Three Essays on the Determinants of Labour Market Outcomes

Ainhoa Osés Arranz

2024

PhD Thesis

Supervised by: Sara de la Rica and Lucía Gorjón

To my parents and Esti

Acknowledgements

First and foremost, I am extremely grateful to Sara de la Rica and Lucía Gorjón for giving me the opportunity to embark on this journey and for their continued support throughout. Sara's dedication and commitment to bridging academia and policy has been an enormous source of inspiration to me. Lucía has always been incredibly supportive and generous with me, making easy what seemed tough. I feel very lucky to have been supervised by them.

I am immensely grateful to Eddie Casey for his trust in me and his unwavering support during my time in IFAC and beyond. I will always carry with me everything I have learned from him: from his technical excellence to his outstanding personal values. I would also like to express my sincere thanks to Enrique M. Quilis for his generosity and for instilling in me his passion for research. I am very grateful to Elish Kelly for her support and guidance. I am also thankful to Juanan and Txoni for laying the groundwork to academically push me forward.

I carried out this thesis while working at ISEAK. My colleagues Odra, Gonzalo, Imanol, Lucía, David, Iker, Pablo, Ainhoa and Alejandra have been an essential part of this process. Beyond the academic work together, the countless moments of laughter with them have always uplifted me along the way. A special mention goes to David Martínez de Lafuente for his guidance throughout the third chapter of this thesis and his exceptional contribution to it. As part of my PhD, I undertook a research stay at the OECD in Paris. I am deeply grateful to Fabio Manca for hosting me in the Directorate for Employment, Labour and Social Affairs.

I would like to express my heartfelt thanks to my friends from home who have always been there for me, no matter the distance. I feel incredibly lucky to be surrounded by you: knowing that I can always count on you is one of the most comforting feelings. Away from home, I met people that I can proudly call friends. A special mention goes to my friends from AIReF (Laura, Rocío, Cruz and Gonzalo) for their true friendship. I am also grateful to my friends from the master's studies in Carlos III University (especially Adri and Sami), and, of course, to my *pisonys*. I would also like to thank the *cairde* that accompanied me for over two years in Dublin (including Mery, Nere, Dana, Max, Ale, Celia, Ire, Rami, Achim and Tere), as well as my colleagues from IFAC (Killian, Kevin, Niall and Eddie) and the ESRI. I am extremely thankful to Andoni Montes for always being there for me, and to Odra Quesada for her unwavering support. I am also thankful to Alaitz Ayarza for our long-standing friendship. I would also like to extend my thanks to my *Parisians*, Sandra, Lucía, Vanesa and Elena; and to Edu Polo, for his support since our university days.

My most special thanks go to my family, especially to my parents and my sister. Words are not enough to express how lucky I am to have you, supporting me unconditionally and being the best example of teamwork. I thank my parents for giving me the best education and always believing in me. And I thank my sister for being my life companion, my confident, the person with whom I burst into laughter and…for always, always being by my side. I am deeply grateful to Josu for always cheering me up and for always being interested in my research. I would also like to thank *gezin* for their support over the past few years. Finally, I want to express my deepest gratitude to Bram for being one of my strongest pillars. Not only did you believe in me, but you also made me believe in myself. Your unconditional love and support cross borders.

Contents

List of Tables

List of Figures

General Introduction

Over recent decades, labour markets have experienced unprecedented changes. First, the rise of automation and new technologies has permeated and transformed production processes, requiring employees to rapidly adapt their skills to this new reality. The acquisition of these skills has not only impacted the current workforce but has also necessitated profound changes in educational curricula, shaping the trajectory of the future workforce. The integration of digital technology in schools has been transversal to advanced economies, aiming not only to facilitate students' learning processes, but also to provide them with the skills that they may require upon entry to the labour market. Second, global labour markets have experienced sharp rises in female labour force participation in the last few decades. This trend arises as a consequence of changes in social norms, resulting in women outnumbering men in higher education in several countries, including Spain.

Despite the potential economic and social benefits of these paradigm shifts in labour markets, some challenges remain. Job polarisation continues to be a pervasive feature of advanced economies, with direct implications for the job composition of labour markets. The large drops of employment in middle-wage occupations have entailed sharp rises in high- as well as low-wage occupations, accentuating inequality across occupations. In addition, despite the large rise in female participation in the labour market, occupations remain largely segregated by gender, with economic and social implications from an efficiency and equity point of view. Even today, occupational segregation is the most important factor in understanding the gender gap in hourly wages, alongside other detrimental implications, including potential constraints on productivity and economic growth.

In effectively tackling the implications of this rapidly changing landscape, it is important to identify the drivers that may lead to diverging labour market outcomes. The aim of this thesis is to offer a holistic approach on the different elements that affect those outcomes. To do so, each chapter focuses on different stages of individuals' educational and labour market trajectories. The first chapter commences with the future workforce, that is, the youth prior to labour market entry. The chapter focuses on the use of educational technology, a tool that has the potential to equip individuals for the challenges of the future labour market, but that, if not used effectively, can hinder individuals' learning processes and consequently their future labour market outcomes. The second chapter focuses on the subsequent phase of individuals' trajectories, starting with their initial labour market outcomes before extending the analysis over the medium and longer term. The third chapter offers a broader picture by analysing the overall workforce, regardless of their seniority, to then focus on pre-university individuals for further empirical analyses. Overall, the thesis explores some of the drivers that may differentially affect individuals throughout their life cycle.

In particular, Chapter 1 explores the non-linear association between ICT use at school and student performance in a number of OECD countries and assesses the causal impact of ICT overuse on student achievement. The results of this study confirm the existence of a hill-shaped relationship between the frequency of ICT use at school and students' mathematics performance across 22 OECD countries. The study reveals that even in the most advanced countries in terms of ICT integration in schools, individuals who are classified as very intensive users are, on average, at a significant disadvantage in terms of their mathematics performance. Conversely, those categorised as low and medium ICT users exhibit better results compared to those classified as very low users. In a causal framework, the results indicate that the overuse of ICT leads to underperformance in mathematics, of the order of more than half academic year in countries like Estonia, Finland, or Spain. The results of this study contribute to expanding policymakers' and educators' earlier knowledge on the way in which the pervasiveness of technology in classrooms affects students' performance and, potentially, their future labour market outcomes.

Chapter 2 analyses the wage penalty arising from a bad start in the Spanish labour market, using large administrative microdata from the Social Security. Bad jobs are defined according to the European Social Charter: individuals whose annual earnings are below 60% of the Spanish average wage in a given year are considered to be in bad jobs. The results provide evidence of a scarring effect of bad jobs in workers' future trajectories: individuals in bad jobs at the entry year may experience wage penalties that amount, on average, to 50% over the medium term (in the fifth working year). A comparable pattern emerges when analysing the long term. Importantly, the scar does not appear to be driven by non-random sorting, which is tested following an instrumental-variables approach. Exploring the drivers of the scar yields that non-employment spells are key determinants of future wages, followed by low daily working hours. However, hourly wages have a much less prominent role, likely because the existence of minimum wages partly dampens this effect. Lastly, the scarring effects appear to be markedly sensitive to the cycle: individuals starting with bad jobs during the crisis have a higher estimated scar as compared to pre-crisis cohorts. The results contribute to the literature on the effects of low-quality jobs on workers' employment trajectories and underscore the need for policies to address the continuous inflows and outflows into and out of employment.

Chapter 3 explores the determinants of gender segregation in the Spanish labour market and empirically tests for a possible solution to tackle this phenomenon. Using an in-house designed large-scale survey addressed to a representative sample of individuals in Spain, the results show that the field of study is the main driver of occupational segregation. Specifically, engagement in HEAL (health, education, administration and literacy) education significantly increases the likelihood of women pursuing occupations characterised by a notable female presence. For men, STEM (science, technology, engineering and mathematics) education correlates with a higher average probability of engagement in male-dominated occupations, while HEAL studies are associated with an increased probability of involvement in occupations less dominated by men. Further analysis into the determinants of educational choices indicates that math anxiety during adolescence is the factor that more notably discourages the pursuit of STEM-related studies. The second part of the paper draws on a second online survey addressed to pre-university individuals. The aim is to explore if a female role model intervention in mathematics has the potential to modify preconceived perceptions that may discourage adolescents from engaging in math-related careers. The experimental results show that the role model intervention has a positive impact on several outcomes that may affect future choices, including on the idea that mathematics can be a useful tool to be applied in everyday life, as well as to address global challenges. These results contribute to the understanding of the drivers behind gender segregation, emphasising the pivotal role of educational presorting even after accounting for other factors commonly associated with this phenomenon.

The remainder of this thesis is organised as follows. The next three sections present, respectively, Chapter 1 (*The Negative Impact of Information and Communication Technologies Overuse on Student Performance: Evidence From OECD Countries*), Chapter 2 (*The Long-Lasting Effects of Landing a Bad Job*) and Chapter 3 (*The (In)Evitable Effects of Educational Presorting on Gender Segregation in the Labour Market)*. Finally, the last section concludes.

Chapter 1

The Negative Impact of Information and Communication Technologies Overuse on Student Performance: Evidence From OECD Countries

1. Introduction

The rapid process of digitalisation has permeated and transformed a number of aspects of citizens' daily live, from social relationships to labour organisation. The last decade has particularly uncovered that adoption of digital skills is paramount in two key ways. Firstly, because it contributes to enhancing citizen participation due to increasing access to information (Polizzi, 2021). Secondly, because it facilitates the process of reskilling or upskilling in a context where demand for new digital skills has risen steadily. In consolidating the process of digital transformation, education and training play a central role. In this paper, we focus on the role of digital technologies to support the learning process of the youth, the engine behind the future of work.

The integration of digital technology to facilitate students' learning, also known as educational technology, has the potential to create a powerful and engaging environment for collaborative and creative learning (European Commission, 2020; Rubach & Lazarides, 2021). However, in absence of a well-founded pedagogical strategy, the use of digital technology at school risks that individuals lag behind (Comi et al., 2017), a matter at the forefront of the debate during the closure of educational institutions with the advent of the COVID-19 pandemic (Talib et al., 2021). The past two decades have seen firm attempts by policymakers to reduce the so-called "digital gap" (Szeles, 2018), or the unequal access to Information and Communication Technologies (ICT). Over time, the reduction of this gap has been substantial, particularly in technologically and economically advanced societies (Vassilakopoulou & Hustad, 2021).

In this context, one of the key questions now is to what extent the use of ICT—rather than solely the access to them—ultimately impacts on students' performance. This paper explores the non-linear association between ICT usage at school and student performance in a number of OECD countries and it assesses the causal impact of ICT overuse on student performance.

Results from this study contribute to expanding policymakers and educators' earlier knowledge on the way technology, widely present in the classrooms, influences student performance.

2. Literature Review and Contributions

2.1. Literature on the Linear Relationship between ICT Usage and Student Performance

The existing evidence on the linear effects of ICT usage on student performance fundamentally depends on the nature of the data. Results arising from experimental (or quasi-experimental) studies are mixed, while those based on international survey data, such as PISA, generally point to a negative association between ICT use and student performance (OECD, 2015). Focusing on PISA studies assessing the effects of ICT use at school, Hu et al. (2018) find that a one-score increment in the frequency of use is negatively associated with academic performance on mathematics, science and reading in the 44 countries examined with PISA 2015 data (between 10 and 13 points in the three fields of analysis, which is roughly equivalent to a fourth of a full academic year). These findings are consistent with previous studies which make use a number of waves of PISA (Zhang & Liu, 2016 find a negative effect of 9 points on mathematics and science using 2000-2012 PISA data) or which focus on a specific wave of PISA (Petko et al., 2017 find a negative association between educational use of ICT in the classroom and the PISA results using PISA 2012 data). Other authors (Skryabin et al., 2015) question whether this issue differs by grade, and find a negative impact for secondary school students (between 13 and 15 points for the three PISA areas), but a positive impact for primary school students (between 5 and 7 points depending on the area).

Country-specific literature using PISA data mostly points to a negative association between the educational use of ICT and student performance. For Turkey, the use of computers for educational purposes is found to negatively affect students' reading performance (Gumus & Atalmis, 2011). This negative association is also found for Spain (Gómez-Fernández &

Mediavilla, 2021): using PISA 2015 data, the authors find a negative association between the educational use of ICT at school and at home and the performance in all three areas of assessment. With regard to the school effect, the authors suggest that the lack of preparation of teachers in terms of digital competences may explain part of the result. In a recent paper, Fernández-Gutiérrez et al. (2020) find, also for Spain, that the use of ICT at school in an Autonomous Community does not positively affect performance in mathematics and reading (although it does for science). For Italy, it is found that the usage of at least one digital device has a positive impact on students' performance in mathematics (Ferraro, 2018) compared to the absence of usage of digital devices, yet the frequency of use is not captured in the model.

2.2. Literature on the Non-Linear Relationship between ICT Use and Student Performance

Previous literature, hence, mostly focuses on analysing the relationship between technology and academic performance in a linear fashion, disregarding the possibility of non-linearity. This may, however, be paramount as the oftentimes found negative relationship might be capturing an average effect that might be overlooking potentially positive effects related to certain degrees of use. The OECD (2015) already suggested that a limited usage of computers in school may trigger better performance than no use at all, but a high use (above the OECD average) could lead to significantly worse academic results.

Some exceptions which have explored the potential existence of non-linearity are identified below. In particular, Woessmann and Fuchs (2004), using data from PISA 2000, find an inverted U-shaped relationship between Internet connectivity at school and student mathematics and reading performance. For the specific case of the Netherlands, Gubbels et al. (2020) find an inverted U-shaped relationship between ICT use and reading performance using data from PISA 2015. Focusing on Hong Kong, a recent study (Zhu & Li, 2022) finds that the use of ICT at school is negatively associated with student performance in a linear fashion, while the usage for other purposes (e.g., for leisure or for off-school learning) follows a hill-shaped relationship with student performance. Relatedly, Hu and Yu (2021) assess the relation of ICT use at school for communication (chatting online with other students and using email at school)—among other variables—and student performance on digital reading by analysing whether the effect varies depending on the frequency of use. The results indicate that over the past decade, adolescents' frequent use of ICT-based social media at school, including chatting online and using email at school, have negative effects on digital reading performance compared to the peers who seldom do so. Lastly, Borgonovi and Pokropek (2021) identify an inverted U-shaped association between different forms of ICT use—including use at school—and reading achievement by using PISA data for 2009-2018 for OECD countries.

2.3. Literature on the Causal Impact of ICT Use on Student Performance

As outlined above, the literature using large-scale surveys usually establishes a correlation relationship, rather than a cause-effect analysis, given the difficulty to address non-observable features such as student motivation (Fernández-Gutiérrez et. al, 2020; Fariña et al., 2015). In broad terms, (quasi-) experimental studies allow for a deeper understanding of a potential causal effect between ICT usage and student performance when compared to the usage of large-scale surveys, while the drawback is that these results are mostly not generalisable as they focus on a very particular context (Fernández-Gutiérrez et al., 2020). In fact, depending on the type of intervention and the context, the impact of ICT on student performance may vary remarkably.

Some interventions have been undertaken through randomized controlled trials (RCT), which are generally regarded as the strongest research design for quantitatively estimating average causal effects (Angrist & Pischke 2008). A tablet intervention addressed to primary school students in Malawi was assessed by Pitchford (2015) through an RCT. The findings point to a positive impact on the mathematical achievement if the software is carefully designed in terms of content and ability to engage students in the learning process. More recently, Araya and Diaz (2020) carried out an RCT to evaluate the impact of an online math program in Chile and found that the platform had a positive impact on students' performance, which amounted to almost double the results achieved by using the same exercises on a paper version for a whole year. Lastly, the evaluation of a math homework program in a state of the United States through an RCT yielded a positive impact on student's mathematics scores compared to existing homework practices (Roschelle et al., 2016).

Because it is not always possible to conduct RCTs (e.g., as they may be financially costly), additional research designs for causal inference have been developed over the past few years, such as instrumental variables, difference in differences, regression discontinuity designs and propensity score designs, known broadly as quasi-experiments (Escueta et al., 2020). As an early example, Angrist and Lavy (2002) adopted an Instrumental Variables approach to assess a policy of installing computers into Israeli primary schools on a wide-spread basis, and they did not find evidence of any relevant effect on students' test scores. Findings from two regression discontinuity designs (RDD) on subsidized computers for households in Romania and the Netherlands (Malamud & Pop-Eleches 2011; Leuven et al., 2007, respectively) point to negative impacts on achievement outcomes, likely in part a result of the students spending more time playing games (Escueta et al., 2020).

Focusing on large-scale surveys, the literature measuring the causal impact of the frequency of ICT usage on student performance is scarce, although some other variables, such as ICT investment, have been analysed. For example, Cabras and Tena Horrillo (2016) study the impact of ICT investment on student performance in Spain using PISA 2012 data and applying Bayesian Additive Regression Trees (BART). Results suggest a moderate positive causal effect of ICT investment on student performance. Some additional techniques to overcome the potential endogeneity bias arising from ICT usage have been employed by authors. For example, Agasisti et al. (2020) resort to propensity score matching and Instrumental Variable techniques to examine the effect of ICT use at home and find a negative causal impact on student performance in almost all EU-15 countries.

In broader terms, meta-analyses over the last decade have pointed to moderate effects of technology integration on student achievement, with these effects varying significantly by educational technology type. For instance, Cheung and Slavin (2013) suggest that educational technology applications generally offer modest positive effects compared to traditional methods. Consistent with those findings, results from a meta-analysis and research synthesis in Sung et al. (2016) indicate that the application of mobile devices to education has a moderate mean effect size. A systematic review carried out in Crompton and Burke (2018) shows that, of 23 studies analysed, the use of mobile learning in higher education has a positive impact in 16 (70%) of them. Lastly, some other meta-analyses or systematic reviews highlight the existing variability in the magnitude of the impact across different contexts. For instance, the effect of ICT on student learning outcomes is found to vary across grades and subjects (Torgerson & Zhu, 2003; Bayraktar, 2001).

2.4. Rationale for the Present Study

The widespread presence of technology has triggered a vivid debate around its usefulness as a tool to enhance student performance. This debate further gained momentum with the outbreak of the COVID-19 pandemic which followed a forced transition to more technology-based education during the closure of educational institutions and triggered learning losses in a wide range of contexts (see Engzell et al., 2021, for the Netherlands; Maldonado & De Witte, 2022, for Belgium; or Aucejo et al., 2020, for the United States).[1]

In general terms, while the evidence on the impact of technology on academic achievement is not conclusive, as shown in the literature section above, this paper attempts to shed light on

[1] Conversely, Clark et al. (2021) find a positive causal effect of online education on Ninth Graders' performance in China, particularly when online lessons were delivered by higher-quality teachers.

two fundamental points related to the impact of the frequency of ICT use at school on student performance. First, it was not until recently that the possibility of a non-linear relationship was formally considered (Zhu & Li, 2022). This is, however, paramount to policy makers, as the establishment of a linear relationship might be capturing an average effect that might not be reflective of the actual relationship. This would happen when the positive or negative effect might vary depending on the degree of usage. If this were the case, instructors and policy makers would need to aim for the optimal frequency of usage, which requires to be combined with an appropriate implementation of digital devices at school (OECD, 2015). Earlier literature in this context has been either country-specific (e.g., based on the Netherlands or Hong Kong) or has provided average results for OECD countries in an aggregate manner, and this study aims to expand the geographical scope to test for the possibility of non-linearity in a wider sphere of OECD countries. This is tested in a non-parametric way, contrary to the vast majority of previous research, where non-linearity is gauged by means of quadratic models. Relatedly, we argue that the separate, country-specific analysis allows to gauge potential geographical divergences in the effects of ICT use on student performance, as found in Agasisti et al. (2020).

Second, most of the previous literature focused on examining the correlation between ICT use and student performance. This approach, while informative, risks offering a blurred picture of the real cause of the impact on account of confounding variables (Busenbark et al., 2021). For instance, if the frequency of use of technology were found to be negatively associated with student performance, then it could well be the case that this be caused by other non-observable variables that correlate with frequency of use (e.g., if more frequent users happened to lack motivation to excel academically, and this was the cause of their underperformance, then the estimate would not be reflective of the causal impact). Addressing causality is, hence, paramount in the development of well-founded public policy recommendations (Athey & Imbens, 2017). This study aims to explore the potential existence of a cause-effect relationship between frequency of ICT use at school and student performance. To our knowledge, this is the first time that the causal impact of frequency of ICT usage at school is analysed, especially in the framework of large-scale surveys. The ultimate goal is to contribute to the guidance of educational policy choices in a context where technology is playing an increasingly central role in the learning process of students.

3. Method

3.1. Research Context and Sample

The present study feeds from the PISA 2018 microdata, a programme led by the OECD that measures the ability of 15-year-old students to use their mathematics, science and reading skills to meet real-life challenges. The 2018 edition includes participation of 600,000 students from 79 countries, representing about 32 million students (OECD, 2020a). The assessment comprises a number of questionnaires, addressed to a wide range of stakeholders, namely students, teachers, parents and school managers. The key questionnaire for this study is the ICT familiarity questionnaire, which includes detailed information on students' use of ICT and their attitude towards it.

The focus of this paper is particularly placed in Estonia (N = 4,862), Finland (N = 4,898) and Spain (N = 28,319), though Appendix 1.D, as shown later, will extend the empirical results to a number of additional countries in order to test for the robustness of the results.[2] We compare Spain, a relatively low-performing country (OECD, 2019) with limited integration of ICT at school despite its ample ICT infrastructure (Gil-Flores et al., 2017), with the opposite side of the coin: two traditionally top-players in the PISA context and where the education policy has made a firm commitment to integrating ICT into their education system.[3] This comparison aims

[2] This refers to the results on the hill-shaped relationship between use of ICT at school and student performance; the results on the causal impact are undertaken solely for Estonia, Finland and Spain for the sake of simplicity.

[3] The Finnish government launched a two-year programme in 2017 to develop teachers' digital skills. For the Estonian case, since 2014, the government implemented a strategy to develop the digital competencies of both

to quantify whether the problems that may be identified are generalisable across countries or, conversely, whether these potential deficiencies do not apply in countries with advanced policies on ICT integration for educational purposes. In total, 22 countries are analysed—including Spain, Finland and Estonia—those for which the questions related to the key variable of interest (i.e., ICT usage in terms of frequency at school) do exist in the database. Those countries are Australia (N =10,830), Belgium (N = 6,891), Switzerland (N = 5,164), Czech Republic (N = 6.181), Denmark (N = 5,976), the United Kingdom (N = 6,975), Greece (N = 5,641), Hungary (N = 4,717), Ireland (N = 5,049), Island (N = 2,675), Italy (N = 9,484), Lithuania (N = 5,840), Luxembourg (N = 4,706), Latvia (N = 4,630), Poland (N = 5,087), Slovakia (N = 4,997), Slovenia (N = 5,447) and Sweden (N = 4,617).

3.2. Variable Description

Dependent Variable

As outlined above, PISA measures students' skills to solve real-life problems in three main areas: mathematics, reading and science. In the present study, the descriptive analysis is presented for these three areas to test whether the observed patterns apply relatively homogeneously. In fact, after confirming that the functional form to relate ICT usage with student performance is comparable across all three knowledge areas, the empirical section specifically focuses on the mathematics field to simplify the analysis. The reason underpinning this choice is that mathematics fosters mental discipline, logical reasoning, mental rigor, and is a paramount element to understanding the content of other fields, such as science. Mathematics is also the engine to STEM-related careers, which are closely related to jobs that will only gain momentum in the future, such as those related to artificial intelligence, machine learning, automation or robotics (Wang & Siau, 2019).

teachers and students by providing IT-training courses and instructional materials (Gabriel et al., 2021). Since 2018, both countries have implemented training courses to encourage teachers' confidence in using ICT.

In terms of measurement, the OECD quantifies students' grades around an OECD average of 500 points and a standard deviation of 100 points.

Control Variable of Interest

Drawing on the ICT familiarity questionnaire, this study focuses on the module of the questionnaire tackling the frequency of use of ICT by students at school, rather than at home. This choice is central to the interpretation of the results, as there are arguments to consider it as more exogenous than usage in other contexts: there is an external factor (such as the teaching staff or the school's policy concerning the use of ICT) which, in principle, determines the use of ICT made at school. This would contrast with the choice of the variable of educational use at home, which may suffer from greater selection bias because it could be determined by the student's own initiative, their socio-economic background or the family environment. Another reason why the analysis focuses on the use of ICT at school is due to its impact on education policy, which is more straightforward to implement as compared to the use of ICT in the private domain.

In order to measure the frequency of use of ICT at school, the questionnaire includes ten different questions. These reflect the extent to which students use a computer at school to do their schoolwork, use the school's computers to do group work or communicate with other students, or surf the Internet in connection with class work. The remaining questions are specified in Appendix 1.A. The possible answers that students can provide are the following: "never or hardly ever", "once or twice a month", "once or twice a week", "almost every day", or "every day".

To synthesise the frequency of use of ICT at school, we create an index that allows to compare students' frequency of use of ICT at school. This index is benchmarked at the country level. This intra- rather than inter-country comparison is most suitable in the context of the present study, especially since cross-country comparisons show a blurred relationship between average ICT use and the average score in mathematics, as described in Appendix 1.B. More broadly, another argument to support this intra-country comparison relates to the fact that reported variables have intrinsic limitations that might hinder inter-country comparisons (for instance, certain cultural aspects of countries might lead students to overstate or understate some questions).

The index summarises the use of ICT at school for student i ($ICT *$). It is calculated by obtaining each student's mean reported frequency (ICT)—across the ten questions included in the questionnaire—and normalising it by subtracting country c's mean use, $\overline{ICT}$, and dividing it all by the country's standard deviation σ_{ICT}. This index will therefore have a mean value of zero for each country, and a standard deviation of one.

$$ICT_i * = \frac{ICT_i - \overline{ICT_c}}{\sigma_{ICT,c}} \qquad (1)$$

It is important to note that the OECD already offers an index to synthesise the use of ICT at school by students. The index is centered around an OECD mean of zero and a standard deviation of one, and its construction is based on the Item Response Theory (IRT) (see OECD 2017 for further methodological details). While this index is useful for inter-country comparisons, it is not fully suitable for our analysis for the abovementioned reasons. To ensure that the index created here is, however, robust to the OECD's index, we calculated the correlation between the two. In the case of Spain, for instance, the correlation between the ICT index created here and that of the OECD is 0.9406028.

Based on the ICT index ($ICT_i *$), five types of users are created, ranging from the very low frequency user to the very intensive one (Table 1.1). The rationale for the creation of those users is to further explore the possible non-linear relationship between ICT use and academic performance. These users are defined on the basis of the country-specific quintiles of the ICT index created in this analysis. Quintiles are created with the purpose of summarising the average reported frequency of the ten questions on ICT usage, which makes the interpretation more comprehensive than if a continuous variable (i.e., the mean of a variable that was originally categorical) were to be included. As quintiles rely on the country's specific distribution, it is worth noting that users might not be directly comparable across countries.

Table 1.1. Definition of ICT Users at School

Country-specific quintiles of frequency of use of ICT at school	Type of ICT user at school
Quintile 1	Very low ICT user
Quintile 2	Low ICT user
Quintile 3	Medium ICT user
Quintile 4	Intensive ICT user
Quintile 5	Very intensive ICT user

Other Control Variables

As explanatory variables we include a set of student background variables that have been frequently identified as relevant factors in earlier literature (e.g., Gamazo & Martínez-Abad, 2020; Hu et. al., 2018). These variables reflect both the student- and school-level characteristics. At the student level, we include discrete indicators of gender, repetition and immigration status, given their relevance in explaining student performance. In particular, repeaters and migrants are frequently found to be negatively correlated with performance, whereas results by gender are mixed depending on the competence under study (Gamazo et al., 2018). Additionally, a binary variable that captures students' late start in the use of technologies (above nine years of age) is also covered, a feature that is generally more common amongst students from lower socio-economic backgrounds (Rodrigues & Biagi, 2017). Separately, the PISA index of economic, social and cultural status (ESCS) and a PISA index on the degree of bullying suffered are also included as control variables. These are standarised variables centered around an OECD mean of zero and a standard deviation of one. The ESCS index synthesises students' responses regarding their family background (e.g., home possession, parents' occupations and parents' highest educational level), while the bullying exposure index comprises other questions such as whether other students made fun of the respondent. The literature indicates that the socio-economic status index is positively related to student performance, although this relationship is far from deterministic (OECD, 2020b). The ESCS is one of the most used variables in the PISA literature, as it helps address questions about educational opportunity and inequalities in learning outcomes (Avvisati, 2020). Conversely, the inclusion of the bullying index is often overlooked in the literature, while research highlights its negative impact on student performance (Yu & Zhao, 2021).

At the school level, we include the school size (in logarithmic terms), which the literature suggests to be associated with improved student performance (Giambona & Porcu, 2018). The type of school ownership (public or not) is also included as a control variable. In this case, the literature on its effect on academic achievement offers mixed results (Gamazo & Martínez-Abad, 2020). Lastly, the inclusion of a ratio to measure the number of computers per student (as a continuous variable) serves as a proxy of the school's available ICT resources per student.

3.3. Research Model and Procedure

The methodological framework is divided into two parts. First, it assesses—through hierarchical linear models—whether the hill-shaped relationship between ICT usage and student performance still holds after taking into account other student-specific determinants. The second part focuses on the very intensive ICT user and adopts a complementary technique to establish a causal relationship between the very intensive ICT usage at school and mathematical performance. This is done through a widely applied technique in the causality literature: Inverse Probability Weighting.

Hierarchical Linear Models

This first part outlines the empirical strategy to assess the relationship between ICT usage and student performance taking into consideration the nested nature of the data. The fact that students are nested within schools implies that multiple regression analysis is not suitable. Instead, the relation is estimated by means of multilevel models, also known as hierarchical linear models (Bryk & Raudenbush, 1992). This is a form of Ordinary Least Squares that analyses the variance in the dependent variable when the predictor variables are at different levels (Woltman et al., 2012).

The rationale for this estimation procedure is described below and is formally specified in Equations 2 and 3. The first-level specification gauges the relationship between student performance (student *i* attending school *j)* and the *p* different explanatory variables considered (i.e., the set of independent variables outlined in the "Variable Description" subsection). More specifically, the variables ranging from *X1* to *X4* are binary variables that denote the type of ICT user each student can be deemed as depending on the level of usage of ICT (based on the country-specific quintiles of ICT usage): low, medium, intensive and very intensive, respectively, and the very low user is taken the as the reference variable. This allows to estimate the relationship between the frequency of ICT usage and mathematics performance when compared to those students who barely ever (or never) make use of it. This is in contrast with most of previous studies that attempt to gauge the non-linear association between ICT use and academic performance, which usually resort to quadratic models, whereas the specification herein used is more flexible by being non-parametric.

The remaining variables entail other features such as the student's gender or socio-economic status, among others (see "Variable Description" subsection). Lastly, e_{ij} refers to the residuals. The second-level specification shows that the intercept varies across schools; that is, the overall mean intercept includes a school-specific random-effect term. The reminder βs are constant across schools. The combination of the equations at level 1 and level 2 gives rise to the final specification as shown in Equation 3.

Level 1 specification

$$(2)\ Y_{ij} = \beta_{0j} + \beta_{1j}X1_{ij} + \beta_{2j}X2_{ij} + \cdots + \beta_{pj} * Xp_{ij} + e_{ij}$$

Level 2 specification

$$\beta_{0j} = \gamma_{00} + \ u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

$$\dots$$

$$\beta_{pj} = \gamma_{p0}$$

Mixed model specification

$$(3)\ Y_{ij} = \left(\gamma_{00} + u_{0j}\right) + \gamma_{10} * X1_{ij} + \gamma_{20} * X2_{ij} + \cdots + \gamma_{p0} * Xp_{ij} + e_{ij}$$

In all cases, the 10 plausible values for each student are considered simultaneously, and the 80 weights assigned to each student are taken into account to avoid potential bias in the estimated coefficients (OECD, 2017).

In sum, while this methodology allows to isolate the correlation between ICT usage at school and the academic performance, causality cannot be inferred. To address this, the following subsection outlines the methodology underpinning the causality analysis.

Inverse Probability Weighting

The second part of the empirical framework focuses on the very intensive ICT user, who is of particular interest on account of the results, which evidence their differentiated socio-demographic profile and their notorious underperformance in mathematics compared to the rest of users. Those results, both at the descriptive and at the empirical levels (through hierarchical linear models), cannot be deemed as causal, which is to be analysed in this second part of the analysis.

The causality analysis attempts to identify whether the variable of interest (very intensive ICT usage in this case) is actually causing the outcome variable (student performance) to decrease. For example, if some non-observable variables shared across very intensive ICT users were determining the low mathematics performance, then these variables—rather than very intensive ICT use—would be the cause of a low performance. The fundamental rationale of the causality analysis is to ideally compare a situation where an individual uses technology very intensively with a situation where that same individual hardly uses it at all. If the comparison were to lead to a significant gap in mathematical performance in favour of the non-intensive user, it could be concluded that the very intensive use of ICT is the cause of poor mathematical skills acquisition. However, since in reality this comparison is not feasible for the same individual, there are a number of econometric techniques that offer an approach to address this issue.

In this paper, the Inverse Probability Weighting (IPW) method is applied. This methodology is based on the idea that random assignment ensures that the distribution of variables among treated and control individuals is probabilistically equivalent. Nevertheless, when the assignment is not random (and this is the case for being a very intensive ICT user), some students have higher probability of being treated, depending on their characteristics. In order to obtain a pseudo-random sample that guarantees that the distribution of covariates would be probabilistically equivalent, we weight students by the inverse probability of being very intensive users (Author, 2019). The aim of this estimation method is, in turn, to approximate the distribution of the observable variables of the treatment group (very intensive users) and of the control group (the rest of the students), assuming that in this way the distribution of the non-observable variables would also be assimilated (see Wooldridge, 2002 and 2010 for a detailed explanation of this methodology).

The estimation method is based on the following procedure. Firstly, a logit model is defined to estimate the probability that student i is a very intensive ICT user (Pr ($VeryIntensive_i$)) based on a number of explanatory variables reflected in vector X in Equation 4, and include gender, socio-economic level, repetition and immigration status, late introduction to ICT, school size, computer/student ratio, bullying rate, and school ownership (public or not). In addition, the final student weights (sw) are also included as established in the framework proposed by DuGoff et al. (2014), who state that the logit model should not be weighted by sample weights, but that these should be included as an additional variable in the model.

(4) $$\Pr(VeryIntensive_i) = f(X, sw_i)$$

Once the model is estimated, the probability of being a very intensive user is predicted (p_i). These predictions are used to create inverse probability weights (w_i) in the following way:

(5) $$w_i = 1/p_i, \qquad if \quad VeryIntensive_i = 1$$

(6) $$w_i = 1/(1 - p_i), \quad \text{if } VeryIntensive_i = 0$$

These weights enable the over-representation of those individuals who, given their characteristics, are likely to be very intensive users but do not report being so on the basis of the ICT questionnaire. On the contrary, if the student's characteristics lead to a prediction of low probability of being a very intensive user and the student does not report to be one, the weight to be applied to that student will be close to one. Similarly, if the model predicts a high probability of being a very intensive user and this is indeed the case, the weights assigned will also be close to one. Finally, when the user is indeed very intensive but her/his characteristics predict a low probability of being so, this person will also be over-represented. Through the approximation of observable variables between the control and treatment groups, it is assumed that this approximation is also assimilated in the unobservable variables.

The estimation of the model through IPW allows to obtain the Average Treatment Effects (ATE), which measures the potential causal impact of the very intensive usage of ICT on student performance. The ATE requires that the whole population under study is eligible to be treated, given that it compares the whole population were it treated versus were it not treated. To ensure that this is the case in the present paper, we will analyse the distribution of the propensity score (i.e., the predicted probability of being a very intensive ICT user) between the treatment and control groups ("Results" section). If the distribution is comparable, then the estimation of ATE is well founded, as long as extreme values are not present in the distribution (Cunningham, 2021). In fact, the presence of extreme values could bias the estimator and induce excessive variance, given that the weights attained through the IPW methodology (see Equations 5 and 6) could become overly large and could hence give raise to unstable estimates (Avagyan & Vansteelandt, 2018).

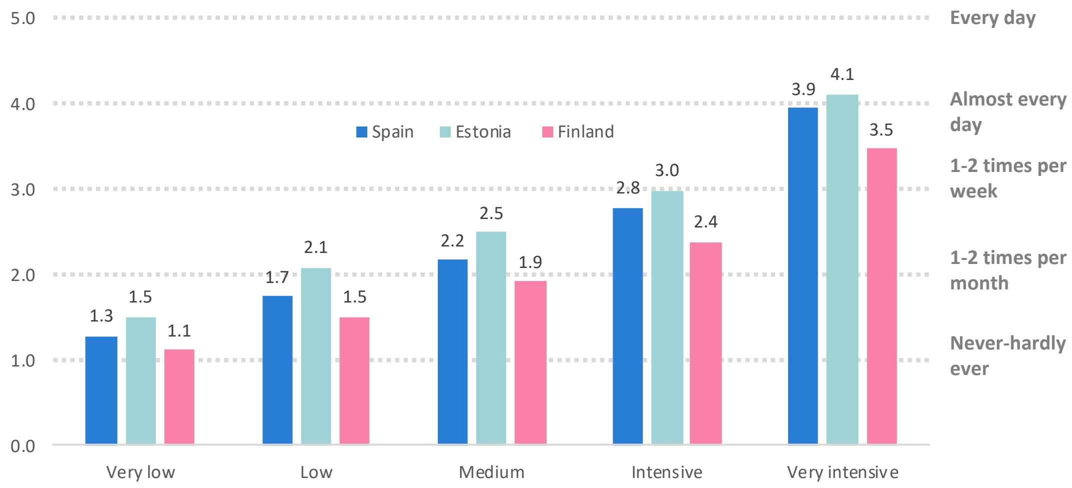
Following the approach proposed by DuGoff et al. (2014), the final weights applied to the model are the product of the sample weights and the IPW weights, calculated as detailed in (4), (5) and (6) above. With these final weights, the average impact in mathematics between the very intensive user and the remainder of the users is estimated, in order to capture whether the existing mathematical gap changes when these weights are applied.

3.4. Data Analysis

The descriptive results show, in first place, the mean use of ICT at school for each of the five types of users herein defined and for the three countries. This allows to infer to what extent a specific ICT user is indeed comparable across the three countries.

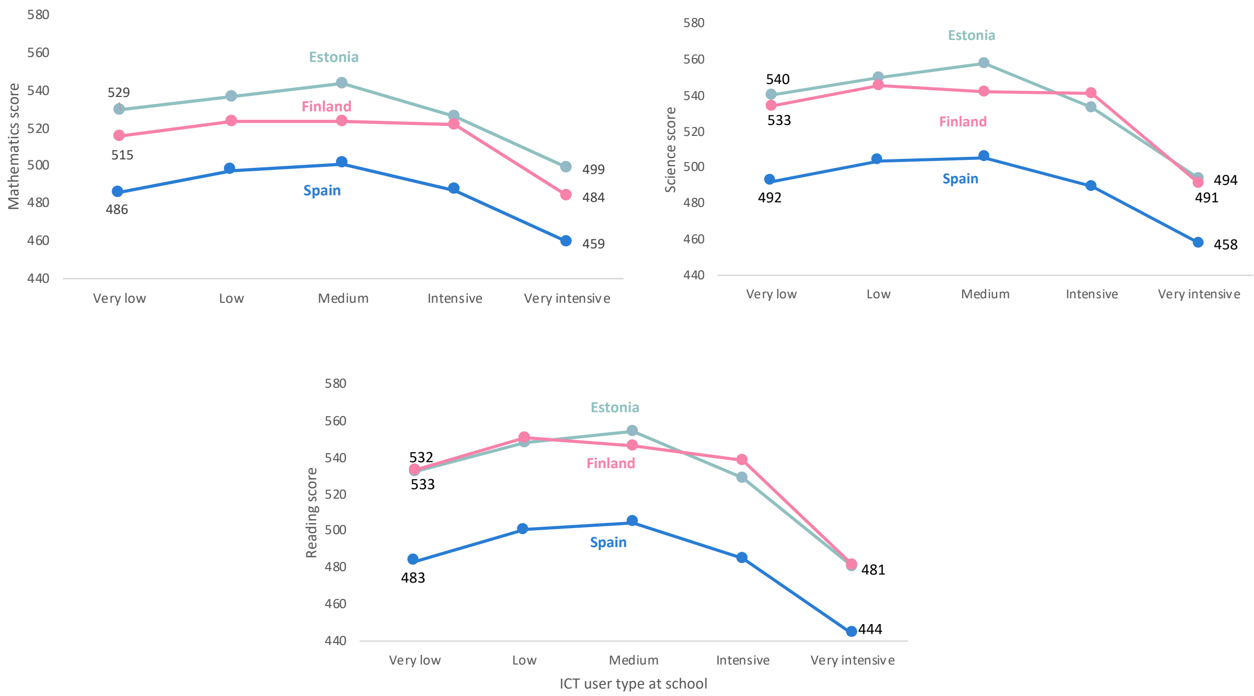
Figure 1.1 shows how the average frequency of ICT use at school varies by type of ICT user in all three countries. For each type of user, the average frequency of ICT use at school is similar in Spain and Estonia, and lower than in Finland. Nevertheless, these differences are relatively small. A clear pattern that emerges is the jump in terms of the frequency reported by the very intensive user in the three countries. While the difference between the four reminder users is relatively stable, the very intensive user reports significantly higher frequency, with the use being close "almost every day", especially for Estonia and Spain. The average frequency reported for each of the ten questions, by country and ICT user type, is provided in Appendix 1.C.

Figure 1.1. Average Frequency of Use of ICT at School in Spain, Estonia and Finland



After identifying the actual average use of ICT at school for each type of user, Figure 1.2 shows the average score of each user in the three main areas of PISA. The results confirm that, in the three countries, the relationship between frequency of ICT use at school and the performance in mathematics, science and reading follows an inverted U shape, where the highest frequency user group (i.e., very intensive users) obtain a significantly lower average grade than the remainder of users. In Spain and Estonia, the maximum peak score in the three knowledge areas is obtained by the medium user (quintile 3, i.e., those who use ICT at school more than 1-2 times per month). In Finland, low users (those that use the ICT 1-2 times per month) and medium users (less than once a week, approximately)—and intensive users, although slightly less so—are the ones who show the strongest mathematical competences, in contrast to the scientific and reading competences, where it is the low intensity user (quintile 2) who obtains the best average competences. Given the fact that the inverted U-shaped relationship holds for the three knowledge areas (mathematics, science and reading), the empirical analysis will focus on the particular case of mathematics.

Figure 1.2. Average Score in Spain, Finland and Estonia by Frequency of Use of ICT at School



Lastly, this section presents a descriptive analysis of the characteristics of the students depending on their frequency of use. Behind each ICT user type, there might be certain socio-demographic profiles that make users perform differently. Table 1.2 aims to compare these features to examine whether patterns emerge depending on the frequency of ICT use.

By gender, male students are more numerous in the extremes of ICT usage: they dominate the groups of analogous users and, more notably, they conform a majority within the group of very intensive ICT users. Immigrants are overrepresented in the group of very intensive users, while they are relatively evenly distributed for the reminder of users. Some other differences can also be observed by country: in Spain, there is a substantial overrepresentation of repeaters within the group of very intensive ICT users. This is also the case in Finland, although the share of repeaters in the country is far lower than in Spain. Additionally, the more frequently students use ICT at school, the higher the proportion of those attending non-public schools is in the

Spanish case. This is despite the fact that performance of Spanish students is typically better in private schools than in public ones (Vega-Bayo & Mariel, 2018), as also found later in the empirical results section (Table 1.3). In Estonia and Spain, there is a positive correlation between ICT usage and the socio-economic profile of students. In those two countries, very intensive users have the largest average ESCS of all the users herein defined. Finally, there is a clear pattern between ICT usage and exposure to bullying. In Spain and Estonia, the average exposure to bullying increases as the frequency of use of ICT at school increases. In fact, analogous users in both countries report, on average, lower exposure to bullying than the average of the OECD, whereas the exposure to bullying for very intensive users is far higher than the country's averages in both cases. Very intensive users, in turn, appear as being much more prone to bullying exposure than the rest of users in those countries, and this applies to the three countries under study.

It is important to note that the results presented in this section are merely descriptive. They do not imply that the hill-shaped relationship is necessarily attributed to the frequency of ICT usage, as there might be other variables beyond the ICT usage that are driving the effect. This might be particularly the case for very intensive ICT users, who have a very differentiated socio-economic profile (Table 1.2). The following section will attempt to infer whether this relationship remains once students' personal characteristics are taken into account.

Table 1.2. Descriptive Statistics by Type of ICT User

	Very low	**Low**	**Medium**	**Intensive**	**Very intensive**
% female					
Estonia	50.5%	58.3%	57.3%	50.2%	40.2%
Finland	47.3%	60.2%	59.5%	52.0%	33.5%
Spain	47.9%	56.1%	56.4%	50.4%	38.0%
% immigrant					
Estonia	1.3%	0.9%	0.7%	1.1%	3.0%
Finland	3.1%	2.5%	2.9%	4.1%	5.0%
Spain	8.7%	7.6%	7.6%	8.3%	10.3%

	Very low	Low	Medium	Intensive	Very intensive
% repeater					
Estonia	3.0%	1.8%	2.2%	2.1%	2.7%
Finland	2.4%	2.4%	2.5%	2.7%	3.7%
Spain	26.7%	21.8%	18.9%	20.8%	28.2%
% public school					
Estonia	95.9%	95.8%	97.0%	96.9%	95.9%
Finland	96.9%	96.5%	94.4%	94.7%	94.9%
Spain	66.4%	65.5%	63.0%	58.6%	54.5%
ESCS					
Estonia	-0.0031	0.0493	0.1397	0.1439	0.0878
Finland	0.2076	0.2647	0.3771	0.4263	0.3678
Spain	-0.1819	-0.0904	-0.0899	-0.0455	-0.0597
Bullying index					
Estonia	-0.0681	0.0203	0.0301	0.0947	0.3262
Finland	-0.0899	-0.1258	0.0033	-0.0118	0.0536
Spain	-0.2990	-0.2767	-0.2974	-0.1691	0.0315

4. Results

This section presents the empirical results and is divided in two parts. The first one shows the estimated relationship between each type of ICT user and their mathematical performance to identify whether the hill-shaped relationship found in the descriptive analysis holds when considering students' characteristics. Results are then disaggregated by gender, ESCS and ICT-related activities. The methodology underpinning these results is based on Hierarchical Linear, or Multi-Level, Models. The second part of this section presents the causal estimates for the very intensive ICT user by applying the IPW framework.

4.1. Results of the Multi-Level Analysis

The results derived from the hierarchical linear model allow to compare the over- or under-performance of low, medium, intensive or very intensive ICT users when compared to very low users, taking into account their socio-demographic features.

Main results

The estimated coefficients of the hierarchical linear model are summarised in Table 1.3 for the three selected countries and expanded to the 22 OECD countries herein considered in

Appendix 1.D. On the one hand, the results show that, for practically all the countries under study (including those in the appendix), the low ICT user status (quintile 2, i.e., average ICT usage slightly below once a month for Spain and Finland; and once or twice a month in Estonia) is related to better results than the very low user status (quintile 1). The medium user (quintile 3, i.e., those with an average use between 1-2 times per month and 1-2 times per week in Spain and Estonia, and 1-2 times per month in Finland) also tends to be related with more positive results than the very low user, although this variable is not significant for an important part of the countries. For Spain and Estonia, on the other hand, positive and significant effects of 10 and 12 points, respectively, are found in relation to the less frequent user.

On the other hand, for the intensive user of ICT at school, a clearly negative trend is observed in most of the countries analysed. However, the coefficient associated to this variable is not significant in Spain, Finland and Estonia, among others. The strong and very significant impact in all the countries lies on the very intensive users, i.e., those who use ICT almost every day. In this group of very intensive users of ICT at school (last quintile), a unanimous pattern is observed in all the countries examined: compared to very low frequency users and, broadly, to the reminder of users, very intensive users score significantly lower in mathematics. In order to interpret such results, it is important to recall that a difference of 40 points is roughly equivalent to a full academic year. This means that very intensive users in Spain or Estonia underperform by more than half year compared to non-ICT users, and by three-quarters of year when compared to low or medium ICT users. In the case of Finland, very intensive ICT users perform a full academic year worse than their low ICT user counterparts.

To expand the analysis, Equation 3 is estimated separately by gender and socio-economic status (higher or lower than the median). The results, shown in Table 1.E1 of Appendix 1.E, confirm that the existence of the inverted U-shaped relationship between ICT usage at school and students' performance in mathematics still holds for all the four groups herein considered.

As a robustness exercise, Appendix 1.F estimates the results for the 22 OECD countries by using the ICT frequency index as a continuous variable—as opposed to the user-specific dummies—and compares it to the results found in Hu et al. (2018), undertaken with the PISA 2015 wave. Results are similar in magnitude when using PISA 2018 and PISA 2015 data, and they point to a negative—and highly significant—relationship between ICT usage and mathematical performance in all the countries analysed.

Table 1.3. Estimated Association between the Mathematics Achievement and the ICT Usage and Other Covariates

	Spain	**Estonia**	**Finland**
Low ICT user	10.21***	6.030	8.759***
	(2.865)	(3.749)	(3.147)
Medium ICT user	10.03***	11.41***	4.503
	(3.050)	(4.061)	(3.639)
Intensive ICT user	-2.963	-4.206	0.245
	(2.965)	(4.537)	(4.009)
Very intensive ICT user	-22.45***	-24.65***	-32.37***
	(3.447)	(4.317)	(3.642)
ESCS (socio-ec. index)	10.36***	19.01***	29.76***
	(0.944)	(1.995)	(1.818)
Immigrant	-19.31***	-25.43**	-25.96***
	(3.663)	(12.95)	(7.091)
Repeater	-86.20***	-48.51***	-59.99***
	(2.853)	(10.47)	(8.595)
Female	-17.57***	-12.30***	-3.310
	(2.313)	(2.832)	(2.817)
Public school	-7.017***	-17.66***	-17.44***
	(1.698)	(4.439)	(2.750)
Number of students at school (log)	3.661***	4.349***	2.358
	(0.774)	(1.552)	(1.515)
Computer-student ratio	-0.708	4.936***	2.263**
	(0.683)	(1.742)	(1.085)
Late ICT users (>9 years old)	-21.19***	-26.93***	-28.40***
	(1.952)	(4.291)	(5.080)
Bullying (index)	-4.015***	0.264	1.164
	(1.322)	(1.679)	(1.324)
Intercept	512.7***	521.3***	512.9***
	(6.138)	(11.72)	(10.98)

Note. Standard errors in parentheses: *** p < 0.01, ** p < 0.05, * p < 0.1. The reference for the different types of ICT users refers to the category of the very low ICT user.

Results by ICT Activity

The analysis so far has focused on the overall use of ICT at school. However, a relevant question to policy makers and instructors is to disentangle which specific ICT-related activities might be associated with improved or poorer performance of students (see Agasisti et al., 2020 for a preliminary analysis on the effects of ICT activities at home). To this end, Equation 3 is now estimated by including the ten specific ICT-related activities listed in Appendix 1.A, which replace the ICT index as the control variable of interest.[4] Each of these activities is introduced in the model as a categorical variable that reflects the reported frequency by students, with five possible answers that range from "never or hardly ever" to "every day" (see the "Variable Description" section), taking the earlier as the reference category.

The results (shown in Appendix 1.G) suggest that for most of the activities, there is a negative association between excessive use and student achievement—particularly the instruction-related ones, to a larger extent than those related with homework—with very few exceptions. The most noteworthy exception relates to the activity of browsing the internet for schoolwork: spending time in this activity is associated with improved performance in mathematics, compared to never or hardly ever doing so, for the three main countries under study. For Spain and Estonia, this association is hill-shaped, whereas it is linear for Finland. Conversely, some other activities such as playing simulations at school, posting work on the school's website or practicing and drilling yield either negative effects (in the first case) or non-significant effects (in the latter activity). These results support the findings of Agasisti et al. (2020), Hori and Fujii (2021), Luu and Freeman (2011), Odell et al. (2020) and OECD (2021) when studying the effects of ICT-related activities on different outcomes. Lastly, a hill-shaped relationship is found between the use of school computers for group work and communication

[4] An alternative option would be to include these variables in separate models and compare their magnitudes. However, this is not optimal since there is a risk that the estimated coefficients are biased due to omission of relevant variables. In addition, the correlation between these variables was tested, and the values are low enough to ensure that the simultaneous inclusion of the variables is plausible.

with other students—among other activities—in Spain and Finland, where the monthly or weekly usage has positive effects on academic performance.[5] In sum, this analysis suggests that the use of ICT at school is more helpful in some school activities than in others, which may hamper student achievement, and that the use of digital devices might be displacing other instructional activities (Falck et al., 2018; OECD, 2021).

As the overall conclusions on the negative effects of ICT overuse hold when analysing the activities separately, the reminder of the paper turns to the usage of the aggregate index. The following section will delve deeper into whether causality can be inferred regarding the negative impact of a very intensive use of ICT on students' mathematical performance.

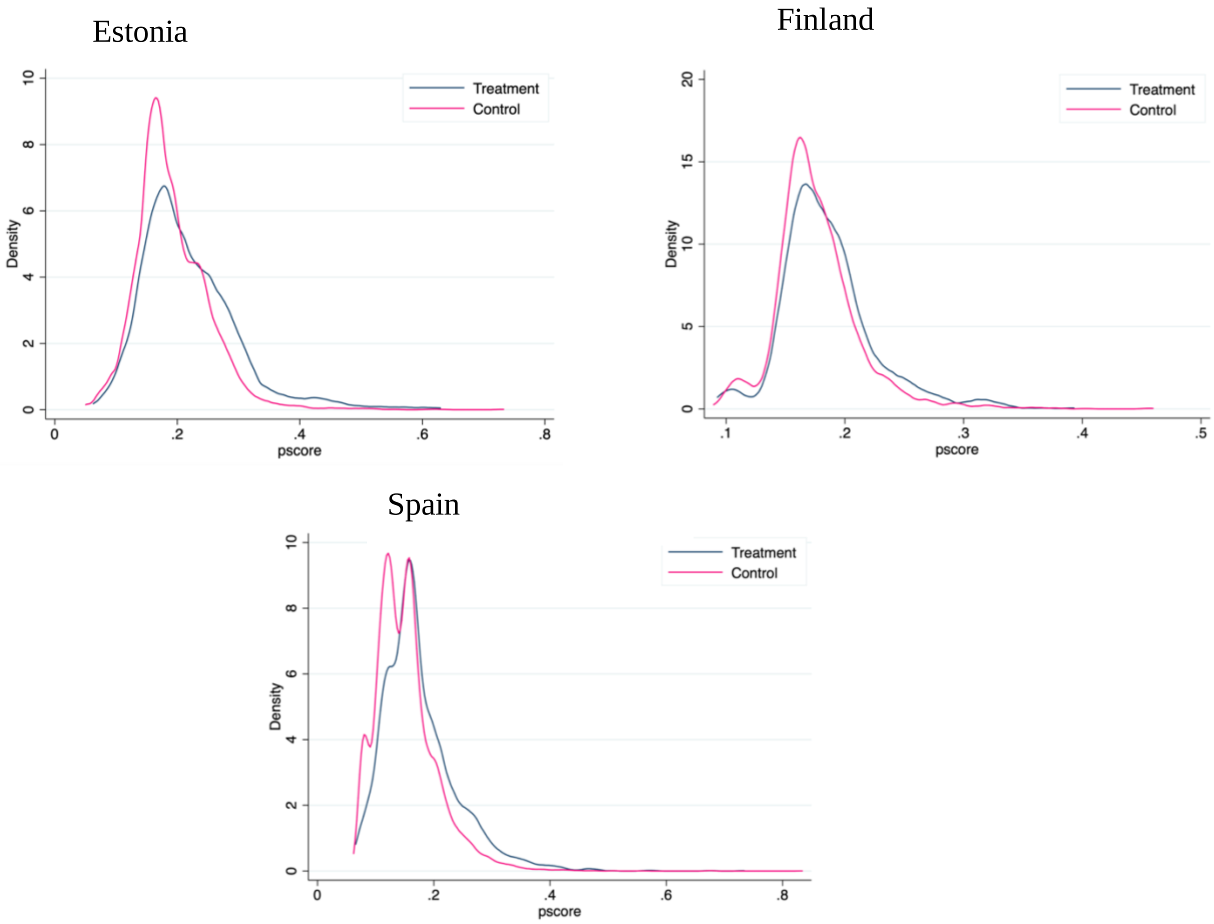
4.2. Results of the Inverse Probability Weighting Analysis

This subsection focuses on the causal impact of the very intensive user to assess whether the underperformance related to the very intensive ICT user seen in the previous subsection can actually be attributed to the very frequent use of ICT.

Before presenting the causal estimates, it is important to first ensure that the distribution of the propensity score (i.e., the predicted probability of being a very intensive ICT user) between the treatment and control groups are comparable such that the ATE is well founded, as long as extreme values are not present in the distribution (Cunningham, 2021). Figure 1.3 shows the distribution of the propensity scores to ensure that the application of IPW would not lead to biased estimates. In the three countries under study, the propensity scores are comparably distributed across treatment and control groups. In addition, extreme cases are uncommon, ensuring that the ATE estimation through IPW is justified.

[5] Additional examples that account for a hill-shaped relationship between the increased use of ICT for an activity and student performance are chatting online at school and using email at school in Estonia (where a daily use yields positive effects relative to no use at all or hardly any use), or using learning apps or learning websites in Finland (where a monthly or weekly use yields positive effects relative to the lowest level of use).

Figure 1.3. Propensity Score Distribution of Treatment and Control Groups



Note. The figure presents the predicted probability of very intensive ICT user for the treatment and control groups, attained through a logit model as set out in Equation 4.

The Inverse Probability Weighting estimates are presented in Table 1.4. In particular, the table shows the Predicted-outcome means (Pomeans) and the ATE, that is, the estimated mean difference in performance of the very intensive user compared to the performance if she/he were to use ICT at school less frequently. For context, the observed gap in the math mean score is also included in the table.

The results shown in Table 1.4 confirm that very intensive ICT usage causes significant underperformance in mathematics. That is, after approximating the observed variables between the treatment and control group (and assuming that the unobserved features are also assimilated), there is evidence that a very intensive usage of ICT causes substantial underperformance in mathematics. The usage of ICT at school more than 1-2 times per week reduces very significantly students score in mathematics. This penalty is equivalent to more than half an academic year for very intensive users in Spain, and ¾ of an academic year for the homologous users in Finland and Estonia.

Table 1.4. Inverse Probability Weighting Estimates of Very Intensive ICT Usage

	Spain	**Estonia**	**Finland**
Inverse probability weighting estimates			
Pomeans	491.8***	533.5***	516.7***
	(0.873)	(1.332)	(1.158)
ATE	-26.28***	-32.38***	-32.83***
	(2.015)	(2.821)	(2.858)
Observed statistics			
Observed math mean score not very intensive ICT user	485.3	529.0	512.0
Observed gap math mean score very intensive user vs not	-25.9	-30.4	-27.9

Note. Standard errors in parentheses: *** p < 0.01, ** p < 0.05, * p < 0.1. Pomeans refers to the predicted-outcome means; and ATE, to the average treatment effect.

As done earlier, the analysis is extended by socio-economic status and gender, with the aim of assessing whether the causal impact of the very intensive ICT usage on the mathematical performance holds across these groups.[6] These results, shown in Appendix 1.E (Table 1.E2), confirm this fact and show that the relative impact of very intensive ICT usage is more negative for female students in Estonia and Finland, and for male students in Spain. In Spain and Estonia, the relative impact is more negative for students from high socio-economic profiles, while the opposite is found for Finland.

5. Discussion

In sum, after confirming the existence of an inverted U-shaped relationship between frequency of use of ICT at school and academic performance, which is significantly negative for very intensive users, the present study has confirmed that the penalty associated to the very

[6] This exercise is, however, not undertaken by ICT activity, as done in the "Results of the Multi-Level Analysis" section. The reason is that the prediction of an intensive usage of ICT would need to be done separately for each activity, and there would be serious concerns on the model's plausibility particularly for students whose frequency across the ten activities has strong variability.

intensive ICT usage is causal, rather than explained by the particular socio-demographic features of this student subgroup. Below, we discuss the results arising from this study and put them into context based on related literature.

5.1. The hill-shaped relationship between ICT usage and student achievement in OECD countries

The results on the inverted U-shaped relationship between ICT usage and student performance are aligned with Gubbels et al. (2020), who focused on the specific case of the Netherlands. While the setting of the study is slightly different—particularly regarding the field of study (reading), the measurement of the frequency of use of ICT at school (a continuous variable based on the OECD index) and the quadratic functional form of the model—results are still comparable. After controlling for similar covariates as in the present study, a hill-shaped relationship is found, and the difference in the mean predicted performance between the least and the most intensive user amounts to the equivalent of over an academic course. This is similar to the difference in the predicted mean in the reading performance found by the OECD (2015). Again, these magnitudes are not directly comparable to the present study, but the overall conclusions do concur. Similarly, the broad conclusions are in line with Borgonovi and Pokropek (2021) and Hu and Yu (2021), while they cover a number of countries in an aggregate manner, as in OECD (2015), and assess the relation with regard to students' reading performance. In contrast, recent findings by Zhu and Li (2022) for Hong Kong are slightly different. While accounting for non-linearity, the authors assess a linear and negative relationship between ICT use at school and student performance in reading. However, the relatively less time available to efficiently use ICT tools compared to OECD peers might partly explain this divergence (Zhu & Li, 2022).

5.2. The negative impact of a very intensive ICT usage on student performance

Concerning the causal negative impact of very intensive use of ICT at school on the mathematical performance found in this paper, to our knowledge there is no directly comparable paper to contrast the results with. However, the study by Agasisti et al. (2020) would constitute a close example assessing causality through analogous econometric techniques. The findings reveal that the intensive use of ICT at home has a negative causal impact on all subjects in most EU-15 countries. The present study focuses on the use of ICT at school, and hence further studies are needed to further delve into the direction of the impact, as well as the factor that might be driving the results.

Overall, the reasons underpinning the negative impact between the very intensive ICT usage on the mathematical performance are beyond the scope of the paper, but some potential hypotheses are explored here. On the one hand, students could possibly get distracted by using ICT at school for activities unrelated with the educational purpose of the usage of these devices. This might lead them to over-report the amount of time spent using technology at school (Agasisti et al., 2020). The possibilities that ICT offers students for "multitasking", i.e., performing a large number of tasks at the same time, can prove detrimental to students' ability to capture information (OECD, 2018; Vedechkina & Borgonovi, 2021; Borgonovi & Pokropek, 2021). On the other hand, deficiencies in training teachers towards digitalisation have also been identified by the OECD (2018) and other authors (e.g., Hu et al., 2018) as an obstacle to successfully foster student learning through digital devices. This might be the case when teachers' ICT knowledge is not regularly updated, although since the outbreak of the COVID-19 pandemic—which is not gauged in this study—many teachers and educators were forced to rapidly develop and learn ICT skills to optimise their instruction (Vedechkina & Borgonovi, 2021).

In view of the results, some policy implications can be drawn. First, stakeholders should exercise caution when integrating ICT at school to ensure that educational technology does not interfere with students' learning processes. To this end, it is important to particularly monitor the very frequent users of technology at school—as the study has shown that this overuse has a strong negative impact on students' achievement in mathematics—together with other key stakeholders in the integration of ICT at school, notably, teachers. If factors largely beyond the reach of instructors were found to explain students' underperformance (e.g., distraction as a result of the large amount of time spent using ICT), as the study suggests, then policies should be addressed in order to limit what could be deemed as an excessive frequency of use. The study also highlights the need for teachers, educators, school principals and policy makers to carefully identify their context-specific deficiencies, which the paper has shown to differ both geographically and across student subgroups. Preliminary results suggest that the negative effects of ICT overuse at school be primarily associated with instruction-related activities, in contrast with other activities such as browsing the internet for schoolwork. Investigating the specific activities that contribute to improved achievement through rigorous studies is paramount for a well-founded implementation of ICT at school. The quality of integration of ICT at school could be improved through channels claimed in the literature, such as computer-assisted instruction or technology-based curricula (Hu et al., 2018).

6. Conclusions, Limitations and Future Research

The present study contributes to the field in two key ways. First, this study captures the varying effects of ICT at school in the performance in mathematics depending on the intensity of use for a number of OECD countries. Second, it applies the Inverse Probability Weighting technique to gauge the potential causal impact of ICT overuse on student performance, while most of previous studies using large-scale surveys limit the results to the correlation sphere.

The results of this study confirm the existence of a hill-shaped relationship in explaining the frequency of ICT usage at school and students' performance in mathematics in 22 OECD countries, with varying magnitudes across countries and types of ICT activities. The study reveals that even in the most advanced countries in terms of ICT integration at school—such as Finland or Estonia—the group of very intensive frequency users experiences a significant penalty in terms of their performance in mathematics, while the low and medium ICT user status is related to better results than the very low user status.

However, these very intensive have a very differentiated socio-economic profile compared to the rest of users: they report above-average levels of bullying, and they are over-represented by male, repeaters and immigrant students. Given this, the study further explores whether the observed underperformance is attributed to such differentiated profiles or, conversely, whether the penalty can be attributed to the excessive use of ICT at school. Results indicate that the overuse of ICT causes an underperformance in mathematics, which is of the order of more than half academic year in Estonia, Finland and Spain.

The present study is not without limitations, which could be addressed in future research. The first relates to the measurement and definition of the main variable of interest, use of ICT at school, which is made on the basis of quantity of time spent, as opposed to quality of usage (Petko et al., 2017). In fact, the definition of the variable of interest is broad, which limits the interpretation of the results. Although we have attempted to disentangle the role of each ICT-related activities on student performance, the rationale for the differing effects—as well as the causal impact of the separate activities—remains a question to be addressed. In addition, future research could further explore whether the impact differs when looking at other computer-related activities that might be more specifically addressed to improving student performance, such as computer thinking applied to digital devices. The PISA data to date did not cover information of this nature; however, future editions intend to include activities such as the use of digital resources to solve equations or for coding purposes (Lorenceau et al., 2021). The second limitation refers to the absence of other covariates that might be of relevance to the model. One question that arises in view of the results is how the performance of students clustered in the same classroom where ICT is very intensively used might vary. This is currently unattainable with the PISA database due to the lack of an identifier that links students with classrooms. The availability of this data would also allow to identify whether specific ICT methodologies implemented by teachers, as well as ICT training received, entail a differential impact on student performance, which has been noted to be paramount in this research context (Pérez-San Agustín et al., 2017). The third limitation relates to the cross-sectional nature of the database: the usage of panel data (or quasi-experimental studies) would further enrich the analysis and, notably, the causality analysis. A fourth limitation related to the causality analysis lies on the assumption that the unobserved features between treatment and control groups are assimilated, which might not always be the case if unobserved variables proved relevant in either of the two groups (whether treatment or control). Lastly, although this paper covers a wide range of countries, ICT activities and population subgroups, results cannot necessarily be generalised to other contexts, whether geographic or temporal. Despite the gaps that are yet to be overcome, the present paper has intended to further contribute to the exploration of the way ICT—which is increasingly present in schools—affects student performance, a paramount topic for instructors and policy makers in their search for an optimal use of technology that enhances students' learning processes.

References

Agasisti, T., Gil-Izquierdo, M., & Won Han, S. (2020). ICT Use at home for school-related tasks: what is the effect on a student's achievement? Empirical evidence from OECD PISA data. *Education Economics, 28(6),* 601-620.

Angrist, J., & Lavy, V. (2002). New evidence on classroom computers and pupil learning. *The Economic Journal, 112 (482)*, 735-765.

Angrist, J., & Pischke, J.S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Araya, R., & Diaz, K. (2020). Implementing Government Elementary Math Exercises Online: Positive Effects Found in RCT under Social Turmoil in Chile. *Education Sciences, 10(9)*, 244. https://doi.org/10.3390/educsci10090244.

Athey, S., & Imbens, G.W. (2017). The State of Applied Econometrics: Causality and Policy Evaluation. *Journal of Economic Perspectives, 31 (2)*, 3-32. https://doi.org/10.1257/jep.31.2.3.

Aucejo, E.M., French, J., Araya, M.P.U., & Zafar, B. (2020). The impact of COVID-19 on student experiences and expectations: Evidence from a survey. *Journal of Public Economics, 191*, 104271. https://doi.org/10.1016/j.jpubeco.2020.104271.

Author (2019). To be added following blind review.

Avagyan, V. & Vansteelandt, S. (2018). Stable inverse probability weighting estimation for longitudinal series. *Scandinavian Journal of Statistics, 48(3),* 1046-1067. https://doi.org/10.1111/sjos.12542.

Avvisati, F. (2020). The measure of socio-economic status in PISA: a review and some suggested improvements. *Large-Scale Assessments in Education, 8*. https://doi.org/10.1186/s40536-020-00086-x.

Bayraktar, S. (2001). A meta-analysis of the effectiveness of computer-assisted instruction in science education. *Journal of Research on Technology in Education, 34(2)*, 173–188. https://doi.org/10.1080/15391523.2001.10782344.

Borgonovi, F. & Pokropek, M. (2021). The evolution of the association between ICT use and reading achievement in 28 countries. *Computers & Education Open, 2*. https://doi.org/10.1016/j.caeo.2021.100047.

Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical Linear Models: Applications and data analysis methods*. Sage Publications, Inc.

Busenbark, J.R, Yoon, H.E., Gamache, D.L, & Withers, M.C. (2021). Omitted Variable Bias: Examining Management Research With the Impact Threshold of a Confounding Variable (ITCV). *Journal of Management, 48(1),* 17-48. https://doi.org/10.1177/01492063211006458

Cabras, S., & Tena Horrillo, J. D. (2016). A Bayesian non-parametric modeling to estimate student response to ICT investment. *Journal of Applied Statistics*, *43(14)*, 2627–2642.

Cheung, A.C.K, & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review,* *9*, 88-113. https://doi.org/10.1016/j.edurev.2013.01.001.

Clark, A.E., Nong, H., Zhu, H., & Zhu, R. (2021). Compensating for academic loss: Online learning and student performance during the COVID-19 pandemic. *China Economic Review*, *68*. https://doi.org/10.1016/j.chieco.2021.101629.

Crompton, H., & Burke, D. (2018). The use of mobile learning in higher education: A systematic review. *Computers & Education, 123,* 53-64. https://doi.org/10.1016/j.compedu.2018.04.007.

Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press. https://doi.org/10.2307/j.ctv1c29t27.

DuGoff, E.H., Schuler, M., & Stuart, E.A. (2014). Generalizing observational study results: applying propensity score methods to complex surveys. *Health Services Research, 49,* 284-303.

Engzell, P., Frey, A., & Verhagen, M.D. (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences*, *118(17)*. https://doi.org/10.1073/pnas.2022376118.

Escueta, M., Nickow, A.J., Oreopoulos, P., & Quan, V. (2020). Upgrading Education with Technology: Insights from Experimental Research. *Journal of Economic Literature, 58(4),* 897-996. https://doi.org/10.1257/jel.20191507.

European Commission (2020). *Digital Education Action Plan 2021-2027. Resetting education and training for the digital age*. Retrieved from: https://ec.europa.eu/education/sites/default/files/document-library-docs/deap-communication-sept2020_en.pdf.

Falck, O., Mang, C., & Woessmann, L. (2018). Virtually No Effect? Different Uses of Classroom Computers and their Effect on Student Achievement. *Oxford Bulletin of Economics and Statistics, 80(1),* 1-38. https://doi.org/10.1111/obes.12192.

Fariña, P., San Martín, E., Preiss, D.D, Claro, M., & Jara, I. (2015). Measuring the relation between computer use and reading literacy in the presence of endogeneity. *Computers & Education, 80,* 176-186.

Fernández-Gutiérrez, M., Gimenez, G., & Calero, J. (2020). Is the use of ICT in education leading to higher student outcomes? Analysis from the Spanish Autonomous Communities. *Computers & Education, 157*, 103969.

Ferraro, S. (2018). Is information and communication technology satisfying educational needs at school? *Computers & Education, 122,* 194-204. https://doi.org/10.1016/j.compedu.2018.04.002.

Gabriel, F., Marrone, R., Van Sebille, Y., Kovanovic, V., & de Laat, M. (2021). Digital education strategies around the world: practices and policies. *Irish Educational Studies, 41(1),* 85-106. https://doi.org/10.1080/03323315.2021.2022513.

Gamazo, A., Martínez-Abad, F., Olmos-Migueláñez, S., & Rodríguez-Conde, M. J. (2018). Assessment of factors related to school effectiveness in PISA 2015. A Multilevel Analysis. *Revista de Educación, 379,* 56–84. doi: 10.4438/1988-592X-RE-2017-379-369.

Gamazo, A., & Martínez-Abad, F. (2020). An Exploration of Factors Linked to Academic Performance in PISA 2018 Through Data Mining Techniques. *Frontiers in Psychology*. doi: 10.3389/fpsyg.2020.575167.

Giambona, F., & Porcu, M. (2018). School size and students' achievement. Empirical evidences from PISA survey data. *Socio-Economic Planning Sciences, 64*, 66-77. https://doi.org/10.1016/j.seps.2017.12.007.

Gil-Flores, J., Rodríguez-Santero, J., Torres-Gordillo, J.J. (2017). Factors that explain the use of ICT in secondary-education classrooms: The role of teacher characteristics and school infrastructure. *Computers in Human Behavior*, *68,* 441-449. https://doi.org/10.1016/j.chb.2016.11.057.

Gómez-Fernández, N. & Mediavilla, M. (2021). Exploring the relationship between Information and Communication Technologies (ICT) and academic performance: A multilevel analysis for Spain. *Socio-Economic Planning Sciences, 77*. https://doi.org/10.1016/j.seps.2021.101009.

Gubbels, J., Swart, N.M., & Groen, M.A. (2020). Everything in moderation: ICT and reading performance of Dutch 15-year-olds. *Large-scale Assessments in Education, 8(1)*. https://doi.org/10.1186/s40536-020-0079-0.

Gumus, S., & Atalmis, E.H. (2011). Exploring the Relationship between Purpose of Computer Usage and Reading Skills of Turkish Students: Evidence from PISA 2006. *Turkish Online Journal of Educational Technology, 10(3),* 129-140.

Hori, R., & Fujii, M. (2021). Impact of Using ICT for Learning Purposes on Self-Efficacy and Persistence: Evidence from Pisa 2018. *Sustainability*, *13*(11), 6463. https://doi.org/10.3390/su13116463.

Hu, X., Gong, Y., Lai, C. & Leung, F. K.S. (2018). The relationship between ICT and student literacy in mathematics, reading, and science across 44 countries: A multilevel analysis. *Computers & Education*, *125,* 1–13. https://doi:10.1016/j.compedu.2018.05.021.

Hu, J., Yu, R. (2021). The effects of ICT-based social media on adolescents' digital reading performance: A longitudinal study of PISA 2009, PISA 2012, PISA 2015 and PISA 2018. *Computers & Education, 175*. https://doi.org/10.1016/j.compedu.2021.104342.

Leuven, E., Lindahl, M., Oosterbeek, H., & Webbink, D. (2007). The Effect of Extra Funding for Disadvantaged Pupils on Achievement. *The Review of Economics and Statistics, 89 (4),* 721–736. https://doi.org/10.1162/rest.89.4.721.

Lorenceau, A., Marec, C., & Mostafa, T. (2019). *Upgrading the ICT questionnaire items in PISA 2021* (OECD Education Working Papers No. 202). Organisation for Economic Co-operation and Development. https://doi.org/10.1787/19939019.

Luu, K., & Freeman, J.G. (2011). An analysis of the relationship between information and communication technology (ICT) and scientific literacy in Canada and Australia. *Computers & Education, 56,* 1072-1082. https://doi.org/10.1016/j.compedu.2010.11.008.

Malamud, O., & Pop-Eleches, C. (2011). Home Computer Use and the Development of Human Capital. *The Quarterly Journal of Economics*, *126(2),* 987–1027. https://doi.org/10.1093/qje/qjr008.

Maldonado, J., & De Witte, K. (2022). The effect of school closures on standardised student test outcomes. *British Educational Research Journal, 48(1),* 49-94. https://doi.org/10.1002/berj.3754.

Odell, B., Cutumisu, M., & Gierl, M. (2020). A scoping review of the relationship between students' ICT and performance in mathematics and science in the PISA data. *Social Psychology of Education*, *23*, 1449-1481. 481 https://doi.org/10.1007/s11218-020-09591-x.

OECD (2015). *Students, computers and learning: Making the connection*. Paris: OECD Publishing. http://dx.doi.org/10.1787/9789264239555-en.

OECD (2017). *PISA 2015 Technical Report*. Paris: OECD Publishing. Retrieved from https://www.oecd.org/pisa/sitedocument/PISA-2015-technical-report-final.pdf

OECD (2018). *How has the internet use changed between 2012 and 2015?* Paris: OECD Publishing. Retrieved from https://doi.org/10.1787/1e912a10-en.

OECD (2019). *PISA 2018: Insights and Interpretations.* Paris: OECD Publishing. Retrieved from https://www.oecd.org/pisa/PISA%202018%20Insights%20and%20Interpretations%20FINAL%20PDF.pdf.

OECD (2020a). *PISA 2018 Results (Volume VI): Are Students Ready to Thrive in an Interconnected World?* Paris: OECD Publishing. Retrieved from https://doi.org/10.1787/d5f68679-en.

OECD (2020b). *PISA 2018 Results (Volume II): Where All Students Ready Can Succeed.* Paris: OECD Publishing. Retrieved from https://doi.org/10.1787/b5fd1b8f-en.

OECD (2021). *21st-Century readers: Developing literacy skills in a digital world*. OECD Publishing. https://doi.org/10.1787/a83d84cb-en.

Pérez-Sanagustín, M., Nussbaum, M., Hilliger, I., Alario-Hoyos, C., Heller, R. S., Twining, P., et al. (2017). Research on ICT in K-12 schools-A review of experimental and survey-based studies in computers & education 2011 to 2015. *Computers & Education, 104,* A1–A15.

Petko, D., Cantieni, A., & Prasse, D. (2017). Perceived Quality of Educational Technology Matters: A Secondary Analysis of Students' ICT Use, ICT-Related Attitudes, and PISA 2012 Test Scores. *Journal of Educational Computing Research, 54(8),* 1070-1091. https://doi.org/10.1177/0735633116649373.

Pitchford, N. (2015). Development of early mathematical skills with a tablet intervention: a randomized control trial in Malawi. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00485.

Polizzi, G. (2021). Internet users' utopian/dystopian imaginaries of society in the digital age: Theorizing critical digital literacy and civic engagement. *New Media & Society*. https://doi.org/10.1177/14614448211018609.

Rodrigues, M., & Biagi, F. (2017). Digital technologies and learning outcomes of students from low socio-economic background: An Analysis of PISA 2015. In *Publications Office of the European Union*. https://doi.org/10.2760/415251.

Roschelle, J., Feng, M., Murphy, R.F., & Mason, C.A. (2016). Online Mathematics Homework Increases Student Achievement. *AERA Open, 2(4)*. https://doi.org/10.1177/2332858416673968.

Rubach, C., & Lazarides, R. (2021). Addressing 21st-century digital skills in schools – Development and validation of an instrument to measure teachers' basic ICT competence beliefs. *Computers in Human Behavior, 118*. https://doi.org/10.1016/j.chb.2020.106636.

Skryabin, M., Zhang, J., Liu, L., & Zhang, D. (2015). How the ICT development level and usage influence student achievement in reading, mathematics, and science. *Computers & Education*, *85,* 49-58. https://doi.org/10.1016/j.compedu.2015.02.004.

Sung, Y.T., Chang, K.E., & Liu, T.C. (2016). The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. *Computers & Education*, *94*, 252-275. https://doi.org/10.1016/j.compedu.2015.11.008.

Szeles, M.R. (2018). New insights from a multilevel approach to the regional digital divide in the European Union. *Telecommunications Policy, 42(6)*, 452-463. https://doi.org/10.1016/j.telpol.2018.03.007.

Talib, A.M., Bettayeb, A.M., & Omer, R.I. (2021). Analytical study on the impact of technology in higher education during the age of COVID-19: Systematic literature review. *Education and Information Technologies*, *26,* 6719-6746. https://doi.org/10.1007/s10639-021-10507-1.

Torgerson, C., & Zhu, D. (2003). A systematic review and meta-analysis of the effectiveness of ICT on literacy learning in English. EPPI-Centre, Social Science Research Unit,

Institute of Education, University of London.

Vassilakopoulou, P., & Hustad, E. (2021). Bridging Digital Divides: a Literature Review and Research Agenda for Information Systems Research. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-020-10096-3.

Vedechkina, M., & Borgonovi, F. (2021). A Review of Evidence on the Role of Digital Technology in Shaping Attention and Cognitive Control in Children. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2021.611155.

Vega-Bayo, A., & Mariel, P. (2018). School Choice across Different Regions of Spain. *Hacienda Pública Española / Review of Public Economics, 227*, 11-36. DOI: 10.7866/HPE-RPE.18.4.1

Wang, W., & Siau, K. (2019). Artificial Intelligence, Machine Learning, Automation, Robotics, Future of Work and Future of Humanity: A Review and Research Agenda. *Journal of Database Management, 30(1),* 61-79.

Woessman L., & Fusch, T. (2004). *Computers and Student Learning: Bivariate and Multivariate Evidence on the Availability and Use of Computers at Home and at School*. Munich: Center for Economic Studies and Ifo Institute (CESifo) Working Paper 1321. https://dx.doi.org/10.2139/ssrn.619101.

Woltman, H., Feldstain, A., MacKay, J.C., & Rocchi, M. (2012). An introduction to Hierarchical Linear Modeling. *Tutorials in Quantitative Methods for Psychology, 8,* 52-69.

Wooldridge, J. (2002). Inverse probability weighted M-estimators for sample selection, attrition, and stratification. *Portuguese Economic Journal, 1,* 117-119. https://doi.org/10.1007/s10258-002-0008-x.

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). Cambridge, MA: MIT Press.

Yu, S., & Zhao, X. (2021). The negative impact of bullying victimization on academic literacy and social integration: Evidence from 51 countries in PISA. *Social Sciences & Humanities, 4(1)*. https://doi.org/10.1016/j.ssaho.2021.100151.

Zhang, D., & Liu, L. (2016). How does ICT use influence students' achievements in math and science over time? Evidence from PISA 2000 to 2012. *Eurasia Journal of Mathematics, Science & Technology Education, 12(9),* 2431-2449. https://doi.org/10.12973/eurasia.2016.1297a.

Zhu, J., & Li, S.S.C. (2022). The non-linear relationship between ICT use and academic achievement of secondary students in Hong Kong. *Computers & Education, 2022*, *187.* https://doi.org/10.1016/j.compedu.2022.104546.

Appendix

Appendix 1.A

Student Questionnaire on the Frequency of Use of ICT

This Appendix lists the ten different questions regarding frequency of ICT use in schools as part of the ICT familiarity questionnaire:

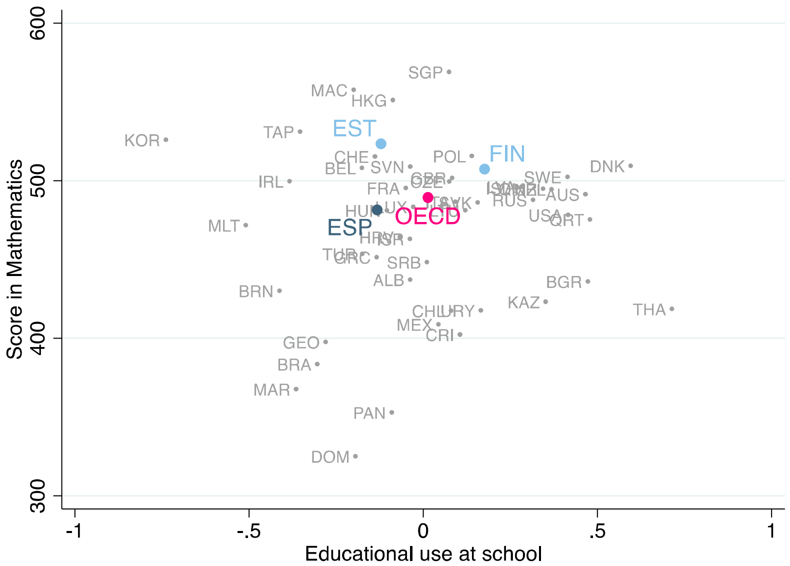
1. Chatting online at school.
2. Using email at school.
3. Browsing the Internet for schoolwork.
4. Downloading, uploading or browsing material from the school's website (e.g. intranet).
5. Posting my work on the school's website.
6. Playing simulations at school.
7. Practicing and drilling, such as for foreign language learning or mathematics.
8. Doing homework on a school computer.
9. Using school computers for group work and communication with other students.
10. Using learning apps or learning websites.

Appendix 1.B

Descriptive Relationship between ICT Usage and Student Performance in Mathematics

In the search for the relationship between the use of ICT at school and students' performance, the first question that arises is how these two factors relate in the different countries participating in the ICT questionnaire. Figure 1.B1 depicts the standard ICT index developed by the OECD in terms of countries' average use of ICT at school and the average score in mathematics. The results do not seem to show a clear association with the average performance of the countries. This is used as the main argument to undertake an intra-country (rather than inter-country) analysis. As with most reported data, certain biases in terms of over/understatement might arise at the country level, hindering the direct comparability across countries.

Figure 1.B1. OECD Indices on ICT Use and their Relation with the Average Score in Mathematics



Note. The educational use at school refers to the OECD-created index that measures the frequency of ICT use at school. It is centered around an OECD mean of zero and a standard deviation of one.

Appendix 1.C

Average Frequency of Use of ICT at School by Activity

Table 1.C1 shows the average frequency reported per country and for each of the five types of users defined in this paper. There exist five possible answers for students to answer in each question listed below, and those are scored as follows: 1, never-hardly ever; 2, once-twice per month; 3, once-twice per week; 4, almost every day; 5, every day.

It is important to note that frequency of use increases uniformly across all questions by type of user. In addition, extreme answers to the individual ICT questions are generally uncommon. These two facts justify the usage of the average, rather than other measures such as the maximum reported frequency.

Table 1.C1. Average Reported Frequency of ICT at School by Activity, Type of User and Country

Very low ICT user			
	Spain	Estonia	Finland
Chatting online at school	1.065	1.008	2.566
Using email at school	1.076	1.149	1.284
Browsing the Internet for schoolwork	1.269	1.313	1.944
Downloading, uploading or browsing material from the school's website (e.g. intranet).	1.018	1.054	1.098
Posting my work on the school's website	1.027	1.011	1.137
Playing simulations at school	1.014	1.021	1.083
Practicing and drilling, such as for foreign language learning or mathematics	1.047	1.045	1.174
Doing homework on a school computer	1.031	1.043	1.074
Using school computers for group work and communication with other students	1.072	1.058	1.358
Using learning apps or learning websites	1.016	1.044	1.189
Low ICT user			
	Spain	Estonia	Finland
Chatting online at school	1.756	1.086	3.789
Using email at school	1.502	1.669	1.858
Browsing the Internet for schoolwork	2.010	2.141	2.741
Downloading, uploading or browsing material from the school's website (e.g. intranet).	1.201	1.300	1.256
Posting my work on the school's website	1.156	1.033	1.290
Playing simulations at school	1.099	1.091	1.207
Practicing and drilling, such as for foreign language learning or mathematics	1.350	1.289	1.408
Doing homework on a school computer.	1.224	1.229	1.170
Using school computers for group work and communication with other students	1.447	1.326	1.743

Using learning apps or learning websites	1.142	1.284	1.495
Medium ICT user			
	Spain	Estonia	Finland
Chatting online at school	2.111	1.417	3.809
Using email at school	1.988	2.135	2.317
Browsing the Internet for schoolwork	2.550	2.597	3.041
Downloading, uploading or browsing material from the school's website (e.g. intranet).	1.574	1.691	1.586
Posting my work on the school's website	1.352	1.134	1.552
Playing simulations at school	1.246	1.178	1.417
Practicing and drilling, such as for foreign language learning or mathematics	1.645	1.592	1.771
Doing homework on a school computer.	1.543	1.440	1.413
Using school computers for group work and communication with other students	1.837	1.563	1.990
Using learning apps or learning websites	1.424	1.598	1.874
Intensive ICT user			
	Spain	Estonia	Finland
Chatting online at school	2.342	2.159	4.101
Using email at school	2.572	2.551	2.828
Browsing the Internet for schoolwork	2.969	2.995	3.488
Downloading, uploading or browsing material from the school's website (e.g. intranet).	2.252	2.348	2.239
Posting my work on the school's website	1.906	1.538	1.956
Playing simulations at school	1.645	1.573	1.602
Practicing and drilling, such as for foreign language learning or mathematics	2.158	2.152	2.160
Doing homework on a school computer.	2.152	1.817	1.774
Using school computers for group work and communication with other students	2.379	2.006	2.369
Using learning apps or learning websites	2.049	2.121	2.369
Very intensive ICT user			
	Spain	Estonia	Finland
Chatting online at school	3.189	3.204	4.146
Using email at school	3.488	3.417	3.478
Browsing the Internet for schoolwork	3.718	3.706	3.927
Downloading, uploading or browsing material from the school's website (e.g. intranet).	3.504	3.605	3.560
Posting my work on the school's website	3.277	3.191	3.461
Playing simulations at school	3.004	3.111	3.159
Practicing and drilling, such as for foreign language learning or mathematics	3.373	3.479	3.480
Doing homework on a school computer.	3.364	3.227	3.342
Using school computers for group work and communication with other students	3.401	3.319	3.562
Using learning apps or learning websites	3.408	3.457	3.610

Appendix 1.D

Estimation Results (Multi-Level Models) for All Countries

This appendix shows the estimation results from the multi-level models for all countries under study.

Table 1.D1. Estimation Coefficients of the Hierarchical Linear Model for All Countries

	AUS	BEL	CHE	CHL	CZE	DNK	ESP	EST	FIN	GBR	GRC
Low ICT user	9.838**	5.636*	15.40***	2.802	3.368	4.189	10.21***	6.030	8.759***	13.63***	17.41***
	(3.849)	(3.077)	(5.193)	(4.406)	(3.879)	(3.868)	(2.865)	(3.749)	(3.147)	(4.033)	(4.506)
Medium ICT user	14.92***	-0.0687	6.325	3.058	-1.127	0.816	10.03***	11.41***	4.503	6.901*	7.431
	(4.043)	(2.967)	(5.033)	(3.738)	(4.161)	(3.948)	(3.050)	(4.061)	(3.639)	(4.006)	(4.606)
Intensive ICT user	-1.188	-2.817	-2.521	-9.813**	-14.05***	-3.410	-2.963	-4.206	0.245	5.749	-15.89***
	(4.050)	(3.230)	(4.952)	(4.333)	(4.177)	(4.532)	(2.965)	(4.537)	(4.009)	(4.119)	(4.478)
Very intensive ICT user	-13.36***	-26.24***	-26.60***	-26.12***	-25.69***	-23.53***	-22.45***	-24.65***	-32.37***	-22.53***	-29.52***
	(4.819)	(3.043)	(4.862)	(4.535)	(3.828)	(4.987)	(3.447)	(4.317)	(3.642)	(4.265)	(4.814)
ESCS (socio-ec. level)	19.26***	14.93***	18.23***	8.751***	15.49***	26.32***	10.36***	19.01***	29.76***	14.92***	15.40***
	(1.511)	(1.635)	(2.449)	(2.230)	(1.816)	(2.541)	(0.944)	(1.995)	(1.818)	(1.674)	(1.779)
Immigrant	1.195	-17.46***	-12.92**	-15.07*	-36.83***	-11.92	-19.31***	-25.43**	-25.96***	-6.406	-22.85***
	(3.537)	(4.591)	(5.441)	(8.280)	(8.207)	(9.108)	(3.663)	(12.95)	(7.091)	(4.871)	(7.195)
Repeater	-34.92***	-59.29***	-46.42***	-51.21***	-53.64***	-47.38***	-86.20***	-48.51***	-59.99***	-48.79***	-35.29***
	(5.761)	(3.272)	(6.214)	(3.718)	(9.850)	(8.771)	(2.853)	(10.47)	(8.595)	(10.33)	(12.60)
Female	-9.203***	-23.94***	-22.20***	-18.47***	-18.91***	-12.64***	-17.57***	-12.30***	-3.310	-12.28***	-17.12***
	(3.094)	(2.597)	(3.425)	(3.008)	(3.295)	(3.100)	(2.313)	(2.832)	(2.817)	(2.632)	(3.286)
Public school	-19.48***		-20.53***	-49.06***	-10.88***	-23.58***	-7.017***	-17.66***	-17.44***	-13.86***	-35.31***
	(1.369)		(5.785)	(2.725)	(3.011)	(2.198)	(1.698)	(4.439)	(2.750)	(4.720)	(3.042)
School size (log)	22.38***	8.028***	19.76***	19.77***	14.79***	9.450***	3.661***	4.349***	2.358	4.872	0.167
	(1.432)	(1.145)	(0.700)	(1.202)	(0.912)	(1.481)	(0.774)	(1.552)	(1.515)	(3.669)	(3.136)
Ratio computer/student	-1.182**	-4.666***	-1.911**	4.130**	2.307	-3.134	-0.708	4.936***	2.263**	-1.024	-46.71***
	(0.509)	(0.985)	(0.861)	(1.945)	(1.438)	(1.924)	(0.683)	(1.742)	(1.085)	(1.966)	(6.643)
Late ICT user (age >9)	-18.78***	-15.55***	-17.39***	-10.33***	-20.40***	-23.67***	-21.19***	-26.93***	-28.40***	-30.41***	-12.77***
	(2.488)	(2.442)	(3.133)	(3.856)	(3.942)	(4.114)	(1.952)	(4.291)	(5.080)	(3.311)	(3.013)
Bullying (index)	-4.86***	-0.133	-4.905**	-1.025	-2.458	0.495	-4.015***	0.264	1.164	-2.442**	-2.153
	(1.132)	(1.545)	(1.950)	(1.362)	(1.531)	(1.674)	(1.322)	(1.679)	(1.324)	(1.152)	(1.669)
Constant	359.8***	506.6***	443.0***	341.5***	451.6***	471.4***	512.7***	521.3***	512.9***	500.6***	514.7***
	(10.70)	(8.527)	(8.291)	(9.205)	(6.051)	(9.846)	(6.138)	(11.72)	(10.98)	(26.62)	(19.40)

	HUN	IRL	ISL	ITA	LTU	LUX	LVA	POL	SVK	SVN	SWE
Low ICT user	3.679	12.86***	14.19**	8.512**	16.21***	11.85***	5.251	5.579	5.534	11.63***	13.80***
	(3.605)	(4.087)	(5.848)	(3.774)	(3.754)	(3.898)	(4.077)	(4.477)	(4.056)	(3.946)	(4.846)
Medium ICT user	4.668	13.25***	17.47**	0.225	-1.966	9.387***	-0.805	-10.01**	-11.62**	3.906	9.534*
	(3.772)	(3.442)	(7.228)	(4.626)	(4.218)	(3.608)	(4.272)	(4.524)	(4.518)	(4.166)	(4.895)
Intensive ICT user	-11.27***	4.418	-4.843	-12.31***	-18.17***	-3.702	-12.11**	-26.40***	-18.05***	-4.992	-6.891
	(3.632)	(3.901)	(7.027)	(3.897)	(3.260)	(3.976)	(4.846)	(4.491)	(4.201)	(4.217)	(4.957)
Very intensive ICT user	-21.67***	-31.42***	-28.63***	-22.18***	-19.97***	-38.51***	-26.63***	-44.73***	-27.72***	-15.91***	-29.14***
	(3.479)	(4.003)	(7.107)	(5.660)	(4.495)	(4.158)	(4.526)	(4.949)	(4.778)	(4.956)	(5.381)
ESCS (socio-ec. level)	8.742***	19.73***	24.07***	6.560***	16.98***	13.05***	17.43***	23.00***	16.95***	6.951***	25.39***
	(1.912)	(1.646)	(2.575)	(2.112)	(1.559)	(1.861)	(1.689)	(1.946)	(2.176)	(2.256)	(1.856)
Immigrant	-20.06	-4.734	-25.56***	-16.37**	-27.00**	-5.290	16.46	-55.94***	-40.86***	-34.64***	-33.55***
	(12.83)	(4.257)	(9.703)	(6.563)	(12.05)	(4.135)	(13.61)	(20.24)	(9.458)	(8.516)	(5.719)
Repeater	-33.29***	-39.81***	-26.23	-41.16***	-68.27***	-57.44***	-69.14***	-83.53***	-102.9***	-60.96***	-42.65***
	(6.164)	(4.988)	(18.26)	(5.329)	(12.19)	(3.179)	(8.016)	(10.69)	(10.14)	(14.48)	(13.44)
Female	-25.77***	-8.420**	2.051	-22.99***	-14.45***	-16.03***	-16.59***	-13.97***	-17.39***	-22.66***	-6.553**
	(2.982)	(4.098)	(4.068)	(3.487)	(3.164)	(2.675)	(2.924)	(3.096)	(3.585)	(3.062)	(3.013)
Public school	-27.48***		-39.64**	-20.60***	-49.75***	-19.01***	-15.28**	-40.66***	-20.56***	-75.35***	
	(3.602)		(17.38)	(3.695)	(5.230)	(5.705)	(7.524)	(3.240)	(2.406)	(7.769)	
School size (log)	35.09***	19.05***	3.593	15.44***	20.52***	10.70***	10.38***	3.119**	11.85***	12.12***	20.49***
	(1.752)	(1.565)	(3.921)	(1.460)	(1.801)	(3.953)	(1.855)	(1.309)	(1.986)	(1.240)	(2.475)
Ratio computer/student	-0.0965	-1.520	-2.131	5.180***	-11.19***	-1.876***	0.611	-12.82***	-9.011***	7.249***	2.582***
	(3.117)	(1.302)	(2.579)	(1.767)	(1.018)	(0.513)	(1.799)	(3.818)	(1.096)	(1.719)	(0.698)
Late ICT user (age >9)	-15.09***	-17.30***	-30.48***	-12.16***	-18.24***	-16.21***	-15.00***	-20.75***	-20.46***	-19.12***	-24.12***
	(2.760)	(2.997)	(4.956)	(2.712)	(3.981)	(3.462)	(3.546)	(4.289)	(3.846)	(3.236)	(4.092)
Bullying (index)	0.373	-0.568	-5.440**	-4.856***	-5.488***	-4.865***	-7.924***	-1.392	-2.289	-1.445	-2.688
	(1.417)	(1.303)	(2.208)	(1.219)	(1.474)	(1.485)	(1.111)	(1.483)	(1.570)	(1.347)	(1.915)
Constant	306.3***	392.7***	505.8***	430.3***	425.6***	468.7***	468.9***	576.8***	478.0***	515.2***	390.9***
	(10.83)	(10.49)	(27.50)	(8.658)	(12.91)	(27.32)	(17.02)	(10.90)	(12.88)	(10.63)	(14.86)

Note. Standard errors in parentheses: *** p < 0.01, ** p < 0.05, * p < 0.1.

Appendix 1.E

Empirical Results by Student Groups

This appendix shows the empirical results disaggregated by gender and socio-economic status. The aim of the first part is to further explore whether the hill-shaped relationship found on an aggregate basis still holds, or, conversely, whether the form changes for certain groups of students (male and female students, and students from high or low socio-economic backgrounds). The second part explores whether the causal impact of very intensive ICT usage is confirmed for those four student groups. The control variables included in these models are the same as those outlined in the "Variable Description" subsection.

The results presented in Table 1.E1 confirm that the inverted-U relationship between ICT usage at school and students' performance in mathematics still holds for all the four groups herein considered. Male students that are categorised as low or medium ICT users perform better than analogous (very low users) male students across the three countries under analysis. However, the effect of male students with a medium ICT usage on the mathematics performance is found to be non-significant. For both male and female students, very intensive ICT users relate with a significantly lower performance than the reminder of users, and this is particularly the case for female students. By socio-economic status, the hill-shaped relationship is also confirmed: low and medium ICT usage entails improved mathematical performance relative to analogous users, and this is the case for students from both high and low socio-economic status. For very intensive ICT users, a negative and significant relationship is, again, found regardless of the socio-economic status of students.

Table 1.E1. The Estimated Association between ICT Users on the Mathematics Score by Gender and Socio-Economic (ESCS) Level (Multi-Level Models)

	Spain		**Estonia**		**Finland**	
	Female	**Male**	**Female**	**Male**	**Female**	**Male**
Low ICT user	4.000	16.96***	0.427	12.17**	2.897	16.16***
	(3.455)	(3.888)	(4.573)	(5.951)	(4.490)	(5.705)
Medium ICT user	4.160	15.31***	4.640	19.83***	0.891	7.231
	(3.999)	(3.789)	(5.903)	(5.677)	(5.222)	(5.036)
Intensive ICT user	-8.226**	2.074	-5.168	-3.543	0.149	-0.919
	(3.744)	(3.652)	(6.008)	(5.881)	(5.439)	(5.511)
Very intensive ICT user	-25.87***	-19.92***	-26.94***	-21.87***	-36.30***	-29.50***
	(5.451)	(3.794)	(6.251)	(5.072)	(5.644)	(5.084)
	High ESCS	**Low ESCS**	**High ESCS**	**Low ESCS**	**High ESCS**	**Low ESCS**
Low ICT user	12.16***	8.058**	4.530	8.438	-1.011	11.41***
	(3.761)	(3.270)	(6.201)	(5.245)	(8.874)	(4.337)
Medium ICT user	9.057**	9.373***	12.32*	12.84**	17.60**	-2.751
	(4.046)	(3.360)	(7.144)	(5.518)	(8.921)	(5.463)
Intensive ICT user	-2.254	-4.938	-2.122	-5.660	8.360	-1.123
	(3.585)	(4.282)	(6.434)	(5.856)	(9.601)	(5.815)
Very intensive ICT user	-19.55***	-26.79***	-27.69***	-22.21***	-28.66***	-31.58***
	(4.535)	(4.191)	(6.442)	(5.661)	(8.576)	(6.118)

Note. Standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The reference for the different types of ICT users refers to the category of the very low ICT user. High and low ESCS refer to the socio-economic index being above or below the country's median, respectively. The usual additional covariates are not shown for simplicity purposes, and full estimation results are available upon request.

Table 1.E2 presents the Average Treatment Effect of very intensive users on the mathematical performance for the four groups considered. By gender, results indicate that the causal impact of the treatment (i.e., the very intensive user) on the mathematical performance for women is -26 points in Spain, and -37 in Estonia and Finland. That is, the very intensive usage of technology at school for women has a more negative impact in Estonia and Finland than it does in Spain, although it is important to note that the two earlier countries have a higher Predicted Outcome Means (in the mathematical performance) than the latter. For men, the ATE amounts to -29 points in Spain, -32 in Estonia, and -31 in Finland. Therefore, a very intensive use of technology causes an underperformance in mathematics that can be equivalent to over half an academic course both for men and women. When looking at the relative impact of the

very intensive ICT usage (the quotient between the ATE and the Predicted Outcome Mean), this is comparable across male and female users in Spain, and it is notably larger (i.e., more negative) for female students in Estonia and Finland compared to their male counterparts.

Analogous to the results by gender, the findings by socio-economic status evidence that the very intensive usage causes an underperformance of close to half an academic year for students, irrespective of their socio-economic status. However, as noted before, the Predicted Outcome Mean in the mathematical performance is larger for students with higher socio-economic status as compared to those with a lower socio-economic background. In fact, when looking at the relative impact (the quotient between the ATE and the Predicted Outcome Mean), this appears relatively comparable for high and low ESCS students in Spain and Estonia, while Finland exhibits a much larger relative impact for lower ESCS students compared to those from higher socio-economic profiles.

The causal impact of a very intensive technology usage at school for students from high socio-economic status amounts to -28 points for Spain and Finland, and -34 for Estonia. For students from lower socio-economic status, the corresponding ATE amount to -24, -30 and -34 points for Spain, Estonia and Finland, respectively. In Spain, the ATE is hence more negative for students with high socio-economic background, in contrast with the findings through hierarchical linear models.

Table 1.E2. Inverse Probability Weighting Estimates of Very Intensive ICT Usage by Gender and Socio-Economic Status

	Spain		Estonia		Finland	
	Female	**Male**	**Female**	**Male**	**Female**	**Male**
Pomeans	487.36***	496.61***	528.47***	539.17***	516.39***	517.09***
	(1.13)	(1.35)	(1.71)	(2.07)	(1.50)	(1.80)
ATE	-25.65***	-28.60***	-37.07***	-31.54***	-36.94***	-30.96***
	(3.13)	(2.68)	(4.16)	(3.84)	(4.81)	(3.65)
Obs. Mean [a]	489.84	481.05	524.68	533.72	514.15	509.61
Obs. Gap [b]	-28.43	-24.95	-37.15	-27.74	-32.74	-24.26
	High ESCS	**Low ESCS**	**High ESCS**	**Low ESCS**	**High ESCS**	**Low ESCS**
Pomeans	516.74***	465.10***	553.45***	512.24***	538.54***	493.88***
	(1.12)	(1.20)	(1.80)	(1.83)	(1.54)	(1.58)
ATE	-28.08***	-23.63***	-34.41***	-30.29***	-28.43***	-34.39***
	(2.49)	(2.90)	(3.92)	(3.80)	(3.94)	(3.71)
Obs. Mean [a]	509.65	460.32	548.88	508.12	533.70	489.72
Obs. Gap [b]	-25.28	-28.15	-31.57	-28.45	-24.29	-33.04

Note. Standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. High and low ESCS refer to the socio-economic index being above or below the country's median, respectively. Pomeans refers to the predicted-outcome means; and ATE, to the average treatment effect. [a] The Obs. Mean refers to the observed mean in mathematics for the non-very-intensive ICT users. [b] The observed gap refers to the difference between the average points of the non-very-intensive ICT users with respect to the very intensive ICT users.

Appendix 1.F

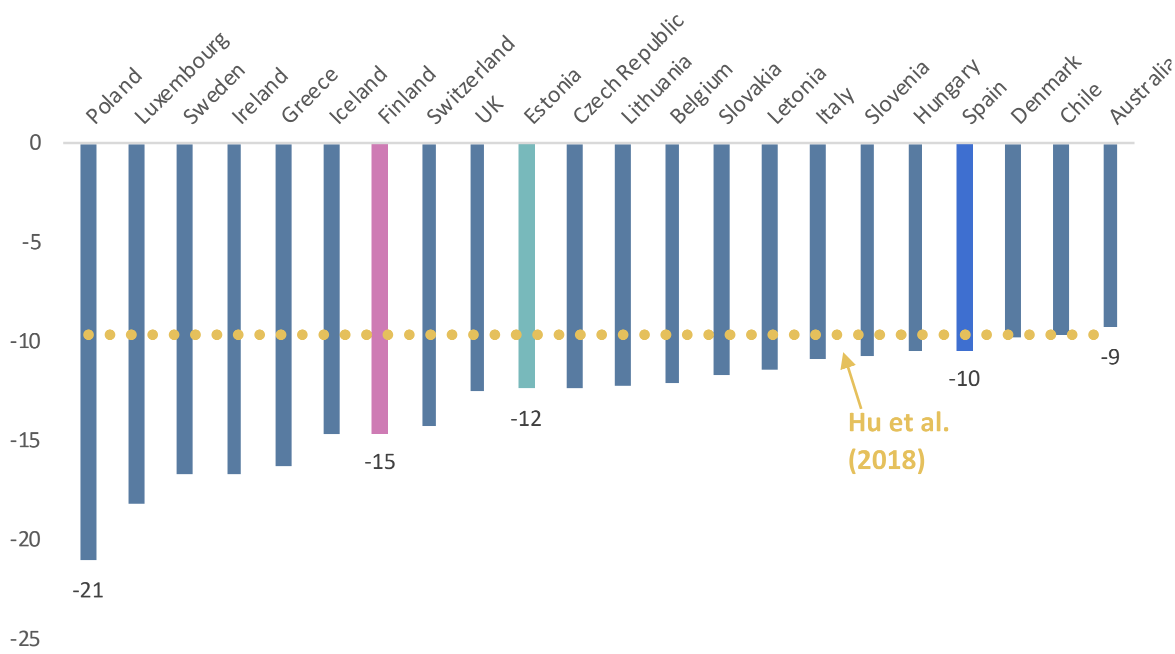
Estimation Results of the Relation between ICT Use and Mathematical Performance When ICT is Defined as a Continuous Variable

In this appendix we estimate Equation 3 by including the ICT variable of frequency of use in the school as a continuous variable, as shown in Identity (1). That is, the standardised variable created in this study is used, which is interpreted as the average impact on the mathematics score if the frequency of use increases by one standard deviation.

Figure 1.F1 shows that a one standard deviation increase in the use of ICTs at school is associated with a negative —and highly significant— effect on the mathematics scores in all the countries analysed. The figure also evidences that the magnitude of the coefficient differs markedly by country: in Poland, an increase in the frequency of use implies a substantially higher penalty than in Australia (an increase in the use of ICT entails penalties of 21 points and 9 points, respectively). Of the countries analysed, Spain is the fourth with the lowest estimated penalty: an increase in use implies an estimated reduction in the mathematical score of around 10 points. In the case of Finland and Estonia, the penalty is higher than that of Spain (with an estimated negative impact of 15 and 13 points, respectively). In summary, this analysis provides very robust results on the negative relationship between the frequency of use of ICT and the performance in mathematics in the 22 countries analysed.

Furthermore, these estimated effects, based on the PISA 2018 data, have similar magnitudes to those found by Hu et al. (2018) for PISA 2015 (with an estimated average effect of -9.67 points). It should be noted that the analysis of Hu et al. (2018) is carried out simultaneously for all countries, with models at three levels, while in this exercise the regressions are carried out for each country separately.

Figure 1.F1. The Estimated Association between an Increase in the Use of ICT at School and the Mathematical Performance



Note. Full estimation results available upon request.

Appendix 1.G

Estimation Results (Multi-Level Models) by ICT Activities at School

Table 1.G1. Estimated Association between the Mathematics Achievement and the ICT Usage by Activity (Multi-Level Models)

	Spain	**Estonia**	**Finland**
Question 1: Chatting online at school			
Once-twice a month	-13.02***	-11.81***	-9.902**
	(2.055)	(4.309)	(5.020)
Once-twice a week	-10.89***	-4.959	-4.187
	(2.075)	(5.306)	(3.995)
Almost every day	-8.897***	11.08**	-6.133
	(2.273)	(4.905)	(3.787)
Every day	-15.08***	9.127*	-5.362
	(2.239)	(5.461)	(3.349)
Question 2: Using email at school			
Once-twice a month	1.939	9.726***	1.755
	(1.671)	(2.531)	(2.648)
Once-twice a week	5.971***	14.47***	2.699
	(1.942)	(3.749)	(3.080)
Almost every day	0.901	14.34***	4.417
	(2.734)	(4.150)	(3.569)
Every day	7.851*	24.41***	-7.114
	(4.052)	(5.765)	(5.592)
Question 3: Browsing the Internet for schoolwork			
Once-twice a month	16.79***	7.386**	17.83***
	(1.932)	(2.922)	(3.617)
Once-twice a week	12.85***	12.81***	26.80***
	(1.848)	(3.376)	(3.772)
Almost every day	11.34***	6.741	31.48***
	(2.656)	(4.239)	(4.345)
Every day	10.91**	1.420	34.47***
	(4.890)	(6.891)	(6.563)
Question 4: Downloading, uploading or browsing material from the school’s website			
Once-twice a month	-3.311	-0.357	-13.50***
	(2.057)	(3.882)	(3.058)
Once-twice a week	-1.581	-8.293**	-22.29***
	(2.469)	(3.903)	(4.162)
Almost every day	-4.495	-8.548*	-34.58***
	(3.308)	(4.465)	(5.736)
Every day	11.76**	-21.91**	-19.37**
	(4.899)	(8.612)	(9.763)

Question 5: Posting my work on the school's website			
Once-twice a month	0.329	-17.15***	-2.220
	(1.967)	(3.752)	(2.786)
Once-twice a week	-1.435	-19.78***	-1.154
	(2.427)	(4.699)	(3.657)
Almost every day	-6.845*	-19.54***	-11.31**
	(3.689)	(6.402)	(5.684)
Every day	-11.12**	-32.86***	-6.770
	(5.073)	(7.536)	(11.49)
Question 6: Playing simulations at school			
Once-twice a month	-10.62***	-10.62***	-9.562***
	(2.327)	(4.094)	(3.324)
Once-twice a week	-11.44***	-13.56***	-24.31***
	(2.868)	(4.994)	(4.001)
Almost every day	-13.07***	-11.05	-23.15***
	(3.618)	(6.861)	(6.601)
Every day	-21.36***	-17.77	-28.51***
	(6.668)	(11.34)	(7.592)
Question 7: Practicing and drilling			
Once-twice a month	1.307	-2.601	5.050*
	(2.020)	(2.908)	(2.582)
Once-twice a week	-1.754	-4.657	-1.907
	(1.992)	(3.994)	(3.809)
Almost every day	2.573	1.861	1.313
	(3.262)	(5.676)	(6.198)
Every day	3.888	9.058	-0.597
	(5.149)	(8.824)	(10.92)
Question 8: Doing homework on a school computer			
Once-twice a month	-6.196***	-2.191	-10.18***
	(1.830)	(2.728)	(2.797)
Once-twice a week	-7.071***	-0.321	-10.92**
	(2.095)	(4.407)	(4.495)
Almost every day	-6.002	-10.55	-11.85*
	(3.845)	(7.677)	(7.003)
Every day	-7.389	-7.734	-10.38
	(5.406)	(10.14)	(12.33)
Question 9: Using school computers for group work and communication with other students			
Once-twice a month	9.669***	4.484	10.72***
	(1.443)	(3.499)	(2.482)
Once-twice a week	5.227**	-5.821	14.54***
	(2.093)	(4.408)	(3.337)
Almost every day	3.571	-0.299	6.512
	(3.717)	(6.498)	(5.720)

Every day	-2.047	-10.41	3.478
	(7.741)	(8.770)	(13.19)
Question 10: Using learning apps or learning websites			
Once-twice a month	-1.746	3.037	7.261***
	(1.865)	(3.384)	(2.391)
Once-twice a week	-8.887***	3.412	7.809**
	(2.554)	(4.051)	(3.626)
Almost every day	-20.33***	-7.007	-0.427
	(3.475)	(6.101)	(5.161)
Every day	-24.02***	-9.388	-15.60
	(5.402)	(7.562)	(13.36)

Note. Standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The reference for the different frequency of ICT use refers to the category “never or hardly ever”. Full estimation results available upon request.

Chapter 2

The Long-Lasting Effects of Landing a Bad Job

1. Introduction

Young inexperienced individuals might sort into jobs of a lesser quality as their lack of specific human capital and relatively lower productivity at the start might be detrimental to employers' hiring decisions (Kahn, 2013; Schönberg, 2007). As a consequence, many entrants start their working lives with temporary or unstable contracts that often involve relatively low wages. As time permits that productivity be revealed, and workers accumulate experience on the job and gain seniority, those initial conditions could improve towards higher wages and more stable contracts. From this perspective, the low quality of jobs at the initial phase of the working career might be regarded as an entrance fee, under the expectation that the situation will eventually change. The question is whether this is the case or whether an initial bad job may become a trap for the future, as suggested by Campbell and Price (2016).

Over the past decades, bad jobs have gained presence in the debate as job polarisation irrupted labour markets, marking a stark and increasing distance between good and bad jobs (Kalleberg, 2011). This growing gap is largely the result of changes in the production and consumption patterns, slower economic growth and notorious changes in social and economic policy. Bad jobs may arise from various channels, including low hourly wages, non-employment spells, low working intensity or overall job insecurity (Olsthoorn, 2014; Rodgers & Rodgers, 1989).

If the presence of such jobs proves persistent throughout employment trajectories, workers may experience adverse effects from a social and economic perspective. On the one hand, low levels of labour income—due to low wages, few hours worked and/or unemployment spells—reduce consumption and saving possibilities, which may severely restrict access to housing and credit (Tridico, 2012) and hamper intertemporal substitution, among other implications that go beyond economic factors, such as fertility, health and well-being (Mauno et al., 2017; Mousteri et al., 2018). On the other hand, non-employment spells or high employer rotation on an involuntary basis affects households' decisions on the future as it hinders the accumulation of work experience, besides incurring substantial transition costs. This gives a bad signal to the market, hurdling potential moves towards better jobs.

In this paper, we use large administrative data on affiliations to Social Security in Spain to investigate the consequences of bad jobs on workers' labour trajectories. In particular, we first, we examine whether involvement in bad jobs have persisting effects on wages over the medium and long term. This is referred to as the scarring effects of bad jobs, a term used by analogy to the notion of the scar of unemployment (e.g., Gangl, 2006; de Fraja et al., 2021). Second, we disentangle the potential labour-market drivers of the scarring effects by analysing the role of non-employment spells, average hours worked daily and average hourly wages. The aim is to shed light on the relative importance of these components on wage trajectories. Third, we study whether entering bad jobs during recession periods has long-standing effects on earnings (von Wachter, 2020).

To define bad jobs, we resort to an operational definition drawing on the European Social Charter. We define bad job as the situation where workers' annual wages are below 60% of the average wage of the country. A key methodological challenge is that involvement in bad jobs might correlate with unobservable factors, such as workers' ability or endogenous timing of entry to the labour market, potentially biasing OLS estimates. To account for this, we resort to an instrumental-variable estimation that leverages the regional nature of the data to capture information on labour market demand, in a similar setting as Arellano-Bover (2022). The instrument captures the incidence of bad jobs for a person's relevant peers, namely individuals with the same educational level and region of birth, as well as the same predicted graduation year (based on age and educational level). This metric is used to predict the probability that an individual holds a bad job, and this prediction is henceforth included as a regressor of future earnings.

We argue that Spain provides a particularly interesting case as its segmented and rigid labour market has proven to entail a more acute scarring effect compared to other more flexible countries (Cocks & Ghirelli, 2016; Fernández-Kranz & Rodríguez-Planas, 2018). In addition, the dual nature of Spain's labour market has significantly intensified labour precariousness in recent years (García-Pérez et al., 2020). This is largely the result of the country's high rates of temporary employment and the wide usage of contracts of very short duration, which applies to a wide range of sectors, notably to construction and hospitality.

Our results show that bad jobs affect most of entrants in Spain and have medium- and long-term effect on workers' wages. In particular, the OLS results point to a medium-term scar of 50% for entrants in bad jobs. The IV estimates point to larger effects, in line with earlier literature (Kahn, 2010). Over the long term, wages are determined by the medium-term involvement in bad jobs to a larger extent than the involvement upon entry. The results reflect that individuals in bad jobs at entry have, on average, 30% lower wages over the longer term. In contrast, individuals in bad jobs over the medium term suffer from an average wage penalty of 41% over the long term. Importantly, the results are robust to an alternative definition of bad jobs. Decomposing the drivers of the scar uncovers that non-employment spells are key for avoiding wage penalties in the medium and long run. In particular, a one-standard-deviation increase in days employed at entry is associated with a 29,4% average increase in medium-term wages. After non-employment spells, working hours per day emerge as the second key driver of the scarring effects. In particular, a one-standard-deviation increase in average working hours per day is estimated to increase medium-term wages by 15,8%, on average. These results show that hourly wages have a less prominent role, as they are partly dampened by the regulated minimum wage. Consistent with prior literature (e.g., Altonji et al., 2016; von Wachter, 2020), we find that involvement in bad jobs during recession periods leaves a deeper scar on future wages, compared to booming periods of the business cycle.

This paper contributes to earlier literature in the following aspects. First, we define adverse labour market conditions at the individual level. This contrasts with most of the earlier research which relies on the macroeconomic environment (notably, on employment) as a central metric for assessing a successful entry to the labour market (Card, 2019). This individual consideration, independently of the macroeconomic environment, is particularly relevant in countries where temporary employment is prominent, a feature that largely relates with an increasing presence of bad jobs. This is the case for Spain (and other EU countries such as Greece or Poland), where the markedly dual nature of its labour market implies that a large proportion of workers hold bad jobs which yield low annual earnings even in expansionary points of the cycle (Font et al., 2015). Second, we explore the labour channels that drive these effects by focusing on work intensity and hourly wages. To our knowledge, this is the first time that this decomposition of the scar is empirically analysed. Earlier literature has explored the long-lasting effects of other factors such as cyclical skills mismatch (Liu et al., 2016), firm size (Arellano-Bover, 2022), type of contract (Garcia-Louzao et al., 2023) or labour market stability (Gardecki & Neumark, 1998) on worker's wage trajectories. Third, we broaden the target group by analysing all young entrants in since 2006, while identifying college graduates. This contrasts with most of earlier research, which focuses on college graduates, both at the national (e.g., Bentolila et al., 2022, except for Fernández-Kranz & Rodríguez-Planas, 2018) and the international level (e.g., Oreopoulos et al., 2012). To our knowledge, this is the first paper that combines these three dimensions to underpin the driving forces of wages arising from bad jobs for the full set of young entrants in the Spanish labour market.

2. Data and Measurement

2.1. Data

In this paper, we use large administrative data on affiliations to Social Security in Spain to analyse the hysteresis of bad jobs in the Spanish labour market for the last decade. The data source is the 2019 wave of the Spanish Continuous Sample of Work Histories (*Muestra Continua de Vidas Laborales*), a microeconomic database that provides full labour market histories of a random sample that represents about 4% of the total affiliations to the Spanish Social Security.

The Continuous Sample of Work Histories contains detailed information on each of the affiliation spells of the individual, including the exact dates of the contract, sector of employment and the average percentage of hours worked daily in each spell as a share of the usual full working day in the employer's enterprise or entity.[7] A complementary dataset also allows to retrieve the monthly contribution bases disaggregated by an anonymised identifier of the employer. In this paper, we use contribution bases as a proxy of labour earnings, as typically done in the literature.[8] Lastly, sociodemographic data include sex, region, educational attainment and date of birth.

The target population comprises nearly 75,000 young workers who first entered the labour market between 2005 and 2013 and whose age ranges between 16 and 30.[9] The working trajectories of these individuals are analysed for their first, fifth and tenth working year. We are unable to observe post-2013 entrants as the last year covered is 2019. We exclude the Covid-19 period to ensure that the analysis is reflective of the structural characteristics of the Spanish

[7] This information is used to calculate hours worked, although certain caveats apply. For instance, the "usual full working day in the enterprise of entity" is not known. Separately, the coefficient may vary over the course of an employment spell, and the database reflects the last value recorded for this datum over the course of that spell. We assume that a usual full working day is equivalent to 8 hours and calculate the total hours worked as the product of the days worked and the daily hours worked based on this assumption.
[8] Contribution bases are top-coded due to regulatory constraints.
[9] These individuals are salaried workers who have been for at least one day during the natural year. Individuals who are self-employed at any point throughout the ten-year period are excluded from the analysis.

labour market. It is worth noting that the first year actually refers to the second one in order to avoid any potential noise underpinning the entry moment, such as working episodes unrelated to future jobs. Appendix 2.A reports summary statistics on the sociodemographic characteristics of the sample.

2.2. Definition of Bad Jobs

To identify bad jobs, an operational definition of the quality of jobs is first needed. We follow prior literature and take annual earnings as a metric to approximate the labour quality of individuals. The earnings threshold, based on the European Social Charter (Council of Europe, 2015), sets the benchmark at 60% of the average wage of all employed workers in each year. This threshold can be viewed as the minimum amount for wages to guarantee a decent standard of living: those individuals whose earnings do not reach that level will be considered to hold "bad jobs".

Annual wages can be decomposed into three fundamental elements: average hourly wages, days worked, and average daily working hours.[10] This disaggregation aims to disentangle the drivers of annual wages and discern the relative importance of work intensity (days worked reflect non-employment spells when these do not reach 365 days, and average daily working hours reflect spells that are not full time when these are lower than eight hours) and hourly wages in determining the quality of jobs.[11,12] While working part time could be a voluntary decision, and hence not negatively affect the quality of jobs, we argue that this is not the case for the vast majority of young Spaniards: only 7% of the workforce employed part-time and

[10] Olsthoorn (2014) provides part of this metric in his indicators on precarious employment. Kalleberg (2009) lists those factors when accounting for job insecurity and precariousness.

[11] Actual days worked are not observed in the database, but rather the duration (in days) of the contract, which most of the times includes weekends and holidays.

[12] Typical empirical studies have used hourly wages and total number of hours worked to analyse total income (Dickinson, 1999). However, the disaggregation of hourly wages herein considered allows to gain deeper understanding on the relative prominence of those two drivers of work intensity.

aged between 16 and 39 report to work part time on a voluntary basis (Spanish Labour Force Survey, 2019, last quarter).[13]

Formally, consider a society S in a period t and associate, to each individual $i \in S$, a function that depends positively on the wages earned on each day d that the person was employed, $d \in e$. Total wages (W) can be calculated as the product of average hourly wages ($\overline{w}$), the total number of days worked (D) and the average number of hours worked per day employed ($\bar{h}$).[14] In sum, workers' total wages in period *t*, which are proxied as the utility that individuals get from a job, are given by Equation 1:

$$W_i(t) \equiv \sum_{d \in e}^{E} W_{i,d} \equiv \frac{\sum_{d \in e}^{E} W_{i,d}}{\sum_{d \in e}^{E} H_{i,d}} * \sum_{d \in e}^{E} d_{i,d} * \frac{\sum_{d \in e}^{E} H_{i,d}}{\sum_{d \in e}^{E} d_{i,d}} \equiv \overline{w_i} * D_i * \overline{h_i} \qquad [1]$$

We then define a bad job at time *t* as one that yields a level of earnings below 60% of the average wage in Spain in each of the given years $q_0(t)$. In particular, if we call $B(S, t)$ the set of workers with bad jobs in society S at time t, we shall have:

$$i \in B(S,t) \Leftrightarrow W_i(t) < q_0(t)$$

Hence, individuals' annual earnings might not reach the threshold for at least one, or a combination of, the following elements: not working full time, having non-employment spells and/or having low average hourly wages. As the latter is a continuous variable, it is important to understand what is meant by low. Even if an individual worked full time for 365 days of the year, that worker would, by definition, still be in a bad job at time t—and hence their average hourly wage would be regarded as low—if $\overline{w_i} * 365 * 8 < q_0$.[15]

[13] Focusing on individuals aged 25-29, only 4% of the part-time workers report this to be voluntary.
[14] When two or more contracts overlap in time, we add up the hours of each contract for the overlapping period, restricting this addition to a maximum of eight hours worked per day.
[15] As noted above, actual days worked are not shown in the database (rather, the duration of the contract) and the expression is therefore a mere illustration to reflect an extreme case where workers' contracts stipulate a total of 365 days and these are full time.

Table 2.1. Thresholds to Define Bad Jobs in Spain and Minimum Wages (Real Euros of 2015)

	Real annual average wage	Annual threshold (60% average)	Hourly threshold (60% of average)	Minimum wage, euros per year	Minimum wage, euros per hour
2006	21,869.37	13,121.62	7.5	6,534.40	3.73
2007	22,221.43	13,332.86	7.62	7,089.71	4.05
2008	23,040.88	13,824.53	7.9	7,762.44	4.44
2009	23,627.47	14,176.48	8.1	8,053.72	4.6
2010	23,454.34	14,072.60	8.04	8,341.32	4.77
2011	22,798.64	13,679.18	7.82	8,704.82	4.97
2012	22,357.39	13,414.43	7.67	8,917.64	5.1
2013	22,244.69	13,346.81	7.63	9,109.18	5.21
2014	22,434.50	13,460.70	7.69	9,091.12	5.19
2015	22,724.40	13,634.64	7.79	9,080.40	5.19
2016	22,613.97	13,568.38	7.75	9,141.61	5.22
2017	22,290.69	13,374.41	7.64	10,075.24	5.76
2018	22,552.72	13,531.63	7.73	10,659.07	6.09
2019	22,815.77	13,689.46	7.82	13,136.76	7.51

Source. Spanish National Statistics Institute, and authors' own calculations.
Note. Data have been annualised to 12 payments. The hourly wage cut-off represents the minimum hourly wage needed to escape bad jobs according to the definition considered in this paper (60% of the country's average wage). The third column shows the necessary, though not sufficient, condition for an individual to not be involved in a bad job.

Table 2.1 shows average wages in Spain since 2006 and the annual thresholds (in real euros of 2015) to determine whether workers are involved in bad jobs. This data is retrieved from the Spanish Labour Force Survey by using the annual distribution of salaries of the whole population.[16] This information is only available as of 2006, which is why we can only analyse information of entrants since 2005 and not earlier, even if administrative data on earlier cohorts does exist. The table shows that average wages in Spain have not recovered since 2010, although an increasing pattern is observed following the post-crisis recovery until 2016. In 2019, average wages were still 3.5% below the peak levels. By contrast, the subsequent

[16] This database is more suitable for the establishment of the threshold as it includes information on actual salaries, in contrast with the Spanish Continuous Sample of Work Histories, which provides information on contribution bases and these are top-coded in Spain, biasing the calculation of average wages.

minimum wage increases undertaken in recent years have resulted in these levels being closer to the threshold, particularly after the sharp minimum wage increase approved in 2019.

3. Empirical Specification

3.1. The Scarring Effects of Bad Jobs

The scarring effect of bad jobs, or the hysteresis of bad jobs, refers to the situation in which a bad job in the present may affect future wages. The following Equations [2] and [3] describe the empirical specification of two linear models for measuring the medium- and long-term scarring effects of bad jobs, respectively:

$$\mathrm{y}_{\mathrm{i},(\mathrm{t}+5)} = \beta_0 + \beta_s * B_{i,(t+1)} + \omega * X_i + e_{i,(t+5)} \qquad [2]$$

$$\mathrm{y}_{\mathrm{i}^{\mathrm{k}},(\mathrm{t}+10)} = \beta_0 + \beta_s * B_{i,(t+k)} + \omega * X_i + e_{i,(t+10)} \qquad [3]$$

Where β_s, reflects the scarring effects of bad jobs: the average change in (log) wages (y) for a person i who held a job in the past ($B_i = 1$), versus someone who did not ($B_i = 0$), given their sociodemographic characteristics X.[17] The covariates X refer to gender, educational level, and fixed effects of sector of activity (lagged to the previous period), region and year of birth. Equation [2] is used for estimating the medium-term effects ($t + 5$) of a bad job at entry ($t + 1$). Equation [3] captures the long-term effects ($t + 10$) of a bad job at entry ($t + 1$), when $k = 1$ or in the fifth working year ($t + 5$), when $k = 5$.[18]

An OLS estimation may give rise to biased β_s estimates given non-random sorting of workers reflected on unobservable variables. Below, we construct an instrumental variable (IV) that aims to alleviate this concern.

[17] The fact that contribution bases are top-coded does not affect the classification of workers in bad jobs: recall that the threshold to classify bad jobs is taken from the Spanish Labour Force Survey. However, it may have certain consequences on the dependent variable, especially over the longer run (as workers gain seniority and access higher wages), imposing a lower bound to our estimates.

[18] Two separate models are estimated for the long run given that the correlation between B_{t+1} and $B_{t+5}$is high enough to discard the simultaneous inclusion of both variables in the same model. Additionally, the IV-TSLS approach explained later makes the inclusion of two instruments for these two variables particularly challenging.

Non-Random Sorting and Instrumental Variables

A causal interpretation of the impact of bad jobs on wages can only be established if holding a bad job is uncorrelated with other determinants of earnings. However, holding a bad job may be endogenous. First, the timing of entry into the labour market may be endogenous: if bad jobs are accentuated during unfavourable economic times, some individuals may delay their decision to enter the labour market by extending their educational career. If timing of entry to the labour market is selected on this basis, the bias could either overstate or understate the true magnitude of the effect. For instance, if individuals with potentially higher earnings select the timing of entry depending on the state of the economy, then one could expect that the effects of adverse initial labour market conditions to be overestimated (Schwandt & von Wachter, 2019). Second, it could be that low-ability individuals systematically sort into this type of jobs because of their condition. If this applied, β_s would simply represent the potential wage penalty that low-ability individuals face, rather than capturing how bad jobs as such could affect individuals, be them of high or low ability. Third, if some individuals strategically migrate to regions where bad jobs are less prevalent, the estimate could suffer from attenuation bias. In particular, previous literature shows that high-ability workers may dodge economic shocks through migration (Kahn, 2010).

To address concerns on endogeneity, the literature on the scarring effects has resorted to two main approaches. One relies on focusing on college graduates (e.g., Bentolila et al., 2022), assuming that this population group exhibits larger overall productivity compared to the less educated workforce (McCall, 2000). While this is relevant, the focus on this group means that an important proportion of the population is left out from the analysis. The inclusion of non-college graduates is relevant as it is well-known that the less educated workforce is particularly prone to suffering from more severe effects (Schwandt & von Wachter, 2019). The second approach used in the literature relies on the usage of instrumental variables (e.g., von Wachter

& Bender, 2006; Kahn, 2010; Schwandt & von Wachter, 2019; Arellano-Bover, 2022; Garcia-Louzao et al., 2023) to account for non-random sorting.

In our setting, both approaches are implemented. First, we re-run Equations 2 and 3 for college graduates and briefly analysing their scarring effects.[19] If those effects still hold when only including, at least potentially, high-productivity workers, these results could be an initial indicative that the scar cannot be mainly attributable to workers' productivity. Instead, this could point to the existence of a mechanism that prevents workers from escaping bad jobs. However, given that some college graduates may still fall under the low productivity workforce, we then resort to an instrumental-variable approach to address concerns on non-random sorting in more detail.

The IV strategy leverages the regional nature of the data to capture information on labour market demand, adapting the strategy developed in Arellano-Bover (2022) to the present setting. The instrument captures the incidence of bad jobs for person i's relevant peers: individuals with the same educational level as i; whose predicted graduation year is the same as i's; and who were born in the same region. This metric is used to predict the probability that an individual holds a bad job, and this prediction is henceforth included as a regressor of future earnings. The instrumented variable ($\overline{B_{-i}^{cer}}$), for a person belonging to cohort c, with an educational level e on a predicted graduation year $\mathrm{t_0(e_i, c_i)}$ and region of birth r, is hence constructed as follows:

$$\overline{B_{-i}^{cer}} = \frac{\sum_{l \neq i} \mathbb{1}\{\, \mathrm{r_l} = r_i, \mathrm{e_l} = e_i, \mathrm{t_l} = \mathrm{t_0(e_i, c_i)}\} \cdot B_{(l)}}{\sum_{l \neq i} \mathbb{1}\{\mathrm{r_l} = r_i, \mathrm{e_l} = e_i, \mathrm{t_l} = \mathrm{t_0(e_i, c_i)}\}} \qquad [4]$$

Where $l = 1, \dots, N$ refers to the workers in the sample, B is a binary variable that indicates whether individuals hold bad jobs, and $\mathbb{1}\{\cdot\}$ is an indicator function. The equation follows a leave-one-out approach (Arellano-Bover, 2022), i.e., we exclude individual i from the

[19] As the educational attainment of workers can change over time, for this analysis we only select workers who hold a college degree since the entry moment to ensure this subgroup is as comparable as possible.

calculation of the instrument whenever individual i started her first job in her predicted graduation year. This is calculated by adding to the birth year the expected age of graduation. This is assumed to be 17 for individuals with high school; 20 for those with vocational education, and 23 for college graduates. A worry that could arise from the choice of region of birth, as opposed to the predicted region of graduation, is that migration flows are not captured. For instance, Arellano-Bover (2022) constructs the instrument by matching the region of birth of the individual to that of the relevant peers who start their first job in that same region. As our sample includes both natives and foreign-born individuals, this approach would imply losing information of a quarter of individuals in the sample. Equally as important, migration in Spain tends to be relatively low: 93% of the individuals in our sample entered the labour market in the region where they were born.

The instrument is constructed on the basis that (1) it is uncorrelated with the error term, i.e., with the unobservable individual characteristics (independence), (2) it affects future earnings solely through the possibility of holding a bad job (exclusion restriction), and (3) it is correlated with the probability of holding a bad job (relevance). In our setting, the instrumental variable is constructed by year of birth, predicted graduation year and educational level. While it is acknowledged that factors affecting the probability of holding a bad job are complex, the instrument focuses on regional and educational traits. These are understood to be unrelated to intrinsic, unobservable variables that affect future earnings, and aims to capture variations in job market conditions at a specific point in time. This temporal alignment enhances the likelihood that the instrument is capturing region-specific economic factors rather than persistent individual characteristics. Regarding the second assumption, individuals with shared educational backgrounds, predicted years of graduation and regions of birth are not expected to experience wage penalties due to these common characteristics. Rather, this effect is expected to be mediated by the fact that individuals are, or not, in bad jobs upon entry. Lastly, the third assumption on relevance is tested in Section 4.

Truncation Bias

By definition, individuals at the entry moment are employed and affiliated to the Spanish Social Security for at least one day. However, as time goes by, a proportion of them no longer have employment affiliations, as shown later in Figure 2.1. The probability of leaving the labour market may hence be partly linked to the quality of the job (Mavromaras et al., 2015): those entering the labour market with worse jobs exhibit higher chances of disappearing from the database. This gives rise to truncation selection bias, as only the outcomes of treated individuals are observable (Wolfolds & Siegel, 2018).

We use the Heckman two-step procedure (Heckman, 1979) to control for non-random selection by first estimating, through probit models, the probability of being treated based on individual and economic characteristics. In the second step, the outcome variable is regressed by adding the fitted values from the selection equation (Inverse Mills Ratio) derived in the first step. A key consideration in running this procedure is the identification of a variable that affects the selection procedure but does not directly affect the outcome variable, except through selection. The identification variable used is the unemployment rate of the region where the first job took place in $t + 4$, both when analysing the medium term and the long term.[20,21] It should be noted that, while the entry moment in the labour market could be endogenous, this instrument covers a larger time horizon that is beyond the scope of individual's choices, particularly considering the changing dynamics of business cycles.

For $t + h$, where $h = (5,10)$, this would yield the outcome equation specified below:

---

[20] Information on region of birth, as opposed to region of first affiliation, cannot be obtained for non-natives, who represent a large share of the total sample. That leads to the choice of region of first affiliation. Concerns on the timing are alleviated for the aforementioned reasons.

[21] Additional tests using alternative time-periods do not yield significant changes in the results.

$$y_{i,t+5} = \beta_0 + \delta\,\overline{B^{cer}_{-i,t+1}} + \beta Z_i + \rho\lambda_{i,t+5} + e_{i,t+5} \quad [5]$$

$$y_{i^k,t+10} = \beta_0 + \delta\,\overline{B^{cer}_{-i,t+k}} + \beta Z_i + \rho\lambda_{i^k,t+10} + e_{i,t+10} \quad [6]$$

Where y refers to the log of wages of individual i, δ captures the estimated coefficient of the instrumental variable for holding a bad job in the lagged period (t+1 in Equation 5; t+1 and t+5, separately, in Equation 6); λ refers to the Inverse Mills Ratio; and Zcomprises the set of sociodemographic covariates (denoted with X in Equations 2 and 3).

The Scarring Effects Throughout the Business Cycle

An additional analysis attempts to study whether the scar varies per year of entry, as the incidence and intensity of bad jobs may become particularly acute in recession periods. This analysis, in turn, aims to examine whether entering in a bad job at a certain moment in the cycle changes the 'depth' of the scar, as suggested in the literature (e.g., Brunner & Kuhn, 2013; Schwandt & von Wachter, 2019). Given that the models are separately estimated by entry year, the sample size for the long-term model is not large enough to make this analysis robust. Because of this, we solely estimate the medium-term model.

Additional Robustness Check: Alternative Definition of Bad Jobs

To test for the robustness of the definition of bad jobs, an alternative threshold is adopted. Instead of considering 60% of the country's average wage as the cut-off point to define bad jobs, we consider 50% of the average wage.[22] The rationale for this choice lies on the recently adopted directive on adequate minimum wages, recently passed in legislation by the European Parliament and the Council (in end-2022), which states that, "to assess the adequacy of the existing statutory minimum wages, Member States may establish a basket of goods and services at real prices, or set it at 60% of the gross median wage and 50% of the gross average wage"

[22] This exercise is applied solely to Equations 2 and 3 as the main purpose is to assess whether the magnitude of the scar varies under an alternative definition of bad jobs.

(European Parliament, 2022). Member States have a two-year limit (until 15 November 2024) to comply with the new rules on adequate minimum wages.

3.2. Decomposing the Scarring Effects of Bad Jobs

After quantifying the effects of bad jobs on wages, we now explore which labour-related factors are most detrimental or beneficial to future earnings based on the three components of annual wages herein considered. Equations [7] and [8] show the medium- and long-term model specification, respectively:

$$\mathrm{y}_{\mathrm{i,t+5}} = \alpha_0 + \alpha_{1,s} * \sum_{s=1}^{3} f_{i,t+1}^{s} + \alpha_2 * X_i + e_{i,t+5} \quad [7]$$

$$\mathrm{y}_{\mathrm{i}^{\mathrm{k}},\mathrm{t+10}} = \alpha_0 + \alpha_{1,s,k} * \sum_{s=1}^{3} f_{i,t+k}^{s} + \alpha_2 * X_i + e_{i,t+10} \quad [8]$$

Where $f = (\bar{h}, D, \bar{w})$ refers to the average daily hours worked, days employed, and average hourly wages ($s = 1, 2, 3$) of individual i. In order to rank the relative importance of the variables of interest ($\alpha_{1,s}$), the three variables ought to be expressed in the same unit. To do so, we standardise them by subtracting their sample mean and dividing it by the standard deviation. The resulting variables have zero mean and one standard deviation. As before, y refers to the (log) wages in the fifth and tenth working years; and $k = (1, 5)$ accounts for initial and medium-term conditions as explanatory variables in the two corresponding long-term models. The set of additional covariates gauged in vector X refers to gender, educational level, and fixed effects of sector of activity (lagged to the previous period), region and year of birth.

4. Scarring Effects of Bad Jobs: Empirical Evidence

4.1. Descriptive Results

To quantify the incidence of bad jobs in the Spanish labour market, Figure 2.1 describes the movements between bad and not-bad jobs in the first, fifth and tenth working years for those individuals entering the labour market between 2005 and 2013.

The first message derived from the flow chart is that three out of four young entrants into the labour market hold bad jobs (73%) upon the moment of entry. Between the first and the fifth working year, 37% of individuals who started with bad jobs moved outside this category. Yet, 63% of them continued holding bad jobs five years later. On the other hand, only 20% of those who did not start with a bad job ended up with a bad job in $t + 5$. This is a first indication that the quality of a job when entering the labour market affects the situation in the medium term.

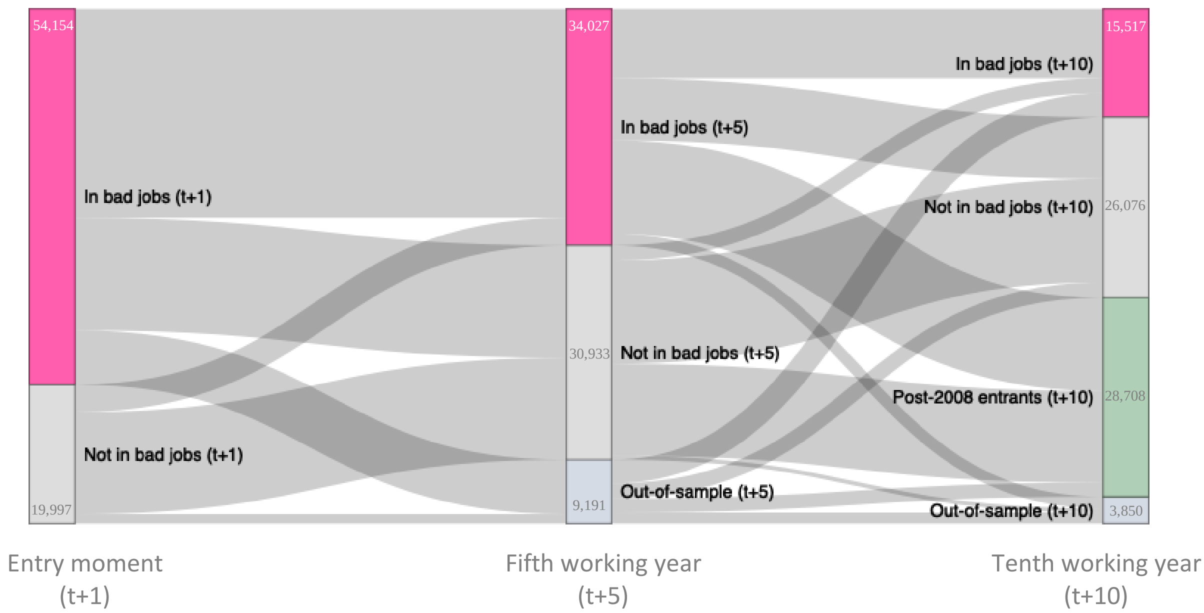
Turning to the longer term, about 38% of the workers with bad jobs in $t + 5$ continued with bad jobs in $t + 10$, whereas only 7% of those not in bad jobs in $t + 5$ transit to bad jobs in $t + 10$. Such dynamics suggests the presence of a scar generated by the involvement in bad jobs, particularly strong in the longer term. Those who exit bad jobs after five working years appear to have much better prospects of escaping from bad jobs in the long run. By gender, Appendix 2.B shows that bad jobs are mostly dominated by women, a share that increases with seniority.[23]

The flow chart also shows that some individuals are no longer present in the database in $t + 5$ or $t + 10$, meaning that they have no employment affiliation with the Social Security. This could be due to factors of a very diverse nature –ranging from emigration to regions outside the country to inactivity in the labour market– which the database unfortunately does not gauge.

[23] Of the three labour factors analysed, they are disproportionately represented within the group of workers holding part-time jobs, a share increases with working time, which is partly related to motherhood (Fernández-Kranz & Rodríguez-Planas, 2011; Webber & Williams, 2008). Similarly, a larger proportion of women than men earn low hourly wages, a gap that also increases with seniority.

The empirical analysis will, however, cover this by resorting to the Heckman correction method.

Figure 2.1. The Dynamics of Bad Jobs in Spain for Labour Market Entrants in the Period 2005-2013



Note. The data refer to job market entrants aged 16-30. Post-2008 entrants are included in a separate category in $t + 10$ given the lack of data availability for this cohort and later ones, since the last available year in the database is end-2019. The values on the bars reflect the total number of selected workers in each category.

As these dynamics may largely be affected by the presence of workers with relatively low productivity, Appendix 2.C (Figure 2.C1) replicates the exercise for college graduates, a population subgroup whose overall productivity is generally higher than that of the general population. A lower proportion of college graduates is involved in bad jobs compared to the general population, but the share is still substantial in magnitude, affecting 40% of graduates at entry (compared to nearly 75% for the general population). This evidences that bad jobs are extended in the overall Spanish youth, including college graduates, at the entry year. In addition, the dynamics suggest the presence of a scarring effect for college graduates: nearly three in every four workers involved in bad jobs in the medium term were also involved in such jobs at the moment of entry. This could suggest that the scarring effects might not be solely attributable to productivity, a matter that is empirically analysed below.

4.2. Empirical Estimates: Scarring Effects

The results in Table 2.2 support the existence of a significant scarring effect of bad jobs in the Spanish labour market. OLS estimates in Column 1 indicate that medium-term wages are, on average, 49.28% lower ($\approx 100 * (\exp(-0.679) - 1)$) for entrants in bad jobs compared to entrants in not bad jobs. This level is only slightly lower for college graduates (Column 2), at 43%, indicating that the scar still exists for this subgroup of potentially high-productivity workers. This fact would suggest that the scar is not solely the result of low productivity insofar as college graduates' productivity is generally high. However, this may not always be the case, as there may also be low-productivity workers within this population subgroup.

To test for this, we add Columns 3-6 to analyse whether the scar varies after tackling potential endogeneity issues in our model specification. Column 3 shows the first-stage results of the IV-TSLS estimation for the probability of holding a bad job at entry. The results suggest that the instrument is a good predictor of bad jobs: the coefficient is significant in explaining the endogenous variable, and the F-statistic is also high. Column 4 shows the IV-TSLS results after accounting for the potential endogeneity bias that the variable of interest might suffer from. The magnitude of the scar doubles in this setting compared to the OLS model.

Although this might seem counterintuitive, it is consistent with earlier literature (Kahn, 2010). For instance, Arellano-Bover (2022) finds that the IV estimate is four times larger than the OLS. In our setting, a possible explanation could be related to the endogenous timing of entry into the labour market, which may undermine the actual effects of entering the labour market in a bad job. For instance, if people extend their education in times where bad jobs are more prominent, and end it when good jobs dominate, the OLS estimates could be attenuated towards zero if this is uniformly distributed across workers (Schwandt & von Wachter, 2019). Lastly, as shown in Column 5, the addition of the Heckman two-step procedure to the IV-TSLS estimation yields a scar of a comparable magnitude to that of the IV-TSLS (fourth column).

The instrument applied to this procedure appears to be a good predictor of selection into the database.[24]

Table 2.2. Estimation Results, Scar of Bad Jobs on Medium-Term Wages

	OLS		IV-TSLS		
	(1) All	(2) College graduates	(3) First stage	(4) IV-TSLS	(5) IV-TSLS, Heckman
Estimated effects					
Bad job (t+1)	-0.679***	-0.561***		-1.462***	-1.481***
	(0.0104)	(0.0215)		(0.0703)	(0.070)
Instrumental variable (t+1)			0.0076***		
			(0.0002)		
F-statistic			339.16		
Observations	61,540	5,496	61,434	61,434	61,434

Note. *Statistically significant at the .10 level; ** at the .05 level; *** at the .01 level. Standard errors in parentheses. The dependent variable captures the (log) wages in t+5. The independent variable of interest is a binary variable that reflects whether individuals held a bad job in their first working year. All regressions control for sex, region of birth, year of birth, three educational levels and economic activity upon entry. Regressions are at the worker level. Column (1) show the results for an OLS regression, while (2) restricts this to college workers. Column (3) shows the results of the first-stage regression estimates for the probability of holding a bad job at entry. The instrumental variable refers to the share of individuals in bad jobs by region of birth, predicted graduation year and educational attainment, following a leave-one-out approach (Equation 4). Results of the IV-TSLS are shown in Column (4). Lastly, Column (5) replicates the results of the preceding column by adding the Inverse Mills Ratio to control for truncation bias in t+5, as defined in Equation 5.

Table 2.3 presents the estimation results for the long-term models. The key takeaway from this estimation is that long-term wages are determined by the medium-term involvement in bad jobs to a larger extent than the involvement upon entry. This conclusion holds both in the OLS and the IV-TSLS models. Taking the more moderate OLS estimates, the results reflect that individuals in bad jobs at entry have, on average, 30% lower wages over the longer term. In contrast, individuals in bad jobs over the medium term suffer from an average wage penalty of 41% over the long term. For college graduates, the magnitude of the scar is again relatively comparable to that of the overall population, although (1) the incidence of bad jobs is lower for this population subgroup, and (2) the low sample of this population subgroup for the long-term model requires certain caution on its interpretation. Turning to the IV-TSLS estimates yields similar conclusions to those of the medium-term analysis, with the estimates of initial

[24] The associated coefficient is significant in all specifications (also in the long-term model) and has the expected sign. Results available upon request.

conditions doubling those of the OLS models. Accounting for truncation bias practically does not change the magnitude of the estimate when compared to IV-TSLS.

Table 2.3. Estimation Results, Scar of Bad Jobs on Long-Term Wages

	OLS				IV-TSLS			
	(1) All (t+1)	(2) All (t+5)	(3) College graduates (t+1)	(4) College graduates (t+5)	(5) All (t+1)	(6) All (t+5)	(7) All, Heckman (t+1)	(8) All, Heckman (t+5)
Estimated effects								
Bad job (t+1)	-0.373*** (0.0108)		-0.429*** (0.0411)		-0.505*** (0.0588)		-0.502*** (0.0589)	
Bad job (t+5)		-0.529*** (0.009)		-0.601*** (0.0422)		-0.691*** (0.0586)		-0.670*** (0.0578)
F-statistic								
Obs.	39,077	35,006	1,546	1,466	38,985	34,764	38,985	34,764

Note. *Statistically significant at the .10 level; ** at the .05 level; *** at the .01 level. Standard errors in parentheses. The dependent variable captures the (log) wages in t+10. The independent variables of interest are binary and reflect whether individuals held a bad job in their first working year (t+1) and in the fifth one (t+5), respectively. All regressions control for sex, region of birth, year of birth, three educational levels, and economic activity on the fifth working year. Regressions are at the worker level. Columns (1) and (2) show the results for the long-term effects of holding a bad job at entry and on the fifth working year, respectively. Columns (3) and (4) replicate this for college graduates. Columns (5) to (8) present the IV-TSLS estimates. The instrumental variable refers to the share of individuals in bad jobs by region of birth, predicted graduation year and educational attainment, following a leave-one-out approach (Equation 5). Columns (7) and (8) add the Inverse Mills Ratio to the standard IV-TSLS regression as per Equation 6.

The Scarring Effects Throughout the Business Cycle

OLS estimates in Appendix 2.D reveal a clear cyclical trend of the estimated scar depending on the entry year. As the 2008 financial crisis approached, the scar of bad jobs became deeper on medium-term wages. For instance, workers on bad jobs in 2010 (i.e., entrants in 2009, in the midst of the recession) experienced a penalty of nearly 52% in their wages five years later. This compares with a lower scar for pre-crisis cohorts, which amounted to around 40%. As the recovery took hold, the magnitude of the scar again decreased. For instance, the estimated scar of bad jobs for entrants in 2013 was below 40%. These results are in line with the consensus in the literature that economic conditions at entry have long-lasting consequences in future labour trajectories (Oreopoulos et al., 2012; von Wachter, 2020).

4.3. Robustness Check: Scarring Effects Using an Alternative Definition of Bad Jobs

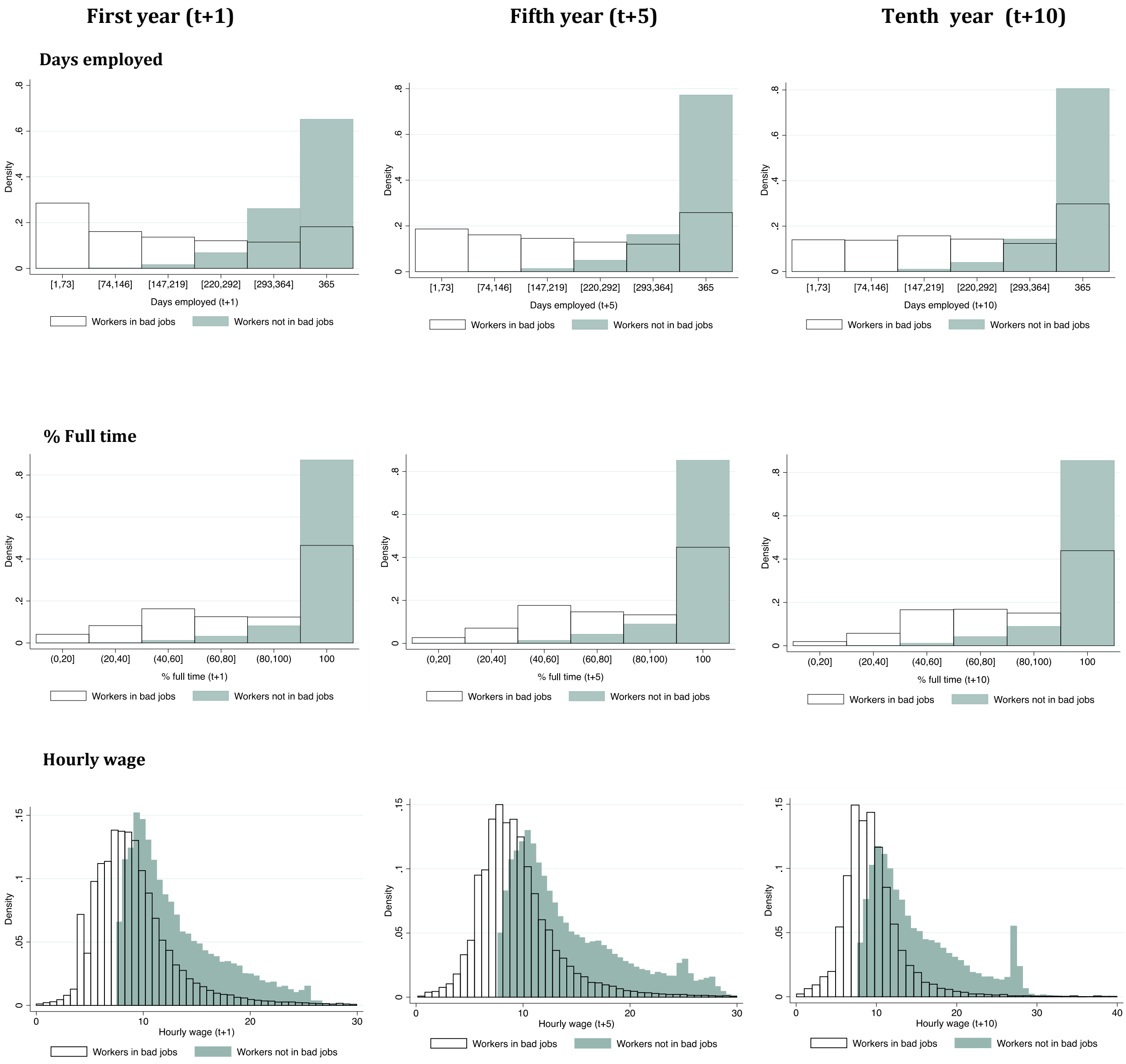
To test for the robustness of the results, we replicate the previous exercise modifying the threshold to define bad jobs by using a benchmark that amounts to 50% of the country's average earnings. Appendix 2.E shows that the overall results are nearly identical to those of the earlier threshold, with the scar being slightly lower in this alternative scenario. This is driven by a lower likelihood of holding bad jobs in the medium term given the relaxation of the threshold, which implies that fewer workers are affected by bad jobs; and those who are affected have a higher likelihood of exiting from them compared to the previous threshold. The long-term results convey a similar message, with the medium-term situation being the key determinant to the quality of jobs in the long term. In sum, the underlying messages remain with this alternative threshold, evidencing the robustness of our results.

5. Decomposing the Scar: Empirical Evidence

5.1. Descriptive Results

Taking separately the three components that determine annual wages, Figure 2.2 presents the distribution of these three factors for workers involved in bad jobs and the rest of workers. The first three panels show the number of days worked in a year, regardless of whether these were full time or not. Days are grouped in quintiles (i.e., 1-73 refers to workers employed 20% of the year). In the entry year, workers in bad jobs are mostly concentrated in the first category, implying that nearly 30% of workers in bad jobs worked at most 2.5 months in the entry period. This contrasts with the distribution of workers not in bad jobs, for whom the modal category refers to a 365-day working period at entry. In fact, over 60% of them are employed during the whole year, and more than 80% for around ten or more months of the year. Over the medium and long term, the amount of workers in bad jobs working during the whole year timidly increases to become the modal category, although the share is still significantly far from that recorded for the rest of workers (close to 80% for the latter, and around 30% for the earlier).

Figure 2.2. Distribution of Days Worked, % of Full-Time Hours Worked Daily and Average Hourly Wages (Entrants in 2005-2013)



Note. Post-2008 entrants are not included in t+10 given the lack of data availability for this cohort and later ones, since the last available year in the database is end-2019. Hourly wages are expressed in real euros of 2015.

Most workers who are not involved in bad jobs are employed full time, regardless of whether they work throughout the whole year. In fact, in the short, medium and long term, over 80% of them work full time, whereas this share of full-time workers decreases to just half for workers in bad jobs. These shares remain relatively constant across the three time periods, evidencing that even after gaining seniority, a large share of workers in bad jobs are still immerse in part-time jobs.

The distribution of hourly wages shows a clearly differentiated pattern between bad-job holders compared to workers not involved in bad jobs. Hourly wages for workers in bad jobs are largely concentrated below 10 euros per hour (in real terms of 2015), in contrast with the rightward distribution of workers not in bad jobs. This pattern applies for the entire time horizon.

Workers in bad jobs fulfil, by definition, at least one of the following conditions: being employed less than 365 days in one year, having part-time contracts, or having below-threshold average hourly wages. To further delve into the ways that these three factors interact, Appendix 2.F shows that workers with at least one non-employment spell are often involved in part-time jobs. This is the case for 34.1% of workers at entry, who have at least one part-time spell and are not employed for at least one day, even though their hourly wages are above threshold.[25] Separately, it is rare that workers in part-time jobs have bad jobs only for that reason: two in every three workers with at least one part time job also have at least one day of non-employment, and one in every five workers combine part-time jobs with low hourly wages and non-employment spells.[26]

5.2. Empirical Estimates: Decomposition of the Scarring Effects

Table 2.4 examines the relative importance of hourly wages, working hours and days in employment in explaining future wages. Given that, as seen before, the IV-TSLS gives rise to a larger scar, we take a conservative approach and resort to OLS in estimating these affects.

[25] In parallel, 21.2% of workers in bad jobs are not employed for at least one day, although all their jobs are full time and their average hourly wages are above threshold.

[26] Over the medium and long term, the labour factors affecting workers in bad jobs are, overall, similar to the entry year, with the exception of part-time working spells, which alone become more prominent. The presence of women in this group becomes larger as they age, which could largely be related to maternity (see Cech & Blair-Joy, 2019 for the role of parenthood in STEM employment).

Over the medium term, the results in Column 1 reveal that the days employed during the entry year is the most prominent variable in magnitude. This implies that earnings in the medium term are mostly marked by the number of days in employment at entry to a larger extent than on average working hours per day or average hourly wages. In fact, hourly wages appear as the least relevant component, with a magnitude that is more than ten times smaller.[27] This yields the first conclusion, which is that working intensity, and particularly the number of days worked, is a key determinant of medium-term wages. In other words, non-employment spells appear to substantially penalise future earnings and, therefore, increase the risk of incurring bad jobs in the medium term.

Over the long term, the results in Columns 3 and 5 show, in line with earlier findings, that the risk of lower earnings is mostly determined by the situation over the medium term, rather than that at entry. The key driver of long-term earnings is again related to the number of days worked, followed by average working hours per day. This implies that non-employment spells, followed by part-time working hours, are detrimental for future earnings to a much larger extent than hourly wages.

Focusing on college graduates yields relatively differentiated results compared to the whole population. The estimates in Columns 2, 4 and 6 reveal that hourly wages are the key determinant of future earnings in the medium and long run for college graduates. Daily working hours are generally high for college graduates: only one fourth does not have full-time jobs for the whole time employed at entry moment, a share that decreases to 10% over the medium and long term. Similarly, non-employment spells are much rarer for college graduates than for the remainder of the population: at entry, 75% work around eleven months, and 90% over the medium and long term. In sum, while average hourly wages for graduates are higher than for

[27] Results show that a one-standard-deviation increase in days employed at entry is associated with a 34.2% ($\approx 100 * (\exp(0.294) - 1)$) average increase in medium-term wages. The figure for a one-standard-deviation increase in average working hours per day amounts to 17.1%. Conversely, working hours per day at entrance do not appear as relevant, with a magnitude of around 1.4%.

the overall population, these are still heterogeneous across graduates. Because of this, hourly wages emerge as an important determinant for this population subgroup.

With respect to the other control variables, our conclusions align with previous analyses. Notably, female workers are more likely to have a low pay in the future, a feature that closely relates to the higher incidence of part-time jobs. Women are disproportionately represented within the group of workers holding part-time jobs. This share increases with time, partly related to motherhood (Fernández-Kranz & Rodríguez-Planas, 2011; Webber & Williams, 2008) or other care-related tasks.

Table 2.4. Estimation Results for the Drivers of Future Wage Prospects

	Effects in t+5		**Effects in t+10**			
	(1) All	(2) College graduates	(3) All	(4) College graduates	(5) All	(6) College graduates
Days worked (t+1)	0.294***	0.215***	0.128***	0.114***		
	(0.00453)	(0.0128)	(0.00503)	(0.0246)		
Working hours per day (t+1)	0.158***	0.189***	0.0701***	0.112***		
	(0.00448)	(0.0127)	(0.00521)	(0.0265)		
Hourly wage (t+1)	0.0138***	0.220***	0.00608*	1.657***		
	(0.00371)	(0.0406)	(0.00320)	(0.210)		
Days worked (t+5)					0.179***	0.178***
					(0.00428)	(0.0212)
Working hours per day (t+5)					0.131***	0.181***
					(0.00442)	(0.0222)
Hourly wage (t+5)					0.0123***	2.429***
					(0.00319)	(0.153)
Observations	61,540	5,496	39,077	1,546	35,006	1,466

Note. *Statistically significant at the .10 level; ** at the .05 level; *** at the .01 level. Standard errors in parentheses. The dependent variable captures the (log) wages in t+5 and t+10 (Columns 1-2 and 3-4, respectively). The independent variables of interest are continuous and reflect the average daily hours worked, days employed, and average hourly wages in t+1 and t+5, respectively. These are expressed in the same unit, as they are standardised by subtracting the sample mean and dividing it by the standard deviation. The resulting variables have zero mean and one standard deviation. All regressions control for sex, region of birth, year of birth, three educational levels and economic activity upon entry. Regressions are at the worker level. Column (1) and (2) show the OLS results for the medium-term wages, first for the whole population and then to college workers. Columns (3) and (4) replicate the exercise for long-term wages as a function of the three aforementioned components in t+1 and t+5, respectively.

Days in employment, in turn, are a key determinant of future earnings for the overall workforce, more so than average working hours per day or hourly wages, which are dampened by the existence of a minimum wage and hence have lower relative importance for the overall workforce compared to work intensity.

6. Conclusions

Using large administrative data on workers affiliated to the Social Security since 2006, we have analysed the scarring effects of bad jobs over their medium- and long-term labour trajectories (five and ten years after entry, respectively). Bad jobs, defined as those that yield annual earnings below 60% of the average as per the European Social Charter, affect three out of four young Spaniards at the moment of entry.

In this paper, we find evidence of a scarring effect of bad jobs in workers' fifth and tenth working years. Importantly, the scar does not appear to be driven by non-random sorting, which is tested following an instrumental-variables approach. The results show that individuals in bad jobs at the entry year may experience wage penalties that amount to 50% over the medium term (in the fifth working year). These results are robust to an alternative definition of bad jobs. A comparable pattern emerges when analysing the long term. In this case, the situation in the medium term determines the probability of holding bad jobs over the long term, and the entry condition is relatively less critical in explaining long-term wages. Exploring the drivers of the scar yields that non-employment spells are key determinants of future wages, followed by low daily working hours. However, hourly wages have a much less prominent role, likely because the existence of minimum wages partly dampens this effect. Lastly, the scarring effects appear to be markedly sensitive to the cycle: individuals starting with bad jobs during the crisis have a higher estimated scar as compared to the pre-crisis cohorts.

The results contribute to the literature on the effects of low-quality jobs on workers' employment trajectories by exploring some labour factors that contribute to those effects. The overall conclusions on the scarring effects concur with earlier literature, but the relative importance of its drivers cannot be compared to previous studies as, to our knowledge, this is the first time that this approach is undertaken. A key implication of the study lies on the importance of focusing on policies that tackle labour intensity and, notably, the continuous inflows and outflows into and out of employment, a feature that characterises the Spanish labour market.

References

Altonji, J.G., Kahn, L.B., & Speer, J.D. (2016). Cashier or Consultant? Entry Labor Market Conditions, Field of Study, and Career Success. *Journal of Labor Economics, 34(S1),* S364-401. https://doi.org/10.1086/682938

Arellano-Bover, J. (2022). Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size. *Journal of Labor Economics.* Advance online publication. https://doi.org/10.1086/723500

Bentolila, S., Felgueroso, F., Jansen, M., & Jimeno, J.F. (2022). Lost in recessions: youth employment and earnings in Spain. *SERIEs*, *13*, 11-49. https://doi.org/10.1007/s13209-021-00244-6

Brunner, B., & Kuhn, A. (2014). The impact of labor market entry conditions on initial job assignment and wages. *Journal of Population Economics, 27*, 705-738. https://doi.org/10.1007/s00148-013-0494-4

Campbell, I., & Price, R. (2016). Precarious work and precarious workers: Towards an improved conceptualisation. *The Economic and Labour Relations Review*, *27*(3), 314–332. https://doi.org/10.1177/1035304616652074

Card, D. (2019). Introduction: A Good Start? Determinants of Initial Labor Market Success. *Journal of Labor Economics*, *37(S1)*, S1–S9. https://doi.org/10.1086/701217

Cech, E.A., & Blair-Loy, M. (2019). The changing career trajectories of new parents in STEM. *Proceedings of the National Academy of Sciences, 116(10),* 4182-187. https://doi.org/10.1073/pnas.1810862116

Cockx, B., & Ghirelli, C. (2016). Scars of recessions in a rigid labor market. *Labour Economics*, *41,* 162-176. https://doi.org/10.1016/j.labeco.2016.05.009

Council of Europe (2015). *European Social Charter. Collected Texts (7th Edition) (updated to 1st January 2015*). Article 455, Scope of the provisions as interpreted by the ECSR. https://rm.coe.int/168048b059

De Fraja, G., Lemos, S., & Rockey, J. (2021). The Wounds That Do Not Heal: The Life-Time Scar of Youth Unemployment. *Economica*, *88(352),* 896-941. https://doi.org/10.1111/ecca.12384

de la Rica, S., & Iza, A. (2005). Career planning in Spain: Do fixed-term contracts delay marriage and parenthood? *Review of Economics of the Household, 3*, 49–73. https://doi.org/10.1007/s11150-004-0979-8

Dickinson, D.L. (1999). An Experimental Examination of Labor Supply and Work Intensities. *Journal of Labor Economics, 17(4),* 638-670. https://doi.org/10.1086/209934

European Parliament (2022). *Proposal for a directive of the European Parliament and the Council*. https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733535/EPRS_BRI(2022)733535_EN.pdf

Fernández-Kranz, D., & Rodríguez-Planas, N. (2011). The part-time pay penalty in a segmented labor market. *Labour Economics, 18(5)*, 591-606. https://doi.org/10.1016/j.labeco.2011.01.001

Fernández-Kranz, D., & Rodríguez-Planas, N. (2018). The Perfect Storm: Graduating during a Recession in a Segmented Labor Market. *ILR Review, 71(2)*. https://doi.org/10.1177/001979391771420

Font, P., Izquierdo, M., & Puente, S. (2015). Real wage responsiveness to unemployment in Spain: asymmetries along the business cycle. *IZA Journal of European Labor Studies, 4(13)*, 1-13. https://doi.org/10.1186/s40174-015-0038-x

Gangl, M. (2006). Scar Effects of Unemployment: An Assessment of Institutional Complementarities. *American Sociological Review*, *71*(6), 986–1013. https://doi.org/10.1177/000312240607100606

Garcia-Louzao, J., Hospido, L., & Ruggieri, A. (2023). Dual returns to experience. *Labour Economics*, *80*, 102290. https://doi.org/10.1016/j.labeco.2022.102290

García-Pérez, C., Prieto, M., & Simón, H. (2020). Multidimensional measurement of precarious employment using hedonic weights: Evidence from Spain. *Journal of Business Research, 113*, 348–359. https://doi.org/10.1016/j.jbusres.2019.09.036

Gardecki, R., & Neumark, D. (1998). Order from Chaos? The Effects of Early Labor Market Experiences on Adult Labor Market Outcomes. *ILR Review, 51(2)*, 299-322. https://doi.org/10.1177/001979399805100209

Heckman, J.J. (1979). Sample selection bias as a specification error. *Econometrica. 47(1)*, 153–161.

Kahn, L.B. (2010). The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics, 17(2)*, 303-316. https://doi.org/10.1016/j.labeco.2009.09.002

Kahn, L.B. (2013). Asymmetric Information between Employers. *American Economic Journal: Applied Economics*, *5(4)*, 165-205. https://doi.org/10.1257/app.5.4.165

Kalleberg, A. (2009). Precarious work, insecure workers: Employment relations in transition. *American Sociological Review, 74(1)*, 1–22. https://doi.org/10.1177/000312240907400101

Liu, K., Salvanes, K.G., & Sørensen, Ø. (2016). Good skills in bad times: Cyclical skill mismatch and the long-term effects of graduating in a recession. *European Economic Review, 84*, 3-17. https://doi.org/10.1016/j.euroecorev.2015.08.015

Mauno, S., Cheng, T., & Lim, V. (2017). The Far-Reaching Consequences of Job Insecurity: A Review on Family-Related Outcomes. *Marriage & Family Review, 53(8)*, 717-743. https://doi.org/10.1080/01494929.2017.1283382

Mavromaras, K., Sloane, P., & Wei, Z. (2015). The scarring effects of unemployment, low pay and skills under-utilization in Australia compared. *Applied Economics, 47(23)*. https://doi.org/10.1080/00036846.2015.1008762

McCall, L. (2000). Gender and the New Inequality: Explaining the College/Non-College Wage Gap. *American Sociological Association, 65(2)*, 234-255. https://doi.org/10.2307/2657439

Mousteri, V., Daly, M., & Delaney, L. (2018). The scarring effect of unemployment on psychological well-being across Europe. *Social Science Research, 72*, 146-169. https://doi.org/10.1016/j.ssresearch.2018.01.007

Olsthoorn, M. (2014). Measuring precarious employment: A proposal for two indicators of precarious employment based on set-theory and tested with Dutch labor market-data. *Social Indicators Research, 119(1)*, 421–441. https://doi.org/10.1007/s11205-013-0480-y

Oreopoulos, P., von Wachter, T., & Heisz, A. (2012). The Short-and Long-Term Career Effects of Graduating in a Recession. *American Economic Journal: Applied Economics, 4(1)*, 1-29.

Raaum, O., & Røed, K. (2006). Do Business Cycle Conditions at the Time of Labor Market Entry Affect Future Employment Prospects? *The Review of Economics and Statistics, 88(2)*, 193-210. https://doi.org/10.1162/rest.88.2.193

Rodgers, G., & Rodgers, J. (1989). Precarious jobs in labour market regulation: The growth of atypical employment in Western Europe. International Labour Organisation. https://www.ilo.org/public/libdoc/ilo/1989/89B09_333_engl.pdf

Schwandt, H., & von Wachter, T. (2019). Unlucky Cohorts: Estimating the Long-Term Effects of Entering the Labor Market in a Recession in Large Cross-Sectional Data Sets. *Journal of Labor Economics, 37(S1)*, S161–S198. https://doi.org/10.1086/701046

Tridico, P. (2012). Financial crisis and global imbalances: its labour market origins and the aftermath. *Cambridge Journal of Economics, 36(1)*, 17-42. https://doi.org/10.1093/cje/ber031

Vignoli, D., Mencarini, L., & Alderotti, G. (2020). Is the effect of job uncertainty on fertility intentions channeled by subjective well-being? *Advances in Life Course Research, 46*, 100343. https://doi.org/10.1016/j.alcr.2020.100343

Von Watcher, T. (2020). The Persistent Effects of Initial Labor Market Conditions for Young Adults and Their Sources. *Journal of Economic Perspectives, 34(4)*, 168-194. https://doi.org/10.1257/jep.34.4.168

Von Wachter, T., & Bender, S. (2006). In the Right Place at the Wrong Time: The Role of Firms and Luck in Young Workers' Careers. *American Economic Review, 96(5)*, 1679-1705. https://doi.org/10.1257/aer.96.5.1679

Webber, G., & Williams, C. (2008). Mothers in “Good” and “Bad” Part-time Jobs: Different Problems, Same Results. *Gender & Society, 22(6)*, 752-777. https://doi.org/10.1177/08912432083256

Wolfolds, S. E., & Siegel, J. (2018). Misaccounting for Endogeneity: The Peril of Relying on the Heckman Two-Step Method without a Valid Instrument. *Strategic Management Journal, 40(3)*, 432-462. https://doi.org/10.1002/smj.2995

Appendix

Appendix 2.A

Summary Statistics of the Sample

Table 2.A1. Summary Statistics of the Sample (Entrants in 2005-2013)

	%	N	
Gender			
Female	53.9	39,962	
Male	46.11	34,189	
Education			
High school	34.88	25,305	
Vocational	32.92	23,883	
College graduates	32.2	23,364	
Year of birth			
1975-1979	8.53	6,330	
1980-1984	21.39	15,870	
1985-1989	42.59	31,581	
1990-1994	26.00	19,287	
1995+	1.46	1,083	
Region of birth			
Andalucía	13.67	9,972	
Aragón	1.87	1,366	
Asturias	1.95	1,423	
Baleares	1.49	1,087	
Canarias	3.45	2,513	
Cantabria	1.06	775	
Castilla y León	4.49	3,272	
Castilla-La Mancha	3.12	2,278	
Cataluña	10.86	7,918	
Comunidad Valenciana	6.47	4,721	
Extremadura	2.06	1,500	
Galicia	5.3	3,869	
Comunidad de Madrid	10.55	7,696	
Murcia	2.06	1,505	
Navarra	1.03	754	
País Vasco	4.00	2,921	
La Rioja	0.37	273	
Ceuta and Melilla	0.3	219	
Foreign	25.88	18,874	
Sector (%)	**t+1**	**t+5**	**t+10**
Agro-fishing	0.69	0.63	0.51
Extractive industry	0.07	0.08	0.09
Manufacturing industry	8.24	9.09	10.83
Energy, gas, water	0.43	0.58	0.75
Construction	7.67	4.95	5.11
Commerce	21.43	21.04	20.4
Transport, storage	2.08	2.71	3.54
Hospitality	14.9	14.27	12.02

ICT	3.1	3.93	4.03
Financial/insurance activities	1.26	1.72	2.24
Real estate activities	0.43	0.46	0.52
Prof.,scientific, tech. activities	5.91	6.37	6.1
Administrative activities	11.09	8.87	9.24
Public admin., Social Security	3.41	3.47	3.69
Education	4.67	5.16	5.13
Health	6.78	8.83	9.13
Arts	3.29	2.55	1.68
Other services	3.16	2.86	2.63
Domestic activities	1.37	2.37	2.32
Extraterritorial activities	0.03	0.03	0.03

Note. The reduction in the sample size in t+10 is the result, on the one hand, of individuals no longer affiliated to the Social Security and employed (as is also the case in t+5) and also due to missing information on post-2008 entrants. The total sample size is 74,151 in t+1; 64,960 in t+5; and 41,593 in t+10.

Appendix 2.B

Profiling Workers in Bad and Not-Bad Jobs

Table 2.B1. Sociodemographic Characteristics of Workers in Bad and Not Bad Jobs (% of the Corresponding Subgroup)

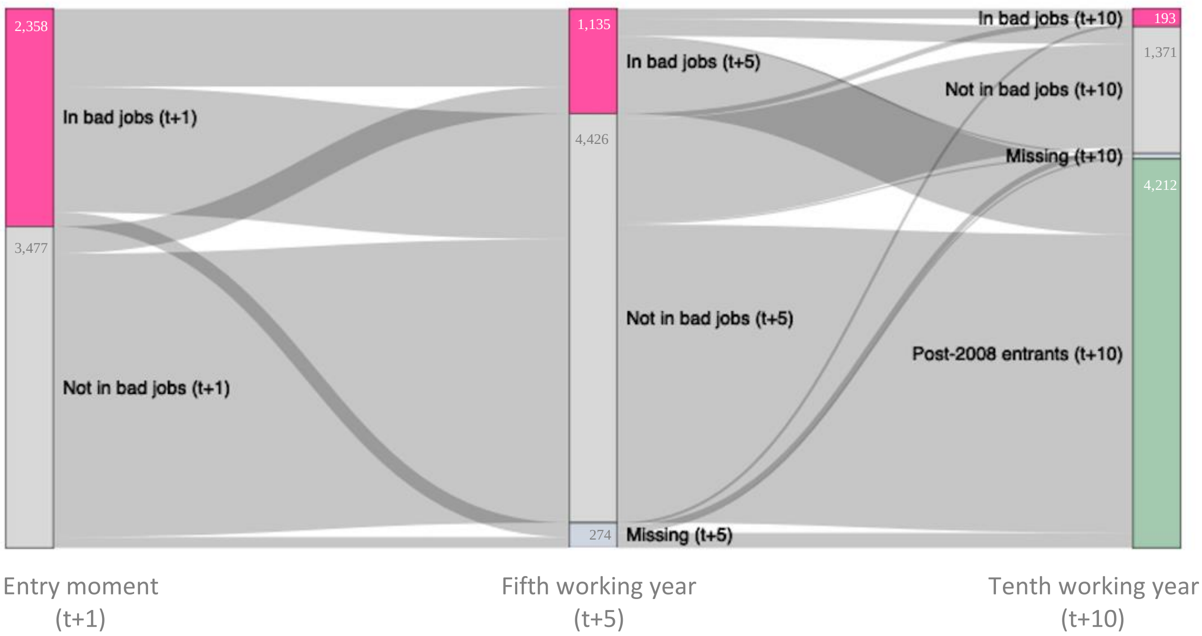
	Total	Not in bad jobs	In bad jobs
Gender (% women)			
t+1	53.89	44.90	57.21
t+5	54.38	49.17	59.11
t+10	52.16	47.32	60.31
Education			
t+1			
High school	34.88	27.43	37.61
Vocational	32.92	28.32	34.6
College	32.2	44.25	27.79
t+5			
High school	33.65	24.87	41.66
Vocational	33.46	31.91	34.86
College	32.89	43.22	23.47
t+10			
High school	36.02	27.96	49.67
Vocational	32.67	32.12	33.59
College	31.32	39.93	16.73
Place of birth (% foreign)			
t+1	25.88	30.96	24.01
t+5	25.53	22.95	27.89
t+10	24.65	20.77	31.2

Note. The table includes all the Spanish entrants aged 16-30 who first accessed the labour market between 2005 and 2013.

Appendix 2.C

Bad Jobs for University Graduates, Descriptive Evidence

Figure 2.C1. The Dynamics of Bad Jobs for College Graduates in Spain for Labour Market Entrants in the Period 2005-2013



Note. The data refer to job market entrants aged 16-30 for the period 2005-2013. Post-2008 entrants are included in a separate category in $t + 10$ given the lack of data availability for this cohort and later ones, since the last available year in the database is end-2019. The numbers on the bars reflect the total number of selected workers in each category.

Appendix 2.D

Scarring Effects of Bad Jobs by Predicted Graduation Year

Table 2.D1. Estimation Results of the Scar of Bad Jobs at Entry on Medium-Term Wages by Year of Entry to the Labour Market

	OLS	**Observations**	**IV-TSLS**	**Observations**
2005	-0.556***	11,984	-1.716***	11,945
	(0.0214)		(0.130)	
2006	-0.667***	9,627	-1.749***	9,596
	(0.0265)		(0.158)	
2007	-0.670***	8,275	-1.535***	8,263
	(0.0304)		(0.167)	
2008	-0.679***	5,946	-1.613***	5,935
	(0.0356)		(0.205)	
2009	-0.720***	4,335	-1.388***	4,332
	(0.0409)		(0.224)	
2010	-0.726***	4,966	-1.799***	4,961
	(0.0380)		(0.250)	
2011	-0.582***	5,227	-1.606***	5,227
	(0.0338)		(0.222)	
2012	-0.548***	4,861	-1.327***	4,860
	(0.0357)		(0.202)	
2013	-0.503***	6,319	-0.737***	6,315
	(0.0296)		(0.189)	

Note. *Statistically significant at the .10 level; ** at the .05 level; *** at the .01 level. Standard errors in parentheses. The dependent variable captures the (log) wages in t+5. The independent variable of interest is binary and reflects whether individuals held a bad job at the moment of entry into the labour market. In the IV-TSLS, this is instrumented with the share of individuals in bad jobs by region of birth, predicted graduation year and educational attainment, following a leave-one-out approach (Equation 4). The models are estimated separately by entry year.

Appendix 2.E

Robustness Test

Table 2.E1. Robustness Test: Estimation Results (OLS) of the Scar Using an Alternative Threshold for Bad Jobs

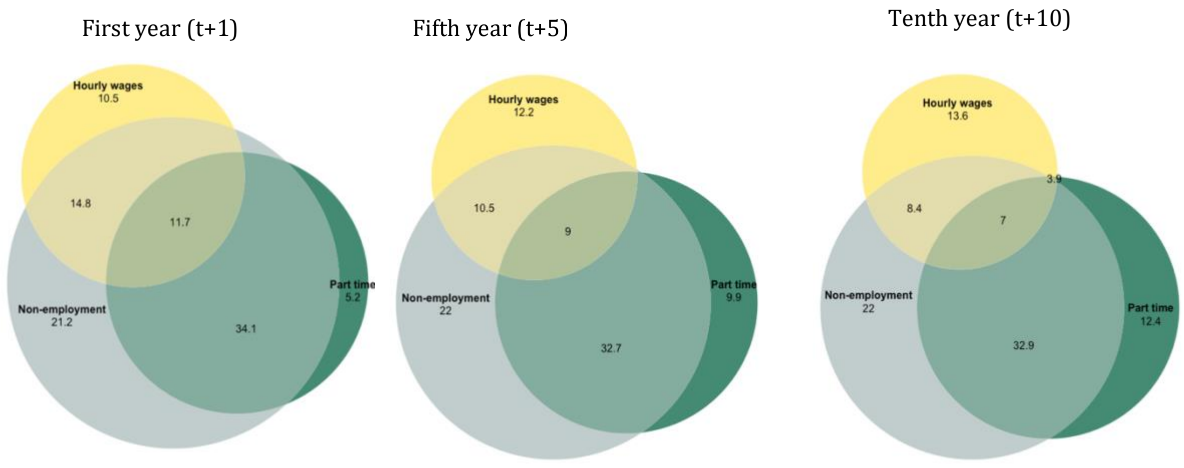
	Bad jobs, t+5		**Bad jobs, t+10**			
	(1) Alternative: Threshold 50% average wage	(2) Baseline: Threshold 60% average wage	(3) Alternative	(4) Baseline	(5) Alternative	(6) Baseline
Bad job (t+1)	-0.659***	-0.679***	-0.337***	-0.373***		
	(0.00978)	(0.0104)	(0.0103)	(0.0108)		
Bad job (t+5)					-0.513***	-0.529***
					(0.00901)	(0.009)
Observations	61,540	61,540	39,077	39,077	35,006	35,006

Note. *Statistically significant at the .10 level; ** at the .05 level; *** at the .01 level. Standard errors in parentheses. The dependent variable captures the (log) wages in t+5 (Columns 1-2) and t+10 (Columns 3-6). The independent variables of interest are binary and reflect whether individuals held a bad job in the past. All regressions control for sex, region of birth, year of birth, three educational levels and economic activity on the fifth working year. Regressions are at the worker level.

Appendix 2.F

Labour Characteristics of Workers in Bad and Not in Bad Jobs

Figure 2.F1. Workers in Bad Jobs: Percentage With at Least One Non-Employment Spell, at Least One Part-Time Job and/or Below-Threshold Hourly Wages (Entrants in 2005-2013)

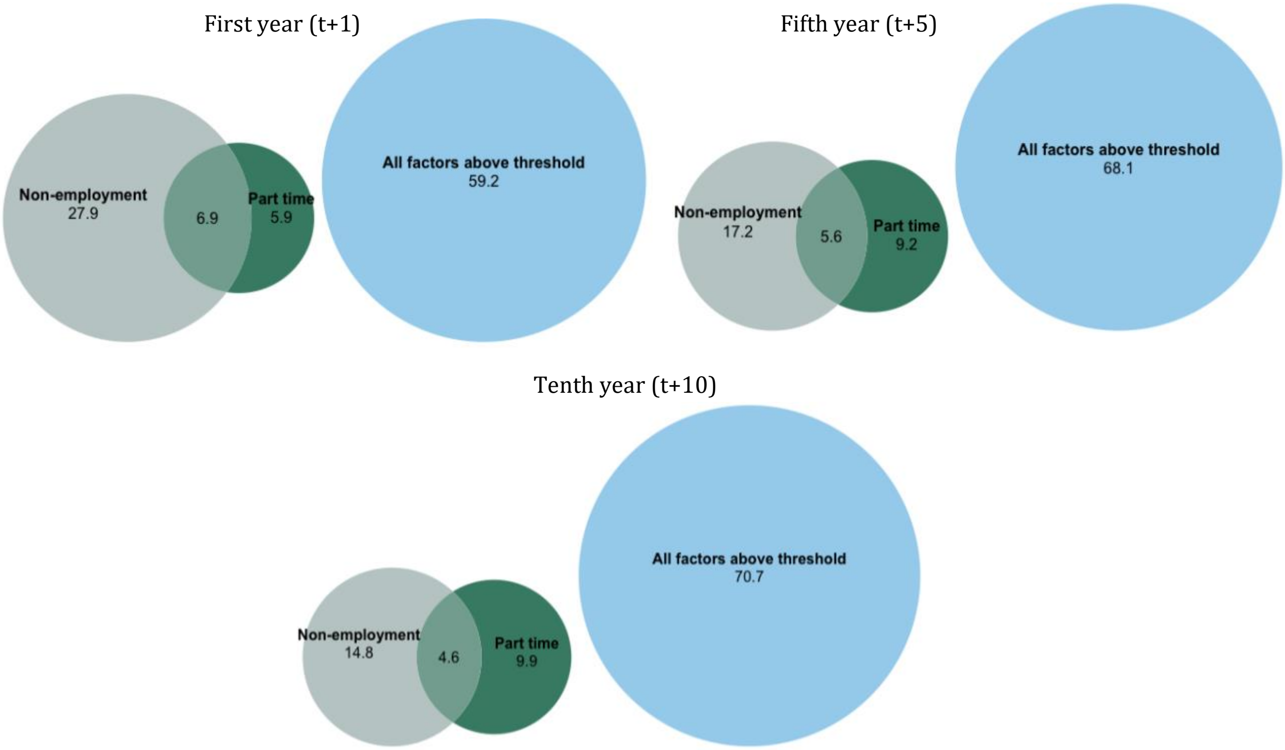


Note. The figure shows the share of workers in bad jobs who fulfil at least one of the following conditions: below-threshold hourly wages, at least one day without employment and at least one day working part time. Thresholds for hourly wages are shown in the third column of Table 2.1. Post-2008 entrants are not included in $t + 10$ given the lack of data availability for this cohort and later ones, since the last available year in the database is end-2019.

Figure 2.F2 shows the share of workers not in bad jobs who have at least one non-employment spell and/or at least one part-time job versus workers employed full-time for the whole year. The figure contains information on entrants between 2005 and 2013, except for the tenth working year, where information on entrants is up to 2008 (as the last available data is for 2019 and hence there is no information available ten years after entry for entrants post-2008). The results evidence that the majority of workers not in bad jobs have full-time contracts and work throughout the whole year. Besides, by definition, their hourly wage is above threshold. In fact, we find that around 59% of those workers fulfil those three conditions, which we refer to as "good jobs". Over time, the proportion of workers in good jobs increases to 68% in the medium term and 71% in the long term. Regarding the remaining workers who are not in good jobs (but also not in bad jobs), either because of non-employment or part-time spells, we find that a larger proportion of them have at least one non-employment spell compared to part-time jobs. However, this gap decreases as workers gain seniority; on the one hand, as workers transit

to good jobs and, on the other, as part-time spells become frequent for some workers, as was the case for workers in bad jobs.

Figure 2.F2. Workers Not in Bad Jobs: Percentage With at Least One Non-Employment Spell and/or at Least One Part-Time Job Versus Workers Employed Full-Time the Whole Year



Note. The figure shows the share of workers not in bad jobs (who entered the labour market between 2005 and 2013) who fulfil at least one of the following conditions: at least one day without employment and/or at least one day working part time, versus workers with all factors above threshold, that is, who work full time for the 365-day period. By definition, workers not in bad jobs have hourly wages above the thresholds shown in Table 2.1.

Chapter 3

The (In)Evitable Effects of Educational Presorting on Gender Segregation in the Labour Market

1. Introduction

Gender segregation remains a pervasive feature of global labour markets, despite changes in social norms and the sharp rise in female labour force participation in recent decades. The concentration of men and women in different occupations and sectors can have severe economic and social consequences. From an economic efficiency standpoint, segregation leads to an under-utilisation of the skills of the population. This is a particularly pressing concern in the current context of polarisation, where labour shortages are exacerbated in segregated occupations (Miller et al., 2004). From an equity perspective, segregation remains the main factor in understanding the gender pay gap (Bishu & Alkadry, 2017). While the consequences of gender segregation have long been studied, there is less consensus on the drivers of this phenomenon. This is because segregation stems from a complex combination of gender differences in multiple, often hard-to-observe factors, such as job skills, preferences, discrimination or social norms (Antecol & Cobb-Clark, 2013; Wang & Degol, 2013; Wiswall & Zafar, 2018).

In response to the negative effects of gender segregation, policy actions have been implemented to target potentially underlying determinants of this phenomenon. In the educational domain, the literature has sought to identify successful interventions aimed at reducing gender segregation in educational choices. These studies encompass diverse methodologies, including randomised controlled trials (Breda et al., 2023), natural experiments (Azmat & Iriberri, 2010; Azmat et al., 2019) or quasi-experimental interventions (Fernández-Cézar et al., 2020). In addition, other policies have targeted later stages of individuals' labour trajectories. These include policies designed to reallocate time within households and other demand-side measures related to the recruitment, selection, hiring, evaluation and promotion of individuals in the labour market (Carranza et al., 2023). However, the persistence of segregation, despite concerted efforts to combat it, underscores the need for further policy action.

The aim of this paper is, first, to shed light on the determinants of gender segregation in the Spanish labour market and, then, to test a possible solution to tackle this phenomenon. We argue that a comprehensive understanding of the underlying drivers of gender segregation is essential for the development of effective interventions. Spain offers a relevant case study as the level of gender segregation has stagnated over the last two decades. This has contributed to the gender pay gap (Amuedo-Dorantes & de la Rica, 2006), which persists even if women already outnumber men in higher education in the country.

To address the first aim of the paper, we design and implement a large-scale online survey aimed at covering information on the drivers of segregation beyond what is typically captured in conventional surveys or administrative databases. The sample consists of nearly 5,000 representative individuals in Spain aged between 18 and 49 years old. Drawing on Eccles' Expectancy Value Theory (1983; 2009), we empirically test the relationship between gender segregation and a wide range of contextual, psychological and aspirational factors. The survey's detailed information on individuals' educational choices allows us to examine the determinants not only of occupational choices, but also of earlier academic choices. In particular, we separately analyse the role of STEM (science, technology, engineering and mathematics) and HEAL (health, education, administration and literacy) studies. This distinction is motivated by the divergent gender composition in these fields. While STEM fields typically have a male predominance, along with above-average salaries and favourable working conditions, HEAL fields have a female overrepresentation and entail lower salaries, higher labour intensity and tasks traditionally associated with femininity, such as caring roles in health and education.

Our results show that the field of study is the main driver of occupational segregation. Specifically, engagement in HEAL education significantly increases the likelihood of women pursuing occupations characterised by a notable female presence. Conversely, participation in STEM education is associated with a lower probability of women working in female-dominated occupations. For men, STEM education correlates with a higher average probability of engagement in male-dominated occupations, while HEAL is associated with an increased probability of occupations less dominated by men. Beyond the field of study, our analysis identifies psychological factors as influential determinants of occupational segregation. In particular, women with higher levels of self-concept are more likely to work in occupations that are not dominated by women. Furthermore, additional analysis on the determinants of educational choices shows that math anxiety during adolescence is the factor that more negatively affects the choice of STEM-related studies.

After identifying the determinants of gender segregation, we test a potential solution based on the findings that educational presorting is the most relevant driver of this phenomenon. To answer this second aim of the paper, we design and conduct a second online survey addressed to 600 pre-university individuals aged between 15 and 18 years old. The survey includes a randomised controlled trial featuring a role model intervention in the field of mathematics. The rationale for this intervention is based on the premise that improving entrenched negative stereotypes about mathematics, which are particularly prominent among girls, can contribute to reducing future segregation in the labour market. The intervention consists of an online video —randomly targeted to half of the sample—and focuses on the work of mathematicians and its importance in the current global context; the social usefulness of mathematics; and the importance of gender balance in mathematics.

The experimental results reveal the potential of role model interventions in changing preconceived perceptions that may discourage adolescents from engaging in maths-related careers. In particular, our results show that the role model intervention has a positive impact on young people's perception of mathematics as a useful tool that can be applied to everyday life and to global challenges. The impact is stronger for women, for whom this idea was less prevalent than men. The intervention also improves individuals' self-projection into math-related jobs. Specifically, treatment effects for female adolescents are a third of an SD higher than those for their male counterparts. This is particularly relevant given that women appear to be particularly reluctant to self-project in such jobs. Separately, the intervention fosters women's idea that math-related jobs allow for a balance between personal and professional life, and it changes women's established view that math-related jobs have little social impact. For men, the role model intervention also increases reported levels of growth mindset. However, the intervention does not significantly change values that may be deeply ingrained in individuals, namely their self-concept in mathematics or the support for promoting female participation in STEM, in line with previous literature (Breda et al., 2023).

This paper contributes to a growing body of literature on gender segregation in several ways. First, this paper adopts a holistic approach to analysing gender segregation through a mix of contextual, psychological and aspirational factors. This contrasts with much of the previous literature, which typically focuses on selective factors (Tandrayen-Ragoobur & Gokulsing, 2022). We overcome some of these limitations by designing a survey addressed to a nationally representative sample that covers factors of a broad nature, allowing us to identify their relative importance in understanding gender segregation in the labour market. Additionally, while much of the previous research focuses on the role of tangible factors such as income, educational attainment or macroeconomic indicators (Borrowman & Klasen, 2019), our study recognises the importance of intangible factors in shaping future occupational choices (Wang & Degol, 2013). Second, to the best of our knowledge, this is the first time that a holistic approach to the causes of gender segregation has been explored for the Spanish case. This contrasts with the national literature, which mainly focuses on the incidence of tangible factors measured in conventional surveys (e.g. Dueñas Fernández et al., 2014). Third, this paper conducts a causal evaluation of a low-cost, easy-to-implement intervention, potentially providing the basis for future public policy in STEM. This novel approach provides valuable evidence in a country where evaluations of STEM education policies are scarce.

2. Context, Theoretical Framework and Literature

Gender segregation in the labour market refers to the uneven distribution of men and women in occupations, sectors, or levels of responsibility (Duncan & Duncan, 1955; Weeden, 1998). Segregation can take two main forms: vertical or horizontal. Vertical segregation refers to the unequal distribution of gender at different levels of responsibility within the same occupation or sector (Levanon & Grusky, 2016). Conversely, horizontal segregation refers to the unequal concentration of men and women in certain occupations or sectors (Kamerāde & Richardson, 2018). The focus of the following subsection is on horizontal segregation in the Spanish labour market.

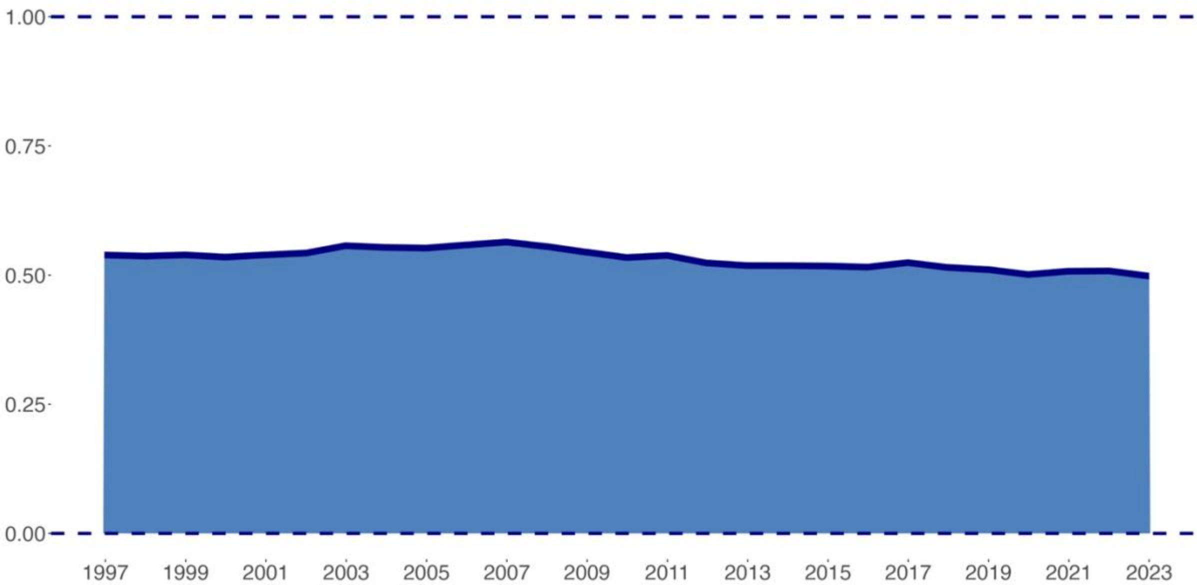
2.1. Gender Segregation in the Spanish Labour Market

In the Spanish labour market, numerous occupations continue to exhibit significant gender imbalances, reflecting the persistent dynamics of gender segregation (Dueñas Fernández et al., 2014). Figure 3.1 shows the gender composition of occupations over the last two decades in Spain, revealing that the level of segregation has remained virtually unchanged in every year between 1997 and 2023. Specifically, the Index of Dissimilarity (Duncan & Duncan, 1955)—which quantifies the intensity of segregation—amounts to 0.50 in 2023, only marginally lower than the value of 0.54 recorded in 1997.[28] These findings underscore the limited reduction in segregation despite changes in social norms and the sharp rise in female participation in the

[28] Formally, let ID denote the Duncan and Duncan (1955) Index of Dissimilarity. The index is calculated by comparing, for each occupation i, the number of men and women employed in that occupation (m_i and f_i, respectively) out of the total number of men and women employed in the labour market (M and F, respectively). The absolute value of this difference is summed for all existing occupations ($i = 1, \dots, N$), and the value is multiplied by $1/2$, so that the index ranges between 0 and 1: $ID = \frac{1}{2}\sum_{i=1}^{N}\left|\frac{m_i}{M} - \frac{f_i}{F}\right|$.

labour market. Compared to the European Union average, Spain's Index of Dissimilarity is slightly above average.[29]

Figure 3.1. Segregation in Spain in 1997-2023, Index of Dissimilarity

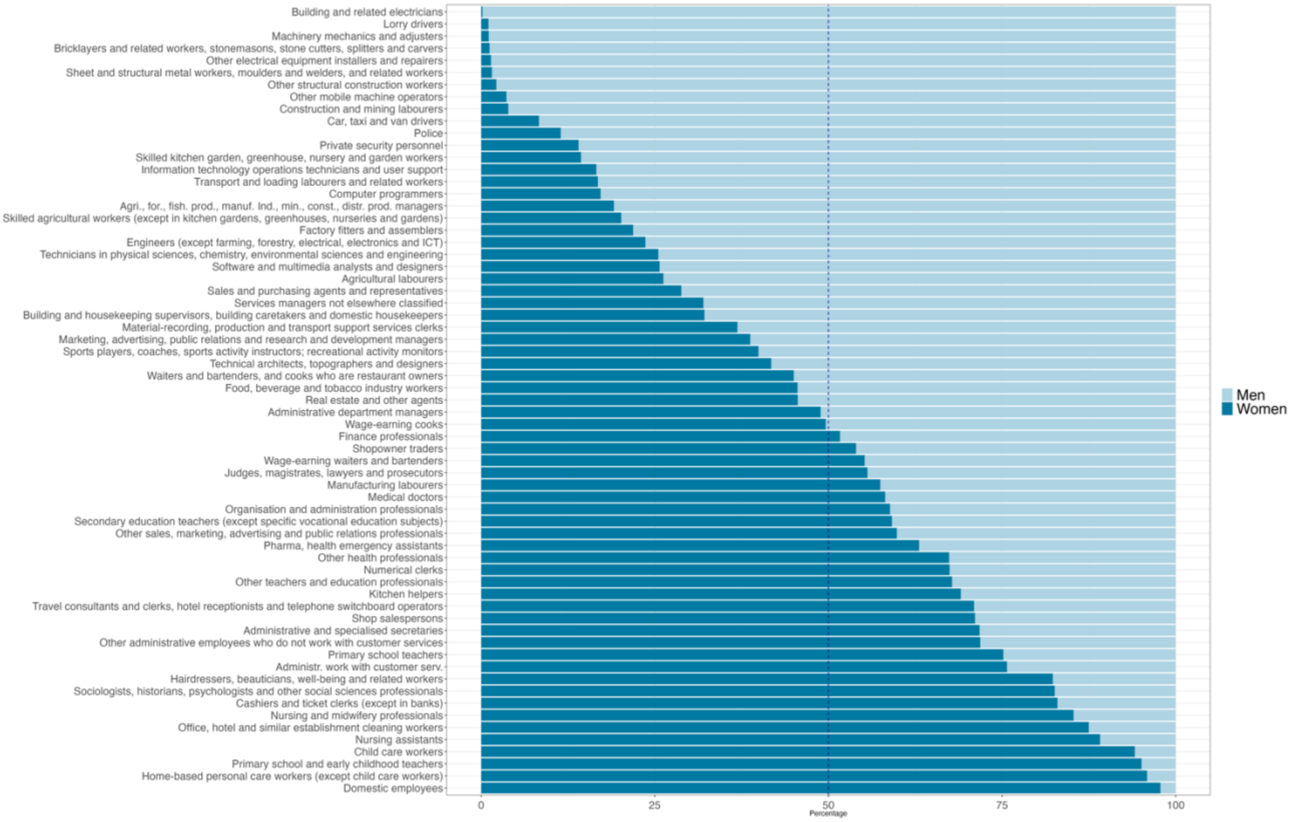


Source. Author's own calculations based on Spain's Labour Force Survey data for 1997-2023 (second quarter).
Note. Occupations are analysed in terms of the NCO-11 classification. For the period 1997-2010 this is achieved by applying the corresponding crosswalk between NCO-94 and NCO-11.

To delve into gender imbalances within specific occupations in more detail, Figure 3.2 shows the sex composition of each occupation at the three-digit level in 2023.[30] The data shows that women still dominate care-related occupations, while manual occupations are predominantly undertaken by men. For instance, 97% of domestic workers are women, comprising around 390,000, compared to around 8,800 men. This occupation accounts for almost 2% of total employment. Conversely, in male-dominated occupations such as bricklayers and related trades, women are significantly underrepresented. Notably, only 1.2% of bricklayers and related trades in Spain in 2023 are women, totalling 4,200 individuals, in stark contrast to the 347,000 men employed in this occupation.

[29] In 2019, the index amounted to 0.51, compared to around 0.49 in the EU on average (Eurofound & Joint Research Centre, 2021).
[30] In this case, when the gender distribution within an occupation aligns with the overall gender distribution in the labour market, it can be regarded as balanced. In 2023, women constituted 46.5% of the employed population in Spain. Thus, any proportion above (below) that level implies an over-representation of women (men).

Figure 3.2. Gender Composition of Occupations in Spain, 2023



Source. Author's own calculations based on Spain's Labour Force Survey data for 2023 (second quarter).
Note. Occupations are shown at the three-digit disaggregation level in terms of NCO-11. Due to the large number of occupations analysed, the graph shows only those whose weight over total employment is at least 0.57% (average value of the weight of each occupation in 2023, second quarter).

In short gender segregation remains deeply rooted in the Spanish labour market. Below, we discuss some of the implications of this phenomenon.

2.2. Implications of Labour Market Segregation

The implications of this gender segregation in the labour market can be analysed along several dimensions, including efficiency, equity, and fairness. From an efficiency point of view, occupational segregation may contribute to a suboptimal allocation of human capital, as labour shortages especially occur in segregated occupations. In particular, the fact that labour supply only comes from only one of the two genders exacerbates this shortage, leading to inefficiencies in the labour market (Miller et al., 2004). The under-allocation of human capital is particularly critical in the current context of labour market polarisation. Human capital shortages have become especially acute in recent decades, particularly in male-dominated skilled occupations and in female-dominated low-skilled occupations (Bettio & Verashchagina, 2009). Among high-skilled occupations with male predominance, the case of STEM (science, technology, engineering, and mathematics) occupations stands out. Although these occupations offer above-average working conditions—in terms of wages and other non-pecuniary attributes, such as flexibility or opportunities for promotion—, they are experiencing labour supply shortages, especially among women (Deming & Noray, 2019; Hanushek et al., 2023). Relatedly, despite women's higher average educational attainment, they are more concentrated in lower-skilled jobs. This contributes to the so-called "skills mismatch", which hampers productivity and economic growth (Carranza et al., 2023).

From an equity and fairness perspective, occupational segregation remains the main factor in understanding the gender gap in hourly wages (Bishu & Alkadry, 2017). In general, feminised occupations offer lower pay and fewer opportunities for career advancement, whereas high-skilled masculinised jobs offer higher wages and prestige (Levanon et al., 2009). However, it is important to note that jobs that are segregated in either direction often involve lower wages (Hegewisch & Hartmann, 2014) than gender-neutral jobs, especially when compared to lower-skilled jobs. Segregation may also perpetuate or reinforce gender roles or stereotypes.[31] This can foster discrimination or biases that may unfairly limit career opportunities and, ultimately, social progress. Occupational segregation, in turn, contributes to significant asymmetries in the labour market.

The negative consequences of labour market segregation, coupled with its prevalence in global labour markets, raises the question of what factors underlie this phenomenon.

[31] Segregation may be the consequence (and not just the cause) of these factors, as examined below.

2.3. Determinants of Segregation: Theoretical Framework

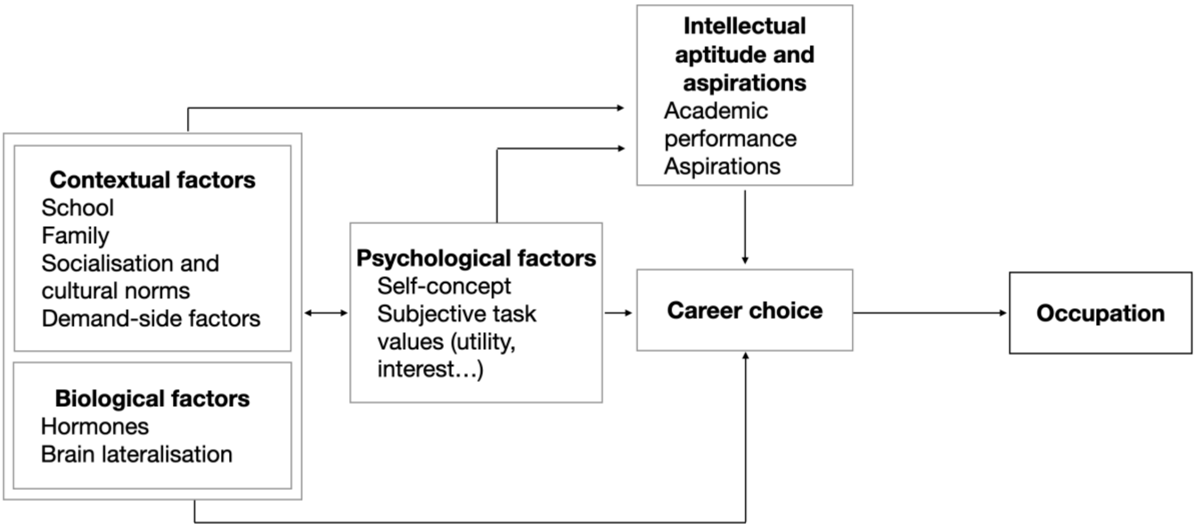
Gender segregation in the labour market emerges from a complex interplay of factors that take place along individuals' lives, some of which operate since early stages. Those factors shape educational choices and can explain the different career choices made by men and women. Men often focus on technical, STEM-related fields—characterised by a higher mathematical content—while women tend to focus on non-technical fields, such as health or education. While some of the factors influencing those choices stem from conscious processes, others operate at a subconscious level. This intricate web of factors not only makes it difficult to quantify the drivers of segregation, but also underscores its multifaceted nature.

To identify the underpinnings of gender segregation in the labour market, we draw on Eccles' Expectancy Value Theory (1983; 2009). This theory offers a comprehensive framework to understand the influence of various factors on decisions related to achievement, such as academic or occupational career choices. Diagram 3.1 outlines the specific, and interrelated, factors that may underlie performance-related decisions, namely contextual, biological, psychological, intellectual, and aspirational factors. The following subsections briefly compile the existing literature on the importance of those mechanisms for understanding gender segregation in the labour market.

Biological Factors. Previous research has sought to investigate how biological sex differences, including hormonal influences and brain lateralisation, contribute to the gender gap in mathematics and ultimately lead to divergent career choices between men and women. In particular, studies have highlighted the significant impact of the social environment on brain development (Blakemore, 2018), suggesting that gender disparities in cognitive outcomes are shaped by mechanisms beyond biological determinants.[32]

[32] A fundamental question relates to the comparative impact of biological versus social determinants that affect this gap, with a lack of consensus in the existing literature. While Stewart-Williams & Halsey (2021) argue that the biological incidence is underestimated in the literature, others show that the gender gap in mathematics largely depends on contextual factors (Ceci et al., 2014; Borra et al., 2023).

Diagram 3.1. Theoretical Diagram on the Determinants of Professional Decisions



Source. Adaption from Wang & Degol (2013).

Contextual Factors. Among the social factors that influence gender differences in labour market careers, contextual factors emerge as relevant drivers. At the school level, experiences with teachers and students can largely determine future career choices (Eccles & Roeser, 2011; Wang & Degol, 2016; Wang et al., 2020; Tandrayen-Ragoobur & Gokulsing, 2022). Outside of the school environment, the family context stands out as paramount in shaping individuals' motivational beliefs (Wang & Degol, 2013). For instance, the literature shows that the gender gap in mathematics achievement widens as families become more entrenched in traditional gender roles (Nicoletti et al., 2022). Beyond the family and school environment, extended socialisation factors and cultural norms can ultimately reinforce gender segregation. In fact, social and cultural beliefs about the role of women in society can help explain the gender gap in mathematics (Nollenberger et al., 2016). Relatedly, labour demand can also be determinant when employers impose barriers that limit job opportunities for a particular gender (Torre &

Jacobs, 2021), due to formal labour market structures or employer biases (Ecklund et al., 2012).[33]

Psychological Factors. Phycological factors also affect subsequent career decisions. Self-confidence, or the belief that one can learn and succeed at certain tasks, encourages individuals to persist and adopt deeper cognitive strategies associated with higher academic achievement (Wigfield & Eccles, 2002). In the field of mathematics, self-confidence can be dampened by math anxiety, a phenomenon that particularly affects women and, overall, discourages individuals from pursuing STEM careers (Ahmed, 2018). In fact, of the different forms of anxiety that can occur in the educational context, math anxiety is particularly widespread (Cassady, 2022). Another psychological factor with a potential impact on career choices relates to competition: women are typically more averse to competition, which leads them to take less risky decisions in general (Combet, 2023). This can partly lead to occupational segregation (Kleinjans, 2009). Importantly, even when people are confident in their own abilities, it does not necessarily follow that their decisions will be related to those abilities, as choices also depend on the subjective task value (Wang & Degol, 2013).[34] In this context, occupational segregation could be driven by diverging labour preferences by gender. The evidence to date, based on discrete choice experiments, shows that women tend to have a higher average willingness to give up salary in exchange for nonpecuniary attributes (Wiswall & Zafar, 2018; Maestas et al., 2022; Osés et al., 2024). However, female-dominated jobs do not necessarily offer those nonpecuniary attributes to a larger extent than other jobs (Osés et al., 2024), with

[33] Carranza et al. (2023) offer an overview of how labour demand operates at different stages of workers' employment processes. Even before the hiring processes, job descriptions containing words generally attributed to the male gender may discourage applications by women (Flory et al., 2015). During the hiring process, Cortina et al. (2021) show, for Spain, that women are particularly discriminated against in relation to jobs involving decision-making, in male and mixed professions, and in jobs requiring both high and low levels of education.

[34] Subjective task value is composed of interest value (the liking or enjoyment of tasks), utility value (the instrumental value of the task), achievement value (the link between the task and the sense of self and identity), and cost (the expected psychological, economic and social costs of different possible tasks or choices). For instance, values-related beliefs predict academic performance and engagement (Schiefele, 2001), but are even stronger predictors of choice behaviours and beliefs, such as career aspirations in STEM (Eccles, 2009; Eccles & Wang, 2012; Wang & Eccles, 2013) or labour aspirations.

some exceptions.[35] This suggests that occupational segregation may often extend beyond pure gender-specific preferences for certain job characteristics.

Intellectual Aptitude and Aspirations. Occupational choices also depend on ability. A large body of literature has explored whether female under-representation in STEM is driven by comparably lower average performance in mathematics compared to their male counterparts. However, recent literature suggests that gender differences in mathematics performance are now moderate and cannot alone explain the strong under-representation of women in mathematics-related fields (Breda & Napp, 2019). Previous research suggests that women's relative advantage in verbal versus mathematical skills—and the fact that people categorise themselves as either numerate or alphabetic, but not both (Marsh & Hau, 2004)—may partly drive women into non-STEM fields.

In sum, the interplay of some of these factors may affect men and women differentially and contribute to differing career choices. The first aim of this paper is to explore the role of certain contextual, psychological, intellectual and aspirational factors, namely those that can be measured by the survey data presented in the following section. Given the nature of the data, some factors, such as the role of biological differences or demand-side factors, are excluded from the analysis.

2.4. Solutions to Gender Segregation: Literature

To implement a solution to gender segregation, which is the second aim of the paper, we first revisit well-established policy interventions aimed at combatting this phenomenon. Policy interventions aimed at addressing gender segregation span from early educational stages to later stages of career trajectories. The latter encompasses policies geared towards redistributing household responsibilities and other demand-side measures related to the recruitment, selection,

[35] For instance, there are gender differences in preferences to avoid physically demanding work: women are willing to sacrifice a large share of the salary to avoid this, and at the same time, they are less concentrated in physically demanding jobs (Maestas et al., 2022).

hiring, evaluation, and promotion of individuals within the labour market (Carranza et al., 2023). While these policies endeavour to mitigate gender segregation among the existing (potential) workforce, educational policies may prove to be an effective tool by preventively curbing future gender segregation through educational presorting strategies. Focusing on educational interventions, multiple initiatives have focused on enhancing students' performance in the STEM fields and counteracting negative stereotypes associated with these disciplines. Such efforts aim to bolster engagement in these fields, a particularly crucial challenge for women, who often exhibit greater reluctance to pursue careers in STEM. Those actions, if well-founded, may hence contribute to preventing future gender segregation within the labour market.

Tutoring programmes in mathematics have proven to be a highly successful and cost-effective tool (Nickow et al., 2020) to improve performance and educational aspirations of low-achieving students (Ramirez et al., 2018; Gortazar et al., 2024), with positive effects on math anxiety (Supekar et al., 2015). Given that females have lower average achievement in mathematics, as well as higher levels of anxiety when faced with mathematics-related issues, such mentoring could ultimately encourage their participation in STEM. Other interventions have specifically targeted the confrontation of math anxiety. For example, some interventions designed to change the mindset by confronting anxiety, rather than avoiding it (the notion of "failure as improvement") have proven effective (Park et al., 2014 provide evidence of the impact of an intervention on expressive writing). Relatedly, interventions aimed at reinforcing the malleability of ability through effort have also been shown to lead to academic improvements (Alan et al., 2019).

Interventions led by role models in science or mathematics may also attract students to the STEM field (González-Pérez et al., 2020; Breda et al., 2023), primarily by fostering a sense of belonging within these disciplines. A large body of literature has found that exposure to role models can effectively counteract students' negative perceptions of these fields. For instance, a large-scale field experiment finds that a brief exposure with female role models in scientific fields influences high school students' perceptions and decisions regarding their choice of undergraduate major. In addition, the intervention improves students' perceptions of science-related jobs (Breda et al., 2023). Besides face-to-face one-off interventions, virtual one-off interventions have also proven effective. In particular, Ashby Plant et al. (2009) find that the use of animated interface agents as social models increased interest, utility beliefs, self-efficacy, and math performance of middle-school students. In a different setting, del Carpio & Guadalupe (2022) find that information provided by a female role model in the application form for a 5-month coding bootcamp was the most influential channel for fostering female participation.

In summary, such interventions hold promise for reversing existing levels of segregation through preventive sorting, particularly through educational presorting. This paper will leverage the existing evidence on the effectiveness of role model interventions to experimentally test a solution to gender segregation.

3. Survey Design, Data Collection and Sample

3.1. Adult Survey

Data Collection and Sample

To answer the first question of this paper on the determinants of gender segregation, we designed, in-house, a large-scale online survey of individuals aged 18-49 in Spain. Participants were recruited through a panel of respondents targeted by the company 40dB. The company was also in charge of the distribution of the surveys, which took place between May 31 and July 3, 2023. The average duration per survey of approximately 20 minutes. [36] After applying multiple quality checks, the final sample contains 4,803 participants.

[36] This includes additional questions beyond the scope of the present analysis.

We imposed quotas on gender, age, educational attainment, and broad regions. Table 3.1 compares the weighted composition of the sample and the Spanish population data. The sample is, by construction, representative of the population along these quota dimensions. However, respondents are more likely to be foreign-born than the average population in Spain.

Table 3.1. Sociodemographic Composition of the Sample, Adult Survey

	Survey data	Population data
Sex		
Woman	49.4	49.4
Man	50.6	50.6
Age		
18-24	17.3	17.3
25-34	26.9	26.9
35-44	35.4	35.4
45-49	20.4	20.4
Educational level		
Compulsory Secondary Education or lower	32.0	32.0
Post-compulsory Secondary Education	26.9	26.9
Higher Education	41.1	41.1
Origin		
Spain-born	86.6	76.3
Foreign-born	13.4	23.7
Observations	4,803	19,591,329

Note. The composition of the survey sample is calculated taking into account population weights. Sex, age and origin population data corresponds to 2022 (Spanish National Statistical Institute). Quotas for the educational level are established as per the Spanish Labour Force Survey (2023, first quarter) and weights are applied through population data for 2021.

Survey Design

The survey is designed such that, after drawing a detailed profile of the background sociodemographic and labour characteristics of respondents, the determinants of gender segregation can be clearly identified. To this end, beyond the background sociodemographic and labour market questions, the survey is mostly structured following Eccles' Expectancy Value Theory (Diagram 1). Below, we provide information on the four blocks composing the survey.

Background socioeconomic questions. The first block includes information on respondents' sociodemographic characteristics (e.g., gender, age, country of birth, educational level, field of knowledge to which the studies correspond).

Background labour market questions. The second block covers information on respondents' labour-related information, including their labour status, occupation, and sector.

Questions on psychological and motivational traits. The third block includes information on STEM-related factors (self-reported mathematics ability, grade repetition, the existence of a role model with a positive impact on their careers), psychological and motivational factors (malleability of their own intellectual ability, goal-achievement, math anxiety during adolescence), and other aspects related to preferences that may impact on career-related choices (preference for individualism versus collectivism, person versus thing orientation, enjoyment for competition, risk aversion). This block is addressed randomly to half of the sample, as the other half responds to separate questions beyond the scope of this analysis.

Questions on contextual factors, stereotypes, aspirations and other. The fourth block includes contextual information related to the education received from the family (type of education, from more conservative to more progressive), parents' educational attainment, questions on gender-related stereotypes (perceptions on gender roles and attitude towards women in the labour market), past professional and personal priorities, and other questions such as respondents' voting intention.

3.2. Adolescent Survey

Data Collection and Sample

After identifying the drivers of gender segregation, the second aim of the paper lies in experimentally testing a potential solution to the segregation phenomenon. This is done through a second online survey, also designed in-house, of 600 pre-university adolescents aged 15-18 in Spain. As in the adult survey, the company 40dB targeted the sample and distributed the surveys. The field work took place between September, 27 and October, 10, 2023. The average duration per survey of approximately 11 minutes.[37]

The imposition of quotas was established at the sex level. Table 3.2 confirms the representativeness of the data to the Spanish population by sex, and it also reasonably adjusts

[37] This includes some additional questions beyond the scope of the present analysis.

to the population data regarding origin and student status. The only exception relates to 15-year-olds, who are underrepresented by construction in those online surveys given their age.

Table 3.2. Sociodemographic Composition of the Sample, Adolescent Survey

	Survey data	Population data
Sex		
Woman	48.5	48.5
Man	51.5	51.5
Age		
15	4.5	25.5
16	20.0	25.3
17	21.7	25.0
18	53.9	24.3
Origin		
Spain-born	82.4	88.6
Foreign-born	17.6	11.4
Currently studying		
No	15.9	13.8
Yes	84.1	86.2
Number of books at home		
0-25	43.2	
26-100	34.4	
101 or more	22.4	
Observations	600	1,941,658

Note. The composition of the sample is calculated taking into account population weights. VET stands for Vocational Education and Training. Population data refers to 2021. The population data referring to the variable "currently studying" is retrieved from the Spanish Labour Force Survey and, due to data availability, only covers individuals aged between 16 and 18.

Survey Design

The first block of the survey comprises questions on individuals' socio-demographic characteristics, including gender, age, country of birth and educational background. At this point, randomly chosen individuals are shown a video treatment—explained in the following subsection—while individuals belonging to the control group access directly the last block of the survey. The last block, which is also responded by treated individuals after watching the video, consists of multiple questions mainly related to student's perceptions about mathematics.

Experimental Design and Randomisation

The experiment consists of a video starring a female mathematician and disseminator with the aim of analysing the potential to malleate students' negative perceptions of STEM. The choice of a female—rather than a male—role model lies in the fact that women are particularly reluctant to this field, and the literature concurs that women's exposure to female role models is, at least, as effective as that of male role models (Drury et al., 2011). Another key consideration in the selection of this fragment relates to the idea that the role model does not follow an overly stereotypical prototype. This is relevant in view that non-stereotypical role models can have more positive effects on self-efficacy than stereotypical role models, with whom treated individuals might not feel as identified (Cheryan et al., 2011). In addition, this type of one-off interventions provides two key benefits. First, from a public policy perspective, these interventions are generally cost-effective, with potentially larger benefits. Second, the evaluation of the impact of those interventions is neater than when analysing other types of interventions, for instance those related to teachers, as these may be influenced by teaching practices (Breda et al., 2023).

Table 3.3. Treatment-Control Balance

	Control mean (1)	Treatment mean (2)	Difference T – C (3)	p-value of difference (4)
Woman	0.492	0.479	-0.013	0.745
	(0.500)	(0.500)	(0.041)	
Age [years]	17.248	17.253	0.005	0.946
	(0.961)	(0.888)	(0.076)	
Foreign-born	0.166	0.186	0.020	0.533
	(0.372)	(0.389)	(0.031)	
Currently not studying	0.170	0.148	-0.023	0.449
	(0.376)	(0.355)	(0.030)	
Books at home				
0-25	0.429	0.434	-0.006	0.889
	(0.495)	(0.496)	(0.041)	
26-100	0.334	0.354	-0.020	0.605
	(0.471)	(0.478)	(0.039)	
More than 100	0.237	0.211	0.026	0.451
	(0.425)	(0.408)	(0.034)	
Math performance				
Very poor	0.091	0.055	0.036	0.096
	(0.287)	(0.228)	(0.021)	
Poor	0.284	0.285	-0.001	0.975
	(0.451)	(0.451)	(0.037)	
Good	0.506	0.499	0.007	0.865
	(0.500)	(0.500)	(0.041)	
Very good	0.120	0.161	-0.041	0.145
	(0.325)	(0.368)	(0.0285)	
Test of joint significance	F-statistic: 1.006 (p-value: 0.4395)			

Note. Each row corresponds to a different linear regression with the dependent variable listed on the left. Columns (1) - (2) show the average value of the control and treatment groups, with the standard deviation below in parenthesis. Column (3) reports the coefficients from the regression of each variable on the treatment group indicator, with the p-value of the difference reported in the last column. The test of joint significance shows the F-statistic and p-values of a regression where treatment condition is estimated as a function of the full set of sociodemographic characteristics.

The intervention covers the following topics: the work of mathematicians and its importance in the current context; the social usefulness of mathematics in issues related to climate change, poverty or care; the importance of gender balance in mathematics given its consequences in the labour market and in the social environment.[38] Half of the sample in the survey was randomly selected ("treatment group") to visualise the video. Table 3.3 shows that random assignment successfully balanced the sociodemographic composition of the treatment and control groups. The math performance indicator is the only variable where the mean differences between the treatment and control groups are close to being significant. In our empirical analysis, we will account for these residual imbalances by controlling for these characteristics.

3.3. Quality of Responses

In both the adult and adolescent surveys, we implement ex-ante and ex-post methods to avoid inattentive and/or careless respondents. Our ex-ante approach aimed to identify careless respondents through "screeners", that is, questions specifically designed to detect inattentive answers (Stantcheva, 2023). In particular, we resort to logical questions (Abbey & Meloy, 2017)—which are sparingly presented to respondents—to filter out potentially unreliable responses. Similarly, the landing page in both surveys attempts to induce respondents' attention before commencement of the survey by (1) warning respondents that researchers can spot careless answers, potentially invalidating participation in the survey, and (2) acknowledging the relevance of their participation in the survey.

After identifying potentially problematic respondents, we also apply ex-post methods. The main filtering method consists of discarding respondents who spend too little time responding to the survey. Specifically, we filter out the top 6% quickest respondents. In parallel, we assess incoherent answers based on contradictory, incoherent or repetitive responses.

[38] The duration of the video is approximately three minutes.

The ex-ante and ex-post quality control measures, in turn, contribute to ensuring the reliability of the data collected.

4. Determinants of Gender Segregation in the Labour Market

4.1. Descriptive Gender Differences

The existence of gender segregation in the labour market necessarily implies that men and women differ on some of the factors shown in Diagram 1. This section explores descriptive gender differences in those potential drivers, before testing empirically whether those factors explain observed levels of segregation.

Gender Differences in Occupational and Career Choices

Using the adult survey data, Table 3.4 categorises occupations as female- or male-dominated by comparing the gender-specific employment absorbed by a given occupation.[39] If female employment in a given occupation surpasses that of males by 2 percentage points, then the occupation is categorised as female-dominated, while if this difference applies to men, the occupation is regarded as male-dominated. The robustness of this definition is empirically tested below.

The nature of occupations dominated by women and men are consistent with the conclusions drawn from Figure 3.2. Female-dominated occupations comprise primarily health- and care-related tasks. Those female-dominated occupations account for 36% of total employment in the sample. For context, half of the women are employed in female-dominated occupations, compared with 24.5% of men. Conversely, technical occupations—such as those related to science or engineering, or occupations requiring manual tasks—are male-dominated and represent 34.6% of total employment in the sample. In this case, 46% of men are employed

[39] The survey does not include a standardised classification due to time constraints: occupations at the one-digit level offer too little detail, while those at the two-digit level would be too long for individuals to identify, potentially compromising not only the time constraints but also the quality of the answers.

in male-dominated occupations, compared to 20% of women. Appendix 3.A (Table 3.A2) provides more detail on the gender composition of the occupations.

Table 3.4. Share of Gender-Specific Employment per Occupational Category

	Female-dominated	Male-dominated
Occupations	Health professionals, teachers. Customer service employees, library staff, postal services. Catering and trade workers. Health care, personal care and aesthetics workers. Domestic servants, cleaning staff, kitchen helpers, porters, waste collectors...	Professionals in science/engineering IT programmers, ICT technicians. Support professionals (supervisors, draughtsmen, real estate agents…). Security and safety services workers. Agriculture, livestock, forestry and fisheries sector. Structural works workers. Construction and installation workers. Workers in manufacturing industries. Drivers and operators of mobile machinery. Agricultural, fishing, construction and transport workers.
Observations	984	943
Share of total employment	36.0	34.5

Note. The categorisation of gender-dominated or not gender-dominated occupations is calculated using the Spanish Labour Force Survey microdata (2023, second quarter) and applied to the occupational classification of this Adult survey. More detail is available in Appendix 3.A.

Focusing on educational choices yields comparable results. Table 3.5 shows the gender distribution of respondents depending on the field of studies undertaken.[40] Our results show that more than half of the men have undertaken STEM education, whereas this is the case only for one in five women in our sample. Conversely, almost 60% of the women undertake HEAL studies—comprising health, education, administration and literacy—in sharp contrast with only 29.9% of men.

Table 3.5. Gender-Specific Composition of Study Choices

	Females	Males
STEM	19.0	50.6
HEAL	59.0	29.9
Other	21.9	19.5
Observations	1,255	1,192

[40] Given the lack of a universal definition of STEM, we follow González-Cervera et al. (2021) for the classification of the different study choices at the VET level. For university careers, we categorise as STEM all the study areas related to science, architecture and engineering. For the HEAL category, we select the study areas related to health, education, administration and literacy.

Gender Differences in Psychological, Contextual, Intellectual and Aspirational Factors

We now document gender differences in psychological traits, socialisation factors, cultural norms, school and family contexts in the past, and factors related to intellectual aptitude and career aspirations. Specifically, we select a set of items that are generally considered in the literature to be important attributes for achievement-related decisions (Wang & Degol, 2013).

Table 3.6 reports the mean and standard deviations by gender of a series of variables, which are constructed following the methodology of Kling et al. (2007). Specifically, we first rescale the original responses to lie between 0 (most negative response) and 1 (most positive). After this, we calculate the *z*-scores by subtracting the (weighted) mean and dividing it by the (weighted) standard deviation, so that each *z*-score has a mean of 0 and a standard deviation of 1. Based on the questions that share a common theme, we create indices by calculating the unweighted means of those z-scores of their components. To further facilitate interpretation, the resulting index is further standardised by subtracting the mean of the control group and dividing by the standard deviation, so that each index has mean 0 and standard deviation 1 (Stantcheva, 2022). In addition to the statistics of the indices and the subcomponents of these, Table 3.6 tests for the significance of the gender differences in these means.

The results in Table 3.6 show that psychological traits vary significantly by gender. The first factor with large and significant gender differences relates to math anxiety in adolescence, which, on average, affects women more than men. In addition, women report lower average levels of self-concept. However, of the three items that make up this composite index of self-concept, the only one on which gender differences are significant relates to enjoyment (aversion) of competition. Conversely, belief in the malleability of intelligence does not appear to affect men and women differently, nor does belief in the ability to achieve one's goals differ significantly by gender.

Another psychological trait relates to the subjective task value, or the value attached to the task in terms of interest, utility and attainment. Women tend to concentrate in person-oriented occupations. Men, on the other hand, are more likely to be concentrated in jobs that involve working with machines. The results in Table 3.6, however, do not indicate significant gender differences in the preference for working with people rather than with machines, in contrast with other findings in the literature (Su et al., 2009; Wang & Degol, 2013). Separately, men prefer being with people than alone to a larger extent than women. Women report more difficulty when having to speak with a stranger, a trait of shyness that is typically not considered "masculine" (Antecol & Cobb-Clark, 2013). The preference for group versus individual objectives characterises women more than men, somewhat counterintuitively, as feminised jobs often require group objectives to be prioritised over individual ones. Lastly, enjoyment for risk is much more present among men, in line with literature, as women avoid risk to a much larger extent (Combet, 2023).

Regarding gender stereotypes, women express stronger opposition to the four dimensions of stereotypes analysed, and in all cases we find significant gender differences in means. First, the idea that men are innately better at mathematics/engineering than women is more prevalent among men. This can lead to so-called 'stereotype threat'—the belief that women are not good at mathematics—which can be detrimental to women's performance in mathematics. However, there is debate about the contribution of this phenomenon to the gender gap in mathematics (Stoet & Geary, 2012). Second, women strongly disagree with the idea that they are solely responsible for doing the housework, even when their husbands are home. This idea is significantly less extended among men. Third, the largest differences in mean are found in the item that directly touches upon gender stereotypes and education: the acceptance that boys play with dolls. In this case, women support more firmly this statement (completely agree) while men express relatively less strong views. Fourth, women are slightly more supportive of the idea that women are as competent as men to be executives in a company.

Table 3.6. Gender Differences in *Z*-Scores of Control Variables

	Females		Males		t-value
	Mean	St. dev.	Mean	St. dev.	
Psychological factors					
Math anxiety during adolescence	0.120	0.978	-0.113	1.01	5.783***
Self-concept	-0.072	0.996	0.068	0.999	-3.458***
Your intelligence can be changed (r)	0.00	0.99	0.00	1.01	0.010
Capable of achieving own goals	0.01	1.02	-0.009	0.978	0.4809
Enjoyment for competition	-0.143	1.01	0.134	0.97	-6.917***
Preference for working with people vs machines	0.0234	0.984	-0.022	1.01	1.122
Socialisation and collectivism	-0.058	0.981	0.054	1.01	-2.769***
Preference for being with people than alone	-0.062	0.988	0.059	1.01	-2.992*
Ease to speak with strangers (r)	-0.013	0.986	0.013	1.01	-0.639
Preference for group objectives vs individual	-0.036	1.00	0.034	0.994	-1.719*
Risk lover	-0.036	1.01	0.0335	0.993	-1.707*
Socialisation factors and cultural norms					
Against gender stereotypes	0.243	0.924	-0.238	1.01	17.179***
Men not innately better in maths/eng. (r)	0.151	0.938	-0.148	1.04	10.488***
Women not the only responsible of household chores (r)	0.141	0.936	-0.138	1.04	9.777***
It is OK for boys to play with dolls	0.269	0.902	-0.263	1.02	19.134***
Women as competent as men to be executives in a company	0.139	0.944	-0.136	1.03	9.632***
Contextual school and family factors					
Female role model	0.012	1.00	-0.011	1.00	0.580
Male role model	-0.066	1.00	0.062	0.995	-3.156***
Progressive family education	0.007	1.00	-0.007	0.996	0.458
Intellectual aptitude and aspirations					
Reported cognitive ability	0.0233	1.01	-0.0220	0.993	1.097
Reported math performance	0.0231	1.01	-0.0218	0.987	1.086
No grade repetition (r)	0.0153	0.998	-0.0144	1.00	0.734
Family vs professional prioritisation in the past	0.0146	1.03	-0.0143	0.972	0.999

Note. The term (r) refers to those items whose answers have been recoded to ensure consistency across the sub-questions that compose the index. For instance, the "ease to speak with strangers" item was originally phrased as "difficulty to speak with strangers", and the answer "totally agree" was hence reverted to "totally disagree", and so on. When bold terms are followed by subitems, those terms refer to indices. Those indices composed by more than one question are stated below each index. Each row corresponds to a different linear regression with the dependent variable listed on the left. Columns (1) - (4) show the average value and standard deviation for males and females. The last column reports the t- value from the regression of each variable on the sex indicator, with the asterisk denoting the significance of the mean difference as follows: *** p<0.01, ** p<0.05, and * p<0.1.

Turning to contextual factors, the difference in the mean incidence of a female role model with a positive impact in the academic/professional live is not significant by gender. However, this mean difference is significant when referring to male role models, who appear more present amongst men individuals than women. Regarding the family context, women report a higher level of a progressive education received from parents, and the difference in the mean is significant by gender.

Lastly, cognitive ability is reported, rather than objectively measured, given the nature of the survey. Table 3.6 documents that mean difference in the reported ability is not significant by gender, although responses may be biased given difficulty to recall the grade, on the one hand, and the fact that the specific grade is phrased from “insufficient” to “exceptional”, with five possible answers, which may give rise to some inaccuracies. Finally, we find no difference in the way men and women prioritised family creation as opposed to the professional environment before taking the decision of following the maximum level of studies undertaken.

4.2. Empirical Estimates

Determinants of Occupational Segregation

We now empirically investigate the determinants of occupational segregation by estimating the following sex-specific probit models:

$$o_{i,s} = \beta_{0,s} + \beta_{j,s}X + \epsilon_{i,s}$$

Where o is a dummy variable that reflects whether the occupation is not dominated by the corresponding sex $s = \{m, w\}$, i.e. $s = m$ in the male equation, and $s = w$ in the female equation (see Table 3.4). X is a vector of sociodemographic, contextual, psychological and other factors potentially relevant to explaining occupational choices, as explained above (e.g., age, birthplace, educational attainment, type of studies, math anxiety, voting intention, parents' education, area of studies, or belief in gender stereotypes). Finally, $\epsilon_{i,s}$ refers to the error term. The model is only estimated for employed individuals (those who were employed in the week before the survey).

Table 3.7. Probability of Being Employed in an Occupation not Dominated by Own Gender, Marginal Effects

	Specification 1		Specification 2	
	Women	Men	Women	Men
	(1)	(2)	(3)	(4)
Field of study				
STEM	0.265***	-0.477***		
	(0.057)	(0.044)		
HEAL			-0.180***	0.339***
			(0.049)	(0.052)
Psychological factors				
Math anxiety during adolescence	-0.020	-0.025	-0.029	-0.001
	(0.027)	(0.028)	(0.027)	(0.027)
Self-concept	0.058**	0.022	0.064**	0.035
	(0.028)	(0.029)	(0.028)	(0.028)
Preference work with people vs machines	-0.012	-0.063**	-0.004	-0.068**
	(0.028)	(0.029)	(0.028)	(0.027)
Socialisation and collectivism	-0.002	0.041	-0.011	0.041
	(0.029)	(0.029)	(0.029)	(0.028)
Risk lover	-0.040	-0.064	-0.057	-0.046
	(0.052)	(0.056)	(0.051)	(0.054)
Socialisation and cultural norms				
Against gender stereotypes	-0.089**	-0.024	-0.095**	-0.006
	(0.032)	(0.026)	(0.031)	(0.025)
Contextual school and family factors				
Female role model	0.054	0.115*	0.053	0.116*
	(0.055)	(0.068)	(0.055)	(0.066)
Male role model	-0.024	0.077	-0.018	0.085
	(0.055)	(0.069)	(0.054)	(0.066)
Progressive family education	-0.025	-0.060**	-0.019	-0.058**
	(0.027)	(0.028)	(0.027)	(0.027)
Intellectual aptitude and aspirations				
Reported cognitive ability	0.032	0.003	0.035	-0.018
	(0.028)	(0.029)	(0.028)	(0.027)
Family prioritisation in past	-0.025	-0.030	-0.025	-0.037
	(0.023)	(0.028)	(0.023)	(0.026)
Observations	479	484	479	484

Note. The coefficients reflect the marginal effects of the probit models. All variables shown in the tables have been transformed into *z*-scores and coefficients can be interpreted as partial correlations. Standard errors in parentheses. *** p<0.01, ** p<0.05, and * p<0.1. All the models control for educational attainment, age, region of birth (foreign or native), voting intention (left versus rest), and mother's and father's educational attainment.

The results in Table 3.7 show that the field of study is the key determinant of future occupational segregation. For women, engagement in STEM largely increases the probability of undertaking non-feminised occupations, while HEAL-related studies are associated with a higher probability of holding female-dominated jobs. Conversely, for men, the STEM field decreases the probability of involvement in jobs not dominated by men, and the opposite applies when it comes to HEAL-related studies. In particular, women whose studies focus on STEM are, on average, 26.5 percentage points more likely to work in an occupation not dominated by women. Similarly, men who undertake STEM studies are 47.7 percentage points more likely to be involved in masculinised jobs. The exclusion of certain variables from the specification—which results in a larger sample size—leaves the overall conclusions unchanged (Appendix 3.B, Table 3.B1).

Besides the field of study, psychological factors also determine occupational segregation. Women with higher levels of self-concept are more likely to carry out jobs not female-dominated. In particular, a one standard deviation increase in the self-concept index increases the average likelihood of women to not work in female-dominated jobs by 5.8-6.4 percentage points. Opposition to gender stereotypes has no significant effect on men's occupational segregation, while, interestingly, it increases women's average probability to work in a female-dominated job. This somewhat counterintuitive finding is discussed in the following subsection. Finally, the presence of a female role model positively affects the average probability that men undertake jobs that are not male-dominated.

To test for the robustness of the definition of occupational segregation, we select two alternative thresholds to categorise occupations as dominated by the corresponding gender or not, one for $t = 0$ and one for $t = (-5, 5)$, respectively, for men and women. The results in Table 3.B2 evidence the robustness of the results to alternative thresholds to define occupational segregation. In addition, we allow for variation in the dependent variable by ranking occupations in the female (male) model from less to more feminised (masculinised)—as done in Antecol & Cobb-Clark (2013)—and then estimate these models through OLS. Our results, shown in Table 3.B3, convey a similar message, being the field of study the most relevant determinant of occupational segregation.

Determinants of Segregation in Educational Choices

The confirmation that the field of study is determinant in understanding occupational segregation leads to the exploration of the drivers of segregation in educational choices. The rationale for this additional specification is twofold. On the one hand, the academic choice tends to be made earlier in life and this may shed light on some of the factors influencing this choice that may not be reflected in the occupational choice. On the other hand, occupational segregation responds to a mix demand- and supply- side factors. Some of these cases may not reflect pure individual choices, for example in cases of skill mismatches between supply and demand.

The specification of the probit model is as follows:

$$s_{i,t} = \alpha_{0,t} + \alpha_{j,t}Z + \mu_{i,t}$$

Where s is a dummy variable reflecting whether or not the field of study is STEM, i.e., $t = STEM$; or whether or not the field of study is HEAL, i.e., $t = HEAL. Z$ is a vector that includes the same individual characteristics as in the previous model, but adds sex as an additional control variable and excludes the educational attainment variables as these are now measured in the dependent variable. Finally, $\mu_{i,t}$ refers to the error term. These models are restricted to individuals whose education is at the VET-level or higher, as earlier educational stages in the data cannot be classified as STEM or HEAL with sufficient levels of accuracy.

Table 3.8 documents the marginal probabilities of undertaking STEM studies, on the one hand, and HEAL studies, on the other. The results indicate that math anxiety during adolescence is the factor that most significantly prevents individuals from selecting studies in the STEM area. In particular, a one standard deviation increase in math anxiety decreases the probability of involvement in the STEM area by 8.9 percentage points. The same increase in math anxiety increases the probability of individuals choosing HEAL-related studies by 4 percentage points.

Table 3.8. Probability of Pursuing STEM and HEAL Education, Marginal Effects

	Baseline Specification 1		Specification 2	
	STEM	HEAL	STEM	HEAL
Psychological factors				
Math anxiety during adolescence	-0.089***	0.042**		
	(0.015)	(0.015)		
Self-concept	0.003	-0.019		
	(0.016)	(0.016)		
Preference work with people vs machines	-0.007	0.051**		
	(0.015)	(0.016)		
Socialisation and collectivism	-0.007	-0.001		
	(0.016)	(0.016)		
Risk lover	-0.016	-0.012		
	(0.030)	(0.031)		
Socialisation and cultural norms				
Against gender stereotypes	-0.060***	0.041**	-0.043***	0.035**
	(0.015)	(0.015)	(0.010)	(0.011)
Contextual school and family factors				
Female role model	-0.033	0.024		
	(0.034)	(0.035)		
Male role model	-0.011	0.027		
	(0.034)	(0.035)		
Progressive family education	0.001	-0.013	0.010	-0.009
	(0.015)	(0.016)	(0.010)	(0.011)
Intellectual aptitude and aspirations				
Reported cognitive ability	0.037**	-0.033**		
	(0.015)	(0.016)		
Family prioritisation in past	0.008	-0.004	0.000	-0.009
	(0.014)	(0.014)	(0.010)	(0.010)
Gender				
Women	-0.304***	0.305***	-0.309***	0.283***
	(0.028)	(0.029)	(0.019)	(0.020)
Observations	1360	1360	2673	2673

Note. The coefficients reflect the marginal effects of the probit models. All variables shown in the tables (except for gender) have been transformed into *z*-scores and coefficients can be interpreted as partial correlations. Standard errors in parentheses. *** p<0.01, ** p<0.05, and * p<0.1. All the models control for age, region of birth (foreign or native), voting intention (left versus rest), and mother's and father's educational attainment.

Discussion

Our findings indicate that educational presorting emerges as the primary driver of occupational gender segregation. Specifically, individuals who pursue studies in fields traditionally associated with their own gender exhibit a higher likelihood of entering gender-typical occupations. This finding aligns closely with the conclusions drawn by Antecol & Cobb-Clark (2013) for the United States and Borghans & Groot (1999) for the Netherlands.

Furthermore, our analysis reveals that higher levels of self-concept among women correlate with a lower probability of entering female-typical occupations. As shown earlier, women's lower levels of self-concept are largely driven by their higher average aversion to situations that require competition, in line with established literature (e.g., Vesterlund & Vesterlund, 2011).

However, the relationship between competitiveness and subsequent professional choices presents mixed results in the literature (Averett et al., 2018). For instance, while Antecol & Cobb-Clark (2013) corroborate our findings by demonstrating that masculine psychosocial traits—such as competitiveness, traditionally associated with masculinity—promote entry into male-dominated fields, Reuben et al. (2017) find no significant effect of competition on career choices (although they do find an effect on earnings expectations).[41]

However, our findings on the positive correlation between opposing gender stereotypes and working in a female-dominated job for women contrast with earlier literature (e.g., He et al., 2019). Several factors in our study's setting may influence this result, including biases introduced by the timing of survey questions and the possibility of stereotype opposition strengthening after entering feminised jobs. Furthermore, unconscious stereotype-related processes may play a significant role, as suggested by recent literature (Cuevas Ruiz et al., 2023). Our study also suggests that female role models increase the probability of men working in non-male-dominated jobs, and the lack of significance for women could also be related to the timing of the survey, a matter that is tackled in the following section by specifically focusing on adolescents.

Further exploration into the determinants of educational presorting yields that math anxiety discourages individuals from pursuing STEM education. These results fully align with previous evidence (Ahmed, 2018; Daker et al., 2021), including from laboratory studies and large-scale international assessments (Foley et al., 2017). While it is beyond the scope of this paper to explore the causes of maths anxiety, the existing literature suggests a possible link between parental and child maths anxiety, which can negatively affect maths performance (Soni & Kumary, 2015). Importantly, research shows that this relation is not due to biological factors alone (Maloney et al., 2015). In addition, teachers with high levels of maths anxiety may

[41] For the Netherlands, Buser et al. (2014) find that competitive students are significantly more likely to undertake the most prestigious tracks in the field of mathematics and science, even when controlling for a proxy of ability.

contribute to students' underachievement in maths (Beilock et al., 2010). Our descriptive results provide evidence that average math anxiety is higher for women than men. This, coupled with the subsequent confirmation that maths anxiety reduces the likelihood of taking up STEM studies, provides insight not only into the general shortage of STEM profiles, but in particular the underrepresentation of women in these fields.

These findings, in turn, underscore the significance of early life factors in shaping educational presorting, such as math anxiety, differentially affecting men's and women's future professional decisions.

5. Solutions to Gender Segregation: an Experimental Analysis

The confirmation that educational presorting shapes occupational segregation underpins the potential effectiveness of providing information to individuals before the moment that they are involved in an occupation, and ideally before making the final educational decision. This section proposes an intervention with a female role model in the field of mathematics with the aim of exploring whether adolescents' perceptions on STEM can improve as a result of the intervention. The intervention is part of the adolescent survey data.

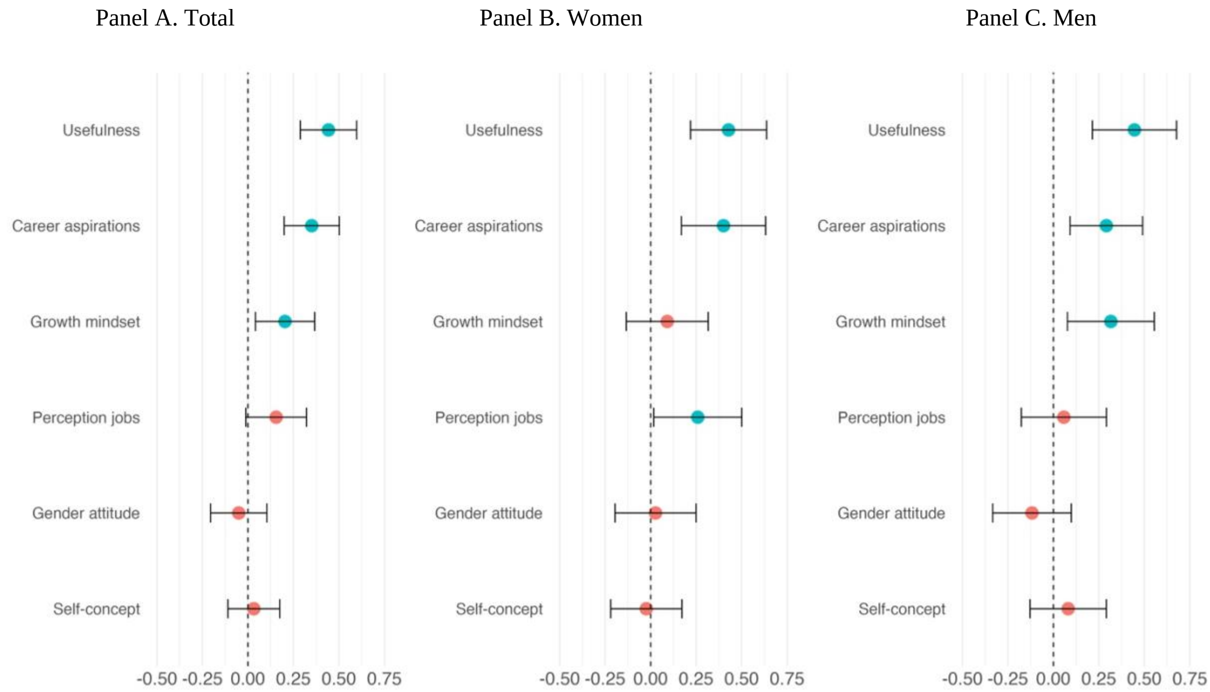
The survey items are designed to measure the effects of the interventions on students' perceptions along six dimensions: (1) usefulness and applicability of mathematics, (2) math-related career aspirations, (3) perceptions about math-related jobs, (4) growth mindset, (5) self-concept, and (6) gender attitudes towards mathematics. These indices summarise a series of sub-questions, shown in Appendix 3.C. The construction of the indices follows a similar methodology as in the previous analysis, with the specificity that the calculation of the *z*-scores of the subquestions that compose each index is based on the mean and standard deviation of the control group (i.e., the group that saw no video treatment). After this, the indices are similarly calculated as the unweighted average of the *z*-scores of its components. Finally, the resulting index is further standardised by subtracting the mean of the control group and dividing by the standard deviation, so that each index has mean 0 and standard deviation 1 for the control group (Stantcheva, 2022).

To empirically assess the impact of the intervention, we estimate the following model through Ordinary Least Squares (OLS):

$$Y_i = \beta_0 + \beta_1 Treatment_i + \beta_j X_i + \epsilon_i$$

Where Y denotes the outcome of individual i. The outcome variable captures, separately, each of the six indices, as well as the *z*-scores of the subquestions that make up each index. The independent variable of interest, $Treatment$, is binary and reflects whether the person belongs to the treatment group (i.e. watches the video) or to the control group. Vector X reflects various sociodemographic and contextual characteristics, namely gender (for the aggregate model), age, whether or not individuals are foreign-born, student status, number of books in the household, and reported skills on mathematics performance. ϵ is the error of the model. Each model is estimated for the total sample, and for females and males separately.

Figure 3.3. Impact of Role Model Intervention on Adolescents' Perceptions



Note. The figures show the treatment effects of the intervention with the 95% confidence interval. Each effect corresponds to a separate regression with the dependent variable corresponding to each index included on the left-hand side. All regressions control for the sociodemographic variables included in Table 3 (except for sex in the sex-specific equations).

Figure 3.3 summarises the main results using the six different indices as dependent variables of the models (Table 3.C1 shows the estimation results in more detail). Adolescents' post-intervention questions reveal that the treatment was effective in challenging their stereotyped views about mathematics, as discussed below.

5.1. Usefulness of Mathematics

Adolescents were asked to what extent they agreed or disagreed with four statements related to the usefulness of mathematics. In particular, they were asked (1) whether they believed that mathematics was abstract and not applicable, (2) whether mathematics help to think logically, (3) whether mathematics is important for tackling global challenges (e.g., inequality or global warming), and (4) whether they believed that they would apply the mathematics learnt in the future as adults. Table 3.9 shows the treatment effects for the *z*-scores of the composite index (in bold), together with the *z*-scores of the individual items that make up the index.

Table 3.9. Treatment Effects on the Usefulness of Mathematics

	Control group mean		**Treatment effects**		
	Women (1)	Men (2)	All (3)	Women (4)	Men (5)
Usefulness of mathematics (index)	-0.0493	0.0478	0.443***	0.428***	0.446***
			(0.079)	(0.107)	(0.118)
Mathematics abstract and not applicable (r) (*z*-scores)	-0.073	0.071	0.145*	0.146	0.129
			(0.087)	(0.124)	(0.125)
Mathematics help think logically (*z*-scores)	0.036	-0.035	0.260***	0.231**	0.283**
			(0.077)	(0.107)	(0.115)
Mathematics helps address global challenges (*z*-scores)	-0.088	0.086	0.425***	0.501***	0.354**
			(0.079)	(0.114)	(0.113)
I will apply the mathematics I learnt as an adult (*z*-scores)	-0.002	0.002	0.318***	0.231*	0.390***
			(0.080)	(0.118)	(0.110)
Observations	149	145	596	299	297

Note. The first two columns show the mean values of the z-scores for the control group. The last three columns reflect the coefficient of the treatment binary variable, with the dependent variable in each model being the one shown on the left-hand side. The item in bold reflects the index, which is composed of the *z*-scores of the sub-items below. (r) means that the order of responses to the questions was reverted so that a higher value of the index indicates a higher perceived level of usefulness about mathematics. All the models control for gender (only for the aggregate model), age, whether or not individuals are foreign-born, student status, number of books in the household, and reported skills on mathematics performance. Standard errors in parentheses. *** p<0.01, ** p<0.05, and * p<0.1.

One of the main aims of the intervention is to precisely foster the idea that mathematics constitutes a useful tool that can be applied to day-to-day and global challenges. As shown in

Table 3.9, we find that boys' and girls' perceptions on the (composite index) of the usefulness of mathematics significantly improves after treatment: the impact amounts to 44% of an SD. By gender, the effects are slightly stronger for men than women (44.6% versus 42.8%), yet they remain significant for genders.

The detailed results for the different components of the index reveal that a significant impact of the role model intervention is observed for almost all the components of the index. The largest impact relates to the statement "without mathematics it would be impossible to fight global challenges such as inequality or climate change". The impact is stronger on women, for whom this baseline idea was less prevalent than men, as shown in the first two columns of Table 3.9. The second largest effect corresponds to the statement "I will apply the mathematics I have learnt so far in the future as an adult". In this case, the effect is larger for boys, for whom the baseline idea was also more spread than for their female counterparts.

These findings are relevant as the literature shows that the perception that mathematics are useful and applicable fosters the sense of belonging to STEM. This notion—which reflects the extent to which individuals perceive their identity as a woman or man to fit with their identity in the STEM field (Shin et al., 2016)—has been shown to be a significant predictor of intention to pursue math careers in the future (Good et al., 2012), partly explaining women's underrepresentation in the field. These findings provide novel takeaways as most of the role model interventions on sense of belonging tackle issues beyond the usefulness of mathematics despite the evidence that when STEM education is oriented towards real-life problems, students show greater interest (Dare et al., 2021).

5.2. Maths-Related Career Aspirations and Perceptions of Maths-Related Jobs

The intervention also aimed to promote adolescents' self-projection into mathematics-related jobs. To assess the impact of the intervention on this measure, respondents were asked about their level of agreement or disagreement with three statements related to career aspirations, namely whether (1) they find jobs in the field of mathematics interesting, (2) they could see themselves working in a job related to mathematics later in life, and (3) career and income prospects play an important role in the choice of study.[42] Our results on the aggregate index confirm that the role model intervention was effective in fostering students' aspirations, with an average impact of 35% of an SD in the composite index. The effect is found to be particularly strong for women, in line with Breda et al. (2023) for France. Compared to men, female adolescents have a treatment effect that surpasses that of men by a third of an SD. This is relevant given that women appear particularly reticent to self-projecting into such jobs, as shown in the first two columns of Table 3.10.

The specific results for the three components of the index reveal that the strongest impact is observed regarding the statement "I could see myself working in a job related to mathematics later in life". The impact is positive for both genders, and slightly larger for women. This is relevant, as women have lower baseline levels of self-projection in those jobs than men (as shown in the first two columns of Table 3.10).

Relatedly, to measure the impact of students' perceptions on maths-related jobs, we use a composite index combining responses to five questions: (1) whether mathematics-related jobs are more monotonous or (2) lonely, (3) whether mathematics-related jobs pay higher salaries, (4) whether, compared to other jobs, mathematics jobs have little social impact, and (5) whether it is difficult to reconcile personal and professional life when working in mathematics-related jobs. The intervention has a positive impact on the composite impact of maths-related jobs (15.5% of an SD), although only at the 10% significance level (Table 3.10).

[42] Although the latter attribute does not specifically refer to the field of mathematics, adolescents largely believe that this attribute is more present in math-related fields than in other fields (see the third item of the index on perceptions of maths-related jobs in Table 3.10).

Table 3.10. Impact of Intervention on Math-Related Career Aspirations and Perceptions of Jobs

	Control group mean		Treatment effects		
	Women	Men	All	Women	Men
	(1)	(2)	(3)	(4)	(5)
Maths-related career aspirations (index)	-0.0841	0.0814	0.351***	0.400***	0.291**
			(0.077)	(0.118)	(0.102)
Interesting math-related jobs	0.041	-0.039	0.231**	0.212*	0.266**
			(0.076)	(0.108)	(0.108)
Self-projection in math-related jobs	-0.136	0.132	0.363***	0.362**	0.357**
			(0.079)	(0.115)	(0.110)
Importance career/income prospects	-0.085	0.082	0.157*	0.283**	0.001
			(0.083)	(0.119)	(0.117)
Perceptions of maths-related jobs (index)	-0.0386	0.0374	0.155*	0.259**	0.058
			(0.085)	(0.123)	(0.119)
Math jobs monotonous (r)	0.008	-0.008	0.060	0.075	0.071
			(0.080)	(0.110)	(0.116)
Math jobs lonely (r)	-0.089	0.086	-0.056	0.061	-0.177
			(0.082)	(0.116)	(0.119)
Math jobs pay higher salary	0.058	-0.056	0.148*	0.045	0.240**
			(0.081)	(0.120)	(0.112)
Low social impact of math jobs (r)	-0.065	0.063	0.095	0.241**	-0.033
			(0.083)	(0.113)	(0.123)
Difficult personal-professional balance in math jobs (r)	-0.01	0.009	0.145*	0.234**	0.045
			(0.082)	(0.117)	(0.116)
Observations	149	145	596	299	297

Note. The first two columns show the mean values of the z-scores for the control group. The last three columns reflect the coefficient of the treatment binary variable, with the dependent variable in each model being the one shown on the left-hand side. The item in bold reflects the index, which is composed of the *z*-scores of the sub-items below. (r) means that the order of responses to the questions was reverted so that a higher value of the index indicates a higher value on the maths-related career aspirations, in the first case, and on better perceptions of maths-related jobs, in the second case. All the models control for gender (only for the aggregate model), age, whether or not individuals are foreign-born, student status, number of books in the household, and reported skills on mathematics performance. *** p<0.01, ** p<0.05, and * p<0.1.

Separate specifications by gender reveal that the positive impact applies solely to women, with a magnitude of the treatment effect that fully aligns with Breda et al. (2023): their estimated coefficient amounts to 29.6% of an SD, while ours amounts to of 25.9%. This positive impact for women is driven by two key determinants. First, the intervention fosters women's idea that math-related jobs are compatible with a fulfilling family life (Breda et al., 2023). Second, the intervention has a positive effect on changing women's established mindset that math-related jobs have little social impact. These two elements are crucial in understanding women's later occupational choices (Wang & Degol, 2017), underscoring the relevance of the intervention in potentially reverting those stereotypes among girls.

5.3. Growth Mindset

While the intervention did not explicitly aim to foster a growth mindset, we examine whether it had any side effects in this regard. For this purpose, we asked adolescents about their level of agreement or disagreement with two statements. The first statement relates to the belief that intelligence remains relatively fixed, while the second statement addressed the belief that skills and intelligence can be developed through hard work and dedication.

The intervention positively affects individuals' growth mindset, although the gender breakdown shows that these effects are only statistically significant for men (Table 3.11). For context, male adolescents baseline levels of growth mindset are lower than their female counterparts, making this result especially relevant. The detailed results from the components of the index reveal that the intervention is only effective in fostering men's idea that skills and intelligence can be developed through hard work and dedication.

Table 3.11. Impact of Intervention on Growth Mindset

	Control group mean		Treatment effects		
	Women (1)	Men (2)	All (3)	Women (4)	Men (5)
Growth mindset (index)	0.193	-0.187	0.204**	0.091	0.316**
			(0.083)	(0.115)	(0.122)
Your intelligence can't change much (r)	0.116	-0.112	0.049	-0.041	0.136
			(0.085)	(0.114)	(0.127)
Your abilities/intelligence can be developed through hard work/dedication	0.153	-0.148	0.234**	0.168	0.304**
			(0.076)	(0.105)	(0.110)
Observations	149	145	596	299	297

Note. The first two columns show the mean values of the z-scores for the control group. The last three columns reflect the coefficient of the treatment binary variable, with the dependent variable in each model being the one shown on the left-hand side. The item in bold reflects the index, which is composed of the *z*-scores of the sub-items below. (r) means that the order of responses to the questions was reverted so that a higher value of the index indicates higher levels of growth mindset. All the models control for gender (only for the aggregate model), age, whether or not individuals are foreign-born, student status, number of books in the household, and reported skills on mathematics performance. Standard errors in parentheses. *** p<0.01, ** p<0.05, and * p<0.1.

5.4. Gender Attitudes

A key aim of the role model intervention was to reinforce the importance of female engagement in the field of mathematics. To measure its potential impact, we asked respondents whether they believed that (1) women and men are born with different brains, (2) men are innately better at mathematics or engineering than women, (3) there are fewer women in mathematics because women actually prefer to work in care-related jobs (education, health, etc.), (4) it is important to achieve a balanced ratio of women and men in mathematics in order to reduce the gender pay gap. Strikingly, the intervention is not effective in changing those gender attitudes (Table 3.C2). This may not have been anticipated given the prominent role of this topic during the intervention. A possible explanation is that, as observed in the previous subsection, the entrenchment of this type of stereotypes is substantial and therefore difficult to change, especially given the limited duration of the intervention.

The detailed results show that, for men, the intervention has no significant effect on any of the four questions. This is relevant taking into account that male adolescents reinforce traditional gender stereotypes to a larger extent than women (first two columns of Table 3.C2). Conversely, for women, the role model has a positive impact on the perceived idea that a balanced ratio of women and men in mathematics is important in order to reduce the gender pay gap

5.5. Self-Concept

Finally, to measure the impact of the intervention on adolescents' self-concept in mathematics, we use a composite index combining the responses to four questions: (1) whether adolescents feel lost when trying to solve a maths problem, (2) whether they get very tense when having to do maths problems, and (3) whether they believe that if they try hard enough, they can do well in mathematics. As the intervention did not touch upon individual-specific self-concept, the intervention does not have an effect on the composite index of self-concept in mathematics. However, the detailed results presented in Table 3.C3 show that the intervention has moderate effects in one of the three questions composing the index, although only for men. Specifically, the role model fosters the idea that, through effort, male respondents can do well in mathematics.

In sum, this analysis shows that a role model intervention has the potential to debias stereotyped views in the field of mathematics. The effects of these actions, if sustained over time, may help foster engagement of potential workforce in STEM, addressing the longstanding underrepresentation of women in this area.

6. Conclusions

This paper identifies the determinants of gender segregation and proposes a potential solution to address this phenomenon. The novelty of the paper is twofold. First, we explore the determinants of gender segregation using a holistic approach that captures potentially relevant information beyond that contained in traditional surveys. This is achieved through a large-scale survey designed and conducted on a representative sample of the Spanish population. Second, this paper contributes to a growing body of literature on the experimental testing of solutions to prevent gender segregation at the educational stage. To this end, we conduct a second survey among adolescents in Spain and test whether a female role model intervention can debias ingrained stereotypes in STEM.

The results show that educational presorting is the main driver of occupational segregation. Specifically, STEM (science, technology, engineering and mathematics) education reduces the probability of women holding female-dominated jobs. For men, HEAL (health, education, administration and literacy) education reduces their average probability of holding male-dominated jobs. Looking further into the drivers of educational presorting, we find that the main reason why individuals avoid STEM studies is related to maths anxiety, a common type of anxiety that is particularly prevalent among women.

After identifying the prominent role of educational presorting on gender segregation in the labour market, we test for a potential solution by focusing specifically on pre-university-age individuals. The results show that the role model intervention improves students' perceptions of mathematics as a useful and applicable tool, promotes their self-projection in math-related jobs, helps to reverse women's negative perceptions of the attributes offered by these jobs, and increases men's growth mindset.

In conclusion, this paper sheds light on the importance of addressing gender segregation prior to labour market entry. Previous literature and the findings of this paper have shown that informed educational interventions can help to reverse attitudes towards mathematics and science that particularly discourage women from making academic decisions choices in these fields. The pivotal role of STEM jobs in driving economic growth and national competitiveness (Deming & Noray, 2019), as well as the importance of non-segregated occupations in reducing the gender wage gap (Kahn & Ginther, 2017; Jiang, 2021), highlights that tackling gender segregation in labour markets should be at the forefront of the policy agenda.

References

Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management*, *53–56*(1), 63–70. https://doi.org/10.1016/j.jom.2017.06.001

Ahmed, W. (2018). Developmental trajectories of math anxiety during adolescence: Associations with STEM career choice. *Journal of Adolescence*, *67*(1), 158–166. https://doi.org/10.1016/j.adolescence.2018.06.010

Alan, S., Boneva, T., & Ertac, S. (2019). Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit. *The Quarterly Journal of Economics*, *134*(3), 1121–1162. https://doi.org/10.1093/qje/qjz006

Amuedo-Dorantes, C., & De La Rica, S. (2006). The Role of Segregation and Pay Structure on the Gender Wage Gap: Evidence from Matched Employer-Employee Data for Spain. *Contributions in Economic Analysis & Policy*, *5*(1). https://doi.org/10.2202/1538-0645.1498

Antecol, H., & Cobb-Clark, D. A. (2013). Do psychosocial traits help explain gender segregation in young people's occupations? *Labour Economics*, *21*, 59–73. https://doi.org/10.1016/j.labeco.2012.12.005

Ashby Plant, E., Baylor, A. L., Doerr, C. E., & Rosenberg-Kima, R. B. (2009). Changing middle-school students' attitudes and performance regarding engineering with computer-based social models. *Computers & Education*, *53*(2), 209–215. https://doi.org/10.1016/j.compedu.2009.01.013

Averett, S. L., Argys, L. M., & Hoffman, S. D. (Eds.). (2018). *The Oxford Handbook of Women and the Economy* (1st ed.). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190628963.001.0001

Azmat, G., & Iriberri, N. (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics*, *94*(7–8), 435–452. https://doi.org/10.1016/j.jpubeco.2010.04.001

Azmat, G., Bagues, M., Cabrales, A., & Iriberri, N. (2019). What You Don't Know…Can't Hurt You? A Natural Field Experiment on Relative Performance Feedback in Higher Education. *Management Science*, *65*(8), 3714–3736. https://doi.org/10.1287/mnsc.2018.3131

Borghans, L., & Groot, L. (1999). Educational presorting and occupational segregation. *Labour Economics*, *6*(3), 375–395. https://doi.org/10.1016/S0927-5371(99)00009-3

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences*, *107*(5), 1860–1863. https://doi.org/10.1073/pnas.0910967107

Bettio, F., & Verashchagina, A. (2009). *Gender segregation in the labour market: Root causes, implications and policy responses in the EU*. Publications Office of the European Union, https://data.europa.eu/doi/10.2767/1063

Blakemore, S-J. (2018). *Inventing Ourselves: The Secret Life of the Teenage Brain*. PublicAffairs.

Bishu, S. G., & Alkadry, M. G. (2017). A Systematic Review of the Gender Pay Gap and Factors That Predict It. *Administration & Society*, *49*(1), 65–104. https://doi.org/10.1177/0095399716636928

Borrowman, M., & Klasen, S. (2020). Drivers of Gendered Sectoral and Occupational Segregation in Developing Countries. *Feminist Economics*, *26*(2), 62–94. https://doi.org/10.1080/13545701.2019.1649708

Breda, T., Grenet, J., Monnet, M., & Van Effenterre, C. (2023). How Effective are Female Role Models in Steering Girls Towards STEM? Evidence from French High Schools. *The Economic Journal*, *133*(653), 1773–1809. https://doi.org/10.1093/ej/uead019

Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, Competitiveness, and Career Choices. *The Quarterly Journal of Economics*, *129*(3), 1409–1447. https://doi.org/10.1093/qje/qju009

Carranza, E., Das, S., & Kotikula, A. (2023). *Gender-based employment segregation: Understanding causes and policy interventions* (Jobs Working Paper 26). World Bank.

Cassady, J.C. (2022). Anxiety in the Schools: Causes, Consequences, and Solutions for Academic Anxieties. In: Gonzaga, L.R.V., Dellazzana-Zanon, L.L., Becker da Silva, A.M. (eds) *Handbook of Stress and Academic Anxiety* (pp. 13-30). Springer, Cham. https://doi-org.ehu.idm.oclc.org/10.1007/978-3-031-12737-3_2

Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in Academic Science: A Changing Landscape. *Psychological Science in the Public Interest*, *15*(3), 75–141. https://doi.org/10.1177/1529100614541236

Cheryan, S., Siy, J. O., Vichayapai, M., Drury, B. J., & Kim, S. (2011). Do Female and Male Role Models Who Embody STEM Stereotypes Hinder Women's Anticipated Success in STEM? *Social Psychological and Personality Science*, *2*(6), 656–664. https://doi.org/10.1177/1948550611405218

Combet, B. (2023). Women's aversion to majors that (seemingly) require systemizing skills causes gendered field of study choice. *European Sociological Review*, jcad021. https://doi.org/10.1093/esr/jcad021

Cortina, C., Rodríguez, J., & González, M. J. (2021). Mind the Job: The Role of Occupational Characteristics in Explaining Gender Discrimination. *Social Indicators Research*, *156*(1), 91–110. https://doi.org/10.1007/s11205-021-02646-2

Cuevas Ruiz, P., Sanz, I., & Sevilla, A. (2023). Explaining the Mathematics Gender Gap: The Role of Stereotypes. In P. Cuevas Ruiz, I. Sanz, & A. Sevilla, Oxford Research Encyclopedia of Economics and Finance. Oxford University Press. https://doi.org/10.1093/acrefore/9780190625979.013.892

Daker, R. J., Gattas, S. U., Sokolowski, H. M., Green, A. E., & Lyons, I. M. (2021). First-year students' math anxiety predicts STEM avoidance and underperformance throughout university, independently of math ability. Npj *Science of Learning*, *6*(1), 17. https://doi.org/10.1038/s41539-021-00095-7

Dare, E. A., Keratithamkul, K., Hiwatig, B. M., & Li, F. (2021). Beyond Content: The Role of STEM Disciplines, Real-World Problems, 21st Century Skills, and STEM Careers within Science Teachers' Conceptions of Integrated STEM Education. *Education Sciences*, *11*(11), 737. https://doi.org/10.3390/educsci11110737

Del Carpio, L., & Guadalupe, M. (2022). More Women in Tech? Evidence from a Field Experiment Addressing Social Identity. *Management Science*, *68*(5), 3196–3218. https://doi.org/10.1287/mnsc.2021.4035

Deming, D., & Noray, K. (2018). *STEM Careers and the Changing Skill Requirements of Work* (w25065; p. w25065). National Bureau of Economic Research. https://doi.org/10.3386/w25065

Dueñas Fernández, D., Iglesias Fernández, C., & Llorente Heras, R. (2014). Occupational segregation by sex in Spain: Exclusion or confinement? *International Labour Review*, *153*(2), 311–336. https://doi.org/10.1111/j.1564-913X.2014.00206.x

Drury, B. J., Siy, J. O., & Cheryan, S. (2011). When Do Female Role Models Benefit Women? The Importance of Differentiating Recruitment From Retention in STEM. *Psychological Inquiry*, *22*(4), 265–269. https://doi.org/10.1080/1047840X.2011.620935

Duncan, O. D., & Duncan, B. (1955). A methodological analysis of segregation indices. *American Sociological Review*, *20*, 200–217.

Eccles, J. S., & Roeser, R. W. (2011). Schools as developmental contexts during adolescence. *Journal of Research on Adolescence*, *21*(1), 225–241. https://doi.org/10.1111/j.1532-7795.2010.00725.x

Ecklund, E. H., Lincoln, A. E., & Tansey, C. (2012). Gender Segregation in Elite Academic Science. *Gender & Society*, *26*(5), 693–717. https://doi.org/10.1177/0891243212451904

Eurofound & European Commission Joint Research Centre (2021), European Jobs Monitor 2021: Gender gaps and the employment structure, European Jobs Monitor series, Publications Office of the European Union, Luxembourg.

Fernández-Cézar, R., Garrido, D., & Solano-Pinto, N. (2020). Do Science, Technology, Engineering and Mathematics (STEM) Experimentation Outreach Programs Affect Attitudes towards Mathematics and Science? A Quasi-Experiment in Primary Education. *Mathematics*, *8*(9), 1490. https://doi.org/10.3390/math8091490

Flory, J. A., Leibbrandt, A., & List, J. A. (2015). Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job Entry Decisions. T*he Review of Economic Studies*, *82*(1), 122–155. https://doi.org/10.1093/restud/rdu030

Foley, A. E., Herts, J. B., Borgonovi, F., Guerriero, S., Levine, S. C., & Beilock, S. L. (2017). The Math Anxiety-Performance Link: A Global Phenomenon. *Current Directions in Psychological Science*, *26*(1), 52–58. https://doi.org/10.1177/0963721416672463

González-Cervera, A., González-Arechavala, Y., Martín-Carrasquilla, O., Santaolalla, E., & Cubiles, M. (2021). *Estudios STEM en España y participación de la mujer. La Formación Profesional STEM, una oportunidad de futuro*. Cátedra para la Promoción de la Mujer en vocaciones STEM en la Formación Profesional para la Movilidad Sostenible.

González-Pérez, S., Mateos De Cabo, R., & Sáinz, M. (2020). Girls in STEM: Is It a Female Role-Model Thing? *Frontiers in Psychology*, *11*, 2204. https://doi.org/10.3389/fpsyg.2020.02204

Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology*, *102*(4), 700–717. https://doi.org/10.1037/a0026659

Gortazar, L., Hupkau, C., & Roldán-Monés, A. (2024). Online tutoring works: Experimental evidence from a program with vulnerable children. *Journal of Public Economics*, *232*, 105082. https://doi.org/10.1016/j.jpubeco.2024.105082

Hanushek, E., Jacobs, B., Schwerdt, G., Van Der Velden, R., Vermeulen, S., & Wiederhold, S. (2023). *Where Do STEM Graduates Stem From? The Intergenerational Transmission of Comparative Skill Advantages* (NBER Working Paper No. 31186). https://doi.org/10.3386/w31186

He, J. C., Kang, S. K., Tse, K., & Toh, S. M. (2019). Stereotypes at work: Occupational stereotypes predict race and gender segregation in the workforce. *Journal of Vocational Behavior*, *115*, 103318. https://doi.org/10.1016/j.jvb.2019.103318

Hegewisch, A., & Hartmann, H. (2014). *Occupational Segregation and the Gender Wage Gap: A Job Half Done*. Institute for Women's Policy Research, https://doi.org/10.1037/e529142014-001

Jiang, X. (2021). Women in STEM: Ability, preference, and value. *Labour Economics*, *70*, 101991. https://doi.org/10.1016/j.labeco.2021.101991

Kahn, S., & Ginther, D. (2017). *Women and STEM* (NBER Working Paper No. w23525). https://doi.org/10.3386/w23525

Kamerāde, D., & Richardson, H. (2018). Gender segregation, underemployment and subjective well-being in the UK labour market. *Human Relations*, *71*(2), 285–309. https://doi.org/10.1177/0018726717713829

Kleinjans, K. J. (2009). Do gender differences in preferences for competition matter for occupational expectations? *Journal of Economic Psychology*, *30*(5), 701–710. https://doi.org/10.1016/j.joep.2009.03.006

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental Analysis of Neighborhood Effects. *Econometrica*, *75*(1), 83–119.

Levanon, A., England, P., & Allison, P. (2009). Occupational Feminization and Pay: Assessing Causal Dynamics Using 1950-2000 U.S. Census Data. *Social Forces*, *88*(2), 865–891. https://doi.org/10.1353/sof.0.0264

Levanon, A., & Grusky, D. B. (2016). The Persistence of Extreme Gender Segregation in the Twenty-first Century. *American Journal of Sociology*, *122*(2), 573–619. https://doi.org/10.1086/688628

Lim, J., & Meer, J. (2017). The Impact of Teacher–Student Gender Matches: Random Assignment Evidence from South Korea. *Journal of Human Resources*, *52*(4), 979–997. https://doi.org/10.3368/jhr.52.4.1215-7585R1

Maestas, N., Mullen, K. J., & Powell, D. (2022). The Value of Working Conditions in the United States and Implications for the Structure of Wages. *American Economic Review*, *113*(7), 2007–2047. https://doi.org/DOI:10.1257/aer.20190846

Maloney, E. A., Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2015). Intergenerational Effects of Parents' Math Anxiety on Children's Math Achievement and Anxiety. *Psychological Science*, *26*(9), 1480–1488. https://doi.org/10.1177/0956797615592630

Marsh, H. W., & Hau, K.-T. (2004). Explaining Paradoxical Relations Between Academic Self-Concepts and Achievements: Cross-Cultural Generalizability of the Internal/External Frame of Reference Predictions Across 26 Countries. *Journal of Educational Psychology*, *96*(1), 56–67. https://doi.org/10.1037/0022-0663.96.1.56

Miller, L., Neathey, F., Pollard, E., & Hill, D. (2004). *Occupational segregation, gender gaps and skill gaps* (Institute for Employment Studies Working Paper 15).

Nickow, A., Oreopoulos, P., & Quan, V. (2020). The Impressive Effects of Tutoring on PreK-12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence (NBER Working Paper No. 27476). https://doi.org/10.3386/w27476

Nicoletti, C., Sevilla, A., & Tonei, V. (2022). Gender Stereotypes in the Family. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4294398

Nollenberger, N., Rodríguez-Planas, N., & Sevilla, A. (2016). The Math Gender Gap: The Role of Culture. *American Economic Review*, *106*(5), 257–261. https://doi.org/10.1257/aer.p20161121

Osés, A., Martínez de Lafuente, D., & de la Rica, S. (2024). *Segregación de género en el mercado laboral: causas y soluciones* (Informe ISEAK 2024/1).

Park, D., Ramirez, G., & Beilock, S. L. (2014). The role of expressive writing in math anxiety. *Journal of Experimental Psychology: Applied*, *20*(2), 103–111. https://doi.org/10.1037/xap0000013

Ramirez, G., Shaw, S. T., & Maloney, E. A. (2018). Math Anxiety: Past Research, Promising Interventions, and a New Interpretation Framework. *Educational Psychologist*, *53*(3), 145–164. https://doi.org/10.1080/00461520.2018.1447384

Reuben, E., Wiswall, M., & Zafar, B. (2017). Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender. *The Economic Journal*, *127*(604), 2153–2186. https://doi.org/10.1111/ecoj.12350

Shauman, K. A. (2006). Occupational sex segregation and the earnings of occupations: What causes the link among college-educated workers? *Social Science Research*, *35*(3), 577–619. https://doi.org/10.1016/j.ssresearch.2004.12.001

Soni, A., & Kumari, S. (2017). The Role of Parental Math Anxiety and Math Attitude in Their Children's Math Achievement. *International Journal of Science and Mathematics Education*, *15*(2), 331–347. https://doi.org/10.1007/s10763-015-9687-5

Stantcheva, S. (2022). *Understanding of Trade* (NBER Working Paper No. 30040). https://doi.org/10.3386/w30040

Stantcheva, S. (2023). How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible. *Annual Review of Economics*, *15*, 205–234. https://doi.org/10.1146/annurev-economics-091622-010157

Stewart-Williams, S., & Halsey, L. G. (2021). Men, women and STEM: Why the differences and what should be done? *European Journal of Personality*, *35*(1), 3–39. https://doi.org/10.1177/0890207020962326

Stoet, G., & Geary, D. C. (2012). Can Stereotype Threat Explain the Gender Gap in Mathematics Performance and Achievement? Review of General Psychology, 16(1), 93–102. https://doi.org/10.1037/a0026617

Supekar, K., Iuculano, T., Chen, L., & Menon, V. (2015). Remediation of Childhood Math Anxiety and Associated Neural Circuits through Cognitive Tutoring. *Journal of Neuroscience*, *35*(36), 12574–12583. https://doi.org/10.1523/JNEUROSCI.0786-15.2015

Tandrayen-Ragoobur, V. and Gokulsing, D. (2022). Gender gap in STEM education and career choices: what matters? *Journal of Applied Research in Higher Education*, *14*(3), 1021-1040. https://doi.org/10.1108/JARHE-09-2019-0235

Torre, M., & Jacobs, J. A. (2021). The Gender Mobility Paradox: Gender Segregation and Women's Mobility Across Gender-Type Boundaries, 1970–2018. *Gender & Society*, *35*(6), 853–883. https://doi.org/10.1177/08912432211046328

Wang, M.-T., & Degol, J. (2013). Motivational pathways to STEM career choices: Using expectancy–value perspective to understand individual and gender differences in STEM fields. *Developmental Review*, *33*(4), 304–340. https://doi.org/10.1016/j.dr.2013.08.001

Wang, M.-T., & Degol, J. L. (2016). School Climate: A Review of the Construct, Measurement, and Impact on Student Outcomes. *Educational Psychology Review*, *28*(2), 315–352. https://doi.org/10.1007/s10648-015-9319-1

Wang, M.-T., & Degol, J. L. (2017). Gender Gap in Science, Technology, Engineering, and Mathematics (STEM): Current Knowledge, Implications for Practice, Policy, and Future Directions. *Educational Psychology Review*, *29*(1), 119–140. https://doi.org/10.1007/s10648-015-9355-x

Wang, M.-T., L. Degol, J., Amemiya, J., Parr, A., & Guo, J. (2020). Classroom climate and children's academic and psychological wellbeing: A systematic review and meta-analysis. *Developmental Review*, *57*, 100912. https://doi.org/10.1016/j.dr.2020.100912

Weeden, K. A. (1998). Revisiting occupational sex segregation in the United States, 1910–1990: Results from a log-linear approach. *Demography*, *35*, 475–487. doi:10.2307/3004015

Wiswall, M., & Zafar, B. (2018). Preference for the Workplace, Investment in Human Capital, and Gender. *The Quarterly Journal of Economics*, *133*(1), 457–507. https://doi.org/10.1093/qje/qjx035

Appendix

Appendix 3.A

Classification of Dependent Variables for the Gender Segregation Models

Table 3.A1. Gender Composition of Occupations in Spain in 2023

Code	Label	Men	Women	(1) Male prop. (%)	(2) Female prop. (%)	(2)-(1)
	Total	11,257.1	9,799.6			
11	Legislators and senior officials; General Government and social-interest organisation executives Executive directors	26.0	15.8	0.23	0.16	-0.07
12	Administrative and commercial department managers	145.2	113.0	1.29	1.15	-0.14
13	Production and operations managers	176.7	69.4	1.57	0.71	-0.86
14	Accommodation, catering and trade managers	119.4	62.5	1.06	0.64	-0.42
15	Services managers not elsewhere classified	91.3	43.0	0.81	0.44	-0.37
21	Health professionals	236.5	591.8	2.10	6.04	3.94
22	Early childhood, primary, secondary and post-secondary teaching professionals	310.8	640.5	2.76	6.54	3.78
23	Other education professionals	73.7	176.5	0.65	1.80	1.15
24	Physical science, chemistry, mathematics and engineering professionals	462.9	241.6	4.11	2.47	-1.65
25	Legal professionals	104.7	134.3	0.93	1.37	0.44
26	Business, administration and marketing professionals	235.3	311.0	2.09	3.17	1.08
27	Information technology professionals	183.3	64.4	1.63	0.66	-0.97
28	Social sciences professionals	63.3	168.1	0.56	1.72	1.15
29	Cultural and entertainment professionals	83.5	66.7	0.74	0.68	-0.06
31	Science and engineering technicians	253.8	97.3	2.25	0.99	-1.26
32	Mining, manufacturing and construction supervisors	93.3	13.0	0.83	0.13	-0.70
33	Alternative therapy health technicians and professionals	55.5	113.2	0.49	1.16	0.66
34	Financial and mathematical support professionals	43.1	45.6	0.38	0.47	0.08
35	Representatives, sales and purchasing agents and related professionals	457.7	244.4	4.07	2.49	-1.57
36	Clerical support workers; security force technicians	106.5	233.3	0.95	2.38	1.43
37	Legal, social, cultural, sporting and similar services support professionals	186.3	166.3	1.65	1.70	0.04
38	Information and communications technology (ICT) technicians	337.4	63.2	3.00	0.64	-2.35
41	Accounting and finance services employees, and production and transport support services employees	267.1	314.2	2.37	3.21	0.83
42	Library, mail carrier and related clerks	22.5	38.4	0.20	0.39	0.19
43	Other administrative employees who do not work with customer services	142.2	363.3	1.26	3.71	2.44
44	Travel consultants and clerks, receptionists and telephone switchboard operators; window employees and the like (except ticket sellers)	104.6	242.2	0.93	2.47	1.54
45	Administrative employees who work with customer services and are not elsewhere classified	144.1	449.4	1.28	4.59	3.31
50	Waiters and bartenders, and cooks who are restaurant owners	134.8	110.4	1.20	1.13	-0.07
51	Catering services wage-earning workers	432.0	496.2	3.84	5.06	1.23

52	Shop salespersons	242.5	571.3	2.15	5.83	3.68
53	Shopowner traders	148.7	174.5	1.32	1.78	0.46
54	Sales workers (except in shops and department stores)	70.6	66.4	0.63	0.68	0.05
55	Cashiers and ticket clerks (except in banks)	30.2	147.2	0.27	1.50	1.23
56	Health services personal care workers	101.2	402.4	0.90	4.11	3.21
57	Other personal care workers	18.9	377.5	0.17	3.85	3.68
58	Personal service workers	189.3	315.4	1.68	3.22	1.54
59	Protective and security services workers	414.2	65.8	3.68	0.67	-3.01
61	Skilled agricultural workers	245.4	51.3	2.18	0.52	-1.66
62	Skilled livestock workers (including poultry, beekeeping and related livestock)	63.7	24.3	0.57	0.25	-0.32
63	Skilled mixed agricultural workers	16.1	3.2	0.14	0.03	-0.11
64	Skilled forestry, fishery and hunting workers	30.7	4.4	0.27	0.04	-0.23
71	Structural construction workers and related workers	592.6	9.8	5.26	0.10	-5.16
72	Building and installation finishers (except electricians), painters and related workers	303.9	7.4	2.70	0.08	-2.62
73	Welders, sheet-metal workers, structural-metal preparers and erectors, blacksmiths, toolmakers and related trades workers	212.3	5.6	1.89	0.06	-1.83
74	Machinery mechanics and adjusters	320.7	3.5	2.85	0.04	-2.81
75	Electrical and electronic trades workers	360.4	5.5	3.20	0.06	-3.15
76	Metal precision mechanics, ceramists, glass workers, handicraft and printing workers	56.2	22.5	0.50	0.23	-0.27
77	Food, beverage and tobacco industry workers	112.7	94.3	1.00	0.96	-0.04
78	Woodworking, textile, garment, fur, leather, footwear and other trade workers	66.6	36.6	0.59	0.37	-0.22
81	Stationary plant and machine operators	307.8	144.3	2.73	1.47	-1.26
82	Factory fitters and assemblers	111.2	31.2	0.99	0.32	-0.67
83	Locomotive engine drivers, agricultural machine and mobile heavy equipment operators, and seamen	212.0	9.4	1.88	0.10	-1.79
84	City or road transport vehicle drivers	794.0	44.0	7.05	0.45	-6.60
91	Domestic employees	8.8	392.2	0.08	4.00	3.92
92	Other cleaning workers	121.1	633.2	1.08	6.46	5.39
93	Food preparation assistants	61.5	126.6	0.55	1.29	0.75
94	Urban garbage workers, street vendors and other elementary services occupations	118.1	53.0	1.05	0.54	-0.51
95	Agrarian, forestry and fishery labourers	260.9	83.1	2.32	0.85	-1.47
96	Construction and mining labourers	144.0	5.9	1.28	0.06	-1.22
97	Manufacturing labourers	83.5	112.8	0.74	1.15	0.41
98	Transport, loading and stocking labourers	253.7	72.5	2.25	0.74	-1.51
00	Military occupations	90.0	13.6	0.80	0.14	-0.66

Source. Author's own workings based on the Spanish Labour Force Survey (2023, second quarter).
Note. The third and fourth columns show the proportion of men and women over the total employment of each sex, respectively. The last column calculates the difference between the proportion of women and that of men for each occupation. Occupations follow Spain's NCO-11 classification.

Table 3.A2. Crosswalk between the Occupational Classification in the Survey and the National Classification of Occupations (NCO-11)

Code, Survey	Label, Survey	Codes, NCO-11	Diff. Share Women and Men
1	Members of executive and legislative bodies	11	-0.07
2	Executive personnel in public administration		
3	Executive personnel in commerce, advertising, public relations, production, services, and hospitality	12-15	-1.79
4	Health professionals, teaching staff	21-23	8.86
5	Professionals in sciences and engineering	24, 27	-2.62
6	Legal professionals, organisational specialists	25-26	1.52
7	Analysts, economists, sociologists, journalists, writers, artists...	28-29	1.09
8	Support professionals (drafters, technicians in sciences and health, navigation professionals, quality control, supervisors, real estate agents...)	31-35	-2.79
9	Technicians of security forces and bodies	36	1.43
10	Athletes, coaches, sports activity instructors...	-	-0.85
11	Computer programmers and ICT technicians	31, 38	-3.61
12	Accountants, financial professionals, materials registration employees, production and transportation support	41	0.83
13	Library employees, postal services, customer service	42-45	7.48
14	Workers in hospitality and commerce	50-55	6.57
15	Workers in health services, care for individuals, and aesthetics	56-58	8.43
16	Workers in protection and security services	59	-3.01
17	Agricultural, livestock, forestry, and fishing sector	61-64	-2.31
18	Structural works workers	71	-5.16
19	Construction and installation workers	72	-2.62
20	Workers in manufacturing industries (welders, metal workers, mechanics, electricians, materials, food, beverages, tobacco, skilled laborers...)	73-78, 97	-7.90
21	Assemblers, miners, drillers...	81-82	-1.93
22	Drivers and mobile machinery operators	83-84	-8.39
23	Domestic employees, cleaning personnel, kitchen assistants, porters, collectors...	91-94	9.55
24	Agricultural, fishing, construction, and transportation laborers	95-96, 98	-4.20
25	Officers, non-commissioned officers, armed forces…	00	-0.66

Note. For the specific NCO-11 labels, see Table 3.A1. The fourth column is calculated based on the Spanish Labour Force Survey microdata for 2023 (second quarter) by applying the crosswalk as shown in this table. For occupational group 10 of the survey, the last column is calculated as per the survey data given the inability of matching it with an NCO-11 occupation.

Table 3.A3. Classification of Study Areas by STEM and HEAL

	STEM	HEAL
VET		
Agriculture	0	0
Maritime-fisheries	0	0
Food industry	1	0
Chemistry	1	0
Personal image	0	0
Health	0	1
Security and environment	0	0
Mechanical manufacturing	1	0
Installation and maintenance	1	0
Electricity and electronics	1	0
Vehicle transportation and maintenance	1	0
Extractive industries	1	0
Building and civil works	1	0
Glass and ceramics	0	0
Wood, furniture and cork	0	0
Textile, clothing and leather	0	0
Graphic arts	0	0
Image and sound	0	0
Information and communication	1	0
Administration and management	0	1
Commerce and marketing	0	1
Sociocultural services and services for the community	0	1
Hospitality and tourism		0
Physical and sports activities	0	0
Other training cycles in Plastic Arts and Design	0	0
Arts and humanities		
Arts (other)	0	1
Social and behavioral sciences	0	1
Languages	0	1
Humanities	0	1
Other	0	1
University studies		
Sciences		
Life sciences	1	0
Chemical, physical and geological sciences	1	0
Manufacturing industry and production	1	0
Mathematics and statistics	1	0
Environment	1	0
Other	1	0

	STEM	HEAL
Social and legal sciences		
Social and behavioral sciences	0	0
Management and administration	0	1
Law	0	0
Economics	0	0
Education (other)	0	1
Primary education teachers	0	1
Secondary education teachers	0	1
Business and administration	0	1
Journalism and documentation	0	1
Psychology	0	1
Audio-visual techniques and communication media	0	0
Social services	0	1
Travel, tourism and leisure	0	0
Other	0	0
Health sciences		
Physical and sports activities	0	0
Nursing	0	1
Medicine	0	1
Veterinary	0	1
Health (other)	0	1
Other	0	1
Engineering and architecture		
Agriculture and livestock	1	0
Architecture and construction	1	0
IT	1	0
Engineering	1	0
Manufacturing industry and production	1	0
Forestry	1	0
Other	1	0

Appendix 3.B

Alternative Specifications and Robustness Checks

Table 3.B1. Alternative Specifications: Exclusion of Certain Variables

	Specification 1		Specification 2	
	Women	Men		
Field of study				
STEM	0.285***	-0.497***		
	(0.037)	(0.031)		
HEAL			-0.171***	0.397***
			(0.032)	(0.035)
Psychological factors				
Math anxiety during adolescence				
Self-concept				
Preference work with people vs machines				
Socialisation and collectivism				
Risk lover				
Socialisation and cultural norms				
Against gender stereotypes	-0.069***	0.007	-0.077***	0.016
	(0.020)	(0.018)	(0.020)	(0.017)
Contextual school and family factors				
Female role model				
Male role model				
Progressive family education	-0.023	-0.041**	-0.020	-0.043**
	(0.017)	(0.020)	(0.017)	(0.019)
Intellectual aptitude and aspirations				
Reported cognitive ability				
Family prioritisation in past	-0.005	0.014	-0.007	0.014
	(0.015)	(0.019)	(0.015)	(0.018)
Observations	1071	899	1071	899

Note. The coefficients reflect the marginal effects of the probit models. All variables shown in the tables have been transformed into *z*-scores and coefficients can be interpreted as partial correlations. Standard errors in parentheses. *** p<0.01, ** p<0.05, and * p<0.1. All the models control for educational attainment, age, region of birth (foreign or native), voting intention (left versus rest), and mother's and father's educational attainment.

Table 3.B2. Alternative Specifications: Threshold Variations on Definition of Occupational Segregation

	Baseline		Alternative (t = 0)		Alternative (t =(-5, 5))	
	Women	Men	Women	Men	Women	Men
Field of study						
STEM	0.265***	-0.477***	0.418***	-0.449***	0.265***	-0.061**
	(0.057)	(0.044)	(0.062)	(0.044)	(0.057)	(0.020)
Psychological factors						
Math anxiety during adolescence	-0.020	-0.025	-0.024	0.011	-0.020	-0.008
	(0.027)	(0.028)	(0.025)	(0.024)	(0.027)	(0.010)
Self-concept	0.058**	0.022	0.008	-0.007	0.058**	-0.003
	(0.028)	(0.029)	(0.026)	(0.026)	(0.028)	(0.012)
Preference work with people vs machines	-0.012	-0.063**	0.018	0.043*	-0.012	0.001
	(0.028)	(0.029)	(0.026)	(0.025)	(0.028)	(0.010)
Socialisation and collectivism	-0.002	0.041	0.026	0.009	-0.002	0.001
	(0.029)	(0.029)	(0.026)	(0.025)	(0.029)	(0.011)
Risk lover	-0.040	-0.064	0.051	-0.111**	-0.040	-0.031*
	(0.052)	(0.056)	(0.047)	(0.052)	(0.052)	(0.019)
Socialisation and cultural norms						
Against gender stereotypes	-0.089**	-0.024	-0.072**	0.014	-0.089**	-0.001
	(0.032)	(0.026)	(0.027)	(0.024)	(0.032)	(0.010)
Contextual school and family factors						
Female role model	0.054	0.115*	0.051	0.148**	0.054	0.041*
	(0.055)	(0.068)	(0.052)	(0.061)	(0.055)	(0.024)
Male role model	-0.024	0.077	-0.032	0.040	-0.024	-0.046*
	(0.055)	(0.069)	(0.051)	(0.062)	(0.055)	(0.026)
Progressive family education	-0.025	-0.060**	-0.028	-0.015	-0.025	0.000
	(0.027)	(0.028)	(0.025)	(0.024)	(0.027)	(0.011)
Intellectual aptitude and aspirations						
Reported cognitive ability	0.032	0.003	0.013	0.033	0.032	0.014
	(0.028)	(0.029)	(0.026)	(0.025)	(0.028)	(0.010)
Family vs professional prioritisation in the past	-0.025	-0.030	-0.015	0.012	-0.025	-0.010
	(0.023)	(0.028)	(0.022)	(0.024)	(0.023)	(0.010)
Socio-demographic and other factors (exc. education)						
Age	0.005	-0.004	-0.002	-0.002	0.005	0.001
	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.001)
Foreign-born	-0.166**	-0.067	-0.122**	-0.010	-0.166**	0.054***
	(0.077)	(0.095)	(0.059)	(0.081)	(0.077)	(0.015)
Left wing	0.017	0.060	-0.041	-0.030	0.017	-0.050**
	(0.050)	(0.056)	(0.046)	(0.048)	(0.050)	(0.025)
Mother's education	0.067**	-0.050	0.041	-0.001	0.067**	0.001
	(0.031)	(0.036)	(0.029)	(0.031)	(0.031)	(0.014)
Father's education	-0.037	-0.022	0.011	-0.054*	-0.037	0.010
	(0.031)	(0.036)	(0.029)	(0.031)	(0.031)	(0.014)
Observations	479	484	479	484	479	484

Note. The coefficients reflect the marginal effects of the probit models. All variables shown in the table, except for the socio-demographic ones, have been transformed into *z*-scores and coefficients can be interpreted as partial correlations. Standard errors in parentheses. *** p<0.01, ** p<0.05, and * p<0.1. All the models also control for educational attainment.

Table 3.B3. Alternative Specifications: OLS estimation, Ranking of Occupations by Descending Degree of Segregation

	Specification 1		Specification 2	
	Women	Men	Women	Men
Field of study				
STEM	-2.492***	3.281***		
	(0.305)	(0.303)		
HEAL			1.568***	-2.913***
			(0.256)	(0.363)
Psychological factors				
Math anxiety during adolescence	0.272	-0.481	0.332	-0.734**
	(0.265)	(0.344)	(0.273)	(0.359)
Self-concept	-0.170	0.029	-0.284	-0.145
	(0.265)	(0.351)	(0.272)	(0.368)
Preference work with people vs machines	-0.295	0.159	-0.456	0.392
	(0.269)	(0.351)	(0.281)	(0.371)
Socialisation and collectivism	0.513*	0.184	0.682**	0.113
	(0.274)	(0.349)	(0.282)	(0.366)
Risk lover	0.222	0.423	0.335	0.302
	(0.249)	(0.326)	(0.256)	(0.342)
Socialisation and cultural norms				
Against gender stereotypes	0.334	-0.239	0.393	-0.494
	(0.296)	(0.321)	(0.305)	(0.335)
Contextual school and family factors				
Female role model	-0.086	-0.403	-0.078	-0.506
	(0.264)	(0.403)	(0.272)	(0.424)
Male role model	-0.063	-0.529	-0.147	-0.627
	(0.266)	(0.402)	(0.273)	(0.421)
Progressive family education	0.025	0.443	-0.039	0.465
	(0.266)	(0.338)	(0.274)	(0.355)
Intellectual aptitude and aspirations				
Reported cognitive ability	-0.230	-0.438	-0.324	-0.076
	(0.273)	(0.345)	(0.280)	(0.359)
Family prioritisation in past	-0.189	-0.075	-0.160	0.032
	(0.227)	(0.322)	(0.233)	(0.338)
Observations	479	484	479	484

Note. The dependent variable for women (men) consists of a ranking of occupations from more feminised (masculinised) to less. This is calculated by comparing the gender-specific share of employment of each occupation for men and women, respectively. All the models are estimated through OLS and the coefficients can be interpreted as partial correlations. Standard errors in parentheses. *** p<0.01, ** p<0.05, and * p<0.1. All the models control for educational attainment, age, region of birth (foreign or native), voting intention (left versus rest), and mother’s and father’s educational attainment.

Appendix 3.C

1. Usefulness of mathematics (four items): “Mathematics are very abstract and not very applicable to day-to-day life”, “Mathematics help to think logically”, “Without mathematics it would be impossible to fight global challenges such as inequality or climate change”, “I will apply the mathematics I have learnt so far in the future as an adult”.
2. Maths-related career aspirations (three items): “Some jobs in the field of mathematics are interesting”, “I could see myself working in a job related to mathematics later in life”, “Career and income prospects play an important role in my choice of study”.
3. Perceptions of maths-related jobs (six items): “Mathematics-related jobs are more monotonous”, “Mathematics-related jobs are quite lonely”, “Mathematics-related jobs pay higher salaries”, “Compared to other jobs, mathematics jobs have little social impact”, “It is difficult to reconcile personal and professional life when working in mathematics-related jobs”.
4. Growth mindset (two items): “Your intelligence is something about you that you can't change too much”, “Your skills and intelligence can be developed through hard work and dedication”.
5. Gender attitudes (four items): “Women and men are born with different brains”, “Men are innately better at mathematics or engineering than women”, “There are fewer women in mathematics because women actually prefer to work in care-related jobs (education, health, etc.)”, “It is important to achieve a balanced ratio of women and men in mathematics in order to reduce the gender pay gap (the fact that women have lower average salaries than men)”.
6. Self-concept (three items): “I feel lost when I try to solve a maths problem”, “I get very tense when I have to do maths problems”, “If I try hard enough, I can do well in mathematics”.

Table 3.C1. Impact of Intervention on Perceptions about Mathematics

	Control group mean			Treatment effects		
	All	Women	Men	All	Women	Men
Utility of mathematics (index)	0.00	-0.0493	0.0478	0.4427***	0.4279***	0.4457***
				(0.0789)	(0.1067)	(0.1181)
Maths-related career aspirations (index)	0.00	-0.0841	0.0814	0.3507***	0.4004***	0.2911***
				(0.0774)	(0.1180)	(0.1016)
Growth mindset (index)	0.00	0.1926	-0.1866	0.2039**	0.09101	0.3157***
				(0.0830)	(0.1149)	(0.1218)
Perceptions of maths-related jobs (index)	0.00	-0.0386	0.0374	0.1550*	0.2590**	0.0577
				(0.0851)	(0.1234)	(0.1194)
Gender attitudes on mathematics (index)	0.00	0.1815	-0.1758	-0.0506	0.0268	-0.1175
				(0.0789)	(0.1136)	(0.1102)
Self-concept in mathematics (index)	0.00	-0.1002	0.0970	0.0325	-0.0239	0.0814
				(0.0726)	(0.0998)	(0.1070)
Observations	294	149	145	596	299	297

Note. The first two columns show the mean values of each standardised index for the control group. The last three columns reflect the coefficient of the treatment binary variable, with the dependent variable in each model being the one shown on the left-hand side. All the models control for gender (only for the aggregate model), age, whether or not individuals are foreign-born, student status, number of books in the household, and reported skills on mathematics performance. Standard errors in parentheses. *** p<0.01, ** p<0.05, and * p<0.1.

Table 3.C2. Impact of Intervention on Gender Attitudes and Self-Concept

	Control group mean		Treatment effects		
	Women (1)	Men (2)	All (3)	Women (4)	Men (5)
Gender attitudes (index)	0.1815	-0.1758	-0.051	0.027	-0.117
			(0.079)	(0.114)	(0.110)
Women and men are born with different brains (r)	0.004	-0.004	-0.140*	-0.165	-0.128
			(0.084)	(0.119)	(0.120)
Men innately better at maths/engineering (r)	0.161	-0.156	-0.026	0.071	-0.089
			(0.079)	(0.109)	(0.116)
Women underrepresented because they prefer care-related jobs (r)	0.108	-0.104	-0.153*	-0.136	-0.171
			(0.081)	(0.121)	(0.110)
Important female inclusion to reduce gender pay gap	0.178	-0.172	0.193**	0.297**	0.096
			(0.075)	(0.103)	(0.110)
Self-concept (index)	-0.1002	0.0970	0.032	-0.024	0.081
			(0.073)	(0.100)	(0.107)
I feel lost when doing maths problems (r)	-0.136	0.131	0.001	-0.028	0.015
			(0.075)	(0.102)	(0.110)
I get very tense when doing maths problems (r)	-0.101	0.098	-0.084	-0.131	-0.043
			(0.078)	(0.109)	(0.113)
If I try hard, I can do well in mathematics	0.030	-0.029	0.150**	0.109	0.196*
			(0.075)	(0.100)	(0.113)
Observations	149	145	596	299	297

Note. The first two columns show the mean values of each standardised index for the control group. The last three columns reflect the coefficient of the treatment binary variable, with the dependent variable in each model being the one shown on the left-hand side. The item in bold reflects the index, which is composed of the *z*-scores of the sub-items below. (r) means that the order of responses to the questions was reverted so that a higher value of the index indicates higher importance to women-related subjects in mathematics, in the first case, and on higher self-concept levels, in the second case. All the models control for gender (only for the aggregate model), age, whether or not individuals are foreign-born, student status, number of books in the household, and reported skills on mathematics performance. Standard errors in parentheses. *** p<0.01, ** p<0.05, and * p<0.1.

General Conclusion

This thesis empirically explores the potential factors contributing to divergent labour market outcomes for individuals at different stages of their careers. The first chapter focuses on adolescent students and presents new evidence on the differential effects of the use of educational technology on students' mathematical achievement. This has implications for labour market trajectories, as achievement in mathematics is a strong predictor of future career choices. The second chapter examines the determinants of wage trajectories of new entrants in the labour market by using large administrative data of the Spanish social security system. The third chapter broadens the analysis to the overall workforce to explore the drivers of occupational gender segregation by implementing a novel large-scale survey designed in-house. This analysis is complemented by a second survey which assesses the causal impact of a cost-effective randomised controlled trial aimed at addressing adolescents' misconceptions about mathematics. While similar initiatives are proliferating in Spain, little attention has been paid to causally analysing their effectiveness in achieving desired outcomes.

The results in Chapter 1 shed light on the existence of an inverted U-shaped relationship between the frequency of usage of technology at school and performance in mathematics. The paper applies causal inference techniques to show that prolonged exposure to educational technology has a large, negative impact on students' performance in mathematics. In the case of Spain, the overuse of ICT leads to an average underperformance equivalent to half an academic year compared to the rest of the students. This chapter contributes to the exploration of the way ICT—which is increasingly present in schools—affects student performance, a paramount topic for instructors and policy makers in their search for an optimal use of technology that enhances students' learning processes.

The second chapter moves on to the following stage of individuals' trajectories by focusing on young entrants to the Spanish labour market. This chapter assesses how initial conditions upon entry may affect wage trajectories in the following five and ten years, respectively. Both the OLS and IV results show that individuals whose yearly earnings fall below 60% of the national average experience significant wage penalties in the medium and long term, a phenomenon commonly referred to as the “scarring effect”. Further exploring the labour-related factors behind the scarring effects yields that work intensity and, in particular, non-employment spells represent the key factor in explaining the scarring effects, more so than hourly wages. This chapter contributes to the literature by examining adverse labour market conditions at the individual level, diverging from the conventional focus on macroeconomic indicators, notably employment. A key implication of this chapter lies in the importance of policies focusing on addressing labour intensity and, notably, the continuous inflows and outflows into and out of employment, a feature that has long characterised the Spanish labour market.

The first part of Chapter 3 expands the analysis to the overall workforce. The results show that educational presorting largely accounts for occupational segregation in the labour market. At the same time, educational presorting stems from a multifaceted interplay of factors, including contextual and psychological traits. Math anxiety emerges as a key factor in understanding discouragement from STEM-related careers (science, technology, engineering and mathematics). This phenomenon particularly affects women, partly explaining the gender gap in the pursuit of technical careers. Finally, this chapter concludes by conducting a causal evaluation of a female role model intervention addressed to pre-university individuals. The results show that the intervention improves individuals’ perceptions of mathematics. This chapter, in turn, sheds light on the importance of addressing gender segregation prior to labour market entry. The central role of STEM jobs in driving economic growth and national competitiveness, as well as the importance of non-segregated occupations in reducing the gender wage gap, highlights that tackling gender segregation in labour markets should be at the forefront of the policy agenda.

In conclusion, this thesis has made progress in understanding the drivers that shape individuals' different labour market outcomes. However, there are avenues for future research. Concerning the first chapter, further research could explore the ways in which the quality of technology use, rather than just the quantity, influences mathematical performance. The second chapter focuses on the determinants of wage trajectories, while some other outcome variables warrant exploration. For instance, while certain factors may be detrimental to future wages (e.g., part-time jobs), these factors may have positive effects on other non-pecuniary outcomes, such as well-being. The third chapter could benefit from the use of longitudinal data to provide a more comprehensive understanding of the drivers of gender segregation in the labour market. Moreover, an assessment of the medium- and long-term effects of the randomised controlled trial could provide valuable insights into the potential sustainability of these effects. All in all, I hope that the findings of this thesis will pave the way for new avenues of research on the determinants of labour market outcomes.