ownership status for children 3 to 5 years old. Psycho-educational theories such as constructivism (put forward by Jean Piaget, and Lev Vygotski) and progressive education (based on Célestin Freinet and Ovide Decroly) served as a guideline for the design of the curriculum. There was a strong emphasis on play-based education, group play, through experiencing the environment, problem solving and critical thinking. The particular objectives of preschool education focused on: (1) personal development where the child masters its own body and understands its own movement possibilities; (2) emotional development where the child interacts with others in a variety of context and communication modes; (3) social development where the child forms good relationships with adults and peers and understands that people have different needs, views, cultures and beliefs; and (4) personal autonomy in the child's usual activities (LOGSE; 3 October 1990). While the pedagogical movements behind the LOGSE are close to those in Scandinavian countries, they have been viewed as an alternative to the test-oriented instruction legislated by the *No Child Left Behind* educational funding act in the US or the reception class in the UK.

In addition, the LOGSE established the maximum number of students per class to be 20 for 3-year olds and 25 for 4- and 5-years old. It is important to point out that classes are grouped based on the year children were born and thus, are not mixed in ages.[12] Finally, the LOGSE required preschool teachers to have a college degree in pedagogy – a requirement previously only enforced for teachers of 4- and 5-year olds. The quality improvements affected all children enrolled in preschool (that is, 3-, 4- and 5- year olds). As a consequence, our analysis should be able to isolate the effect of the expansion of high-quality public childcare from overall quality improvements of preschool education of 4- and 5-year olds (all cohorts under study were affected by the improved quality of preschool for 4- and 5-year olds).

Despite being a national law and being financed nationwide, the responsibility of implementing the expansion of public preschool slots was transferred to the states. The timing of such implementation expanded over ten years and varied considerably

[12]As a consequence, we ought not to worry about potential spillover effects from incoming 3-year old children on 4-year old children. This point is important as age-mixed groups would represent a threat to our estimation methodology.

across states frequently for arbitrary reasons. Implementation lags arose largely due to a scarcity of qualified teachers and constraints on classroom space in existing primary schools (El País, October 3rd 2005). In fact, the ratio children per classroom in childcare centers (public and private for children age 0 to 5 years old) was 24.1 among treatment states in 1990 versus 27.2 among control states.[13] Moreover, an initially higher level of private childcare facilities in treatment states in comparison to control states might have provided the necessary know-how to implement childcare facilities at a faster speed.

Our empirical strategy (described in the next section) exploits the differences in the timing of implementation across states.

3.3. Empirical Specification

The aim of this paper is to explore whether the preschool component of the early 1990s educational reform in Spain had an impact on children's educational achievement 12 years later when they were 15 years old, and just about to finish mandatory education in Spain. To estimate the effect of the large-scale expansion in publicly available childcare slots for 3 year olds we follow a DiD strategy. Thus, our identification strategy is based on the fact that supply shocks induced by the childhood component of the reform were larger in some states than in others.

The DiD framework basically compares the cognitive development at age 15 of children who were 3 years old before and after the reform in states where childcare expanded a lot (the treatment group) and of those who were in states where the increment in childcare coverage was less pronounced immediately after the reform (the control group). To determine the cut-off coverage-rate needed to define which states belong to the treatment and the control group, we follow Havnes and Mogstad (2011), and order states according to their percentage point increase in *public* childcare coverage of 3 years old from 1990 to 1993. We then separate the sample at the median. Those states that experienced an increase in public childcare coverage above the median belong to our treatment group whereas those with an increase under the median belong to our control group.

---

[13] Calculated by the authors based on statistics from the Spanish Ministry of Education.

Figure 1 displays the average increase in public childcare coverage for 3-year olds for the treatment and the control groups from 1987/88 to 2002/03. Prior to the reform, there are few differences, on average, between treatment and control groups: in 1990/91 the enrollment rate of publicly available preschool places for 3-year olds amounted to 9.9 percent in the treatment group and to 7.4 percent in the control group. Yet, families living in treatment states experienced a much stronger initial increase in preschool places than families living in control states. For instance, among states in the treatment group, the public enrollment rate for 3-year olds rose from 9.9 percent in school year 1990/91 to 44.0 percent in the school year 1993/94 and 57.1 percent in school year 1996/97. In comparison, the public enrollment rate for 3-year olds in the control group increased from 7.4 percent to 15.3 percent in 1993/94 to 29.4 percent in school year 1996/97. Figure 1 also shows that while there are dramatic differences in the initial expansion of public childcare, the control states fully catch up within a decade. Figure 2 provides evidence that, in contrast with the observed differences in public childcare, trends in private childcare are remarkably similar across the treatment and control group. As a result, our study compares states that differ distinctly in terms of initial changes in public childcare coverage, not, however, in terms of long-run trends or potential demand for childcare.

Our basic DiD model, estimated by OLS over the sample of children from PISA 2003, 2006 and 2009, can be expressed as follows:[14]

$$\begin{aligned} Y_{ijt} &= \gamma_1 Treat_i + \gamma_2 Cohort90 + \gamma_3 Cohort93 + \theta_1 (Treat_i * Cohort90) + \\ &+ \theta_2 (Treat_i * Cohort93) + X_i^{'} \beta + \delta_j + \varepsilon_{ijt} \end{aligned} \tag{1}$$

where Y_{ijt} measures the educational outcome a child i achieves in year t living in state j, $Treat_i$ is a binary variable indicating whether or not child i lives in one of the fast-implementing states. *Cohort90* and *Cohort93* are cohort-specific dummies equal to 1 if the child is tested in PISA 2006 or in PISA 2009, respectively. Children from PISA 2006 and 2009 were born in 1990 and 1993, respectively, and thus they were fully affected by the early childhood component of the LOGSE at age 3 *if they lived in a state that rapidly expanded the supply of public childcare slots* (that is, in a treatment

[14]We use OLS for all of our estimations. For our limited-dependent-variable outcomes we replicate our analysis using logit models, which yield similar results.

state). Children from PISA 2003 were born in 1987. They were 4 years old when the LOGSE was first implemented during the school year 1991/92, and thus, they were unaffected by the expansion of publicly subsidized childcare.

The coefficients θ_1 and θ_2 belonging to the interaction terms between treated states and the cohort dummies for 1990 and 1993, respectively, measure the average causal effect of the increase in childcare places for 3-year olds in the treatment states relative to the control states between 1990/91 and 1993/94 as well as 1990/91 and 1996/97, respectively, on different outcomes measuring children's cognitive development at age 15.

The vector X_i includes *only* time-invariant individual features that are expected to be correlated with educational achievement: gender and immigrant status. Since all additional socio-demographic characteristics that we observe at age 15 are time variant and thus potentially endogenous to our treatment, we decided not to include them in our main specification. However, our results are robust to sensitivity analysis where we sequentially add these additional variables to equation (1).[15] In addition, we allow for pre-reform heterogeneity within the group of treatment states by estimating a specification where we include state-specific fixed effects, denoted by δ_j.

The DiD strategy controls implicitly for unobservable differences between children born in different locations (by including a dummy for the treatment areas) and in different years (by including a dummy for the different cohorts). Yet, it assumes that in the absence of the reform children residing in the treatment states would have experienced the same change in outcomes than children residing in the controls states. Thus, we need to assume that there are no pre-existing differences in the features of the area (neither in the institutions nor in the population) which may systematically relate to the determinants of the local growth in childcare slots and that may explain differential development in children's cognitive development, biasing our results. The next section explores whether such systematic pre-reform differences between treated and control states exist. In addition, in the Robustness Section we present a battery of alternative specifications to test the sensitivity of our results. As

[15] Results of the specification including these control variables are shown in Appendix Table A.1.

this assumption is not testable, we also conducted a placebo test where we exploit the child's month of birth, which is directly related to cognitive development.[16]

The errors in equation (1) vary at the children, state, and year level, and therefore they might be correlated across time and state. (Bertrand *et al.* 2004) point out that the standard errors in DiD regressions may be underestimated if there is serial correlation in the state-time shocks. Although we reduce this problem substantially by using only three years of data (one before and two years after policy), in the robustness section we have estimated the models with standard errors clustered at the state-period level.

3.4. Data

The data used for this study stem from the *Programme for International Student Assessment* (PISA), an internationally standardized assessment that was jointly developed by participating economies and administered to 15 year olds in schools. The purpose of PISA is to test whether students, near the end of compulsory education have acquired the knowledge and skills essential for full participation in society. In particular, it administers specific tests to assess whether students can analyze, reason, and communicate effectively.

For our purpose, we rely on the 2003, 2006 and 2009 PISA datasets for Spain. Thus, our sample consists of children belonging to the birth cohorts 1987, 1990, and 1993. We exclude immigrant children who arrived to Spain *after* their 3rd birthday as well as children residing in the Basque Country, Navarra and Ceuta and Melilla. The reason for excluding children from the two first states is that both the Basque Country and Navarra have had greater fiscal and political autonomy since the mid-1970s and, as a consequence, their educational policy has differed from that of Spain as a whole. Ceuta and Melilla are excluded as they are only available in the PISA datasets from 2006 onwards.

Our analysis focuses on reading literacy and mathematics as performance in these domains are fully comparable across PISA cycles from 2003 forward. In contrast, we are unable to use the scientific scores for our analysis as the

[16] We were unable to pretend that the child-care expansion took place before the LOGSE because of data restrictions: state of residence is not available in the 2000 PISA.

questionnaires are not comparable before and after 2006 (OECD, 2006). Tests scores are standardized, implying that the estimated coefficient represents the percentage increase (or decrease) in standard deviations (henceforth sd).[17] We also estimate the effect of the reform on two additional variables, available only in the 2003 and 2009 PISA waves. One of these variables measures whether the child fell behind one grade or more during primary school, the other variable measures whether the child fell behind one grade or more during high-school.

PISA gathers information on the students' demographic characteristics (such as gender, age, immigrant status and age of arrival to Spain for immigrants), students' socio-economic background (including mother and father education level, home educational resources, other home possessions and incomes), and school characteristics (such as ownership, and whether it is rural or urban). Unfortunately, we do not have information on whether the child was enrolled in school when they were 3 years old implying that—as in Baker *et al.* (2008); Fitzpatrick (2008), or Havnes and Mogstad (2011) - our estimates are intention-to-treat (ITT) estimates.

Another limitation is that PISA only provides information about the socio-economic background at the time of the test, that is, when children are 15 years old. As discussed above, these variables are potentially endogenous to treatment. As a consequence, we decided to include in our main specification *only* controls that are time invariant (such as gender and immigrant status). In the Results Section, we explore the sensitivity of our results to sequentially adding other individual characteristics (potentially endogenous), such as parents' level of education or home possessions (including a desk for study, a room of your own, a computer, internet, classic literature, books, works of art, dishwasher, among others), type of school, and population density of the area of residence.

A related concern is that we do not observe the state the child lived when she was 3 years old, but instead the state he lives when he is 15 years old. Because migration across states in Spain is surprisingly low (Jimeno and Bentolilla, 1998, Bentolilla, 2001), there is little concern that the policy may have induced families to

[17] Standardization is done by subtracting the mean to each individual test score and dividing by the standard deviation for the whole sample. We have conducted sensitivity analysis where the test scores have been standardized at the year level. In this case θ_1 and θ_2 are estimating the causal effects of the policy on the relative position of treatment states versus control states within a year, eliminating any potential problems with testing differences across years. Results are very similar to those from our preferred specification and available from authors upon request.

move from slow implementing states to fast implementing states. In the Results Section we present further evidence that migration is unlikely to be a concern.

Descriptive Statistics

Panel A in Table 1 presents outcome means before and after the reform. We observe that children in the treated states outperform those in the control states prior to the reform. After the reform, the performance gap across treatment and control groups widens further. This improvement is suggestive that the childcare reform in Spain may have increased children's cognitive development. The last row of Panel A of Table 1 shows the means for relative age, our placebo outcome.[18] In this case, we observe no statistically significant differences pre- or post-reform.

Panel B.1 of Table 1 presents baseline summary statistics for time-invariant children's socio-demographic characteristics, which may be correlated with children's cognitive development. If there are changes in the composition of pre-reform socio-demographic characteristics over time, the DD estimates may be biased. Nevertheless, we perform balancing test, and can reject any statistically significant difference across the three cohorts at the 95 percent level.[19]

Although not necessary for our estimation strategy because of the inclusion of state fixed effects, it would be useful if the intensity of implementation of the law across states were uncorrelated with any regional differences. In Panel B.2 of Table 1, we display regional characteristics of the treatment and control groups prior to the reform to better understand the determinants of the expansion across states. We observe that overall control states are poorer and have higher unemployment rate than treatment states. Notice, that these differences may in part be responsible for the pre-reform differences in children's cognitive performance. In the Robustness Section we estimate a DD using states that were very similar prior to the reform in terms of these observables.

[18] Following Bedard and Dhuey (2006), the relative age is defined as the difference between the month of birth and the cut-off date for children to begin school. As in Spain the cut-off date is January 1, the relative age is equal to 0 for students born in the December (the youngest) and equal to 11 for students born in January (the oldest).

[19] We do find some differences when using as the dependent variable maternal education, type of school and home possessions across the three cohorts, which is not that surprising considering that these variables are measured ex-post when the child is 15 years old. That is why we decide to not include them in the main specification.

Another concern is the potential endogeneity of our policy. For example, we may worry that the increase in public-preschool seats for 3 years old in a particular state was a response to the increasing incidence of working mothers, and that it is the increase in maternal employment (not the increase supply of childcare slots) that affected children's cognitive development. Underlying trends are rather difficult to capture. Yet, to the extent that they are driven by observable features we can control for them by including interactions between the cohort dummies and pre-reform states socio-economic characteristics (as in Duflo 2001). However, if treatment and control areas are characterized by unobservable diverging trends, our estimates are biased even when controlling for diverging trends in observables. Thus, to further increase confidence in our estimates we conduct a placebo test as described earlier.

3.5. Results

Effects on Private Childcare Enrollment and Maternal Employment

The impact of expanding public childcare on children's long-run cognitive development depends on whether the expansion in public childcare led to a crowding out of alternative care modes. The first essay finds a modest effect of the reform on maternal employment and no evidence of crowding out of private childcare. Given that in the first essay we use a different identification strategy, we first re-estimate the results of the first essay but adjust the identification strategy to be comparable to the baseline strategy of this one (for details please refer to the Appendix).

Results are shown in Table 2. It shows that children residing in treatment states were offered substantially more public childcare than children residing in control states: this differential increase amounted to 26.1 percentage points for the 1993/94 cohort and to 25.6 percentage points for the 1996/97 cohort. Yet, the reform did not lead to the crowding out of private childcare enrollment (Panel A). While this result may come as a surprise, it is important to highlight that preschool for 3-year olds was implemented within primary school regardless of school ownership. As a consequence parents who wished to enroll their children in private school would now enroll their 3-year old to the private school as soon as preschool for that age group was offered (to guarantee a space thereafter).

Table 2 also shows that the effect of universal childcare on maternal employment is much smaller than the increase in the enrollment in childcare (Panel

B). A 1 percentage point increase in enrollment among 3-year olds led to between 0.06 and 0.09 percentage points increase in maternal employment.[20,21]

Finally, in contrast to other studies, the expansion in public childcare did not lead to a crowding out of informal care arrangements. Most of the mothers of 3-year olds who worked prior to the reform had their child already enrolled in either public or private childcare. Prior to the reform, 35.7 percent of mothers of 3-year olds worked in treated states while 32.5 percent of 3-year olds were enrolled in formal care (9.9 percent in public childcare and 22.6 percent in private childcare).[22] Thus, the Spanish reform mainly implied that mothers took their children to full-time (9 am to 5 pm) childcare even though they continued not to work.

Therefore, our findings have to be interpreted as the effects of an expansion in public childcare that mainly led to a crowding out of maternal care, but not to a crowding out of private or informal care arrangements. Moreover, any potential income effect from an increase in maternal employment caused by the reform is modest at most.

Effects on Children Cognitive Development

Table 3 shows our main results, the impact of the expanding public childcare on all children living in the treated area – the so called intention to treat effect (ITT) - and on the average child placed in public childcare following the expansion of public childcare – denoted by the treatment effect (TT). Unfortunately, our dataset does not contain individual usage of public childcare. To obtain the TT estimates we therefore follow Baker *et al.* (2008), and divide the ITT estimates by the probability of treatment. Hence, in 2006, we adjust the ITT estimates by dividing them by the increase in childcare coverage between 1990/91 and 1993/94 in treatment states relative to the controls states (26.0 percentage points), and in 2009, we divide the ITT estimates by the increase in childcare coverage between 1990/91 and 1996/97 in treatment states relative to the controls states (25.2 percentage points).

---

[20]This estimate is the ratio between the percentage point increase in maternal employment rate (0.024 and 0.016) and the percentage point increase in 3-year olds' public childcare enrollment due to the reform (0.261 and 0.256).

[21]Due to a different identification strategy this estimate is slightly different to that of the first essay of this dissertation.

[22]This pre-reform situation contrasts with that of Havnes and Mogstad (2011) as in their study, childcare coverage (10 percent) was half the size of maternal employment (20 percent).

Table 3 displays the results for four alternative outcome variables: reading and math tests, and the likelihood of falling behind one grade in primary school or in high-school. All regressions are estimated first without any control variables and then including states fixed-effects and the set of *invariant* control variables (gender and immigration status). Comparing unconditional and conditional estimates in Table 3 reveals no significant differences. We therefore focus our discussion on this last specification.

Focusing first on the effects of the reform on children's standardized reading tests scores at age 15, the direction of the effect of the expansion in public childcare for 3 year olds is positive and statistically significant at any conventional significance level. The expansion of public childcare slots by 26.2 percentage points leads to an increase in reading test scores by 0.13 sd for the 1990 cohort living in one of the treated states and of 0.10 sd for the 1993 cohort living in one of the treated states. Considering children who actually attended public childcare following the reform, the effects are substantial: the TT estimates indicate that the reform implied an improvement in reading scores of 0.48 sd for the children born in 1990 who attended public childcare cohort and of 0.42 sd for children born in 1993 who attended public childcare.

The reform also improved children's math performance, yet to a slightly lower extent. Children who were born 1990 and lived in one of the treated states outperform children who lived in one of the control states in the math test by 0.08 sd. This translates into an improvement among children who actually attended childcare by 0.20 sd. Yet, among the 1993 cohort the estimate is considerably smaller and no longer statistically significant.

How do these effects compare to the established evidence? There is extensive evidence on the impact of early childhood programs targeted to disadvantaged children. For instance, evaluations of Head Start, a federal program for preschool children from low-income families in the US, reveal significant improvements of children's oral comprehension at 1st grade by up to 0.13 sd, not however, on children's math abilities (Almond and Currie 2011). Regarding universal childcare, the evidence regarding short run-effects is mixed: while Loeb *et al.* (2007) find that pre-primary education in the US is associated with improved reading and math skills, Magnuson *et al.* (2007) state that these effects dissipate at the end of first grade.

Berlinski *et al*. (2009), however, still find a substantial improvement of cognitive skills (by 0.23 sd) at the end of third grade. To the best of our knowledge, Fitzpatrick (2008) is the only paper analyzing medium-run effects of universal child-care provision on cognitive test scores. Nevertheless, she only reveals a significant impact on the population of disadvantaged children, defined as those living in rural areas: gains from the childcare reform range between 0.07 and 0.12 sd on reading scores, and between 0.06 and 0.09 of sd on math scores.

Moving to the effects of the reform on the likelihood of falling behind a grade, we also find beneficial effects of the reform. More specifically, we observe that the reform reduced the incidence of falling behind a grade by 2.5 percentage points in primary-school and 3.1 percentage points in high-school (this effect is significant at the 95 and 90 percent level, respectively). Given the initial likelihood of falling behind a grade among children in the treated states of 5% in primary school and 23 % in high-school, the effect of the reform represents a substantial decrease in the incidence of retention (50 % in primary school and 13 % in high-school).

Our results are comparable to the ones found for special intervention programs, such as the Carolina Abecedarian program, where grade retention is more than 30% lower among participating children than among non-participants. Moreover, analyzing universal childcare in Uruguay, Berlinski *et al.* (2008) find that children having attended preschool possess at age 15 of 0.8 years more education. Cascio (2009), however, evaluating the impact of universal public Kindergarten in the US, cannot detect any significant impact on grade retention. Nevertheless, as she points out the program is of low-intensity and it did not only lead to a significant crowding out of participation in federally-funded programs, such as Head Start, but also to cut-backs in state expenditure on schools to fund kindergartens.

Heterogeneity by gender and parental education

Table 4 presents subgroup analysis by gender or parents' education level. It is important to point out – as discussed already in Baker *et al.* (2008) and Havnes and Mogstad (2011) – that the lack of childcare usage for the different subgroups leads to important limitations for the interpretation of the results of the subgroup analysis. The reason is as follows. The underlying assumption when comparing the different subgroups is that the take-up rates are identical among them. Depending on the characteristic we stratify upon, this assumption is more or less credible. Parents´ education, for instance, might be likely to correlate positively with the take-up rate and thus we might underestimate the effect of the preschool component of LOGSE on children from the most disadvantaged families.

Estimates from Panel A reveal that universal preschool provision had large, positive and significant effects on girls' cognitive development. For boys, we can only observe an improvement in their reading skills. Yet, while the improvements in reading is comparable between boys and girls in the cohort 1990, the effect reduce by more than half and is no longer statistically significant for boys in the 1993 cohort. Regarding math skills, we can only detect a significant effect of preschool attendance among girls. The reform increased math test scores by 0.09 and 0.10 sd for the 1990 and 1993 female cohorts, respectively. Our results, thus speak to previous findings regarding the gender gaps in reading and math skills. Our estimates suggest that the early preschool exposure can help closing the gender gap in math scores – girls fare generally worse in math — but not in reading – boys are generally worse in reading. Finally, we also find positive and significant effects (at the 90 percent level) on grade retention among girls, not, however among boys. Children in the treated states are 2.5 percentage points less likely to repeat a grade during primary school and 4.5 percentage points during high-school.

The gender asymmetry in returns to public childcare has already been stated in previous studies. Gathmann and Sass (2012), for instance, find that attending public childcare is beneficial for girls´ early development of social, motoric and daily skills, not, however, for their language skills. In a similar vein, in the study by Havnes and Mogstad (2011), improved labor market outcomes due to an expansion of public childcare are also only present among girls. Yet, boys and girls benefit similarly in terms of educational outcomes, such as high school completion or college attendance. Our results may therefore indicate that public pre-school may help girls to improve upon their relative disadvantage in certain cognitive skills, such as math, that may help to close the gender earnings gap (gender differences in tests scores is discussed in the next essay).

Panel B in Table 4 presents results by parents' education level. Average gains in cognitive performance due to universal childcare seem to be completely driven by children of low skilled parents, defined as those for whom neither parent has a high-school degree.[23] The underlying explanation for this might be the counterfactual care mode for this subgroup. As we can see in Panel B, Table 5, low educated mothers – in contrast to high educated mothers - increase their employment following the LOGSE substantially (by 7.7 percentage points). Thus, in case the quality of public childcare might outperform the quality of care provided by a low educated mother, children might benefit from attending public care.

Previous studies have already pointed to the importance of the counterfactual care mode, when interpreting the effects of public childcare. Felfe and Lalive (2011), for instance, show that selection into childcare based on family background (mainly on parental education) can lead to a substantial underestimation of the effect of public childcare on children's cognitive and non-cognitive development. Datta-Gupta and Simonsen (2010) find that children attending family day care fare clearly worse in terms of non-cognitive skills than children attending high quality public care. Finally, Havnes and Mogstad (2011) have shown that high-quality childcare provision has positive long-run effects on income distribution and equity.

3.6. External Validity and Sensitivity Analysis

Attrition Bias

Although PISA interviews students when they are 15 years old, and thus at a time in which school is still mandatory, dropout rates are high in Spain. For example, the average dropout rate at age 16 was 18% in 2003, 20% in 2006, and 15% in 2009.[24]

Given our previous findings, it is possible that LOGSE also had an effect on the dropout rate. If this is true, our baseline data is plagued by attrition and our previous results are biased, in particular biased towards zero (a differential reduction in dropout rates in treated and control states would lead to a differential representation of children from the lower tail of the ability distribution in treated and control states). We explore the issue of attrition using an alternative data set, the Labor Force Survey (LFS), which is representative of the Spanish population and contains information on

[23] Because our measured is self-reported by the child (not the parent) and measured at aged 15, we felt that measuring parents' education skills in this way was the most appropriate to minimize endogeneity and measurement error problems. Moreover, this classification divides the sample by about half.

[24] Estimates are calculated by the authors using the Spanish LFS. Note that they are considerably lower to OECD estimates presented in the introduction. The difference is due to the fact that our estimates are measured at age 16, whereas those of the OECD include older individuals (up to age 24).

dropout.[25] Using the same birth cohorts as in our baseline data (1987, 1990, and 1993), we re-estimate equation (1) but using high-school dropout as the dependent variable (See Panel A, Table 5). Results from this estimation do not indicate any significant impact on high-school dropout up to the age of 16 and thus, let us conclude that attrition is not a major threat to the validity of our estimates

Selective migration

Another potential source of bias might be selective migration: families might have moved from slow implementing states to fast implementing states Since PISA only provides information on the state of residence at age 15 (but not at age 3), we again rely on the LFS (now on years 2003, 2006 and 2009) to assess the concern of selective migration. We first assess the likelihood of living at age 15 in a different state than the state of birth. This probability is, however, negligible (4.6 percent in 2003, 5.2 percent in 2006, and 4.9 percent in 2009). Second, we estimate the likelihood of having migrated from a control state to a treated state (and vice versa). The results do not indicate an increased migration into treated states, but if anything a small decrease (by 1 percentage points) among the 1990 cohort (see Panel B, Table 7) after the reform. Thus, the existing evidence does not point to a problem of selective migration.

Placebo Test

Unfortunately, we do not possess of sufficient cohorts unaffected by the reform to perform a placebo test where we pretend that the childcare expansion took place earlier. Moreover, pre-reform outcomes are rather rare in our dataset. Thus, we rely on one available measure, which is directly related to cognitive development,[26]

---

[25] To construct high school dropout we use information on whether the individual is attending high-school at age 16 in 2004, 2007 and 2010 LFS. Notice, that the LFS only collects information on completed education and employment status only for individuals 16 years old or older.

[26] The impact of date of birth on cognitive test scores is well documented across many countries, with the youngest children in each academic year performing more poorly, on average, than the older members of their cohort (see, for example, Bedard and Dhuey, 2006; or Puhani and Weber, 2005). In fact, our dataset shows a significant correlation between PISA scores and relative age of the child. We estimated test scores on a variable reflecting the relative age when classes began, based on school cut-off dates and assuming that the rule is strictly followed. We also controlled for other individual characteristics, such as gender, immigrant status, parents' level of education, place of residence and type of school (public or private). We find that to be one month older when school begins increases

but should not be correlated with the policy change, to perform a placebo test: child's month of birth. Hence, we re-estimate equation (1) but using as dependent variable the relative age of the child (defined as the difference between the month of birth and the cut-off date for children to begin school).

A significant effect of the reform on month of birth would cast some doubt on our results, as it would suggest that unobserved heterogeneity correlated with cognitive development might be driving our findings. Panel C in Table 5 displays the results from this placebo test. The estimates are not statistically significantly different from zero, providing further support that our estimates are true policy impacts.

Alternative Specifications

We have additionally tested the robustness of our results using different specifications of our baseline regressions (for an overview of the results please refer to table 6): i) we assume a more flexible specification in which all covariates are interacted with the treatment dummy, allowing for differential effects of individual characteristics and state of residence across treatment and comparison groups; ii) given economic differences between treatment and control states (see Panel B.1 of Table 1), we perform our analysis using *only* states that are very similar in terms of these observables; iii) we estimate a specification in which cohort FE are interacted with pre-reform states socio-economic characteristics to allow for different underlying trends in children's potential outcomes depending on the states' pre-reform characteristics; iv) we run our regression while clustering at the state-year level; v) we experiment with different definitions of treatment and the control groups, such as a definition based on the growth of enrollment rates from 1990/91 to 1996/97 (instead of growth from 1990/91 to 1993/94) to identify the treatment states as well as vi) a definition where treatment states are those above the 67th percentile in public enrollment growth, and control states are those below the 33th percentile.

Despite some partial loss of precision and some variation in the magnitude of our coefficients, our baseline finding that the reform improved educational outcomes for children growing up in the treated states relative to those growing up in the control states remains.

---

both the reading and math tests scores in 2 sd. Both coefficients are statically significant at 99 percent level

Alternative Identification Strategy

We have also tested the robustness of our results to assuming a different identification strategy. We follow Berlinski *et al.*, 2009, and estimate the effect of offering one additional child-care slot estimating the following equation and employing OLS (see Table 7):

$$Y_{ij(t+13)} = \theta Seats_{ijt} + \delta_j + \lambda_{(t+13)} + X'_{ij(t+13)}\beta + \varepsilon_{ij(t+13)} \quad (3)$$

where $Seats_{ijt}$ is the number of public preschools seats per 100 for children from 3 to 5 years old in the state j in year t where child i lives.[27] This specification has the advantage that it does not rely on the definition of treatment status. However, it assumes a constant effect of offering one further child-care slot across the whole offer distribution, thus, offering an 11th seat for every hundred children is assumed to have the same effect as every 91st seat per hundred children.

According to the results presented in Table 7, offering one more slot per hundred children leads on average to an improvement in children's reading test' scores of 0.008 sd. But again we find no statistically significant effects on children's math test scores.

3.7. Conclusion

Should governments provide childcare? Universal provision of childcare is one of the family policies present in the policy and academic debate. Supporters argue that this policy has the advantage to serve a dual purpose, providing children with valuable skills and parents with an implicit subsidy for their care. However, there is still limited consensus in the literature about the effect of childcare and maternal employment partly because the effects of universalizing childcare depends on the quality of both public childcare and the counterfactual care mode (Blau and Currie 2006). A major concern among deterrents of public childcare is its high costs for a

[27] Following Berlinski and Galiani (2007), we estimate the proportion of public preschool seats offered in each state as the number of public preschool units available for 3-5 year olds in each region times the average size of the classroom divided by the population of 3- to 5-years old in each state. Unfortunately, these data are not available by detailed age group. However, as enrollment rate of 4- and 5-years old was already close to 95 percent in the late 1980s, and as fertility remained stable over that period, basically all observed increase should be driven by 3-year-old children.

non-mandatory service for which the short- and long-term gains on the children's development relative to other forms of early childcare (such as parental, informal, or private care) remain uncertain. This essay contributes to this debate by providing quasi-experimental evidence for the impact of shifting hours of care provided by mother to hours of care provided by high-quality public preschools. Exploiting a natural experiment, we analyze whether universal childcare for 3 year olds has an effect on children's educational achievement at the end of compulsory education (age 15). We find sizable improvements in reading and math test scores, as well as sizable reductions in grade retentions in primary school and high-school. The results are, however, driven mostly by girls and children from disadvantaged backgrounds, pointing toward the relevance of pre-school for closing the gender and equity gap.

REFERENCES

Adserà, Alicia (2004). "Changing Fertility Rates in Developed Countries. The Impact of Labor Market Institutions." *Journal of Population Economics*, 17: 17-43.

Almond, Douglas, and Janet Currie (2011). "Killing Me Softly: The Fetal Origins Hypothesis." *Journal of Economic Perspectives*, 25(3): 153-172.

Amuedo-Dorantes, Catalina, and Sara de la Rica (2009). "The Timing of Work and Work-Family Conflicts in Spain: Who Has a Split Work Schedule and Why?" IZA Discussion Paper 4542.

Baker, Michael, Jonathan Gruber, and Kevin Milligan (2008). "Universal child care, maternal labor supply, and family well-being." *Journal of Political Economy*, 116(4): 709–745.

Baker, Michael, and Kevin Milligan (2012). "Maternity Leave and Children's Cognitive and Behavioral Development". NBER Working Paper 17105.

Bedard, Kelly, and Elizabeth Dhuey (2004). "The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects." *The Quarterly Journal of Economics*, 121(4): 1437-72.

Bentolilla, Samuel. "Las migraciones interiores en España." In *Mercado de Trabajo, Inmigración y Estado de Bienestar*, by J.A. Herce and J.F. Jimeno. Madrid: FEDEA and CEA, 2001.

Berlinski, Samuel, and Sebastián Galiani (2007). "The effect of a large expansion in pre-primary school facilities on preschool attendance and maternal employment." *Labour Economics*, 14: 665-680.

Berlinski, Samuel, Sebastian Galiani, und Marco Manarcorda (2008) "Giving children a better start: Preschool attendance and school-age profile." *Journal of Public Economics*: 92(5-6): 1416-1440.

Berlinski, Samuel, Sebastián Galiani, and Paul Gertler (2009). "The Effect of Pre-Primary Education on Primary School Performance." *Journal of Public Economics*, 93(1-2): 219-34.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics* 119, no. 1 (2004): 249-75.

Bettio, Francesca, and Paola Villa (1998). "A Mediterranean perspective on the breakdown of the relationship between participation and fertility." *Cambridge Journal of Economics*, 22(2): 137-171.

Blau, David, and Janet Currie (2006). *Preschool, Day Care, and After School Care: Who's Minding the Kids?* Edited by Eric Hanushek and Finis Welsh. Handbook of Economics of Education.

Carneiro, Pedro, Katrine Vellesen Løken, and Kjell G. Salvanes (2010). "A Flying Start? Long Term Consequences of Maternal Time Investments in Children During Their First Year of Life". IZA Discussion Paper 5362.

Cascio, Elizabeth (2009). "Do Investments in Universal Early Education Pay Off? Long-term Effects of Introducing into Public Schools." NBER Working Paper 14951.

Currie, Janet, and Duncan Thomas (1995). "Does Head Start Make A Difference?" *The American Economic Review*, 85(3): 341-364.

Datta Gupta, Nabanita, and Marianne Simonsen (2010). "Non-cognitive child outcomes and universal high quality child care." *Journal of Public Economics*, 94(1-2): 30-42.

de Laat, Joost, and Almudena Sevilla-Sanz (2011). "The Fertility and Women's Labor Force Participation puzzle in OECD Countries: The Role of Men's Home Production." *Feminist Economics*, 17(2): 87-119.

Drange, Nina, Tarjei Havnes, and Astrid M. J. Sandsør (2012). "Kindergarten for all: Long run effects of a universal intervention." Statistics Norway Discussion Paper 695.

Duflo, Esther (2001). "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *Amercian Economic Review*, 91(4): 795-813.

Dustmann, Christian, Anna Raute, and Uta Schönberg (2012). "Does Universal Child Care Matter? Evidence from a Large Expansion in Pre-School Education." Mimeo.

Dustmann, Christian and Uta Schöenberg (*forthcoming*). "The Effect of Expansions in Maternity Leave Coverage on Children Long-term Outcomes". *American Economic Journal: Applied Economics.*

El País (October 3rd 2005). "La LOGSE, 15 años después" by Elena Martín Ortega.

Felfe, Christina, and Rafael Lalive (2011). "Child Care and Child Development. What Works for Whom?. Mimeo.

Fitzpatrick, Maria D (2008). "Starting School at Four: The Effect of Universal Pre-Kindergarten on Children’s Academic Achievement." *The B.E. Journal of Economic Analysis & Policy*, 8(1): 1-38.

Gathmann, Christina, and Björn Sass (2012). “Taxing Childcare: Effects on Family Labor Supply and Children.” IZA Discussion Paper 6640.

Gormley Jr, William T., and Ted Gayer (2005). "Promoting School Readiness in Oklahoma. An Evaluation of Tulsa's Pre-K Program." *Journal of Human Resources*, 3: 533-558.

Gutierrez-Domenech, Maria (2005). "Employment Transitions after Motherhood in Spain." *Review of Labour Economics and Industrial Relations* 19: 123–148.

Havnes, Tarjei, and Magne Mogstad (2011). “No Child Left Behind: Subsidized Child Care and Children’s Long-Run Outcomes.” *American Economic Journal: Economic Policy*, 3: 97-129.

Jimeno, JF, and S Bentolilla (1998) "Regional Unemployment Persistence (Spain, 1976-1994)." *Labour Economics* 5, 1: 25-51.

Liu, Quian, and Oskar Nordstrom Skans (2010). "The Duration of Paid Parental Leave and Children's Scholastic Performance." *The B.E. Journal of Economic Analysis & Policy*, 10 (1): 1935-1682.

Loeb, Susanna, Margaret Bridges, Daphna Bassok, Bruce Fuller, and Russell Rumberger (2007). "How Much is Too Much? The Influence of Preschool Centers on Children's Social and Cognitive Development." *Economics of Education Review*, 26(19): 52-66.

Magnuson, Katherine, Christopher Ruhm, and Jane Waldfogel (2007). "Does prekindergarten improve school preparation and performance?" *Economics of Education Review*, 26(1): 33-51.

OECD (2006). "PISA 2006 Science Competencies for Tomorrow's World.".

OECD (2009). "PISA Data Analysis Manual: SPSS® Second Edition".

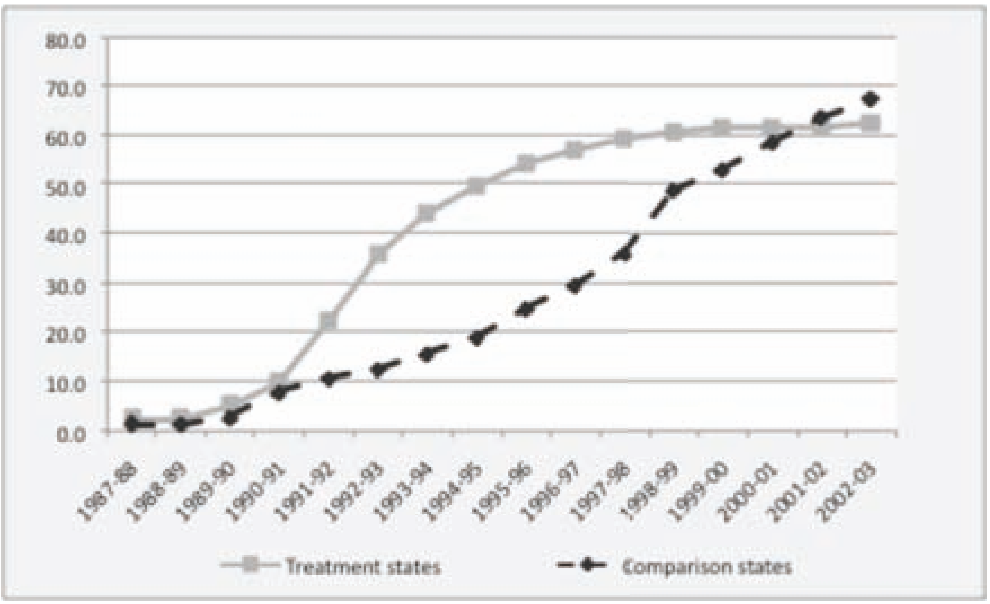
Puhani, Patrick A, and Andrea Weber (2008). "Does the early bird catch the worm?" *The Economics of Education and Training*, 105-132.

Rasmussen, Astrid Würtz (2010). "Increasing the Length of Parents' Birth-Related Leave: The Effect on Children's Long-Term Educational Outcomes." *Labour Economics,* 17(1): 91-100

Ruhm, Christopher J (1998)."The Economic Consequences of Parental Leave Mandates: Lessons from Europe." *The Quarterly Journal of Economics*, 113(1): 285-317.

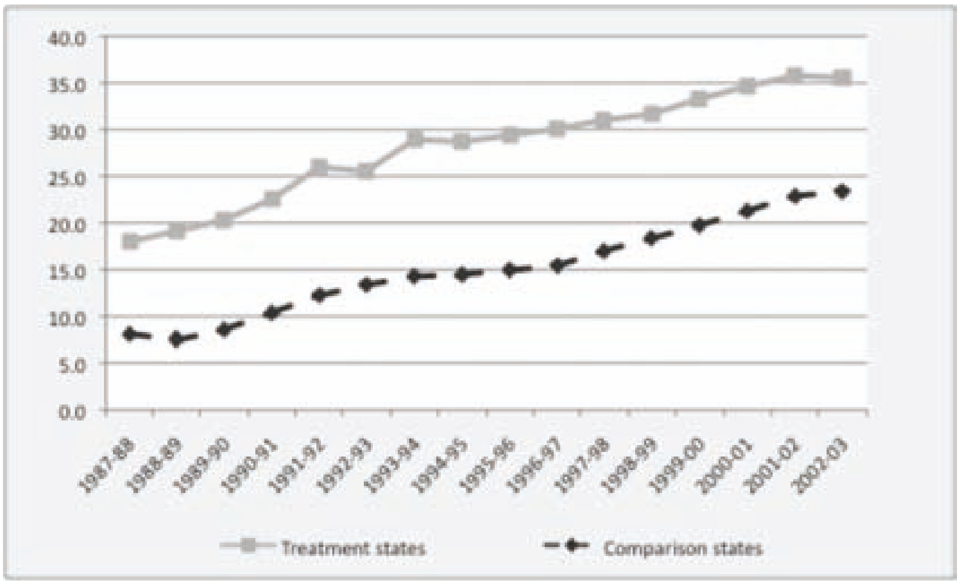
TABLES AND FIGURES

Figure 1: Enrollment rates in public childcare among 3 years old



Notes: Displayed numbers are weighted averages of *public* enrollment rates for the treatment (Galicia, Cataluña, Asturias, Rioja, Castilla y León, Cantabria, Madrid and Castilla-La Mancha) and control states (Extremadura, Aragón, Baleares, Valencia, Andalucía, Murcia and Canarias). Weights reflect the population of each state (CCAA).

Figure 2: Enrollment rates in private childcare among 3 years old



Notes: Displayed numbers are weighted averages of *private* enrollment rates for the treatment and control states. See figure 1 for list of states in treatment and control groups. Weights reflect the population of each state.

Table 1. Descriptive Statistics

	Treated States		***Differences between Treated and Control States***		
	Pre-Reform		Pre-Reform	Cohort90	Cohort93
A.1 Outcomes variables					
Standardized Reading Scores	0.071	[0.957]	0.269***	0.381***	0.363***
Standardized Math Scores	0.008	[0.938]	0.304***	0.381***	0.333***
Falling behind inprimary school	0.053	[0.224]	-0.010	n.a.	-0.036***
Falling behind in secondary school	0.230	[0.421]	-0.059***	n.a.	-0.089***
A.2 Placebo variable					
Relative age (placebo outcome)	5.418	[3.423]	-0.098	-0.093	-0.011
B.1 Individual Characteristics					
Gender (Male=1)	0.471	[0.499]	-0.029	0.002	-0.019
Born in Spain	0.991	[0.093]	-0.003	-0.009	-0.003
B.2 Regional Characteristics					
GDP (Euros/capita)	10,559	[1,935]	811	930	1,057
Unemployment rate –Males	0.095	[0.022]	-0.034	-0.060	-0.024
Unemployment rate- Females	0.209	[0.048]	-0.058	-0.063	-0.042
Employment rate- Females	0.261	[0.043]	0.018	0.010	0.001
Years of education – Males	8.620	[0.289]	0.302	0.137	0.082
Years of education – Females	8.234	[0.243]	0.201	0.179	0.074
Total population (in millions)	2.479	[2.080]	0.151	0.135	0.107
0-6 years old (percentage)	0.068	[0.007]	-0.015***	-0.015	-0.013
Population density (inhab. per km^2)	150.0	[195.8]	45.6	44.9	43.3
Sample sizes					
Treated States			4,116	7,456	7,276
Control States			2,040	3,196	5,404

Notes: The table displays mean and standard deviation in brackets. The asterisks indicate statistically significant differences between treatment and control states. In the case of individual characteristics, the asterisks indicate statistically significant differences from carry out balancing test in covariates. *Significant at 10 percent level; ** Significant at 5 percent level; *** Significant at 1 percent level. Regional characteristics are calculated by the authors based on Spanish LFS microdata (unemployment, education, female employment rate) and on data at regional level available in www.ine.es (GDP, Population, 0-6 years old, Population density). The displayed sample sizes correspond to PISA datasets and are not weighted. The relationship between the treated and control states' sample varied across time because different states expanded their samples in different waves. For this reason, in all of our estimates we use the final student weights.

Table 2: Crowding out

A. Childcare Coverage	*Pre-treatment means*				
	Treatment States	*Control States*	***ITT***	***Se[ITT]***	***% increase***
Public Childcare					
Treat*1993	0.099	0.074	0.261***	[0.060]	264%
Treat*1996	0.099	0.074	0.256***	[0.065]	259%
Private Childcare					
Treat*1993	0.226	0.102	0.021	[0.038]	9.3%
Treat*1996	0.226	0.102	0.020	[0.029]	8.9%
B. Maternal Employment					
Effect up to 1995	0.357	0.289	0.024*	[0.014]	6.7%
Effect up to 1997	0.357	0.289	0.016*	[0.009]	4.5%

Notes: Robust standard errors in brackets. * Significant at 10 percent level; ** Significant at 5 percent level; *** Significant at 1 percent level. Panel A displays the results from estimating equation (1) using as the LHS variable the enrollment rate of 3-years old in public (private) schools. In this case we use annual data from the Spanish Ministry of Education. Sample size: 45 (15 states, 3 years). Panel B displays the results from estimating the effects of LOGSE on maternal employment using Spanish LFS data. Sample sizes: up to 1995 78,123 mothers, up 1997 to 105,748 mothers. Please refer to the Appendix for a thorough explanation of the methodological approach. We control for the pre-reform regional characteristics shown in the panel B.2 of Table 1, except the initial level of childcare coverage when the LHS variable is the enrollment rate of 3 years old. Results are really similar when instead we include state fixed-effects.

Table 1. Main Results

	TT (1)	ITT (2)	Se[ITT] (3)	Controls and states FE (4)
Children outcomes				
Standardized Reading Scores				
Treated*Cohort90	0.454***	0.119***	[0.042]	NO
	0.501***	0.130***	[0.040]	YES
Treated*Cohot93	0.392**	0.099**	[0.040]	NO
	0.395***	0.099***	[0.038]	YES
Standardized Math Scores				
Treated*Cohort90	0.314**	0.082**	[0.040]	NO
	0.291*	0.076*	[0.039]	YES
Treated*Cohot93	0.119	0.030	[0.039]	NO
	0.056	0.014	[0.038]	YES
Falling behind a grade at primary school				
Treated*Cohot93	-0.105**	-0.027**	[0.011]	NO
	-0.098**	-0.025**	[0.011]	YES
Falling behind a grade at high school				
Treated*Cohot93	-0.121*	-0.031*	[0.018]	NO
	-0.124*	-0.031*	[0.018]	YES

Notes: Robust standard errors in brackets. * significant at 10% level; ** significant at 5% level; *** significant at 1% level. ITT/TT=0.260 in 2006 (the increase in childcare coverage between 1990/91 and 1993/94 in treatment states relative to the controls states) and ITT/TT=0.252 in 2009 (the increase in childcare coverage between 1990/91 and 1996/97 in treatment states relative to the controls states). Sample sizes: for Readings and Math scores 34,725; for likelihood of falling a grade (only available to 2003 and 2009) 21,439. The specification with states fixed effects also includes cohort effects, a gender dummy and immigration status.

Table 5. Heterogenous Effects

Panel A By Gender:	*Boys*		*Girls*	
Standardized Reading Scores				
Treated*Cohort90	0.148**	[0.061]	0.120**	[0.052]
Treated*Cohort93	0.064	[0.057]	0.151***	[0.051]
Standardized Math Scores				
Treated*Cohort90	0.060	[0.059]	0.093*	[0.052]
Treated*Cohort93	-0.069	[0.056]	0.105**	[0.051]
Falling behind a grade at primary school				
Treated*Cohort93	-0.024	[0.017]	-0.025*	[0.015]
Falling behind a grade at high school				
Treated*Cohort93	-0.013	[0.026]	-0.045*	[0.024]

Panel B: By education	***Neither of them have a High School degree***		***At least one of them have a High School degree***	
A) Children outcomes				
Standardized Reading Scores				
Treated*Cohort90	0.1399**	[0.0681]	0.0762	[0.0490]
Treated*Cohort93	0.1136*	[0.0680]	0.0578	[0.0456]
Standardized Math Scores				
Treated*Cohort90	0.0439	[0.0663]	0.0486	[0.0486]
Treated*Cohort93	-0.0451	[0.0685]	0.0101	[0.0448]
Falling behind a grade at primary school				
Treated*Cohort93	-0.0398*	[0.0240]	-0.0076	[0.0122]
Falling behind a grade at high school				
Treated*Cohort93	-0.0394	[0.0351]	-0.0173	[0.0204]

Notes: The table reports the ITT parameter. Robust standard errors in brackets. * significant at 10% level; ** significant at 5% level; *** significant at 1% level. Sample sizes are for those with parents of low education: Test scores 14,658, Grade repetition 8,801; for those with at least one parent of high education: Test scores 18,907, Grade repetition 11,487.

Table 7. Robustness checks

	ITT	SE[ITT]	Sates FE and Controls
Panel A) Effect on probability to dropout high school			
Treated*Cohort90	-0.035	[0.025]	NO
	-0.038	[0.025]	YES
Treated*Cohort93	-0.015	[0.024]	NO
	-0.016	[0.024]	YES
B) Effect on probability of migrating across states			
B.1 From control to treatment states			
Treated*Cohort90	-0.010**	[0.005]	NO
	-0.010**	[0.005]	YES
Treated*Cohort93	-0.003	[0.005]	NO
	-0.004	[0.005]	YES
B.2 From treatment to control states			
Treated*Cohort90	0.005	[0.004]	NO
	0.004	[0.004]	YES
Treated*Cohort93	-0.001	[0.004]	NO
	0.000	[0.004]	YES
C) Placebo test: Effect on Birth month			
Treated*Cohort90	-0.020	[0.137]	NO
	-0.055	[0.137]	YES
Treated*Cohort93	0.097	[0.132]	NO
	0.071	[0.132]	YES

Notes: Robust standard errors in brackets. * significant at 10% level; ** significant at 5% level; *** significant at 1% level. In Panel A), the LHS variable is a dummy equal to one if the individual is attending to secondary school at 16 years old. We restrict the sample to natives and immigrants that arrived to Spain before the 3 years old and we use only the 1st and 2nd quarter of each LFS. The total sample size is of 9,927 observations. As covariates we include the sex and immigration status. In Panel B.1 (B.2), the LHS variable is a dummy equal to one if the individual has migrated form a control to a treatment state (o vice versa). We restrict the sample to natives and we use the all quarters of the 2003, 2006 and 2009 LFS. The total sample size is of 19,731 observations. In Panel C), the LHS variable is the relative age of the child (defined as the difference between the month of birth and the cut-off date for children to begin school). Sample size: 34,725. The specification with states fixed effects also includes cohort effects, a gender dummy and immigration status.

Table 8. Alternative Specifications

	Baseline	Treat*Covar	Without richest and poorest	Flexible	Cluster SE	Using growth of public child care in 90-96	Comparing 66th/33th
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Reading Scores							
Treat*Cohort90	0.130***	0.132***	0.103**	0.231***	0.130*	0.188***	0.143***
	[0.040]	[0.040]	[0.051]	[0.075]	[0.07]	[0.039]	[0.042]
Treat*Cohort93	0.099***	0.101***	0.056	0.082	0.099	0.088**	0.082**
	[0.038]	[0.038]	[0.051]	[0.072]	[0.070]	[0.037]	[0.041]
Math Scores							
Treat*Cohort90	0.076*	0.076*	0.143***	0.134*	0.076	0.062	0.036
	[0.039]	[0.039]	[0.050]	[0.073]	[0.059]	[0.039]	[0.042]
Treat*Cohort93	0.014	0.015	0.094*	-0.016	0.014	0.034	-0.007
	[0.038]	[0.038]	[0.050]	[0.071]	[0.064]	[0.037]	[0.041]
Falling behind a grade at primary school							
Treat*Cohort93	-0.025**	-0.024**	-0.018	-0.030*	-0.025***	-0.013	-0.032***
	[0.011]	[0.011]	[0.014]	[0.017]	[0.009]	[0.011]	[0.011]
Falling behind a grade at high school							
Treat*Cohort93	-0.031*	-0.030*	-0.043*	-0.030	-0.031	0.005	0.004
	[0.018]	[0.018]	[0.023]	[0.034]	[0.025]	[0.017]	[0.018]

Notes: We report the intent to treatment effect (ITT) including covariates and states and cohorts fixed effects. Column (1) presents our preferred specification. In column (2), we allow for different effect of covariates between control and treatment states interacting the covariates with the treatment dummy.Column (3) shows the estimates dropping the richest and the poorest states within treatment and control groups.In column (4), the cohort fixed effects are interacted with pre-reform states socio-economic characteristics. In column (5), standard errors are clustered to account for serial dependence of the errors within state-period groups. In column (6), we use as treatment group those states with a growth in the enrollment rate above median between 1990 and 1996. And in column (7) treatment states are above 67th percentile in public enrollment growth, while control states are below the 33th. Robust standard errors in brackets. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 9. Alternative Identification Strategy

	ITT	Se[ITT]	Controls and states FE
	(1)	(2)	(3)
Using the public childcare seats offered from 3 to 5 years old			
Standardized Reading Scores			
Post-LOGSE	0.008***	[0.002]	NO
	0.008***	[0.002]	YES
Standardized Math Scores			
Post-LOGSE	0.0014	[0.002]	NO
	0.0014	[0.002]	YES
Falling behind a grade at primary school			
Post-LOGSE	-0.002***	[0.001]	NO
	-0.002***	[0.001]	YES
Falling behind a grade at high school			
Post-LOGSE	0.001	[0.001]	NO
	0.001	[0.001]	YES

Notes: Robust standard errors in brackets. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

APPENDIX

Estimating the effects of the reform on maternal employment

As in the first essay, we follow a Difference-in-Difference-in-Difference approach exploiting the fact that the legislation affected children of 3 years old but not mothers of 2 years old. We therefore estimate the following equation:

$$Y_{ijt} = \gamma_1 Treat_j + \gamma_2 Mom3_i + \gamma_3 Post_t + \gamma_4 (Treat_j * Mom3_i) + \gamma_5 (Treat_j * Post_t) + \\ + \gamma_6 Mom3_i * Post_j + \theta \, (Treat_j * Mom3_i * Post_t) + \delta_j + \lambda_t + X'_{ijt}\beta + \varepsilon_{ijt}$$

where Y_{ijt} is the outcome of interest (employment or weekly hours worked) for a sample of mothers whose youngest child is 2 or 3 years old, $Treat_j$ is equal to one if the mother lives in a treatment state and zero otherwise; $Mom3_i$ is equal to one for mothers whose youngest child is 3 years old and zero for mothers whose youngest child is 2 years old; the variable $Post_t$ is equal to one after LOGSE implementation began (that is, from 1991/92 onwards). The coefficient θ capture any difference in the likelihood of being employed for mothers of treated children (3 year olds) relative to control children (2 year olds) living in treated states after the childcare expansion. The vector X_{ijt} includes the same individual and regional controls as in the first essay, namely age squared, dummies indicating the number of other children, a dummy for being foreign-born, educational attainment dummies (high-school dropout, high-school graduate, and college), a dummy for being married or cohabitating. We also include state and year fixed effects. We estimate this equation by OLS using data from the Spanish Labor Force Survey from 1987 to 1994 and also from 1987 to 1997.

Table A 1. Sensitivity Analysis of Covariates Included

	Unconditional	+ State FE	+ Individ. Characteristics	+ Fahter's Education	+ Mother's education	+ Home Possessions	+ Type of school	+ Pop. density of place of residence
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Reading score								
Treat*Cohort90	0.118***	0.121***	0.130***	0.112***	0.101***	0.105***	0.119***	0.125***
	[0.042]	[0.041]	[0.040]	[0.039]	[0.039]	[0.038]	[0.038]	[0.038]
Treat*Cohort93	0.099**	0.095**	0.099***	0.078**	0.065*	0.096***	0.113***	0.106***
	[0.040]	[0.039]	[0.038]	[0.037]	[0.037]	[0.036]	[0.037]	[0.036]
Maths score								
Treat*Cohort90	0.082**	0.082**	0.076*	0.056	0.044	0.048	0.071*	0.075**
	[0.040]	[0.040]	[0.039]	[0.038]	[0.038]	[0.037]	[0.037]	[0.037]
Treat*Cohort93	0.03	0.026	0.014	-0.009	-0.023	0.01	0.031	0.026
	[0.039]	[0.038]	[0.038]	[0.037]	[0.036]	[0.035]	[0.036]	[0.036]
Falling behind a grade at primary school								
Treat*Cohort93	-0.027**	-0.023**	-0.025**	-0.022**	-0.021*	-0.027**	-0.031***	-0.029***
	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]
Falling behind a grade at high school								
Treat*Cohort93	-0.031*	-0.030*	-0.031*	-0.024	-0.020	-0.033*	-0.036**	-0.034**
	[0.018]	[0.018]	[0.018]	[0.017]	[0.017]	[0.017]	[0.018]	[0.018]

Notes: Individual characteristics: male, immigrants; Family Background: level of education of parents (hs dropout - omitted, hs graduate, college); Home possessions: index derived from students' responses to the following items: do you have: a desk for study, a room of your own, a computer, internet, classic literature, books, works of art, dishwasher, among others; Type of school: public-omitted; private; Population density of place of residence: Village, Small Town, Town, City, Large City, Metropoli -omitted.

4. Essay 3: The role of culture in explaining the educational gender gap

4.1 Introduction

It has been widely documented that whereas girls perform slightly better than boys at school at younger ages in most subjects, they start to diverge when they become teenagers, with boys performing better in math, and girls performing better in reading (Hyde *et al.* 1990, Hedges and Nowell 1995, Hausmann *et al.* 2008). This divergence in tests scores between sexes is of high interest since it might to explain the persistence of a substantial gender inequality at the top of the income distribution (Brown and Corcoran 1997; Weinberger 1999).

The source of these differences is a matter of controversy. One strand of the literature points to biological explanations based on genetic, hormonal or cerebral differences between sexes (see Wilder and Powell 1989 and Penner 2008 for reviews). Others highlight the relevance of societal factors, claiming that the variation across countries cannot been explained by biological differences (Baker and Jones 1993; Penner 2008). Among the studies analyzing the role of societal factors, some researchers have focused on the institutional or environmental explanations.[1]

In a recent *Science* article, Guiso *et. al*. (2008) examine the relationship between cross country differences in gender gap in math scores and cultural attitudes toward women. Using data from 2003 PISA tests, they find that *"gender gap in math scores disappears in countries with a more gender-equal culture*" (pp 1165). Note that since Guiso *et al.* (2008) focus on measures of gender equality from the country where the individual lives and uses the whole population (both natives and migrants), their approach makes it difficult to disentangle the effect of culture versus that of institutions, and does not identify a causal relationship. They attempt to correct the last concern

[1] For instance, Booth *et al.* (2011); Booth and Nolen (2012); Billger (2002), analyze the role of learning environments and find that when girls are taught separately from boys they often do better in subjects such as math than if classes are mixed. Others put forward on differential treatment by teachers (Heller and Parsons 1981; Leinhardt *et al.* 1979; Parsons *et al.* 1982) or parents (Muller 1998; Bouffard and Hill 2005; Bhanot and Jovanovic 2005), stereotype threat (Spencer *et al.* 1999; O'Brien and Crandall 2003), and other environmental factors (Gneezy *et al.* 2003; Levine *et al.* 2005).

with the use of country fixed-effects, nonetheless omitted variable bias is likely to affect their estimates[2].

The distinction between norms and institutions is crucial because the policy implications of these two interpretations are very different. If, according to the norms interpretation, girls in countries with rigid gender roles were socialized into believing that it is not a girl's role to study math, then policies that alter the beliefs about gender roles early in life will be most appropriate (see for example work on the transmission of preferences from mothers to daughters and sons: Fernández *et al.* 2004, Farré and Vella *forthcoming*, González de San Román and de la Rica 2012). In contrast, it could be that girls in different countries hold the same beliefs about what boys and girls ought to study, yet girls in countries with rigid gender roles face different sort of institutional constraints. For example, it may be that less gender-equal countries may also exhibit very poor labor market prospects for women in traditionally male dominated professions. Thus, girls may decide not to invest in education (or male-dominant educational choices) if they believe that they will not be able to find a job later on. The institutional constraint interpretation may thus imply policies aimed at reducing the wage gender gap (see for instance Dolado *et al.* 2012).

Following the epidemiological approach fully developed in Fernández and Fogli (2009) and reviewed in Fernández (2011), this essay goes a step further at examining the relationship between gender differences in math test's performance. In particular, we look at how second-generation immigrant girls and boys perform in school according to the prevailing gender roles in their parents' country of origin. This approach enables us to disentangle the interrelation between institutions and cultural norms. Immigrants in our sample have lived under the laws, institutions, and markets of the receiving country. However, since their preferences are likely to reflect the attitudes of their parents and ethnic communities, differences in gender roles in their country of origin may be interpreted as evidence of the importance of beliefs. For example, if institutions (such

[2] This approach have been also applied by other researchers. Using Second International Mathematics Study (SIMS) data, Backer and Jones (1993) identify a correlation between the cross-country gender differences in mathematic performance and several measures of gender stratification of opportunities, such as the percentage of females in higher education and the percentage of females in the labor force. They interpret their results as evidence of how different opportunity structures can determine different performances. More recently, Gonzalez San Roman and de la Rica (2012) replicated the Guiso et al.' relationship (both at country and individual level) using the wider sample of countries participating in the 2009 PISA wave and additional measures of gender-equality. The authors also find evidence of the same relationship across regions within a country, which imply that the same is found within a more homogenous institutional setting.

as labor market discrimination) were the only explanation for why girls perform much worse than boys in say, Italy, than say in Norway, then when we remove differences in laws by examining Italians and Norwegians living in the same country, all Italian-Norwegian differentials should be eliminated. Instead, if home country gender roles can explain educational choices of childhood migrants who have spent most of their lives exposed to the receiving country's culture and norms, this may be interpreted as evidence that cultural variation is at least a partial explanation for the differences in educational achievement.

In our empirical analysis, we use multi-country data from 2003, 2006, and 2009 PISA data sets to estimate the relationship between gender educational gap of second-generation immigrants and gender roles in their country of origin. We use alternative data sets to construct several measures to capture gender roles. In particular, we use the Gender Gap Index and the Political Empowerment Index, both elaborated by the World Economic Forum, as well as Female Labor Force Participation rates.

We find that the higher degree of gender equality in the country of origin the higher the tests scores of second-generation immigrant girls relative to boys. Our results are robust to a battery of sensitivity tests to sample selection and we also rule out alternative explanations for the result such as immigrants from countries of origin with more gender-equal cultures living in host countries with lower gender gap in math scores.

Our work complements a growing literature on the effect of culture on socio-economic outcomes (see Fernández 2011 and Guiso *et al.* 2006 for a review). Using methodologies similar to ours, studies have examined the effect of culture on savings rates (Carroll *et al.* 1994), fertility and female labor force participation (Antecol 2000; Fernández and Fogli 2006; Fernández 2007; Fernández and Fogli 2009), living arrangements (Giuliano 2007), unemployment rates (Brügger *et al.* 2009), preferences for a child's sex (Almond *et al.* 2009) and divorce decisions (Furtado *et al.* 2012). We add to this body of knowledge by examining the role of culture on educational attainment.

The essay is organized as follows. Section 4.2 presents the empirical strategy and Section 4.3 describes the data. Baseline results and robustness checks are discussed in Section 4.4. Section 4.5 briefly discuss the mechanisms through culture could operate and Section 4.6 concludes.

4.2 Empirical Strategy

Our empirical approach exploits the fact that the sons and daughters of immigrants are, and have been, exposed to the same markets, norms, and institutions, i.e., those of the host country. However, their parents grew up under a different institutional framework and therefore their beliefs are likely to reflect the values of their country of origin rather than the ones of the destination country. Thus, evidence that home country gender-equality can explain gender gaps in educational attainment of second-generation immigrants might be interpreted as suggestive of the role of culture.

This approach assumes that parents transmit their cultural beliefs to their children and that the cultural beliefs vary systematically across groups such that cultural beliefs of immigrants effectively reflects the average culture of their country of origin. Even more, it assumes that individuals who live in the same country face similar economic and institutional environments. These quite restrictive assumptions imply that one is more likely to rule out the effect of culture than to find it when there is not. Therefore, the power of this test of culture is asymmetric and the absence of a significant coefficient on the cultural proxy does not imply that only the host economic and institutional setting matters (Fernández 2011 and Fernández and Fogli 2009). In addition, despite studying second-generation rather than first-generation immigrants reduce the effect of usual shocks resulting from immigration (e.g., language barriers), the fact that they are second-generation is bound to dilute the strength of cultural effects on economic actions since cultural transmission is restricted mostly to parents and ethnic social networks rather than operating in society at large (e.g., throughout schools, mass media, etc.).

The baseline specification is as follows:[3]

$$E_{ijkt}=\alpha_1 female_i + \alpha_2(female_i * GE_j)+\lambda_t+\lambda_j+X'_{ijk}\beta + \varepsilon_{ijk} \qquad (1)$$

where i is the individual of origin j who lives in country of destination k. E_{ijkt} indicates individual's educational attainment, in our case, math (or reading) tests scores. To identify the differences in educational attainment between sexes, the variable $female_i$ is a dummy equal to one if the individual is a girl and zero otherwise. The vector X_{ijk} contains socio-demographic characteristics that may affect educational attainment for

[3] We assume a linear relationship and estimate this equation by OLS.

reasons unrelated to culture. Descriptive statistics of these socio-demographic variables are described in detail in Appendix Table A1. In order to have the larger variation possible in terms of both host countries and country of origin, we pool the 2003, 2006 and 2009 PISA waves and include year-fixed effects λ_t in all of our specifications. In order to take into account that PISA samples students in two stages (first schools are sampled and then students are sampled in the participating schools), we follow the OECD recommendations and compute the standard errors by applying Balanced Repeated Replication--see the details in OECD, 2009.

As in Guiso *et al.* (2008) we interpret the degree of gender equality in a country as an indicator of the beliefs about the role of women in that country. The variable GE_j is a measure of gender equality in country of origin j with higher values of GE_j representing a more gender-equal society in country j. We use several indicators of gender equality in the country of origin, which have been used in the literature (Guiso *et al.* 2008, Gonzalez de San Roman and de la Rica 2012): (1) the Gender Gap Index (henceforth GGI) elaborated by the World Economic Forum, which synthesizes the relative position of women in a society taking into account the gap between men and women in several areas; (2) the Political Empowerment Index (henceforth PEI), from the same source, which measures the gap between men and women in political participation; (3) the total Female Labor Force Participation Rate (henceforth FLFP), and (4) the Female Labor Force Participation Rate (henceforth FLFP35-54) for women between 35 and 54 years old, from the International Labour Organization. As Gonzalez de San Roman and de la Rica (2012) explain, the interest in this cohort of women is that it coincides with the age interval of the mothers of the PISA students.[4] Appendix Table A.1 presents a detailed description of each of these measures.

Ideally, we would like to use past values of gender equality indicators, either when the parents left their country of birth or when they were young. As the children in our sample are 16 years old in 2003, 2006 and 2009 and were born in the host country, their parents must have emigrated during the 1986-1993 period or earlier. Data limitations prevent us from using all the indicators for those years. For instance, the Gender Gap Index is only available from 2006 onwards. Consequently, in our main

---

[4] Guiso *et al.* (2008) and Gonzalez de San Roman and de la Rica (2012) also use as proxy of gender equality an index of cultural attitudes towards women, which is elaborated based on the World Value Survey--for details about this index see the Supporting Online Material of Guiso *et al.* 2008 paper in www.sicencemag.org/cgi/full/320/5880/1164/DC1. We do not use this indicator as only 27 out of 41 countries in our sample participate in that survey.

specification we use the average home country's measures of gender equality for the period 2003-2009. This is a common practice in the literature and ought not represent a problem for our estimates. If countries' aggregated preferences and beliefs about the role of women in society change slowly over time, these variables should also have explanatory power despite being measured contemporaneously (Fernández, 2011). In addition, as Fernández and Fogli (2009) point out, one could argued that the values that parents and society transmit are best reflected in what their contemporaneous counterparts are doing in the country of ancestry. Rather than trying to solve this debate on theoretical grounds, we take an empirical approach and also check the robustness of our results with proxies of gender-equality when data are available for the 1990s.

The interaction between the female dummy and the gender equality measure (GE_j) in country of origin j, α_2, captures the role of gender roles in explaining the gender differences in the educational attainment of second-generation immigrants. A positive and significant estimation of α_2, the interaction between female and the measure of gender equality in the country of origin would suggest that more egalitarian attitudes would be correlated with a higher relative performance of second generation immigrant girls over boys, and thus smaller gender differences in children educational attainment. In other words, a positive α_2 means that for two immigrant girls living in the same host country under the same markets and institutions, a girl's math score from a country of origin *j with more egalitarian gender roles* (higher GE_j) is higher than a girl's score from a country of origin with less egalitarian gender roles.

Since there are many other sources of heterogeneity across countries in addition to their cultural beliefs affecting the educational achievements of second-generation immigrants, we include country of origin fixed effects, denoted by λ_j. By doing so, we also control for differences other home country characteristics that may affect the educational performance of male and female immigrants.

In a second specification, we also include host-country fixed effects (λ_k). Host country fixed effects λ_j are included to account for regional variation in host country's educational gender gaps that may arise from cross-country differentials in culture or institutions.

$$E_{ijkt}=\alpha_1 female_i + \alpha_2(female_i * GE_j)+\lambda_t+\lambda_j+\lambda_k+X'_{ijk}\beta +\varepsilon_{ijk} \qquad (2)$$

In this case we exploit uniquely the cultural variation *within* the same host country. In this way we control for regional variation in the gender attainment gap that might arise from differences across host countries in the institutional setting. For example, if immigrants from countries of origin with less gender-equal cultures tend to also settle in host countries with less gender-equal cultures and institutions it might lead to an upward bias in our coefficient of interest as the gender-equality proxy may be capturing the effect of the host country laws and institutions rather than the effect of the culture of origin.

In a third specification, we further include country of origin GDP per capita (in logarithms) interacted by the female dummy.

$$E_{ijkt}=\alpha_1 female_i + \alpha_2(female_i*GE_j)+ \alpha_3(female_i*lgdp_j)+\lambda_t+ \lambda_j+\lambda_k+X'_{ijk}\beta +\varepsilon_{ijk} \qquad (3)$$

Our measures of gender-equality might be reflecting differences in the level of development of the countries rather than differences in cultural attitudes towards women. Including the interaction between the female dummy and the level of development of the country of origin, we ensure that our coefficient of interest, α_2, is capturing the effect of culture and not just the influence of the level the development of the country of origin.

Controlling for host-country fixed effects our estimate is identified through within-country of destiny changes in migrant's cultural background. Such a specification may be seen as too restrictive given that it does not take into account how immigrants are distributed across our sample of host countries. For instance, it does not take into account whether, for example, the educational attainment of Chinese children who go to Australia differs to those whose parents migrate to other countries. This could be relevant given that countries differ in their immigration admission policies. For instance, traditional immigrant-receiving countries such as Australia, Canada and New Zealand have instituted skills-based ‘point systems' that reward certain socio-economic traits in the admission formula (Borjas, 2001). Indeed, previous research have found that educational achievements of second-generation immigrants is quite heterogeneous across countries, and that countries where immigration is common such as Australia or Canada seem to do well in absorbing immigrant children, with test scores gaps between immigrants and natives disappearing after conditioning on parental characteristics (see Dustmann *et al., forthcoming*).

Thus we estimate a fourth specification that replaces the host-country fixed effect by a proxy of immigration policies in the host country instead of host country fixed-effects as follows:

$$E_{ijkt}=\alpha_1 female_i + \alpha_2(female_i*GE_j)+ \alpha_3(female_i*lgdp_j)+\lambda_t+ \lambda_j+X'_{ijk}\beta +Z'_k\delta+\varepsilon_{ijk} \qquad (4)$$

where Z_k includes host country proxies of the degree of immigrant's assimilation, such as the native-immigrant gap in tests scores, and the level of development of the country.

4.3 Data and Sample Selection

We use data from the *Programme for International Student Assessment* (PISA), an internationally standardized assessment that was jointly developed by participating economies and administered to 15-year olds in schools. The purpose of PISA is to test whether students, near the end of compulsory education have acquired the knowledge and skills essential for full participation in society. In particular, it administers specific tests to assess whether students can analyze, reason, and communicate effectively. PISA uses imputation methods, denoted as plausible values (hereinafter PV), to report student performance. In all of our analysis we use PV and follow the OECD recommendations that involve estimating one regression for each set of PV (there are five PV to each domain) and then report the arithmetic average of these estimates. The PISA program scaled the scores to have a mean of 500 and a standard deviation of 100 in the OECD student population.

Our analysis focuses on mathematics test scores and we pool the 2003, 2006 and 2009 PISA waves. To check if the relationship between cross-country differences in gender equality and gender gaps in math scores is also observable in other domains, we carry out the same analysis on reading scores. The student's performance in both domains are fully comparable across PISA cycles from 2003 onwards.

Our sample consists of second-generation immigrants who were born and reside in a host country within PISA but whose parents (both of them) were born in another country. To determine a pupil's country of origin, we need specific information on the country of birth of the parents. We restrict our sample to those participating countries providing detailed information about the parents' birth place. These are Australia, Austria, Belgium, Denmark, Finland, Germany, Greece, Latvia, Liechtenstein, Luxembourg, New Zealand, Norway, Portugal, Switzerland and Scotland in 2003, 2006, and 2009 PISA, and Argentina, Czech Republic, Israel, Netherlands and Qatar in 2009 PISA. When parents come from different countries we use the mother's birthplace to assign the country of ancestry. In the robustness section, we test the sensitivity of our results to this decision using the father's birthplace.

Several sample restrictions were applied as described below. First, to ensure a minimum number of observations by country of origin and destiny (and then to be able to do comparisons across averages), we eliminated second-generation immigrants from countries of origin with fewer than 10 observations within the same host country.[5] Second, we checked the proportion of males and females from the same country of origin within each host country and removed the single sex cases. Third, since our estimate is identified through within host country changes in migrant's cultural origin, we need to ensure some degree of variation within host countries. Thus, we drop those host countries with less than three home-country groups of immigrants.[6]

We also exclude from our analysis both second-generation German migrants and migrants to Germany because we are unable to identify Germans from East and West Germany. As these children were born during or right after the reunification period and their parents came regions within a country with great socio-economic and cultural differences, we decided to exclude them from our sample. Similarly, as we could not identify sons of East Germans and clearly East German migrants affected the arrival of other ethnicities at the time of the reunification, we dropped Germany from the analysis. We also lose immigrants from Occupied Palestine Territory due to the lack of internationally comparable macro indicators.

Our final sample constitutes 83 percent of the original 2003, 2006 and 2009 PISA sample of second-generation migrants, namely 11,177 second-generation immigrants from 43 different countries of origin and living in 12 different host countries. Appendix Table A2 presents our final sample by host country and countries of origin. Host countries are mainly OECD countries participating in PISA (with the

---

[5] Since our regressions are all run at the individual level, including or not these small numbers of observations does not affect our results.

[6] The first restriction implied losing 174 observations (1.2 percent of the sample); the second restriction implied losing 22 additional observations, and the third restriction implied losing a total of 42 observations.

exceptions of Argentina, Latvia, Liechtenstein and Qatar), whereas countries of origin are from various continents and levels of development. The largest sample of immigrants come from Portugal, Turkey, Serbia-Montenegro or United Kingdom (they represent the 53 percent of the sample). Host countries with the highest sample of second generation immigrants are Australia, Switzerland and Luxembourg (immigrants living in these countries represent 65 percent of the sample).

Table 1 present summary statistics of the relevant variables by the country of origin of second-generation migrants' parents. We order our sample from the highest to the lowest gender gap in math scores of second-generation immigrants (difference between girls' and boys' scores). Column 1 shows that there is a large variation in the gender gap in math scores across countries of origin, from a negative gap above 100 points for second-generation migrants whose parents' were born in Macedonia to a positive gap of 66 points for those whose parents came from The Netherlands. The following columns in Table 1 display the different gender-equality measures and the GDP per capita in the parent's country of origin, which suggest a positive relationship between gender equality and girls' math scores in relation to boys'. In Figure 1 we plot the average second-generation migrants' gender gap in math scores by country of origin (column 1 of Table 1) versus two measures of gender equality: the Gender Gap Index and the Female Labor Force Participation (columns 2 and 4 of Table 1). The raw data shows a positive correlation between the gender equality in the country of origin and the relative performance in math scores of second generation immigrant girls with respect to boys.

Gender differences in tests scores vary over the test scores distribution, such that among high-achieving students the relative disadvantage of girls in math scores is higher and the relative advantage of girls on reading scores is lower (Gonzalez de San Roman and de la Rica 2012). Table 2 looks at whether the relative position of second-generation immigrants in the test scores distribution may be biasing our results. Table 2 displays the raw average gender gap on math and reading scores for all individuals and for second-generation natives estimated on both the full sample of individuals participating in 2003, 2006 and 2009 PISA waves and on our sample of host countries. Second generation immigrants are not at the bottom of the distribution. On the contrary, their scores in both math and reading tests are, on average, slightly higher than those of all individuals participating in PISA tests. Overall, second-generation children in our sample also perform slightly better than other children. It is also important to note that girls underperform with respect to boys in math and over perform with respect to boys in reading, regardless of whether they are second-generation migrants or natives, and regardless of the sample of countries used. It thus seems that test scores of our sample of second-generation immigrants compare well with native children and that our immigrant sample is not a biased sample.

4.4 Results

Replicating Guiso et al. (2008)

Prior to presenting our analysis, we first replicate Guiso *et al.* (2008) analysis at the student level using the pool of 2003, 2006 and 2009 PISA waves to verify that Guiso and co-authors' results can be replicated when pooling the three PISA waives--as opposed to only using the 2003 wave. Note that since identification of the coefficient of interest, α_2, in Guiso *et al.* (2008) comes from variation across countries, we use all sample of countries available in PISA as opposed to our restricted sample, which is smaller due to the more demanding data sources needed.[7] Table 3 shows that the results from Guiso *et al.* (2008) hold when adding the 2006 and 2009 waves. Namely, we find a strong positive association between several measures of gender-equality and the relative performance of girls on both math and reading scores.

As explained earlier, estimates from Table 3 only capture correlations. In the next section we apply the epidemiological approach to test whether such relationship could be interpreted as causal evidence of the role of culture in explaining the educational gender gaps.

Culture and Educational Gender Gap

Table 4 presents our main results when estimating the different specifications presented in Section 4.2. Four estimates are presented for each measure. First, we present the baseline specification in Equation (1), which only includes year fixed-effects and country of origin fixed effects. Second, we add host-country fixed effects as in

[7] Because of data restrictions, our sample has 12 host countries. In contrast, when focusing the analysis on all individuals, Guiso *et al.*'s sample contains 39 countries, and our three waves of PISA dataset contains 63 countries. When doing Guiso et al.'s estimation with the 12 countries we obtain the coefficient of interest being of similar size and sign as for the whole sample, but we lack precision as the cross-country variation is greatly reduced.

Equation (2). Third, we add the interaction between the GDP per capita of the country of origin and the female dummy as an additional control as indicated in Equation (3). Finally, we present an alternative specification in which we control for a proxy of the immigrants assimilation and GDP level of the host countries instead if host country fixed effects. Column 1 displays the coefficient on the female dummy and Column 2 the interaction between the female dummy and the different measures of gender equality in the parent's country of origin. Panels A, B, C, and D indicate which measure of gender equality is used in each case.

Results from our baseline specification in Table 4 show that our coefficient of interest on the interaction between the gender-equality measure in the parents' country of origin and the female dummy is positive and statistically significant at conventional levels for all the gender equality indicators. This result suggests that parents' culture matters in explaining second-generation migrants' gender gaps. In particular, the more gender-equal the parents' country of origin is, the higher the math scores of second-generation immigrant girls relative to those of boys. In general, the coefficient estimated is greater and of opposite sign than the coefficient on the female dummy (only in the case of FLP we find an interaction coefficient of smaller size than the one on the female dummy), implying that gender differences in mathematics decrease for those immigrants whose parents come from more gender-equal countries.

As we explained in Section 4.2 if immigrants from more gender-equal countries prefer settling in countries with similar cultural beliefs about gender roles or laws and institutions (labor market policies) that motivate women to do better at school in male dominated subjects than those of their country of origin, our coefficient of interest might be capturing the effect of the host-country norms rather than the effect of the culture of the parents' country of origin. To address this concern, our second specification adds host-country fixed effects. Identification of α_2 now comes from the comparison of second-generation immigrant children from different cultural origins *within* a given host country. The magnitude and statistical significance of α_2 remains practically unchanged, suggesting that our main result is not driven by immigrants migrating into countries that resemble their country of origin. The results from our third and preferred specification when adding the GDP level interacted by the female dummy suggest that our measures of gender-equality is indeed capturing the effect of cultural attitudes towards women and not just the effect of the level of development of the country of origin. The interaction with all our measures of gender-equality increases in size and remain statistically significant.

To test to what extent differences in the way that countries deal with, or select immigrants, could explain our results we estimated Equation (4) controlling for a proxy of immigration policies in the host country instead of host country fixed-effects. As proxy of immigrant's children assimilation we use the gap between natives and second-generation immigrants on PISA tests scores (specifically, in math scores). We also control for the GDP level of the host country. Results, displayed in Table 3 (last row within each panel) show that the effect of culture from the parents' country of origin on females math scores remains positive and statistically significant when we control for the immigrant-natives gap in math scores and the GDP per capita instead of host-country fixed effects.

Results from estimating Equations (1) through (4) when the dependent variable is the gender gap in reading scores are shown in Table 5. The interaction between the female dummy and the measures of gender-equality is always positive, although statistically significant in two out of four gender equality measures, suggesting that girls from more gender-equal countries gain an absolute advantage over boys in scores.[8] These results are consistent with the findings from the broader literature analyzing the effect of single-sex classes on girls and boys attainment. This strand of literature generally finds that girls in single-sex classes gain in "male" subjects but there are no differences in attainments in traditionally "female" subjects for male students--see Mael *et al.* 2005 for a review of this literature.

We show that our results are not sensitive to sample selection by carrying out several robustness checks on our preferred specification –the specification in Equation (3). The results are presented in Table 6. For simplicity, we only present the results on math scores but the general conclusions are the same for reading scores (see Appendix Table A 3). Panels A in Table 6 shows that our results are not driven by the main group of immigrants (Portuguese). Excluding Switzerland, the host country with the larger sample of immigrants, does not change the results either (Panel B). We also test the robustness of our results when using the most recent 2009 PISA data. In this case, the

[8] Given that our sample consists of second-generation of immigrants, of whom 47 percent speak at home other than the language test, we reran the regression dividing the sample according to the language spoken at home and find the results are driven by those who speak the test language at home.

coefficients estimated are quite similar but we lose precision due to the smaller sample size (see Panel C of Table 6). The main result does not change when we assign the father's country of birth (instead of the mother's) or when we restrict the sample to children whose parents come from the same foreign country (Panels D and E of Table 6, respectively). We also test if our results remain when we use measures of gender equality from 15 years earlier instead contemporaneous measures. As measure of political empowerment we use the proportion of seats held by women in national parliaments during 1990, available in the World Bank Statistics. We also test the robustness of our results to using female labor participation rates of 1990 instead of the 2003, 2006 and 2009 average. As can be seen in Panel F of Table 6, the effect is higher for the measure of political empowerment but it decreases and loses significance for one of the two measures of female labor force participation rates measured in 1990.

In summary, using individuals who lived and grew up under the same institutional framework but who have different cultural background, we find evidence that culture explains the gender gap in tests scores. As previously suggested by Guiso *et al.* (2008), our results indicate that in more gender-equal cultures, girls perform as well as boys in mathematics and much better than them in reading.

4.5 The Transmission of Culture

There are several sources of heterogeneity across individuals other than cultural beliefs that may affect their educational attainment. Since many socio-demographic characteristics may well be influenced by culture and are thus endogenous, we use these characteristics to test whether culture affects educational outcomes directly or indirectly in later specifications. For instance, one might expect parents (and mothers in particular) of those children from countries of origin with more gender-equal cultures to have higher levels of education. In a similar way, it is reasonable to think that the type of school immigrants send their children to (single-sex schools and/or private schools) could be influenced by their culture. In this section, we sequentially add a set of control variables available in PISA in order to test whether culture affects educational outcomes beyond the ways in which it is already reflected in family decisions. To allow different effects of each control variable on girls and boys, we also include the interaction between each variable and the female dummy. Finding that after the inclusion of some of these variables our coefficient of interest (the interaction between culture proxy and the female dummy) changes will be indicative of which channels culture is being transmitted.

There is extensive information about individual characteristics available in PISA datasets, which show that many of them vary considerably across the individuals in our sample (see the Appendix Table A 4). For instance, the education levels of parents vary widely across countries of origin, with Paraguay, Bolivia, Bosnia-Herzegovina and Portugal having the lowest proportion of immigrants whose parents have a college degree and Egypt, Korea and United States the highest. A similar pattern emerges when looking at the index of amount of books at home. In addition, while all immigrants from Bolivia, Chile and Switzerland speak the test language at home, all immigrants of Macedonia speak a foreign language. There are also noticeable differences in the type of school they attend and also in the place they tend to settle within the host country (large versus small cities or towns).

Results on math scores are shown in Table 7.[9] It is first important to highlight that the coefficient of interest, α_2, remains significant and positive and for some indicators the size of the coefficient remains quite large. This result suggests that culture affects the gender gap in quite a direct way. Nonetheless, the introduction of some controls, namely parental education or language spoken at home, seem to matter indicating that culture may operate indirectly through them. The higher variation is observed when we add an indicator of whether the individual speaks a foreign language at home. In this case, the estimated direct effect of culture decreases substantially and even looses statistically significance in 2 out of 4 measures of gender equality.

4.6 Conclusion

This paper aims to rigorously disentangle the effects of markets and institutions from the effects of culture in determining gender differences in educational attainment. Because second-generation immigrants live in the host country they absorb home country culture from their parents and ethnic communities but are exposed to the host country's laws and institutions. We interpret the positive estimated effect of parents' home country measures of gender equality on their educational attainment as evidence of the role of culture.

[9] For simplicity we do not show the coefficients estimated for each covariate included, but the complete table is available upon author request.

We find that the higher degree of gender equality in the country of origin improves the performance of second-generation immigrant girls relative to boys. In particular, the gender gap disappears among immigrants from more gender-equal cultures. Our results are robust to a battery of sensitivity tests to sample selection. We also rule out alternative explanations for the result such as immigrants from countries of origin with more gender-equal cultures living in host countries with lower gender gaps in math scores. We also exploit the channels through which culture is transmitted, and find that language spoken at home and parental education are important channels.

Our findings that a more gender-equal culture affects girls' academic achievement with respect to boys yield support policies that alter the believes about gender roles early in life (see for example work on the transmission of preferences from mothers to daughters and sons: Fernández *et al.* 2004, Farré and Vella *forthcoming*, González de San Román and de la Rica 2012). Our findings may also explain why similar education institutions may have a different impact on the gender gap in test scores and point to the interplay between culture and institutions as the missing factor.

REFERENCES

Almond, D.Jr, Edlund, L. and Milligan, K. (2009). "Son Preference and the Persistence of Culture: Evidence from Asian Immigrants to Canada." NBER Working Paper No. 15391.

Antecol, H. (2000). "An Examination of Cross-Country Differences in the Gender Gap in Labor Force Participation rates." *Labour Economics*, 7(4), 409-426.

Baker DP, Jones DP (1993) "Creating gender equality: Cross-national gender stratification and mathematical performance". *Sociology of Education* (American Sociological Association), 66: 91–103.

Bhanot, R., and J. Jovanovic (2005) "Do Parents' Academic Gender Stereotypes Influence Whether They Intrude on Their Children's Homework?" *Sex Roles*, 52(9/10): 597–607.

Billger, S.M. (2002) "Admitting men into a women's college: a natural experiment." *Applied Economics Letters,* 9 (7): 479-483.

Booth, A. L., & Nolen, P. (2012). "Gender Differences in Risk Behaviour: Does Nurture Matter?"*The Economic Journal, 122*, 56-78.

Booth, A. L., Cardona Sosa, L., & Nolen, P. J. (2011). "Gender Differences in Risk Aversion: Do Single-Sex Environments Affect their Development?"IZA Discussion Paper , 6133.

Borjas, G.J. (2001). Immigration Policy: A Proposal. Pp. 17–20 in Blueprints for an Ideal Legal Immigration Policy, edited by R. D. Lamm and A. Simpson. Washington, DC: Centre for Immigration Studies.

Bouffard, S. M., and N.E. Hill. (2005) “Maternal Perceptions of Competence and Children’s Academic Adjustment: Longitudinal Relations across Early Elementary School.” *Social Psychology of Education*, 8(4): 441–63.

Brown, C and Cororan, M (1997)S "Sex-Based Differences in School Content and the Male-Female Wage Gap.', *Journal of Labor Economics*, 1997, pp. 431-465.

Brügger, B., Lalive, R. and Zweimüller, J. (2009). “Does Culture Affect Unemployment? Evidence from Röstigraben.” IZA Discussion Papers 4283.

Carroll, Ch. D., Rhee, B. and Rhee, Ch. (1994). “Are There Cultural Effects on Saving? Some Cross-Sectional Evidence.” *Quarterly Journal of Economics*, 109(3), 685-699.

Dolado, J.J., García-Peñalosa, C. and de la Rica, S. (2012). "On gender gaps and self-fulfilling expectations:An alternative approach based on of paid-for-training"*Economic Inquiry*.doi: 10.1111/j.1465-7295.2012.00485.x

Dustmann, Ch; Frattini, T and Lanzara, G (*forthcoming*). "Educational Achievement of Second Generation Immigrants: An International Comparison" *Economic Policy.*

Farré, L., & Vella, F. (*forthcoming*). "The Intergenerational Transmission of Gender Role Attitudes and its Implications for Female Labor Force Participation". *Economica*.

Fernández, R. (2011). Does culture matter? (M. O. J. Benhabib, Ed.) *Handbook of Social Economics, 1A*.

Fernández, R. (2007). “Women, Work, and Culture.” *Journal of the European Economic Association*, 5(2-3), 305-332.

Fernández, R., and Fogli, A. (2009). "Culture: An Empirical Investigation of Beliefs, Work, and Fertility". *American Economic Journal: Macroeconomics, 1* (1), 146–177.

Fernández, R. and Fogli, A. (2006). “Fertility: The Role of Culture and Family Experience.” *Journal of the European Economic Association*, 4(2-3), 552-561.

Fernández, R., Fogli, A., and Olivetti, C. (2004). "Mothers and Sons: Preference Formation and Female Labor Force Dynamics". *Quarterly Journal of Economics* , 1249-99.

Furtado, D; Marcem, M and Sevilla-Sanz, A (forthcoming) "Does Culture Affect Divorce Decisions? Evidence from European Immigrants in the US". *Demography.*

Giuliano, P. (2007). “Living Arrangements in Western Europe: Does Cultural Origin Matter?” *Journal of the European Economic Association*, 5(5), 927-952.

Gneezy, U., M. Niederle, and A. Rustichini (2003) “Performance in Competitive Environments: Gender Differences.” *Quarterly Journal of Economics*, 118(3): 1049–74.

González de San Román, A., and S. de la Rica (2012). "Gender Gaps in PISA Test Scores: The Impact of Social Norms and the Mother's Transmission of Role Attitudes". IZA Discussion Papers , 6338.

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). "Culture, Gender and Math"*Science, 320* (5880), 1164-1165.

Guiso, L., Sapienza, P., & Zingales, L. (2006). "Does Culture Affect Economic Outcomes?"*Journal of Economic Perspectives, 20* (2), 23-48.

Hausmann, R., Tyson, L. D., & Zahidi, S. (2008). The Global Gender Gap Report 2008. *World Economic Froum* .

Hedges, L., and Nowell, A. (1995). "Sex differences in mental test scores, variability, and numbers of high-scoring individuals". *Science, 269 (41)*.

Heller, K. A., and J.E. Parsons (1981) “Sex Differences in Teacher’s Evaluative Feedback and Students’ Expectancies for Success in Mathematics.” *Child Development*, 52(3):1015–19.

Hyde, J., Fennema, E., & Lamon, S. J. (1990). "Gender Differences in Mathematics Performance: a Meta-Analysis". *Psychological Bulletin, 107* (2), 139-155.

Leinhardt, G., A. M. Seewald, and M. Engel (1979) “Learning What’s Taught: Sex Differences in Instruction.” *Journal of Educational Psychology*, 71(4): 432–39.

Levine, S.C., M. Vasilyeva, S.F. Lourenco, N.S. Newcombe, and J. Huttenlocher (2005) “Socioeconomic Status Modifies the Sex Difference in Spatial Skill.” *Psychological Science*, 16(11): 841–45.

Mael, F., Alonso, A., Gibson, D., Rogers, K., & Smith, M. (2005). *“Single-sex versus coeducational schooling: a systematic review.”* . US Department of Education.

Muller, Ch. (1998) “Gender Differences in Parental Involvement and Adolescents’ Mathematics Achievement.” *Sociology of Education*, 17(4): 336–56.

O’Brien, L.T., and Ch.S. Crandall (2003) “Stereotype Threat and Arousal: Effects on Women’s Math Performance.” *Personality and Social Psychology Bulletin*, 29(6): 782–89.

OECD (2006). “PISA 2006 Science Competencies for Tomorrow's World”

OECD. (2009). "PISA Data Analysis Manual. Second Edition"

Parsons, J. E., T.F. Adler, and C.M. Kaczala (1982) “Socialization of Achievement Attitudes and Beliefs: Parental Influences.” *Child Development*, 53(2): 310–21.

Penner, A.M. (2008) "Gender differences in extreme mathematical achievement: An international perspective on biological and social factors". Am J Sociology 114:S138–S170.

Spencer, S.J., C.M. Steele, and D.M. Quinn (1999) “Stereotype Threat and Women’s Math Performance.” *Journal of Experimental Social Psychology*, 35(1): 4–28.

Weinberger, CJ (1999) "Mathematical College Majors and the Gender Gap in Wages", *Industrial Relations*, 1999, pp. 407-13.

Wilder, Gita Z. and Kristin Powell (1989). "Sex differences in test performance: a survey of the literature". College Board Report, No. 89-3.

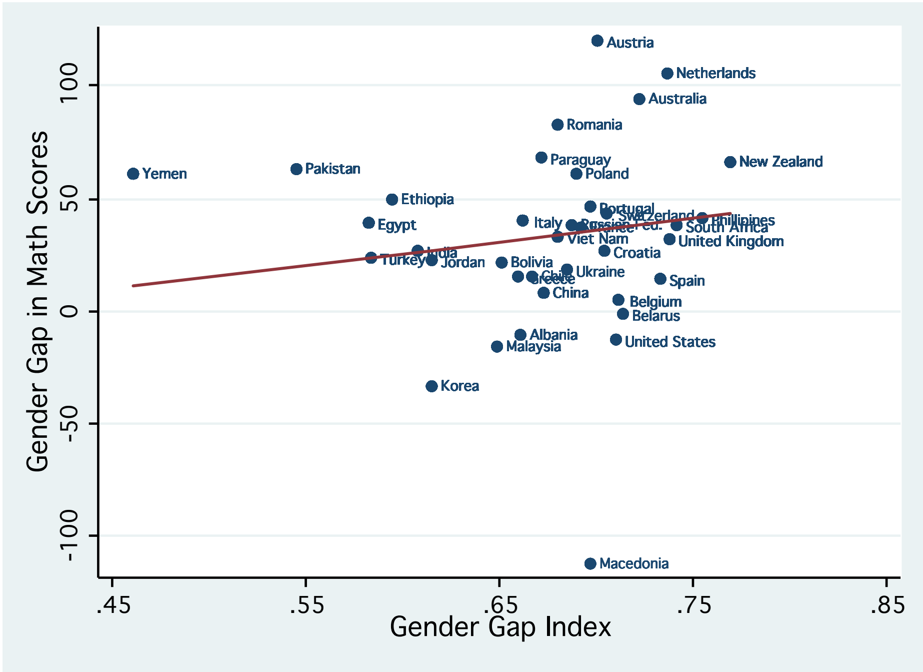
TABLES AND FIGURES

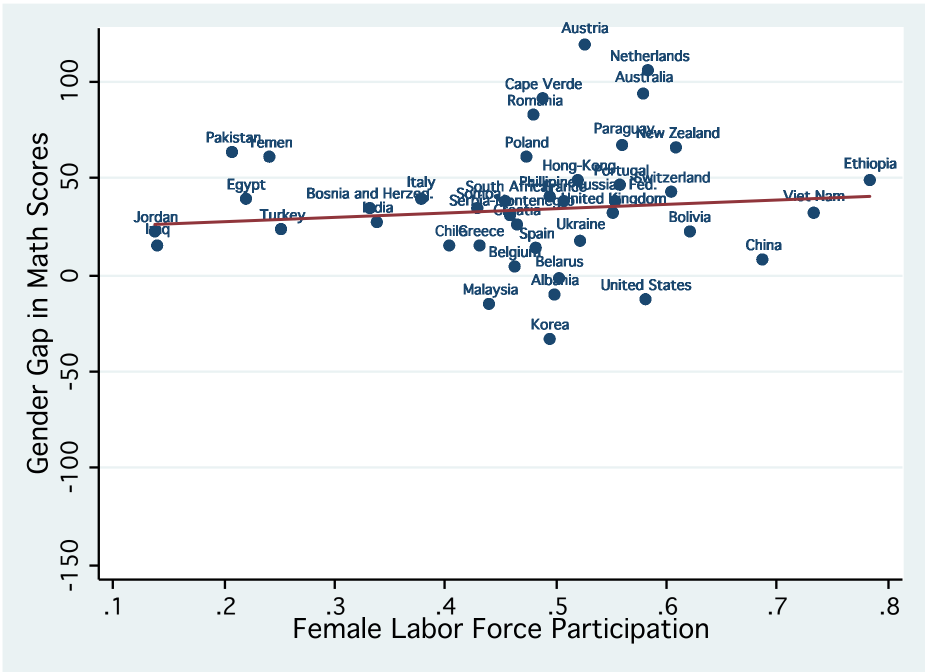
Table 2: Descriptive Statistics by Country of Origin – Gender Gap in Math, Gender Equality and GDP

Country	Math Gender Gap	GGI	GGI PE	FLFP (+15)	FLFP (35-54)	GDP per capita
Macedonia	-124.63	0.70	0.17	.	.	7,148
Korea	-82.08	0.62	0.07	0.49	0.62	23,928
Belarus	-65.12	0.71	0.16	0.50	0.85	11,094
Croatia	-46.74	0.70	0.20	0.46	0.75	14,597
United States	-41.97	0.71	0.12	0.58	0.76	41,841
Belgium	-39.43	0.71	0.24	0.46	0.75	34,173
Albania	-37.44	0.66	0.04	0.50	0.67	5,836
Ukraine	-35.37	0.68	0.06	0.52	0.80	6,222
Spain	-35.32	0.73	0.40	0.48	0.69	27,815
China	-33.62	0.67	0.13	0.69	0.85	5,872
Serbia-Montenegr	-27.74	.	.	0.46	0.73	14,515
Turkey	-25.81	0.58	0.06	0.25	0.27	9,821
Bosnia and Herze	-23.74	.	.	0.33	0.50	6,049
Bolivia	-22.47	0.65	0.12	0.62	0.75	3,610
Malaysia	-22.16	0.65	0.06	0.44	0.51	11,013
India	-20.72	0.61	0.25	0.34	0.44	2,897
Switzerland	-19.91	0.71	0.24	0.60	0.82	38,599
Jordan	-17.69	0.61	0.06	0.14	0.15	4,520
South Africa	-16.76	0.74	0.39	0.46	0.63	7,274
Iraq	-16.01	.	.	0.14	0.19	4,233
United Kingdom	-15.96	0.74	0.29	0.55	0.79	33,796
Poland	-14.87	0.69	0.14	0.47	0.77	14,929
Greece	-14.80	0.66	0.07	0.43	0.67	26,515
France	-14.24	0.69	0.20	0.51	0.83	30,905
Russian Fed.	-13.23	0.69	0.06	0.55	0.88	13,501
Italy	-11.11	0.66	0.12	0.38	0.63	28,542
Phillipines	-10.49	0.75	0.28	0.50	0.62	2,651
Somoa	-9.82	.	.	0.43	0.56	6,605
Egypt	-8.13	0.58	0.02	0.22	0.26	4,619
Portugal	-5.79	0.70	0.15	0.56	0.80	20,001
Pakistan	-5.31	0.54	0.15	0.21	0.25	2,233
Viet Nam	-1.02	0.68	0.12	0.73	0.89	2,595
Hong-Kong	10.75	.	.	0.52	0.65	34,688
Yemen	11.15	0.46	0.01	0.24	0.30	2,411
New Zealand	17.60	0.77	0.36	0.61	0.79	27,663
Ethiopia	17.61	0.59	0.11	0.78	0.85	597
Romania	19.22	0.68	0.06	0.48	0.70	9,044
Chile	20.97	0.67	0.18	0.40	0.54	11,613
Cape Verde	33.34	.	.	0.49	0.58	3,326
Paraguay	45.85	0.67	0.12	0.56	0.66	3,628
Australia	49.49	0.72	0.18	0.58	0.75	39,837
Austria	54.26	0.70	0.28	0.53	0.82	36,522
Netherlands	66.27	0.74	0.33	0.58	0.79	39,429
Mean	-14.59	0.68	0.17	0.48	0.66	17,096
Sd	20.60	0.06	0.11	0.14	0.20	11,691

Note: Countries of origin are ordered by Gender Gap in Math. It was obtained from estimating a linear regression using the Plausible Values provided by the PISA data sets as LHS variable and a female dummy as RHS (we estimated one regression for each PV and present the average of the 5 coefficients estimated). See Appendix Table A.3 for details about the descriptive variables. We display means estimated using our sample of second generation immigrants taken from 2003, 2006 and 2009 PISA data sets.

Figure 1. Gender Gap in MathScores of Second-generation Immigrants and Gender Equality in Countries of Origin





Notes: The average Gender Gap in math scores among second-generation immigrants was obtained from estimating a linear regression using the Plausible Values provided by the PISA data sets as LHS variable and a female dummy as RHS variable. We estimated one regression for each PV for each country and present the average of the 5 coefficients estimated. We use individuals whose both parents were born in a foreign country from the 2003, 2006 and 2009 PISA datasets. Definitions and data sources of GGI and FLFP are presented in Appendix Table A 1.

Table 3: Gender Gap in Test Scores

	All Countries				Countries included in our sample			
	All individuals		Second-generation immigrants		All individuals		Second-generation immigrants	
Math Scores								
Boys	460.29	[105.80]	466.41	[94.59]	478.63	[112.97]	494.32	[104.07]
Girls	447.40	[100.97]	457.10	[92.80]	462.55	[109.22]	475.33	[97.09]
Gender Gap	**-12.89**		**-9.31**		**-16.08**		**-18.99**	
Reading Scores								
Boys	438.90	[104.04]	445.54	[102.26]	445.97	[118.96]	462.22	[106.78]
Girls	470.00	[97.95]	481.48	[95.96]	482.36	[109.87]	491.62	[102.51]
Gender Gap	**31.10**		**35.94**		**36.39**		**29.40**	

Notes: Author's calculations based on 2003, 2006 and 2009 PISA datasets. Mean and Standard Deviation in bracket. See Data section for details about which countries are included in our sample.

Table 4: Replication of Guiso *et al.* (2008) at student level

	Math Scores	**Reading Scores**		
	Gender-Equality*Female	Gender-Equality*Female	Obs.	R-squared
GGI	162.67***	47.28***	891,461	0.29
	[7.90]	[7.99]		
PEI	23.12***	30.66***	891,461	0.29
	[3.06]	[3.23]		
FLFP	43.49***	10.33***	924,782	0.29
	[3.27]	[3.24]		
FLFP (35-54)	43.56***	11.74***	924,782	0.29
	[2.01]	[2.04]		

Notes: Results from estimating the relationship between math (reading) scores and countries' gender equality using 2003, 2006 and 2009 PISA datasets. The specification include year and countries fixed effects, the GDP per capita (in logarithms) interacted by the female dummy and individual controls (a dummy for any students who are in a grade different from the modal one in the country, as in Guiso *et al.* 2008). Standard Errors adjusted using the BRR methodology. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 5: Math Scores of Second-Generation Immigrants and Gender Equality in the Country of Origin

	Female	Gender-Equality* Female	Year FE	Country of Origin FE	Host Country FE	GDPpc* Female	Host Countries Charact.	Obs.	R-squared
A) GGI	-77.24***	94.51**	X	X				9,179	0.32
	[28.32]	[41.97]							
	-75.81***	92.69**	X	X	X			9,179	0.33
	[28.12]	[41.71]							
	-55.01	130.90***	X	X	X	X		9,179	0.33
	[33.58]	[49.95]							
	-55.13	134.37***	X	X		X	X	9,179	0.33
	[33.71]	[50.20]							
B) PEI	-22.56***	52.88**	X	X				9,179	0.32
	[5.32]	[26.07]							
	-22.26***	52.32**	X	X	X			9,179	0.33
	[5.29]	[26.05]							
	9.64	65.17**	X	X	X	X		9,179	0.33
	[32.41]	[27.68]							
	11.28	66.52**	X	X		X	X	9,179	0.33
	[32.47]	[27.79]							
C) FLFP	-35.47***	43.62**	X	X				11,166	0.29
	[8.59]	[18.15]							
	-34.79***	42.70**	X	X	X			11,166	0.30
	[8.50]	[18.01]							
	-29.42	42.62**	X	X	X	X		11,166	0.30
	[29.60]	[17.91]							
	-28.34	41.79**	X	X		X	X	11,166	0.30
	[29.68]	[18.04]							
D) FLFP (35-54)	-28.99***	21.87*	X	X				11,166	0.30
	[8.03]	[11.79]							
	-28.44***	21.39*	X	X	X			11,166	0.30
	[7.92]	[11.65]							
	-11.60	23.47*	X	X	X	X		11,166	0.30
	[29.98]	[12.45]							
	-10.64	23.54*	X	X		X	X	11,166	0.30
	[30.02]	[12.59]							

Notes: Results from estimating the equation 1, 2, 3 and 4 using the sample of immigrants of second generation from 2003, 2006 and 2009 PISA datasets. When parents come from different countries we assign the mother's country of origin. Host Countries Variables include the native-immigrant gap in tests scores in host countries and the host countries' GDP per capita in logarithms. Standard Errors adjusted using the BRR methodology. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 6: Reading Scores of Second-Generation Immigrants and Gender Equality in the Country of Origin

	Female	Gender-Equality* Female	Year FE	Country of Origin FE	Host Country FE	GDPpc* Female	Host Countries Variables	Obs.	R-squared
B) GGI	-35.81	102.39**	X	X				9,179	0.32
	[30.85]	[45.27]							
	-34.96	101.25**	X	X	X			9,179	0.33
	[30.34]	[44.61]							
	-28.3	113.48**	X	X	X	X		9,179	0.33
	[37.87]	[52.35]							
	-28.71	116.16**	X	X		X	X	9,179	0.33
	[38.08]	[52.97]							
C) PEI	22.01***	65.46**	X	X				9,179	0.32
	[6.03]	[27.33]							
	21.95***	66.27**	X	X	X			9,179	0.33
	[5.95]	[27.15]							
	30.92	69.88**	X	X	X	X		9,179	0.33
	[37.02]	[28.08]							
	31.82	71.16**	X	X		X	X	9,179	0.33
	[37.03]	[28.32]							
D) FLFP	22.60**	22.85	X	X				11,166	0.29
	[9.80]	[20.27]							
	23.00**	22.20	X	X	X			11,166	0.30
	[9.64]	[20.04]							
	4.73	22.48	X	X	X	X		11,166	0.30
	[33.19]	[19.92]							
	4.95	22.94	X	X		X	X	11,166	0.30
	[33.27]	[20.08]							
E) FLFP (35-54)	20.70**	19.51	X	X				11,166	0.30
	[8.97]	[12.95]							
	21.11**	19.00	X	X	X			11,166	0.30
	[8.78]	[12.72]							
	13.18	18.03	X	X	X	X		11,166	0.30
	[33.96]	[13.81]							
	13.58	18.35	X	X		X	X	11,166	0.30
	[33.97]	[13.95]							

Notes: Results from estimating the equation 1, 2, 3 and 4 using the sample of immigrants of second generation from 2003, 2006 and 2009 PISA datasets. When parents come from different countries we assign the mother's country of origin. Host Countries Variables include the native-immigrant gap in tests scores in host countries and the host countries' GDP per capita in logarithms. Standard Errors adjusted using the BRR methodology. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 7: Robustness Checks – Math Scores

	GGI	PEI	FLFP	FLFP (35-54)
Baseline	130.90***	65.17**	41.79**	23.47*
	[49.95]	[27.68]	[18.04]	[12.45]
Observations	9,179	9,179	11,166	11,166
R-squared	0.33	0.33	0.30	0.30
A) Without Portuguese				
Culture*female	131.16***	68.06**	41.96**	22.83*
	[50.03]	[27.77]	[18.11]	[12.51]
Observations	7,351	7,351	9,338	9,338
R-squared	0.33	0.33	0.31	0.31
B) Without Switzerland				
Culture*female	175.45**	77.47*	56.53**	27.36*
	[83.88]	[46.62]	[22.28]	[14.05]
Observations	7,065	7,065	9,004	9,004
R-squared	0.27	0.27	0.22	0.22
C) Using only the 2009 PISA dataset				
Culture*female	178.55**	76.45*	43.98*	20.34
	[80.29]	[44.04]	[25.04]	[17.02]
Observations	4,411	4,411	5,582	5,582
R-squared	0.37	0.37	0.34	0.34
D) Giving priority to the father country of origin				
Culture*female	118.45**	49.50*	40.47**	24.22**
	[50.74]	[27.59]	[17.80]	[12.33]
Observations	9,171	9,171	11,174	11,174
R-squared	0.33	0.33	0.30	0.30
E) Keeping only those whose parents come from the same country				
Culture*female	146.38**	58.45*	40.99**	25.00*
	[60.73]	[34.48]	[20.16]	[13.45]
Observations	6,701	6,701	8,515	8,515
R-squared	0.37	0.37	0.33	0.33
F) Using Cultural Proxies from 1990 (#)				
Culture*female	-.-	86.12*	34.14*	13.58
		[45.47]	[20.73]	[14.13]
Observations		8,339	11,093	11,093
R-squared		0.32	0.30	0.30
Year fixed effects	YES	YES	YES	YES
Host Country fixed effects	YES	YES	YES	YES
GDP per capita and GDP per capita*female	YES	YES	YES	YES

Notes: Results from estimating the equation (3) using different sample selection. Standard Errors adjusted using the BRR methodology. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.
(#) As measure of political empowerment we use the proportions of seats held by women in national parliaments in 1990 (source: World Bank Statistics).

Table 8: Mechanisms

	GGI	PEI	FLFP	FLFP (35_54)
Baseline	130.90***	65.17**	41.79**	23.47*
	[49.95]	[27.68]	[18.04]	[12.45]
+ Individual Characteristics (Age, Diff. grade)	132.20**	65.28**	44.03**	24.67*
	[51.56]	[27.99]	[18.47]	[12.79]
+Mom Education (College Mom=1)	148.94***	71.44**	48.88**	30.16**
	[54.73]	[29.04]	[19.95]	[13.61]
+Dad Education (College Dad=1)	167.82***	77.68**	51.75**	31.62**
	[58.12]	[30.56]	[20.69]	[13.97]
+Mom Work (Mom working=1)	176.83***	81.28***	53.05**	32.73**
	[60.12]	[30.72]	[20.99]	[14.36]
+Amount of books at home (PISA index)	171.36***	73.66**	47.65**	30.98**
	[60.08]	[30.02]	[21.23]	[14.41]
+Language at home (foreign language=1)	131.17**	55.71*	30.51	16.98
	[60.22]	[33.24]	[21.49]	[15.13]
+Proportion of girls at school (PISA variable)	126.18**	52.23	25.69	13.1
	[61.54]	[34.24]	[22.02]	[15.49]
+Type of school (Private=1)	127.20**	56.48	25.98	12.96
	[63.27]	[34.80]	[22.56]	[15.80]
+Place of residence (City or Metropoli=1)	128.09**	56.54	26.22	13.2
	[63.28]	[34.80]	[22.56]	[15.84]

Notes: Results from estimating equation (3), adding sequentially individual controls. Each control variables is interacted with the female dummy. Standard Errors adjusted using the BRR methodology. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

APPENDIX

Table A 1: Data Sources and Definition of Variables

Variable	Definition	Source
Measures of Gender Equality		
Gender Gap Index (GGI)	Synthesizes the position of women by considering economic opportunities, economic participation, educational attainment, political achievements, health and well-being. Larger values point to a better position of women in society. We use the average between the 2006 and 2009 indexes.	World Economic Forum, 2006 and 2009 Reports.
Political Empowerment Index (PEI)	Measure women's political participation based on three components: (1) the ratio women to men with seats in parliament; (2) the ratio of women to men in ministerial level and (3) the ratio of the number of years with a women as head of state to the years with a man. Larger values point to a better position of women in society. We use the average between the 2006 and 2009 indexes.	World Economic Forum, 2006 and 2009 Reports.
FLFP	Female Labor Force Participation, from 15 years old. We average the 2003, 2006 and 2009 rates.	International Labour Organization (ILO).
FLFP 35-54	Female Labor Force Participation from 35 to 54 years old. We average the 2003, 2006 and 2009 rates.	
Macro Variables		
GDP per capita	Gross Domestic Product per capita in real terms deflated with Laspeyres price index. We average the 2003, 2006 and 2009 values.	Heston, A., Summers, R. and Aten, B, Penn, World Table Version 7.0, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, May 2011.
Control Variables		
Age	Years and months	PISA data sets – student's questionnaire
Diff.Grade	Dummy equal to 1 if current grade is different from the modal grade at the children age in the host country and 0 otherwise	idem
Mom (Dad) College	Dummy equal to 1 if mother (father) has a college degree and 0 otherwise	idem
Mom Work	Dummy equal to 1 if mother is working and 0 otherwise	idem
Books	Index of amount of books at home: 1=0-10; 2=11-100; 3=101-500; 4=more than 500.	idem
Language at home	1 if language spoken at home is different from the test language	idem
Percentage of girls	PISA index of the proportion of girls enrolled at school derived from school principals' responses regarding the number of girls divided by the total of girls and boys at a school.	PISA data sets – school's questionnaire
Private school	Dummy equal to 1 if school is private and 0 otherwise	idem
Place of residence	Dummyequal to 1 if the school is in a large city or metropolis and 0otherwise	idem

Table A 2: Sample Size by Country of Origin and Destiny

	ARG	AUS	AUT	BEL	CHE	DNK	ISR	LIE	LUX	LVA	NZL	QAT	Total
Albania					94								**94**
Australia											32		**32**
Austria					36			14					**50**
Belgium									87				**87**
Bolivia	41												**41**
Bosnia and Herz.			163										**163**
Belarus										39			**39**
Cape Verde									45				**45**
Chile	11												**11**
China		240									91		**331**
Croatia			49										**49**
Ethiopia							105						**105**
France				99	162		35		161				**457**
Greece		46											**46**
Hong-Kong		48											**48**
India		125											**125**
Iraq						40							**40**
Italy		69			641			12	230				**952**
Jordan												62	**62**
Korea		35											**35**
Malaysia		43											**43**
Netherlands		15		40									**55**
New Zealand		319											**319**
Pakistan						92							**92**
Paraguay	26												**26**
Phillipines		168											**168**
Poland			35										**35**
Portugal					492				1,336				**1,828**
Romania			34										**34**
Russian Fed.							231			97			**328**
Viet Nam		107											**107**
South Africa		66											**66**
Spain					225								**225**
Switzerland								23					**23**
Turkey			393	347	456	428							**1,624**
Ukraine										34			**34**
Macedonia			11										**11**
Egypt												286	**286**
United Kingdom		843							19		158		**1,020**
United States		38					38						**76**
Somoa											258		**258**
Yemen												263	**263**
Serbia-Monten			168		1,016	108			152				**1,444**
Total	**78**	**2,162**	**853**	**486**	**3,122**	**668**	**409**	**49**	**2,030**	**170**	**539**	**611**	**11,177**

Notes: Final sampleof second-generation immigrants from 2003, 2006 and 2009 PISA datasets. ARG=Argentina, AUS=Australia, AUT=Austria, BEL=Belgium, CHE=Switzerland, DNK=Denmark, ISR=Israel, LIE=Liechtenstein, LUX=Luxembourg, LVA=Latvia, NZL= New Zealand, QAT=Qatar.

Table A 3. Robustness Checks – Reading Scores

	GGI	PEI	FLFP	FLFP (35-54)
Baseline	113.48**	69.88**	22.48	18.03
	[52.35]	[28.08]	[19.92]	[13.81]
Observations	9,179	9,179	11,166	11,166
R-squared	0.33	0.33	0.30	0.30
A) Without Portuguese				
Culture*female	113.13**	73.85***	20.93	16.87
	[52.45]	[28.23]	[20.17]	[13.96]
Observations	7,351	7,351	9,338	9,338
R-squared	0.33	0.33	0.31	0.31
B) Without Switzerland				
Culture*female	125.48**	84.38***	16.88	15.99
	[58.25]	[32.38]	[21.98]	[15.75]
Observations	7,065	7,065	9,004	9,004
R-squared	0.27	0.27	0.22	0.22
C) Using only the 2009 PISA dataset				
Culture*female	84.08	45.63	4.88	1.41
	[84.26]	[43.05]	[27.54]	[18.18]
Observations	4,411	4,411	5,582	5,582
R-squared	0.37	0.37	0.34	0.34
D) Giving priority to the father country of origin				
Culture*female	113.43**	56.97**	26.00	20.83
	[54.18]	[27.74]	[20.43]	[13.95]
Observations	9,171	9,171	11,174	11,174
R-squared	0.33	0.33	0.30	0.30
E) Keeping only those whose parents come from the same country				
Culture*female	110.58*	43.21	22.15	21.23
	[63.28]	[32.95]	[22.69]	[14.88]
Observations	6,701	6,701	8,515	8,515
R-squared	0.37	0.37	0.33	0.33
F) Using Cultural Proxies from 1990 (#)				
Culture*female	-.-	79.26*	11.58	8.71
		[45.37]	[22.52]	[14.79]
Observations		8,339	11,093	11,093
R-squared		0.32	0.30	0.30
Year fixed effects	YES	YES	YES	YES
Host Country fixed effects	YES	YES	YES	YES
GDP per capita and GDP per capita*female	YES	YES	YES	YES

Notes: Results from estimating the equation (3) using different sample selection. Standard Errors adjusted using the BRR methodology. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.
(#) As measure of political empowerment we use the proportions of seats held by women in national parliaments in 1990 (source: World Bank Statistics).

Table A 4: Descriptive Statistics by Country of Origin – Socio-economic Characteristics

Country	Math Gender Gap	Individual Charact.			Family Background					School Charact.		Place of Residence
		Boys	Age	Diff. Grade	Mom College	Dad College	Mom work	Books	Lang at home	% girls in school	Private School	City o Metropoli
Macedonia	-124.63	0.60	15.71	0.84	0.00	0.17	0.89	1.77	1.00	0.52	0.20	0.32
Korea	-82.08	0.54	15.74	0.24	0.62	0.77	0.85	3.00	0.56	0.40	0.27	0.26
Belarus	-65.12	0.47	15.78	0.17	0.19	0.29	0.98	2.33	0.31	0.51	0.13	0.41
Croatia	-46.74	0.44	15.84	0.55	0.14	0.19	1.00	1.99	0.66	0.48	0.15	0.39
United States	-41.97	0.41	15.68	0.23	0.60	0.73	0.82	3.01	0.31	0.46	0.18	0.41
Belgium	-39.43	0.40	15.81	0.45	0.49	0.43	0.90	2.82	0.36	0.38	0.08	0.38
Albania	-37.44	0.62	15.73	0.45	0.06	0.20	0.79	1.80	0.60	0.50	0.10	0.46
Ukraine	-35.37	0.45	15.76	0.31	0.31	0.26	0.95	2.20	0.22	0.47	0.16	0.41
Spain	-35.32	0.46	15.83	0.44	0.13	0.09	0.90	2.30	0.50	0.48	0.22	0.38
China	-33.62	0.58	15.79	0.23	0.32	0.40	0.92	2.46	0.62	0.49	0.29	0.29
Serbia-Mont.	-27.74	0.50	15.82	0.41	0.08	0.11	0.86	1.76	0.73	0.49	0.18	0.36
Turkey	-25.81	0.48	15.81	0.55	0.03	0.06	0.70	1.89	0.72	0.49	0.18	0.32
Bosnia-Herz.	-23.74	0.48	15.81	0.54	0.04	0.03	0.93	1.75	0.82	0.55	0.18	0.45
Bolivia	-22.47	0.32	15.80	0.34	0.03	0.00	0.43	1.68	0.00	0.52	0.16	0.28
Malaysia	-22.16	0.56	15.82	0.42	0.38	0.56	0.91	2.41	0.06	0.49	0.20	0.29
India	-20.72	0.48	15.77	0.31	0.49	0.49	0.94	2.44	0.12	0.48	0.24	0.32
Switzerland	-19.91	0.47	15.84	0.26	0.19	0.36	0.83	2.59	0.00	0.43	0.22	0.36
Jordan	-17.69	0.52	15.80	0.48	0.47	0.65	0.54	2.40	0.24	0.53	0.11	0.26
South Africa	-16.76	0.56	15.82	0.25	0.54	0.61	0.94	2.91	0.08	0.44	0.24	0.35
Iraq	-16.01	0.62	15.65	0.21	0.22	0.24	0.78	2.05	0.64	0.50	0.17	0.32
UK	-15.96	0.50	15.76	0.33	0.34	0.35	0.95	2.77	0.01	0.48	0.17	0.32
Poland	-14.87	0.47	15.87	0.40	0.20	0.25	1.00	2.54	0.62	0.49	0.38	0.16
Greece	-14.80	0.55	15.73	0.11	0.07	0.13	0.91	2.41	0.39	0.54	0.30	0.37
France	-14.24	0.48	15.81	0.53	0.35	0.43	0.90	2.69	0.14	0.46	0.18	0.34
Russian Fed.	-13.23	0.50	15.69	0.18	0.52	0.46	0.96	2.28	0.31	0.51	0.23	0.50
Italy	-11.11	0.55	15.82	0.45	0.11	0.13	0.87	2.17	0.39	0.48	0.18	0.35
Phillipines	-10.49	0.40	15.76	0.19	0.56	0.49	0.96	2.38	0.06	0.48	0.21	0.40
Somoa	-9.82	0.49	15.78	0.11	0.11	0.12	0.94	2.12	0.36	0.52	0.16	0.27
Egypt	-8.13	0.49	15.75	0.38	0.66	0.77	0.66	2.48	0.28	0.44	0.17	0.40
Portugal	-5.79	0.48	15.81	0.47	0.05	0.03	0.92	2.07	0.61	0.54	0.15	0.29
Pakistan	-5.31	0.43	15.75	0.19	0.11	0.18	0.76	1.92	0.52	0.52	0.17	0.35
Viet Nam	-1.02	0.41	15.75	0.31	0.09	0.18	0.92	2.20	0.62	0.52	0.12	0.39
Hong-Kong	10.75	0.47	15.73	0.23	0.34	0.50	0.92	2.56	0.51	0.43	0.26	0.36
Yemen	11.15	0.54	15.72	0.33	0.07	0.19	0.18	2.17	0.20	0.48	0.11	0.36
New Zealand	17.60	0.45	15.77	0.35	0.26	0.28	0.91	2.49	0.01	0.47	0.15	0.39
Ethiopia	17.61	0.53	15.71	0.18	0.03	0.06	0.76	1.75	0.18	0.56	0.25	0.44
Romania	19.22	0.55	15.76	0.69	0.10	0.11	0.88	2.29	0.56	0.56	0.16	0.34
Chile	20.97	0.74	15.75	0.45	0.19	0.23	0.69	1.79	0.00	0.48	0.42	0.47
Cape Verde	33.34	0.40	15.76	0.33	0.06	0.12	0.96	2.00	0.89	0.64	0.44	0.34
Paraguay	45.85	0.31	15.79	0.30	0.00	0.00	0.59	1.79	0.27	0.50	0.07	0.23
Australia	49.49	0.43	15.71	0.01	0.33	0.32	0.95	2.90	0.02	0.50	0.09	0.32
Austria	54.26	0.46	15.87	0.38	0.27	0.29	0.92	2.57	0.09	0.53	0.22	0.21
Netherlands	66.27	0.56	15.85	0.28	0.19	0.20	0.89	2.70	0.12	0.49	0.18	0.47
Mean	-14.59	0.49	15.78	0.37	0.21	0.23	0.85	2.23	0.37	0.49	0.19	0.34
Sd	20.60	0.50	0.29	0.48	0.40	0.42	0.36	0.82	0.48	0.20	0.39	0.47

Note: Countries of origin are ordered by Gender Gap in Math. It was obtained from estimating a linear regression using the Plausible Values provided by the PISA data sets as LHS variable and a female dummy as RHS (we estimated one regression for each PV and present the average of the 5 coefficients estimated). See Appendix Table A 1 for details about the descriptive variables. We display means estimated using our sample of second generation immigrants taken from 2003, 2006 and 2009 PISA data sets.

5. General Conclusions and Further Extensions

The three essays that compose this dissertation represent new contributions to our understanding of topics of high relevance on gender equality and education. What is the effect of providing full-time public childcare to children of 3 years old on maternal employment? What is the effect of such a policy on children' cognitive outcomes? What is the role of culture in explaining the gender differences in tests scores at 15 years old?

The results from the first essay suggest that even under a context of low female labor force participation and meager supply of childcare spaces, the provision of childcare alone will not have large effects on maternal employment. Nonetheless, the effects on maternal employment persist over time as the child aged and are driven by mothers with a high-school degree and older mothers. This suggests that the policy reduced the depreciation of human capital accumulated in school and in former jobs, encouraging the accumulation of new human capital acquired on the job. Despite we check whether the policy had an effect on fertility and rule out this possibility, a possible extension of this paper may be to explore this issue more carefully. In addition, the fact we find a modest effect of a very large expansion in childcare on maternal employment even in a context of very low participation of women in the labor market, leaves open several questions for future research. For instance, to what extent this modest effect is explaining by the lack of suitable opportunities in the labor market (as part-time, more flexible jobs)? What is the role of culture, such as the prevalence of a male bread-winner tradition? These questions could be addressed by future extensions of this research.

A related important question, explored in the second essay, is whether universal childcare has short and long-term beneficial or detrimental effects on the cognitive or non-cognitive development of children relative to other forms of early childhood care, such as parental or relative care. The effects of childcare on child development depend on the quality of both public childcare and the counterfactual care mode. The second essay provides quasi-experimental evidence for the impact of shifting hours of care provided by mother to hours of care provided by high-quality public preschools. Results suggest that high-quality public childcare does not only neutralize potentially negative effects of maternal employment, but has even positive effects on children's cognitive development, at least among children with less educated parents and for girls. Given that this essay measures the effect on cognitive outcomes at the end of mandatory school, future extensions could be to explore whether these effects persist in the long run or fade away. On the other hand, since the effects seem to be driven by more disadvantaged children, another interesting extension could be to explore the effects of such a policy on risky behaviors, like early pregnancy, use of drugs, and crime. Finally, the fact that an earlier entry to the educational system seems to had a greater effect on girls than on boys (as showed by this research but also by other previous studies) leaves open the question about whether an earlier entry to educational system may offset other factors affecting the educational gender gap, such as the studied in the third essay of this dissertation.

The third essay use second-generation of immigrants to disentangle the effects of markets and institutions from the effects of culture in determining gender differences in educational attainment. Given that second-generation immigrants grow up under the host country's laws and institutions, but they are likely affected by the culture from their parents, we interpret the effect of parents' home country measures of gender equality on their educational attainment as evidence of the role of culture.We find that the higher degree of gender equality in the country of origin improves the performance of second-generation immigrant girls relative to boys. We rule out alternative explanations for the result such as immigrants from countries of origin with more gender-equal cultures living in host countries with lower gender gaps in math scores. Despite we analyse the channels through which culture is transmitted and find that language spoken at home and parental education are important, future extensions of this research should be also to explore the role of networks in cultural transmission. This could be addressed, for instance, by including a measure of the concentration of immigrants of different origins within each host country. Alternatively, the role of networks in cultural transmission may be explored taking into account the percentage of immigrants within the same school, an information available on 2003 and 2009 PISA datasets.